Ordination, clustering and forecasting of phytoplankton dynamics in the Myponga drinking water reservoir by means of supervised and non-supervised artificial neural networks

[1]Welk, A., [1]Recknagel, F. and [2]M. Burch

[1]CRC for Water Quality and Treatment, School of Earth and Environmental Sciences
University of Adelaide, Adelaide, 5005 Australia
[2]CRC for Water Quality and Treatment, Australian Centre for Water Quality, Salisbury, 5108 Australia

Keywords: *Algal blooms, artificial neural networks, drinking water reservoir, forecasting, ordination and clustering.*

ABSTRACT

In order to minimize water quality problems caused by cyanobacteria blooms, it is desirable to either operationally control or prevent them. Computational modelling using artificial neural networks (ANN) is one way to contribute to both. It enables not only the forecasting of growing algal population abundances several days in advance, but also a better understanding of both processes and environmental conditions that accelerate algal growth and response behaviour of algal populations to different management. Appropriate measures can then be determined and implemented to control the growth of algal populations before they reach bloom proportions.

A new approach using both supervised artificial neural networks (SNN) and non-supervised artificial neural networks (NSNN) was applied as a framework to explore 20 years of water quality time-series from the Myponga Reservoir, South Australia, for knowledge discovery and forecasting of phytoplankton dynamics.

The Myponga Reservoir has a history of summer blooms of the toxic cyanobacteria *Anabaena*. Anatoxin-a, produced and released by *Anabaena,* not only disintegrates aquatic food chains bottom up but also imposes costs for water treatment and restrictions on water consumption. In order to prevent summer algal blooms in the Myponga Reservoir artificial mixing has been implemented as an operational control measure since the 1980s.

In the present study, firstly recurrent SNN demonstrated their capability to perform 7-days-ahead forecasting of timing and magnitudes of chlorophyll-a (Chl-*a)* for two years of unseen data, which were different in management. Secondly, the combination of ordination by NSNN and sensitivity analysis by SNN revealed relationships between water temperature, PO_4-P and NO_3-N concentrations and Chl-*a* dynamics. Finally, ordination and clustering of four 2-year-periods of water quality time series, which differed in intensity and methods of artificial mixing, provided insights into effects of management on chemical and biological water quality properties.

Results of the Chl-*a* forecasting by SNN demonstrated that using PO_4-P, NO_3-N, water temperature and turbidity as input variables only allowed forecasting of basic trends in Chl-*a* dynamics. The integrated application of NSNN and SNN demonstrated that qualitative relationships between water quality parameters discovered by ordination and clustering could be quantitatively determined by sensitivity analyses. It revealed specific temperature and nutrient ranges that favour phytoplankton growth.

NSNN applied to the more complex task of examining water quality changes between periods of years with different management regimes indicated that both seasonal shifts as well as changed magnitudes of nutrients, metals and Chl-*a* occurred in response to periods with stratification and changed mixing strategies.

It can be concluded from this study that the combined applications of SNN and NSNN provide a useful framework not only for forecasting phytoplankton dynamics but also explaining complex ecological relationships driving these dynamics. The so gained information can facilitate early warning and improve causal understanding of algal blooms.

1. INTRODUCTION

Cyanobacteria blooms in drinking water reservoirs impose high water treatment costs and restrictions on water consumption. To prevent or operationally control such events is therefore highly desirable. Computational modelling by using artificial neural networks (ANN) can contribute to this aim. ANN enable not only the forecasting of phytoplankton growth several days in advance but also a better understanding of processes and environmental conditions that accelerate algal growth, and how specific algal populations respond to different management (Recknagel *et al.* 2005a; Recknagel *et al*., 2005b)

The Myponga Reservoir has a history of summer blooms of the toxic cyanobacteria *Anabaena*. Anatoxin-a, produced and released by *Anabaena*, not only disintegrates aquatic food chains bottom up but also imposes public health risks. One ongoing practice to operationally control the biomass of cyanobacteria in the Myponga Reservoir is the application of $CuSO_4$ as algicide up to 5 times during the summer season (Lewis et al. 2003). In an attempt to prevent algal blooms and therefore avoid the controversial $CuSO_4$ treatment, different artificial mixing strategies, including aeration and mechanical mixing have been applied since the 1980s with insufficient success.

This paper features preliminary results from an integrated application of supervised artificial neural networks (SNN) and non-supervised artificial neural networks (NSNN) for water quality ordination, clustering and Chl-*a* forecasting for the Myponga Reservoir, South Australia, based on 20 years of data. Firstly recurrent SNN demonstrated their capability to perform 7-days-ahead forecasting of timing and magnitudes of Chl-*a* for two years of unseen data, which were different in management. Secondly the combination of ordination by NSNN and sensitivity analysis by SNN revealed relationships between water temperature, PO_4-P and NO_3-N concentrations and Chl-*a* dynamics. Finally, ordination and clustering of three 2-years-periods of water quality time series, which differed in intensity and methods of artificial mixing, provided insights into effects of management on chemical and biological water quality properties.

2. STUDY SITE & DATA

Myponga Reservoir is situated approximately 70km south of Adelaide and is fed by the Myponga River. It provides vital water supply to southern metropolitan areas of Adelaide and the Fleurieu Peninsula. The catchment area has an average annual rainfall of 750mm and originates in the Adelaide foothills with dominant land use estimated to be 62% livestock grazing and 24% dairying (Smalley, 1998; Thomas *et al.*, 1999). Table 1 summarises major characteristics of the reservoir.

Every summer Myponga Reservoir undergoes thermal stratification due to the increased energy from radiation and decreased water inflow from the catchment. In the past the stratification combined with the highly coloured and nutrient rich water created optimal conditions for the growth of algae, particularly cyanobacteria (Kelly 1998). Major bloom events are often in mid-January, although management is on alert from December to March annually (Burch, personal communication, 2005). Artificial mixing is in place from October to March annually.

Table 1. Characteristics of Myponga Reservoir

Surface area	2.8 km^2
Max. volume	26,800 ML
Max. depth	36m
Mean depth	15m
Water residence time	Approx. 3 years
Catchment area	124 km^2

Table 2. Myponga Reservoir database details

Variable	Years	Mean/Min/Max
PO_4	1986-2003	0.022/0.005/0.09
NO_3	1986-2003	0.11/0.001/0.37
Chl-*a*	1986-2003	7/0.2/41.6
Turbidity NTU	1986-2003	4.4/1.2/30
Water temp.	1986-2003	16.1/8.4/25
Iron	1986-2003	0.55/0.05/1.11
Manganese	1986-2003	0.025/0.005/0.12

As the measurement intervals of the raw data (see Table 2 for summary) were highly irregular and sampling dates were different for physical, chemical and biological data, it was necessary to interpolate the data to create consistent daily values as required for the development of ANN models.

3. METHODS

A recurrent SNN (Pineda, 1987) was used for the present study, as they have proved very powerful for time-series modelling of phytoplankton dynamics (Recknagel, 2001; Walter *et al*., 2001; Gurbuz *et al*., 2003; Jeong *et al*., 2003). Recurrent SNN are modifications of a typical backpropagation SNN in that when calculating

the output for a given time (*t*), not only external input variables are considered, but also fed back activation weights from the time step before (*t-1*).
A recurrent SNN was developed using NeuroSolutions 4.2 (NeuroDimension, 2003) to forecast Chl-*a* values 7 days ahead, using 16 years of data for training (1986, 1988-1999 and 2001-2003), and two years for testing (1987 and 2000). Data used for training consisted of the input variables water temperature, PO_4, NO_3 and turbidity. The forecasting results were validated with the two years of independent data that were not used for training of the SNN. The SNN was designed with one Hidden Layer of 16 processing elements (nodes), the Tanh Axon (hyperbolic) transfer function and the Momentum learning rule. A comprehensive sensitivity analysis was conducted by means of the recurrent SNN to discover relationships between the input variables water temperature, PO_4, NO_3 and turbidity and the output variable Chl-*a*.
NSNN, specifically Kohonen SOMs (Kohonen, 1982), were developed using Matlab 6.5.1 (The Math Works Inc, 2003) and the SOM toolbox to ordinate, cluster and visualise water quality and Chl-*a* data with respect to seasons, water temperature, ranges of nutrients and artificial mixing strategies. This method allows the discovery of significant patterns in the input data similar to the traditional Principal Component Analysis (Jongman *et al.*, 1987), whilst also having the ability to cope with non-linearities (Boddy & Morris, 1999).

4. RESULTS & DISCUSSION

Figure 1 illustrates the 7-days-ahead forecasting results for Chl-*a* by the recurrent SNN where two independent testing years were selected from periods of different management regimes, 1987 was without artificial mixing and during 2000 surface mixers and a 200m aerator were used.

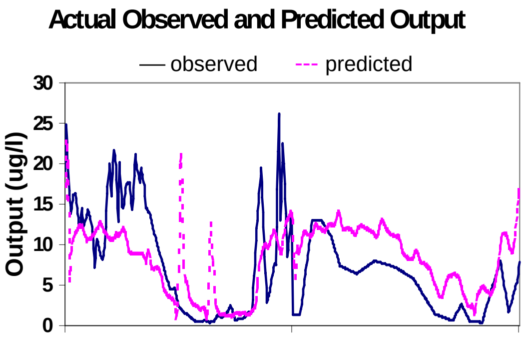

Figure 1. 7 day ahead forecasting of Chl-*a* dynamics using SNN for 2 testing years.

The results show that the recurrent SNN predicts the seasonal dynamics well but underestimates peak events of Chl-*a* concentrations in summer 1987. The testing year 1987 experienced a unique autumn peak which has not been observed in any of the remaining years used for training. The slight overestimation of Chl-*a* in early 2000 may reflect realistic conditions, but algal biomass control by $CuSO_4$ took place, resulting in the drastic drop of Chl-*a* and prediction inaccuracy. Unpredictable events, such as $CuSO_4$ dosing, lead to difficulties in accurate forecasting for highly managed water bodies such as Myponga Reservoir.

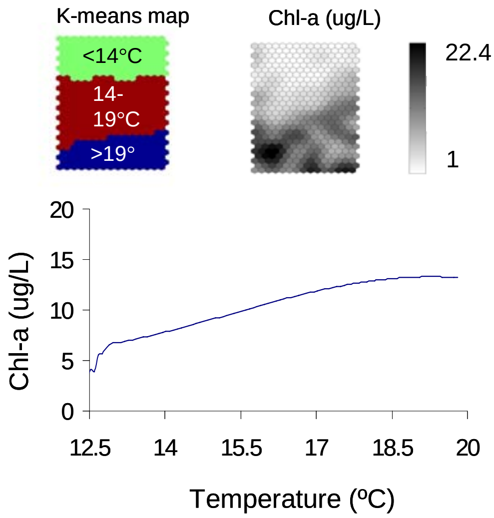

Figure 2. K-means map for temperature ranges (top left) with corresponding component plane for Chl-*a* (top right) in Myponga Reservoir; sensitivity curve for Chl-*a* in response to temperature change.

Sensitivity analyses produced in conjunction with the SNN demonstrated relationships between the input variables and the output (Chl-*a*), and were combined with qualitative ordination and clustering by NSNN. Figure 2 shows the relationship between Chl-*a* and water temperature. As the SNN sensitivity curve (Fig. 2, bottom) indicates, Chl-*a* increases steadily between 13 and 18°C, and levels off at its maximum between 19 and 20°C. These findings are backed up by the NSNN-based ordination and clustering of Chl-*a* regarding 3 temperature ranges (Fig. 2, top). Whilst Chl-*a* concentration is lowest in the temperature range below 14°C, it reaches its maximum in the range above 19°C. As previously explained, cyanobacteria tend to have high abundances in summer and autumn in Myponga Reservoir, and in general, cyanobacteria are known to prefer high water temperatures (Shapiro, 1990; Reynolds, 1984). The stimulating effect of temperature can be both direct and indirect. The indirect effect occurs when the pelagic zone thermally stratifies, typical for Myponga data before 1988. It may not be exclusively the water temperature that attracts the higher algal abundances, it may also be some other conditions which coincide with the warmer weather. Thermal stratification enables buoyancy regulation by some types of cyanobacteria possessing gas vacuoles, allowing them to adjust their vertical position in the water column for access to solar radiation at the surface layers and nutrients near the thermocline (Reynolds, 1984).

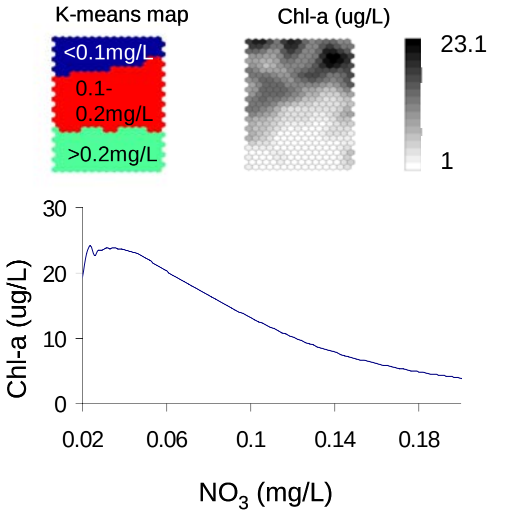

Figure 3. K-means map for NO_3 ranges (top left) with corresponding component plane for Chl-*a* (top right) in Myponga Reservoir; sensitivity curve for Chl-*a* in response to NO_3 change.

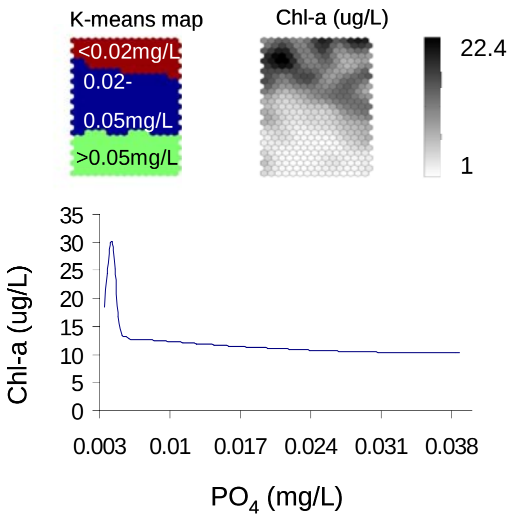

Figure 4. K-means map for PO_4 ranges (top left) with corresponding component plane for Chl-*a* (top right) in Myponga Reservoir; sensitivity curve for Chl-*a* in response to PO_4 change.

The sensitivity analysis in Fig. 3 (bottom) shows that Chl-*a* decreases with increasing NO_3 concentrations, and peaks at concentrations of 0.02 to 0.03mg/l. The corresponding ordination and clustering of Chl-*a* related to 3 ranges of NO_3 concentrations (Fig. 3, top) confirms these findings with Chl-*a* abundance peaking in an area corresponding to the lowest NO_3 concentration range (<0.1mg/L) on the k-means map. There are two possible explanations for this result. On the one hand, it could be postulated that at the time of fast algal growth, NO3 consumption is highest causing inverse effects. Alternatively *Anabaena*, the dominating cyanobacteria in summer, is well known to perform nitrogen fixation by heterocysts and continue to grow during stages of NO_3 deficiency (Reynolds, 1984).

Fig. 4 reflects the relationship between PO_4 and Chl-*a* concentrations. Similar to Fig.3, it reveals that highest Chl-*a* concentrations coincide with low PO_4 concentrations. PO_4 is most likely to be the limiting nutrient in Myponga Reservoir due to minimal nutrient influx from the catchment runoff during summer (Smalley, 1998), so its inverse relationship with Chl-*a* seems to be determined not just be the PO_4 consumption by algae. *Anabaena* often peaks in summer at low PO_4 concentrations, which may be due to their capability to maximise nutrient uptake during times of depletion (Sommer, 1989).

Fig. 5 visualises results of the seasonal ordination and clustering of major water quality variables of the Myponga Reservoir, by means of NSNN, for four periods with different mixing conditions. It shows that during the period from 1986 to 1987 (Fig. 5, left column), when no artificial mixing was used and the reservoir experienced thermal stratification, Chl-a was highest in summer and autumn with a maximum of 15.9ug/L. *Anabaena* reached highest abundance at 1320 cells/ml in autumn and PO_4 concentration was greatest in winter and spring with a maximum of 0.056mg/L. The concentrations of Manganese (Mn), that also have implications for drinking water quality were high in summer and autumn with a maximum of 0.03mg/L.

During the period from 1988 to 1989, when 3 submersible mixers were used (Fig. 5, second column from left), the abundance of Anabaena was highest in spring with a much lower maximum of 852 cells/ml compared to the previous period with stratification. This result suggests that the introduction of artificial mixing succeeded in creating an unsuitable environment for the species, leading it to peak in a different season, at a level of half its abundance in the previous period where no mixing was used. Although Chl-*a* was lower as well, it still peaked in autumn, and was obviously not caused by

Anabaena. The lower *Anabaena* and Chl-*a* abundances in 1988 to 1989 were also reported by Velzeboer et al. (1991), with the conclusion that the mixers were successful in reducing total algal biomass. Interestingly, the results in Fig. 5 indicate a slight decrease of PO_4 concentrations in 1988 to 1989 compared to the previous period with stratification. This result may hint at lower internal PO_4 loading from anaerobic sediments since mixing aims to maintain aerobic conditions at the sediment. In contrast, the Mn concentrations increased in the same period and peaked in summer. During the period from 1990 to 1991 (Fig. 5, third column from left) an aerator was implemented to the Myponga Reservoir, which resulted in an increase of Chl-*a* similar in concentration to the period with stratification, but different regarding seasonality with maximum concentrations in autumn and winter. The PO_4 and Mn concentrations also increased slightly.

Finally, the period from 2000 to 2001 (Fig. 5, right column), where a combination of surface mixers and aerators was implemented, showed very similar patterns as the period with stratification where Chl-a peaked in summer and autumn with higher magnitude, PO_4 peaked in spring and winter at a higher concentration and Mn peaked in autumn with a higher concentration. Mn concentration consistently increased throughout the periods and peaked in autumn in all periods. This is possibly due to sediment release that may occur during the warmer seasons of summer and autumn, when stratification can potentially occur.

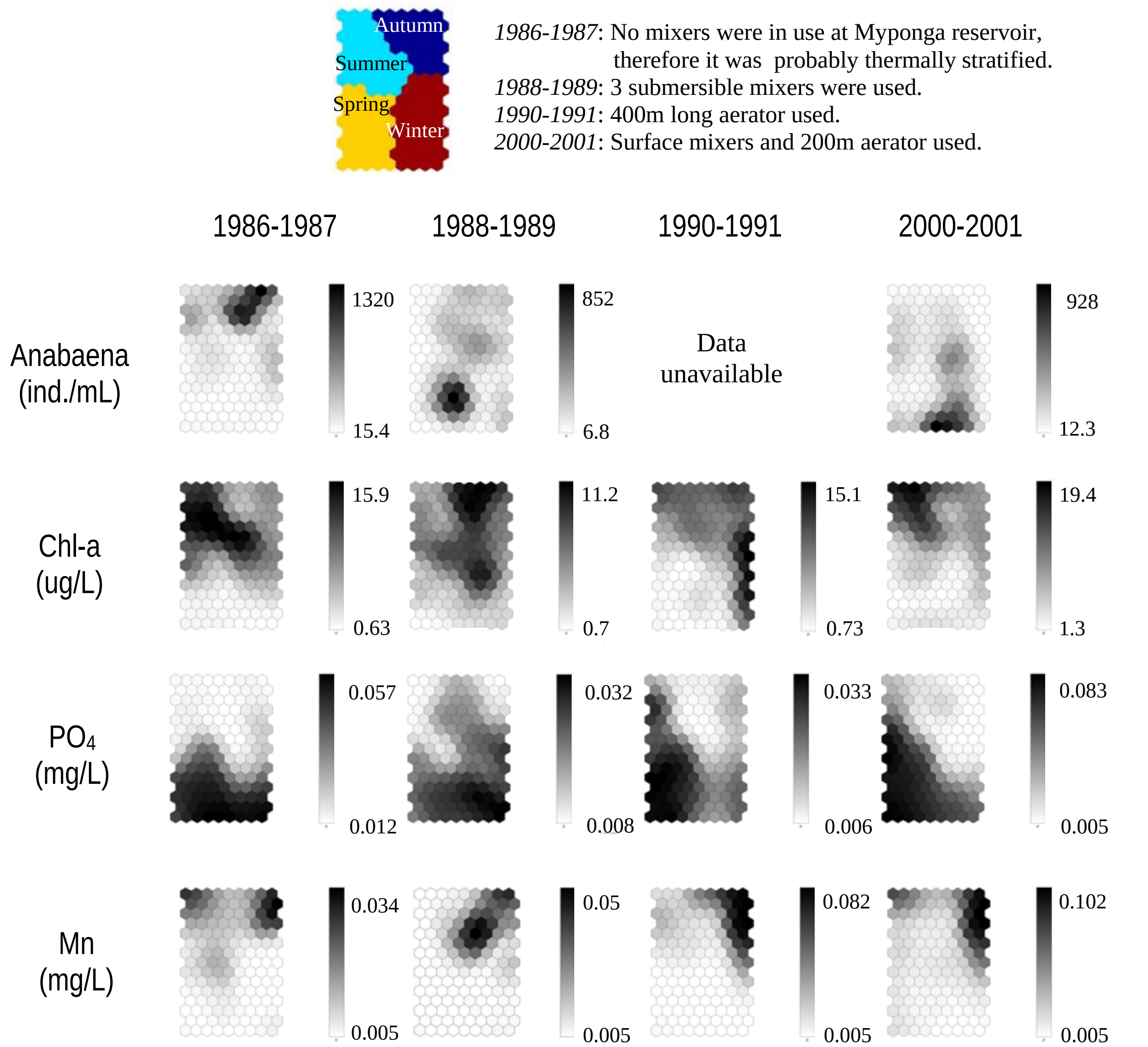

Figure 5. K-means map showing seasons (top) with corresponding component planes for major water quality variables

5. CONCLUSION

The present study used a new approach for the exploration of ecological time-series both qualitatively and quantitatively. It can be concluded that the combined applications of SNN and NSNN provide a useful framework not only for forecasting phytoplankton dynamics but also explaining complex ecological relationships driving these dynamics. A major criticism of ANNs has been that they are ‘black box’ models, which give no indication of the processes involved in the modelled system. The use of sensitivity analyses performed on the training data of SNNs rectifies this issue by revealing how changes in the input variables affect the output, thereby giving SNN an explanatory quality in addition to predictive capabilities. The so gained information can facilitate early warning and improve causal understanding of algal blooms. In future, this method will be linked to online water quality monitoring, which will enable real-time forecasting of phytoplankton dynamics and the development of early warning systems for algal blooms.

REFERENCES

Boddy, L. and C.W. Morris (1999), Artificial neural networks for pattern recognition. In: Machine Learning Methods for Ecological Applications. (Ed. A. Fielding) Kluwer Academic Publishers, Boston, Dordrecht, London. 37-87.

Gurbuz, H., Kivrak, E., Soyupak, S. and S. Yerli (2003), Predicting dominant phytoplankton quantities in a reservoir by using neural networks. *Hydrobiologia* 504: 133-141

Jeong, K-S, Recknagel, F. and G-J. Joo (2003), Prediction and elucidation of population dynamics of the blue-green algae *Microcystis aeruginosa* and the diatom *Stephanodiscus hantzchii* in the Nakdong River-Reservoir system by a recurrent artificial neural network. In: Recknagel, F. (2005). Ecological Informatics. 2nd Edition. Springer-Verlag. New York , 255-273.

Jongman, R.H.G., ter Braack, C.J.F., and O.F.R. van Tongeren (1987), Data analysis in community and landscape ecology. Cambridge University Press, Cambridge.

Kelly, L. (1998), The diversity and abundance of algae in Myponga Reservoir over the 1997-1998 summer season, CRC for Water Quality and Treatment, Adelaide.

Kohonen, T. (1982), Self-organised formation of topologically correct feature maps. *Biological cybernetics* 43: 59-69.

Lewis, D., J.A. Elliot, J.D. Brooks, A.E. Irish, M.F. Lambert and C.S. Reynolds (2003), Modelling the effects of artificial mixing and copper sulphate dosing on phytoplankton in an Australian Reservoir. *Lakes & Reservoirs: Research and Management* 8, 31-40

NeuroDimension (2003) NeuroSolutions: The Neural Network Simulation Environment (Version 4.24 Developers) and NeuroSolutions for Excel (Version 4.21).

Pineda F. (1987), Generalisation of backpropagation to recurrent neural networks. *Phys. Rev. Lett.*19: 2229-2232.

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

Recknagel, F., B. Kim and A. Welk (2005)a. 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., A. Welk, B. Kim and N. Takamura (2005)b. 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 , 325-345.

Reynolds, C.S. (1984), The Ecology of Freshwater Phytoplankton, Cambridge University Press, Cambridge.

Shapiro, J. (1990), Current beliefs regarding the dominance by blue-greens: the case for the importance of CO_2 and pH. *Verh.Int.Verein. Limnol.*, 24, 38-54.

Smalley, T. (1998), Efficiency of Artificial Destratification for Control of Blue Green Algae in the Myponga Reservoir. Honours Thesis, University of Adelaide.

Sommer U. (1989), Toward a Darwinian Ecology of Plankton. In: Plankton Ecology: Succession in Plankton Communities. (Ed. U. Sommer) 1-8 Springer-Verlag: Berlin, Heidelberg, New York, London, Paris, Tokyo.

The Math Works Inc. (2003) Matlab (Version 6.5.1) and SOM toolbox.

Thomas, D., S. Kotz and S. Rixon (1999), Watercourse survey and management recommendations for the Myponga River catchment. Environmental Protection Agency, Adelaide, South Australia, July 1999.

Velzeboer, R.M.A., J.A. Cugley and J.C. Patterson (1991), Modelling Optimum Conditions for Reservoir Destratification using Mechanical Mixers. Urban Water Research Association of Australia, Research Report No.24, April 1991.

Walter, M., Recknagel, F., Carpenter, C. and M. Bormans (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.