Use of Recurrent ANN and Hybrid EA for the Prediction of Phytoplankton Abundance and Succession Before and After Eutrophication Control of Two Shallow Lakes

[1]**Talib, A.** , [1] **F. Recknagel,** [1]**H . Cao and** [2]**van der Molen. D.T.**

[1]School of Earth and Environmental Sciences, University of Adelaide, 5005, Australia
[2]Institute of Inland Water Management, 8200 Lelystad, The Netherlands
E-Mail: anita.talib@adelaide.edu.au

Keywords: *Recurrent supervised ANN; hybrid evolutionary algorithms; rule sets; forecasting; eutrophication control; phytoplankton succession*

EXTENDED ABSTRACT

Eleven years of water quality time-series of the two Dutch lakes Veluwemeer and Wolderwijd were subject to predictive modelling by recurrent supervised artificial neural networks (RNN) and hybrid evolutionary algorithms (HEA). The modeling aimed at forecasting changes of the phytoplankton community in response to the control of external nutrient loadings and fish abundances as consecutively implemented to both lakes since 1979. The water quality time-series of both lakes were structured for the RNN and HEA modeling in order to reflect following three different management periods by both training and validation datasets: no management (1976-1978), lake flushing and waste water treatment (1979 onwards) and lake flushing, waste water treatment and food web manipulation (1991-1993). This approach facilitated a comparative analysis for the two lakes and the three management periods. Firstly recurrent RNN achieved reasonably accurate results for 5-days-ahead forecasting of abundances of blue-green algae *Oscillatoria* and green algae *Scenedesmus* in both lakes. Secondly hybrid evolutionary algorithms (HEA) achieved similar good forecasting results but also provided model representations for both algae species in the form of rule sets. HEA has been designed to evolve both the structure of rule sets as well as the parameter values imbedded in the rule sets by means of a genetic algorithms. With regards to the different approaches for eutrophication management the modeling results have shown that only the combination of external nutrient control with food web manipulation has changed the lakes from hypereutrophic to mesotrophic conditions reflected by the change from the dominance of blue-green algae *Oscillatoria* to the dominance of green algae *Scenedesmus*. Even though both modeling techniques have forecasted the succession of two functional algal groups represented by *Oscillatoria* and *Scenedesmus* only HEA provides rule sets for the explanation of these ecological changes. The results revealed that phosphorus limitation by means of seasonal lake flushing and wastewater treatment in combination with increased zooplankton grazing by food-web manipulation diminished the abundance of the harmful *Oscillatoria* but enhanced the abundance of harmless *Scenedesmus*. These findings consent well with literature findings e.g. by Benndorf (1995) that the eutrophic lakes requires primarily efficient nutrient control that secondarily can be finetuned by food-web manipulation.

1. INTRODUCTION

In the framework of the present study recurrent supervised artificial neural networks (RNN) and hybrid evolutionary algorithms (HEA) were applied to forecasting population dynamics of the blue-green algae *Oscillatoria* and the green algae *Scenedesmus* in the Dutch lakes Veluwemeer and Wolderwijd. RNN have successfully been applied for modelling eutrophication processes in freshwater lakes and rivers based on long-term time series (Jeong *et al.* 2001; Walter *et al*. 2001; Jeong *et al.* 2005; Recknagel, Kim and Welk 2005; Recknagel *et al*. 2005). HEA is a newly emerging technique that not only facilitates time series forecasting but also discovery of explanatory rules (Cao *et al*. 2005). Both methods are applied to eleven years of water quality time-series of the two Dutch lakes in order to forecast changes of the algal populations in response to the control of external nutrient loadings and fish abundances as consecutively implemented to both lakes since 1979. The water quality time-series of both lakes are structured for the RNN and HEA modeling so that it is possible to reflect the following three different management periods by both training and validation datasets: no management (1976-1978), lake flushing and waste water treatment (1979 onwards) and lake flushing, waste water treatment and food web manipulation (1991-1993). This approach facilitates a comparative analysis for the two lakes and the three management periods. It also allows a comparison between the two computational techniques with regards to forecasting accuracy and explanation similar to Recknagel *et al* (2002). The results demonstrate firstly that RNN achieve reasonably accurate 5-days-ahead forecasts of abundances of blue-green algae *Oscillatoria* and green algae *Scenedesmus* in both lakes. Secondly it is shown that hybrid evolutionary algorithms (HEA) achieve similar good forecasting results but also provided model representations for both algae species in the form of rule sets which can be causally interpreted. Thirdly it is revealed that phosphorus limitation by means of seasonal lake flushing and wastewater treatment in combination with increased zooplankton grazing by food-web manipulation successfully diminished the abundance of the harmful *Oscillatoria* but enhanced the abundance of harmless *Scenedesmus* in both lakes.

2. STUDY SITES, MATERIALS AND METHODS

2.1. Lake Veluwemeer

Lake Veluwemeer was created in 1957 and is adjacent to Lake Wolderwijd (Fig. 1). Both lakes have similar geographical and hydrological conditions (Table 2.1). Originally Veluwemeer was a clear water lake with abundant macrophytes. From 1965 onwards the lake became increasingly turbid as a result of rising phosphorus loadings. As a result frequent blooms by *Oscillatoria* agardhii occurred in the mid 1970s. Phosphorus control by a sewage treatment plant and lake flushing in winter was implemented in 1979. Polder water with low concentrations of algae and phosphorus but high concentrations of calcium and nitrate was used for flushing. From 1985 onwards, summer flushing was also implemented. Commercial fishing was introduced to Lake Veluwemeer in the early 1990s, peaking in 1994 (Portielje & Rijsdijk 2003).

2.2. Lake Wolderwijd

Lake Wolderwijd was created in 1968 and ongoing eutrophication processes caused hypertrophic conditions in the 1970’s and early 1980s with high abundances of blue-green algae. Between 1980 and 1983 large amounts of water from upstream lake Veluwemeer have occasionally been flushed through Lake Wolderwijd (van der Molen 1999). From November 1990 to July 1991 a food web manipulation was carried out whereby 75% of bream was removed and young pike introduced. During that period the abundance of blue-green algae decreased by approximately 50%. From 1996 and 1997 the water quality declined again (Van der Molen, 1999).

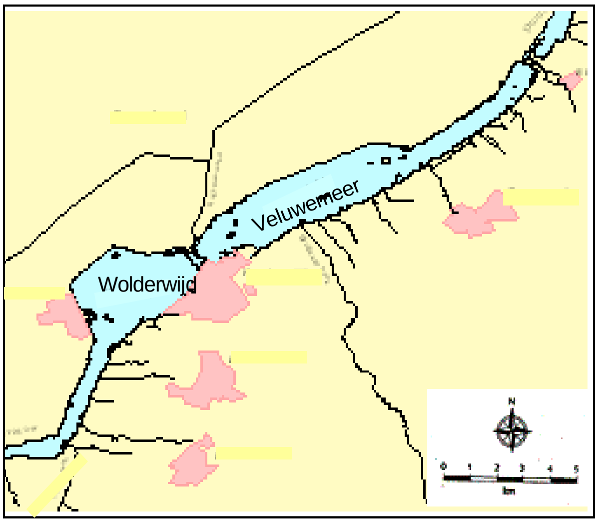

Figure 1 Locations of Lakes Veluwemeer and Wolderwijd in central Netherlands

Table 2.1 Atttributes of Lake Veluwemeer and Wolderwijd

Attributes	Lake Veluwemeer	Lake Wolderwijd
Latitude	52 ° 23'	52 ° 20'
Longitude (-0.2 m below sea level)	5 ° 40'	5° 35'
Area (Km2)	32.4	26.7
Maximum Depth (m)	7.8	5.7
Mean Depth (m)	1.58	1.81
Precipitation (mm)	800	800
Geological Components	Sandy deposits	Sandy deposits

Table 2.2 Analyses of limnological variables of Lake Veluwemeer and Wolderwijd

	Lake Veluwemeer (1976-1993)	**Lake Wolderwijd (1976-1993)**
Limnological variables	**Mean/Min/Max**	**Mean/Min/max**
Nitrate NO_3-N DIN mg/l	0.86/0.001/5.77	0.24/0.001/7.24
Phosphate PO_4-P DIP mg/l	0.04/0.0001/0.42	0.01/0.0001/0.12
Silica Si mg/l	2.6/0.05/7.05	2.06/0.01/19.1
Ammonium NH_4^+ mg/l	0.12/0.001/1.77	0.05/0.001/0.85
pH	8.5/7.3/10.5	8.5/7.1/9.7
Temperature Temp ^oC	10.8/-1.7/25	11.0/0/23.9
Secchi Depth SD m	0.4/0.1/1.7	0.4/0.15/1.3
Chlorophyll-a Chla µg/l	111.8/12.6/459.2	95.6/9/265.3
Oscillatoria cells/ml	17658/25/95850	26970/0/97650
Scenedesmus cells/ml	2216/0/17250	1299/0/12688

2.3. Lake Data

Data of the Lakes Veluwemeer and Wolderwijd were preprocessed by linear interpolation to create two consistent data sets. Table 2.2 lists the water quality variables from both lakes that were considered for the present study.

2.4. Recurrent Artificial Neural Networks RNN

RNN as introduced by (Pineda 1987) mimic deterministic modelling whereby the system state at time t is calculated by means of the system state at time (t-1) (Recknagel 2001) and the copied weights of time (t-1) used as feedback inputs to determine weights of neurons at time t. Eleven years of water quality data of the two lakes from 1976 to 1993 were used for training the 2 recurrent supervised ANN models. The selection of training and testing years was based on data availability to include the years typical for distinct eutrophication management approaches. Testing was based on three years data that represented different management periods: 1978, 1985 and 1993. The output variables tested by the recurrent supervised ANN models for both lakes are 5-days ahead forecasts of *Oscillatoria agardhii* and *Scenedesmus* cell counts. To achieve this aim, the models were trained with input variables time lagged by 5 days. The remaining variables listed in Table 2.2 were considered as input variables.

The RNN were designed with one hidden layer for all applications. All models were trained using 21 nodes. Hyperbolic tangent function was chosen to estimate the activation level for both the hidden and output layers with momentum set at 0.7 for both layers. Training was done for up to a maximum of 1000 iterations. The termination was based on the minimum MSE (mean squared error) of 0.01. Model validation was based on visual comparison of the curve fitting trends and the R^2 value derived from linear regression without intercept for the measured and predicted output data. All models were developed using NeuroSolutions Version 4.24 (NeuroDimension 2003).

2.5. Hybrid Evolutionary Algorithms HEA

HEA has been designed in order to discover predictive rule set. It firstly evolves the structure of the rule sets by using genetic programming (GP) (Koza 1992, 1994; Banzhaf *et al.* 1997), and secondly optimises the random parameters in the rule set by using a general genetic algorithm (Yu et al. 1999). Rules discovered by HEA have the IF-THEN-ELSE structure and allow imbedding complex functions synthesised from various predefined arithmetic operators. The principal framework of HEA for the rule discovery in water quality time-series is represented in Fig. 2.

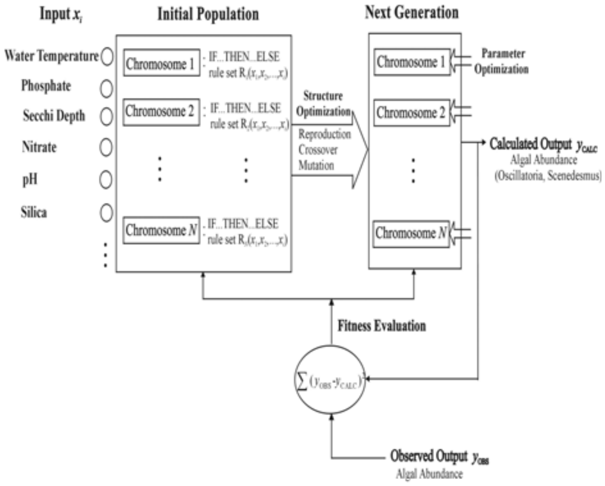

Fig 2 Conceptual diagram of HEA for the discovery of predictive rule sets in water quality time-series

In order to make comparisons with RNN results, we used the same data sets for training and testing. 100 runs were conducted independently for each data set. For simplicity, we set the maximal rule size to be 1 (single rule). All the experiments were performed on a Hydra supercomputer (IBM eServer 1350 Linux) with a peak speed of 1.2 TFlops by using the programming language C.

In order to validate the results of different rules, we define the RMSE (Root Mean Square Error) as

$$\text{Fitness} = \sqrt{\frac{1}{k}\sum_{i=1}^{k}(\hat{y}_i - y_i)^2}$$

the training error and testing error:

where k is the number of training (testing data points) y_i and $\hat{y}_i$ are the ith observed value and the ith predicted value of the output variable.

3. RESULTS AND DISCUSSION

3.1. Forecasting of *Oscillatoria* abundance under different lake management conditions

The measured data of the three independent testing years 1978, 1985 and 1993 showed similar trends of the development of *Oscillatoria* abundances in response to different management for both lakes. The two modeling techniques achieved 5-days-ahead forecasts of the observed trends of *Oscillatoria* in Lake Veluwemeer with R^2 = 0.87 (RNN, see Fig 3a) respective R^2 = 0.92 (HEA, see Fig 3b, and in Lake Wolderwijd with R^2 = 0.66 (RNN see Fig 3e) respective R^2 = 0.65 (HEA see Fig 3f). Both models predicted reasonably well the high abundances of *Oscillatoria* in 1978 as result of highly eutrophic conditions, but consistently overestimated the abundances in 1985 when external nutrient control was implemented in both lakes. In 1993 when nutrient control was complemented by food web manipulation both models the peaks of *Oscillatoria* in summer.

3.2. Forecasting of *Scenedesmus* abundance under different lake management conditions

Both modeling techniques performed 5-days-ahead predictions of the observed trends of *Scenedesmus* in Lake Veluwemeer less accurately compared to *Oscillatoria* with R^2 = 0.42 (RNN, see Fig 3c) respective R^2 = 0.52 (HEA, see Fig 3d), and in Lake Wolderwijd with R^2 = 0.29 (RNN see Fig 3g) respective R^2 = 0.29 (HEA, see Fig 3h). Whilst the RNN model overestimated the abundances of *Scenedesmus* in Lake Veluwemeer in 1985 and matched well the observed abundance in 1993, the HEA model performed more accurately for 1985 but overestimated the abundances of *Scenedesmus* in 1993. Both models had difficulties to predict timing and magnitudes of *Scenedesmus* in in Lake Wolderwijd for the three testing years with reasonable accuracy.

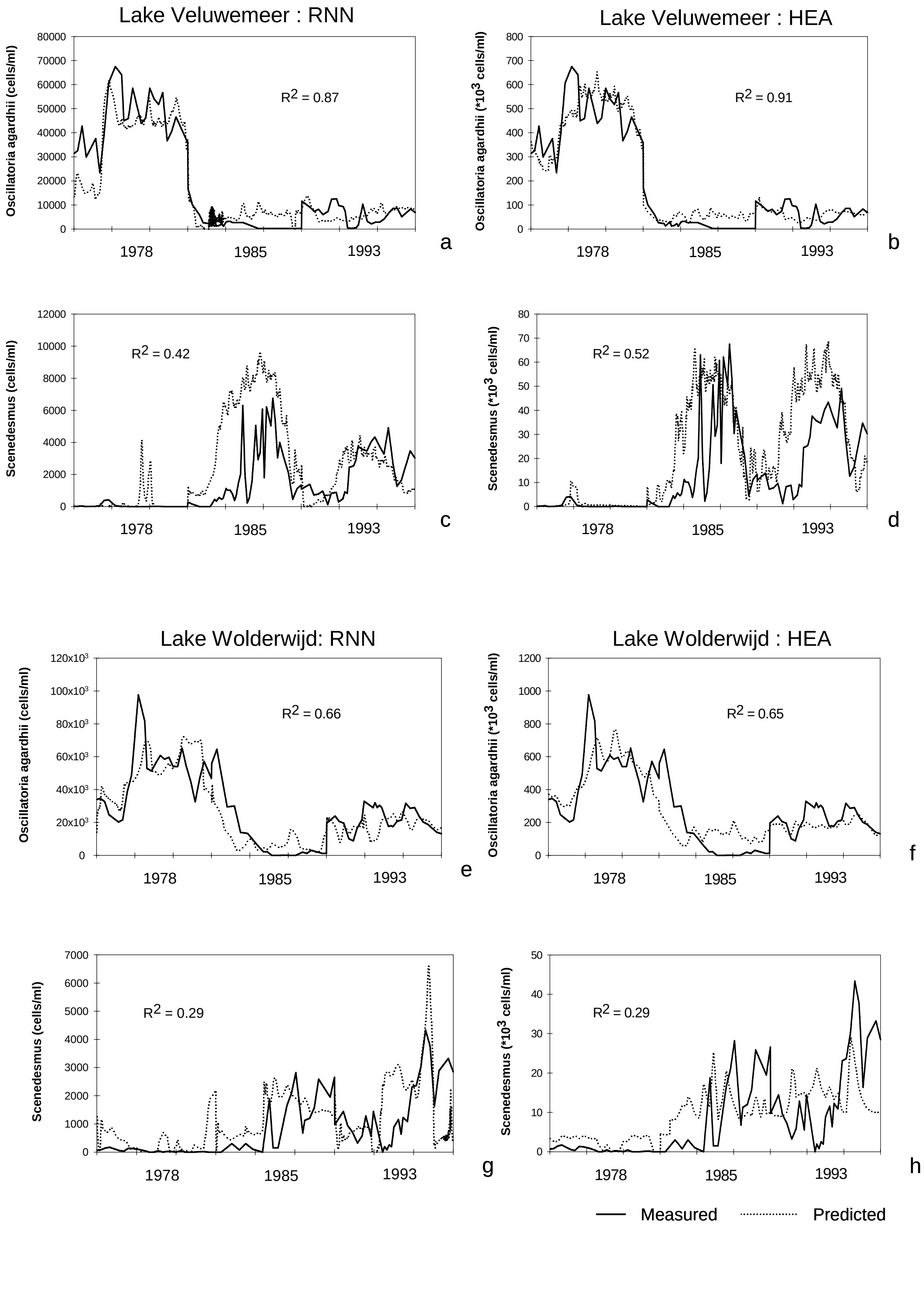

Fig 3: 5- days ahead prediction of *Oscillatoria agardhii* and *Scenedesmus* of Lake Veluwemeer (a-d) and Lake Wolderwijd (e-h) using RNN and HEA tested with 3 years data tested for 1978 (no management), 1985 (nutrient control) and 1993 (nutrient control and food web manipulation).

3.3. Successional patterns of *Oscillatoria* and *Scenedesmus* under different lake management conditions

The hypothesis was that phosphate reduction and intensive flushing during winter may lead to the break-up of *Oscillatoria* bloom with clear-water, P-limited algal growth, reduced pH and therefore lowered P release from the sediments. This should trigger the shift towards a clear-water state with increased abundance of green-algae, *Scenedesmus*. The rule sets that were discovered by HEA for the two algae populations in the two lakes and used for the prediction results in Fig. 3 are documented in Table 3. The rule sets 1 and 2 in Table 3 that were discovered for the 5-days-ahead prediction of *Oscillatoria* in the two lakes are both largely determined by chlorophyll a, Secchi depths, silica and water temperature. By contrast rule sets 3 and 4 in Table 3 for *Scenedesmus* are determined by chlorophyll a, PO_4-P , silica, pH and water temperature . These rule sets give evidence that *Oscillatoria* and *Scenedesmus* abundance are affected by chlorophyll-a and temperature as they are common in all the rule sets. This indicates the role of light-limiting factors in the succession between *Oscillatoria* and *Scenedesmus*. The available underwater light often limits growth of diatoms and green algae in eutrophic lakes when chlorophyll-a concentrations are high as a result of blue-green algae abundance in summer (Reynolds 1984, Chorus & Bartram 1999). Blue-green algae tend to grow faster at low irradiance as was demonstrated by competition experiments for light between *Oscillatoria* and *Scenedesmus* under lab-controlled conditions (Walsby 1992). *Scenedesmus* grow more rapidly and absorb more of the light entering the system, thus the average irradiance in the water column falls (Reynolds, 1984). Buoyancy, is another factor as *Scenedesmus* are non-bouyant while blue-green algae are able to gain bouyancy at low irradiance and lose it at high irradiance (Walsby, 1992). Another common variable discovered from the rule sets is pH. Sudden shifts in pH may alter phytoplankton assemblages which may explain *Oscillatoria* abundance before phosphate reduction and lake flushing was implemented (1978). It was suggested that the high levels of Ca_2^+ and HCO_3^- in the flushing water may contribute to pH reduction (Hosper 1997). The forecasts have shown that *Scenedesmus* were only abundant after this period. During the periods of increased *Scenedesmus* abundance (1985 and 1993), the rule sets discovered that *Scenedesmus* abundance can be explained with the additional variable phosphate. Similarly, the use of non supervised ANN has shown that the years 1985 and 1993 for both lakes corresponded to the periods of P-limitation with increasing abundance of green algae and diatoms as a result of flushing and phosphate reduction measures (Recknagel, Talib and van der Molen 2005). It is interesting to note that the limitation of both algal groups by silica revealed the complex nature of shallow lake dynamics involving multispecies competition and succession. This is an indirect causal link as silica is important for the growth of diatoms. Although not forecasted in this model, our results indicate that forecasting of *Oscillatoria* and *Scenedesmus* are interrelated to the growth and competition from other algal groups including diatoms. Although previously, phosphorus reduction has been suggested as key-factor for controlling the summer dominance of blue-green algal in the Lakes Veluwemeer and Wolderwijd (Reeders *et al.* 1998), results in this study have illustrated that a combination of ongoing phosphorus control and biomanipulation has achieved both, to further diminish *Oscillatoria* abundance and shift an increasing abundance of *Scenedesmus*. We suggest that the long-term successional patterns observed for *Oscillatoria* and *Scenedesmus* are related to a periodic shift between nutrients and light-limitation with decreasing trophic conditions. Complex dynamics involving competition and co-existence may form the basis for the long-term dynamics of the phytoplankton in Lake Veluwemeer and Lake Wolderwijd. This is typical for the transitions of lakes from hypertrophy to mesotrophic conditions as generalised by Reynolds (1984). Further work will involve studying the merged data sets from both lakes to discover improved HEA rules that relate to the successional dynamics between *Oscillatoria* and *Scenedesmus* in both lakes. Knowledge discovery will be made more robust by excluding chlorophyll-a from the inputs. To improve understanding of the phytoplankton successional dynamics, this modeling approach will also be attempted at the phytoplankton functional group level.

Table 3: Best rule sets for *Oscillatoria* and *Scenedesmus* for Lake Veluwemeer and Lake Wolderwijd.

Algal Population	Best Rule Sets	Training Error	Testing Error
Oscillatoria L.Veluwemeer	**RULE SET 1:** IF (((SD>0.37) OR (Chla<97.39)) AND (exp((DIN-303.47))>134.27)) THEN *Oscillatoria* = ((Chla+Si/SD))+ (Si/SD)) ELSE *Oscillatoria* =((Chla +((temp+Si)*9.73+(10.04/SD))	85.53	63.24
Oscillatoria L.Wolderwijd	**RULE SET 2:** IF (Si<=6.31) THEN *Oscillatoria* = (Chla*((104.96-(Chla/5.48))/(4.31-(Temp+Si))) ELSE *Oscillatoria* = (Temp*pH)	129.36	129.30
Scenedesmus L.Veluwemeer	**RULE SET 3:** IF ((((Chla+DIP)/DIP)>362.77) AND (Chla<=103.64)) THEN Scenedesmus = (((ln(|Sd|) +pH)*((Temp/10.38)*ln(|Chla))))-(pH/ln(| Si|)-Si)) ELSE Scenedesmus = (Temp/(Si*pH))	**22.61**	**17.69**
Scenedesmus L. Wolderwijd	**RULE SET 4:** IF ((exp(Si)+Chla)>125.50) THEN Scenedesmus = (4.82-(DIP*73.84)) ELSE Scenedesmus = (pH- ((Temp*(Temp-110.03))*DIP))	**15.32**	**8.95**

4. CONCLUSIONS

The present study has demonstrated that distinct predictable patterns and complex explanatory rule-sets have been revealed of *Oscillatoria agardhii* and *Scenedesmus* as they undergo competition and succession over temperature preferences and pH tolerance as well as long-term changes in phosphate and underwater light limitations. The rule sets discovered that interactions between the supplies of phosphate, nitrate, silica and the effects of temperature and pH can explain the competition leading to a shift in succession between *Oscillatoria* and *Scenedesmus* in Lake Veluwemeer and Wolderwijd over a long-term period. From the assessment of long-term patterns in both lakes it can be concluded that the ongoing control of nutrient regimes towards phosphorus limitation by means of seasonal lake flushing and wastewater treatment up streams of Lake Veluwemeer since the early 1980s has achieved phosphorus limitation, weakened the abundance of *Oscillatoria* but elevated abundances of *Scenedesmus*. The additional eutrophication control by food web manipulation in both lakes since the early 1990s may have contributed to this shift in algal succession. These findings consent with previous recommendations that the control of eutrophic lakes requires primarily efficient nutrient control that secondarily can be finetuned by biomanipulation (Benndorf 1995). Further research will complement this preliminary study. The next attempt is to apply this method for knowledge discovery at the phytoplankton functional group level.

5. REFERENCES

Banzhaf W, Nordin P, Keller RE, Francone FD (1997) Genetic Programming: An Introduction on the Automatic Evolution of Computer Programs and its Applications. Morgan Kaufmann

Benndorf, J., 1995. Possibilities and limits of controlling eutrophication by biomanipulation. Int. Revue ges. Hydrobiol. 80, 4, 519-534.

Cao H, Recknagel F, Kim B, Takamura N (2005) Evolutionary Algorithm Approach to Unravel and Forecast Ecological Complexity of Two Lakes Different in Morphometry and Eutrophication. In: Recknagel F (ed) Ecological Informatics. Scope, Techniques and Applications.2nd Edition. Springer-Verlag, Heidelberg, New York, 347-367.

Chorus I, Bartram J (eds) (1999) Toxic Cyanobacteria in Water. A guide to their public health consequences, monitoring and management, Vol. E & FN Spon, London and New York

Hosper H (1997) Clearing Lakes. PhD, Wageningen University, The Netherlands

Jeong K-S, Joo G-J, Kim H-W, Ha K, Recknagel F (2001) Prediction and elucidation of phytoplankton dynamics in the Nakdong River (Korea) by means of a recurrent artificial neural network. Ecological Modelling 146:115-129

Jeong K-S, Recknagel F, Joo G-J (2005) 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: Recknagel F (ed) Ecological Informatics. Scope, Techniques and Applications.2nd Edition. Springer-Verlag, Heidelberg, New York, 255-273.

Koza JR (1992) Genetic programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA: MIT Press.

Koza, J R (1994) Genetic programming II: Automatic Discovery of Reusable Programs. Cambridge, MA: MIT Press

NeuroDimension (2003) NeuroSolutions: The Neural NetworkSimulation Environment (Vers 4.24 Developers) and NeuroSolutions for Excel (Vers 4.21)

Pineda FJ (1987) Generalization of back-propagation to recurrent neural network. Phys. Rev. Lett. 59:2229-2232

Portielje R, Rijsdijk RE (2003) Stochastic modelling of nutrient loading and lake ecosystem response in relation to submerged macrophytes and benthivorous fish. Freshwater Biology 48:741-755

Recknagel F (2001) Applications of machine learning to ecological modelling. Ecological Modelling 146:303-310

Recknagel, F., Kim, B. and A. Welk (2005). Unravelling and prediction of ecosystem behaviours of Lake Soyang (South Korea) in response to changing seasons and management by means of artificial

neural networks. *Verh.Internat.Verein.Limnol.* (in press)
Recknagel, F., Kim, B. and A. Welk (2005). Artificial Neural Network Approach to Unravel and Forecast Algal Population Dynamics of Two Lakes Different in Morphometry and Eutrophication. In: Recknagel, F. (2005). Ecological Informatics. 2nd Edition. Springer-Verlag. New York , 1-485.
Recknagel F, Talib A, van der Molen DT (2005) Phytoplankton community dynamics of two adjacent Dutch lakes in response to seasons and eutrophication control unravelled by non-supervised artificial neural networks. Ecological Modelling (in press).
Reeders HH, Boers PC, van der Molen DT, Helmerhorst TH (1998) Cyanobacterial dominance in the lakes Veluwemeer and Wolderwijd, The Netherlands. Water Science and Technology 37:85-92
Reynolds CS (ed) (1984) The Ecology of Freshwater Phytoplankton., Vol. Cambridge University Press.
van der Molen DT (1999) The role of eutrophication models in water management. PhD, Wageningen University, The Netherlands
Walsby AE (ed) (1992) The control of gas-vacoulate cyanobacteria, Vol. Freshwater Biological Association, London
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. Ecological Modelling 146:97-113
Yu JX, Cao HQ, Chen YY, Kang LS Yang HX (1999) A new approach to estimation of the electrocrystallization parameters. Journal of Electroanalytical Chemistry, 474(1), 69-73