Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

Modeling of algal blooms in freshwaters using artificial neural networks

M. French[a] & F. Recknagel[b]

[a] *Department of Civil Engineering, University of Louisville, Louisville, Kentucky 40292, USA*

[b] *Department of Environmental Science, University of Adelaide, Roseworthy, S.A. 5371, Australia*

Abstract

The development of a neural network model for predicting algal blooms is described. The neural network consists of a 3 layer structure with input, hidden, and output layers. Training is conducted using back-propagation where the data are presented as a series of learning sets such that the inputs are observable water quality parameters and outputs are the biomass quantities of specific algal groups. Training is conducted using three years of daily values of water quality parameters and validation is performed using two years of independent daily values to predict the magnitude and timing of blooms of 7 different algae groups with a lead time of 1 day using only the current day water quality parameters. The water quality data represent physical and limnological characteristics of a drinking water reservoir in Germany. Results indicate that the neural network model is capable of learning the complex relationships describing the seasonal succession of phytoplankton in freshwaters. The neural network is shown to perform well for predicting both the timing and magnitude of algae blooms for data in used in training set and to accurately predict the timing and typically over- or under-estimate the magnitude of blooms when applied to the independent data.

1 Introduction

Algal blooms are characterized by proliferation of one or more algae species in huge numbers where one of the abundance criteria (number of cells, biomass, chlorophyll-a, dry weight of seston) is violating a stipulated limit value. Among the harmful consequences of algal blooms which may impede the supply of drinking water are taste, odor and color in the raw water, clogging of filters and encrustation of pipes in water works. Factors affecting the growth of algae in freshwaters are not well understood, although the availability of nutrients, light, warm temperatures, calm conditions and grazing by herbivorous zooplankton appear to be important. If the key factors of algal growth, and importantly their relationships, can be clearly described then it may be possible to predict algal blooms.

Different types of algal models have been proposed and developed to describe the development of phytoplankton in freshwaters in a statistical manner [Vollennweider and Kerekes 1], in a deterministic manner [Jorgensen 2; Straskraba und Dvorakova 3; Park et al. 4; Benndorf and Recknagel 5; Recknagel and Benndorf 6] and in a rule-based manner [Reynolds 7; Sommer et al. 8; Recknagel et al. 9]. Until recently, none of the existing approaches have successfully accomplished the task of predicting both the magnitude and timing of occurrence and growth of several various algal groups using a short-term lead time. The ability to predict specific algal blooms in water bodies can be very useful for efficient preventive or operational control of such events. A study by Ruck et al. [10] showed success in the use of a neural network model to predict benthic community populations based on geophysical, limnological, and sedimentological variables from the Canadian Great Lakes; their results were preliminary in nature due to the limited size of the available data base.

In this work, observable water quality conditions controlling and/or indicating a favorable environment for algae growth are utilized to predict blooms of specific algal groups using a 3 layer neural network (NN) model. Each node of the input layer represents one of the physical or limnological variables at a daily time step and each output layer node represents the magnitude (likelihood) of the biomass of an algal group for the following day.

2 Artificial neural network model

Neural networks are mathematical models of theorized mind and brain activity which attempt to exploit the massively parallel local processing and distributed storage properties believed to exist in the brain. A description of NN concepts is given by French et al. [11] and briefly summarized here following that work. A classical comparison of the information processing capabilities of the human and the computer is highlighted by the attempt to mechanize human information processing. The computer can multiply large numbers at high speed yet it cannot recognize unconstrained, speaker independent speech. Human abilities complement those of the computer in that speech is easily recognized, even when slurred and in a noisy environment, yet most persons cannot determine the square root of a prime number without pencil and paper, or a computer. The root of these differences lies in the method each uses to process information. A conventional computer uses algorithm-based programs that operate serially, are controlled by a central processing unit, and store information at addressed locations in memory. On the other hand, the brain operates with highly distributed transformations that operate in parallel, have distributed control through thousands of interconnected neurons, and appear to store information as distributed correlations among connections.

The primary characteristics of the neural network are the distributed representation of information, the local operations, and nonlinear processing. These attributes emphasize the popular application areas of NNs: (1) situations where only a few decisions are required from a massive amount of data, such as a classification problem; (2) applications involving large combinatorial optimization; or (3) situations where a complex nonlinear mapping must be learned. Advantageous characteristics of neural network modeling, design or problem solving include the following: (1) problems may be addressed which are either poorly defined or misunderstood, and observations of the process may be difficult or impossible to perform; application of a neural network does not require apriori knowledge of the underlying process; (2) one may not recognize all of the existing complex relationships between aspects of the process under investigation; through a training procedure the neural network incorporates the role of all necessary relationships controlling the process; (3) a standard optimization approach or statistical model provides a solution only when allowed to run to completion; the neural network always converges to an optimal (or suboptimal) solution and need not run to any prespecified solution condition; (4) neither constraints nor an apriori solution structure is necessarily assumed or strictly enforced in the neural network development.

2.1 Neural network structure

The neural network was developed using a three layer structure, consisting of the input layer, hidden layer and output layer shown in Figure 1. Each layer is made up of several nodes, and layers are interconnected by sets of correlation weights. The nodes receive input from either outside the model (the initial inputs) or from interconnections between layers. The weights function to multiply an incoming firing rate prior to its arrival at the next layer. Nodes operate on the input transforming it through a sigmoid function to produce an analogue output called a firing rate.

The structure of the NN and type of neurons used are specific to this particular approach and not proposed as the only or best solution; other types of NN structures and neuron forms exist which might also be feasible. No well defined algorithm exists for determining the optimal number of hidden nodes, although some guidelines are proposed by Mirchandani and Cao [12]. The multitude of possibilities requires that this discussion be limited to the particular NN used and the reader is referred to Simpson [13] for a more complete discussion of the various neural network forms.

2.2 Neural network training

A learning process, or training, forms the interconnections (correlations) between neurons and is accomplished using known inputs and outputs, by presenting these to the NN in an ordered manner. The strength of the interconnections is adjusted using an error convergence technique (back-propagation), so that a desired output will be produced from a given input. Once formed, the interconnections may remain fixed and the NN is used to carry out the intended task.

The neural network is initialized by assigning random numbers from the interval [-1,+1] to the interconnection weights and defining the sigmoid function parameters, β and θ, for both the hidden layer and output layer nodes. A learning rate, η, is selected to control the incremental change in the interconnection weights during iterative training as a percentage of the difference between the desired or target output and the NN computed output. Proper selection of η is important since a value either too high or too low can cause the training process not to converge, instead either an oscillatory, nonoptimal solution is approached or no recognizable solution is developed. The training method used in this work is known as back propagation; it is an iterative technique commonly used for NN learning [Simpson 13; Vemuri 14; Jones and Hoskins 15]. Briefly, the procedure involves computing the error

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

between the known desired output and the computed NN output, then, based on the magnitude of the error, adjusting the interconnection weights in a backward sweep through the NN. A large number of learning sets, consisting of known inputs and outputs, are usually required to perform successful training. The NN solution improves with each training iteration and training continues until either a desired level of accuracy is achieved or some maximum number of iterations are completed.

3 Application/Validation of the Neural Network

The NN model is applied to a five year (1979, 1980, 1981, 1984, 1985) database collected at the Saidenbach Reservoir in Germany [Horn and Horn 16]. The data are reported at a temporal resolution of 7 to 10 days and are interpolated to produce a daily time series for use in the NN model. Input quantities consist of the 9 physical and limnological variables represented in Figure 1 and outputs were the one day ahead algal species biomass for 7 individual types of algae shown in Figure 1. The years 1979, 1981, and 1985 were used for training and the years 1980 and 1984 for validation. There were 1080 (3 years) learning sets, and 720 (2 years) validation sets available from the data base. In order to briefly investigate the role of the number of hidden nodes on NN performance, two NN models were developed using 10 and 30 hidden nodes; results are reported for the 10 hidden node model and results for the 30 hidden node model were similar. The learning rate, η, was set at 0.1 and the number of training iterations used was 100000.

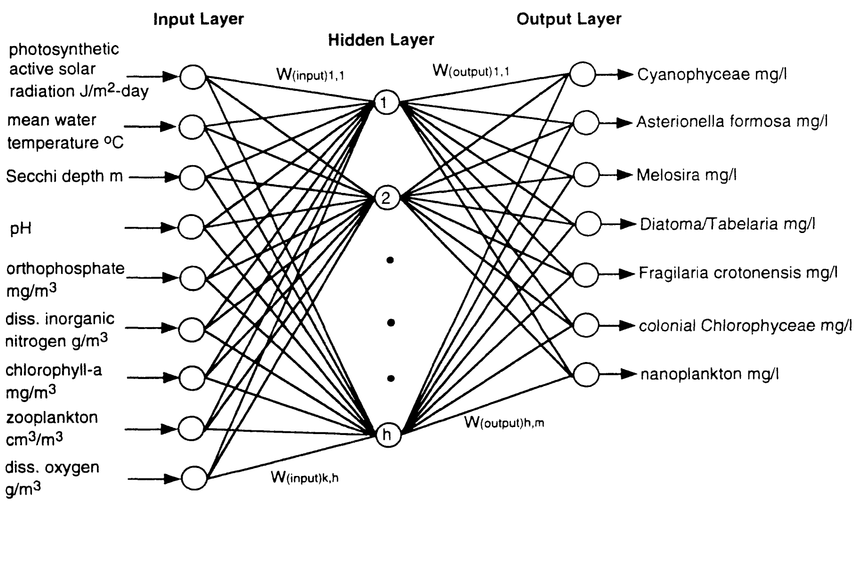

Figure 1: Structure of the neural network

4 Results and closing remarks

Results are provided which illustrate the ability of the NN to predict algal specie dynamics from a timing and magnitude perspective, for two applications: (1) where the input data are those used in training, indicating how well the NN learned the relationship it was trained to recognize; and (2) for input data which are independent from the training process, to illustrate the degree to which the NN can generalize its ability to forecast for events not included in the training process. Model performance on the training data is shown in Figure 2; and Figure 3 shows the performance of the same NN model applied to the independent data. Both figures indicate the observed data by an open circle and the NN predicted value by a solid circle; the horizontal axis shows the daily time step of data used in the study.

Figure 2 shows that the NN model learned the complex interrelationships between the observable water quality and limnological variables and the growth dynamics of the algal species. Additionally, Figures 2 and 3 show that an objective approach was used in selecting the years for training and validation. As a result, some of the particular characteristics of the algal dynamics are not present in the data used for training. This means that while the NN model is capable of learning the variety of conditions represented in the learning sets, not all possible situations were present; the validation results illustrate this point. Figure 3 shows that the NN model in most cases predicted an increase in the growth of the particular algae species at the correct time, however the magnitude of the biomass was typically not correct. The availability of a more extensive data base for training, in which more numerous conditions are represented, could lead to improved model performance.

The utility of the NN approach is useful since once the predictions of the network are validated for a freshwater system using the list of input variables, the input requirements can be optimized while maintaining predictive capability. The procedure would merely require training the NN model using various selected input variables and eliminating those variables that do not affect performance. This allows the identification of key data variables affecting algal growth, and the optimization of data sampling programs and of data collecting costs.

Alternatives to the existing approach include the following: (1) removal of the variable of chlorophyll-a from the input in order to utilize a more completely objective set of inputs to produce a similar set of outputs; and (2) redefinition of training to be either on or off in terms of an algal bloom being likely or unlikely; this would eliminate to some degree the difficulty in forecasting the particular magnitude of algal biomass without a loss of utility of the model results.

In closing, this is a preliminary investigation and the completed study is by no means conclusive as to the utility of a neural network approach to algal bloom forecasting. It provides merely a first step toward understanding and evaluating a role for neural networks in the investigation of complex environmental processes. Many more studies must be performed, using more extensive real data frameworks, before a conclusive statement can be reached. The authors consider this work primarily a first contribution to such a statement.

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

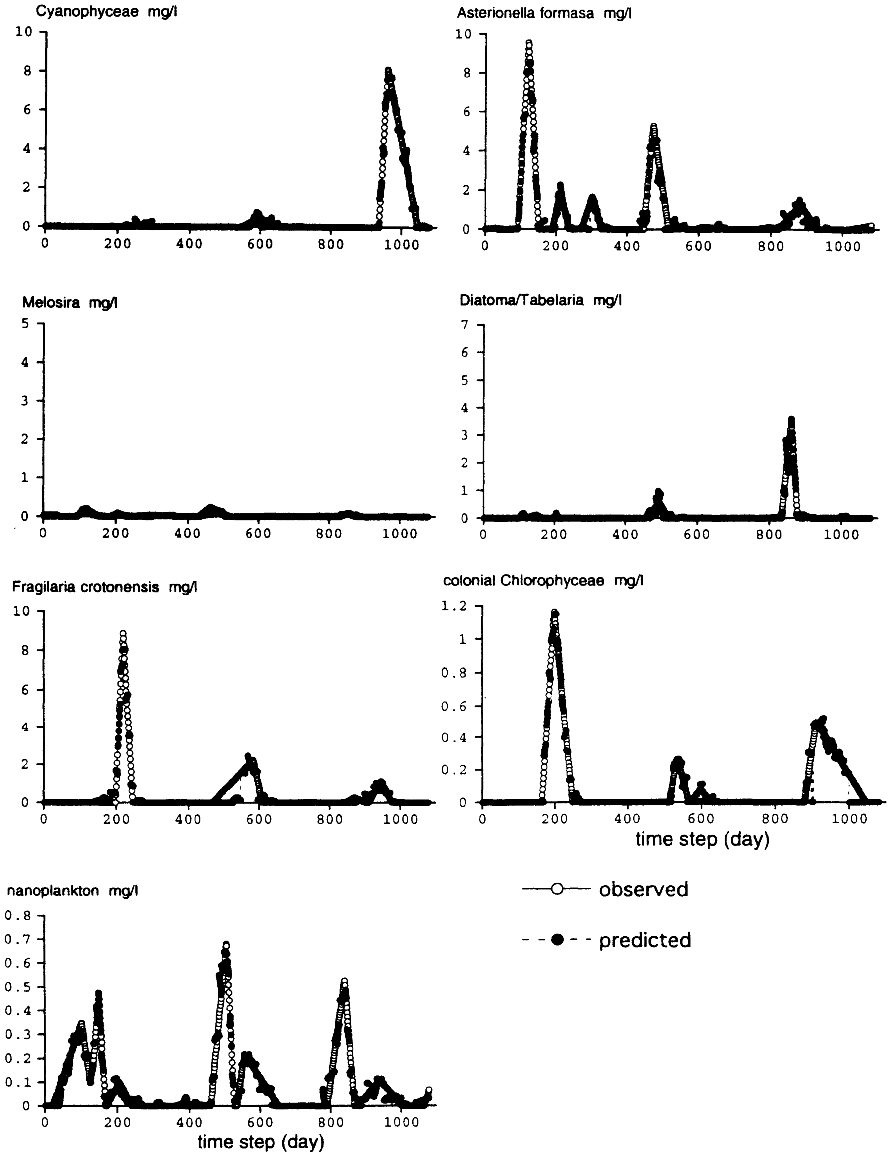

Figure 2: Validation of neural network training using of data from the Saidenbach Reservoir for 1979, 1981 and 1985

Transactions on Ecology and the Environment vol 6, © 1994 WIT Press, www.witpress.com, ISSN 1743-3541

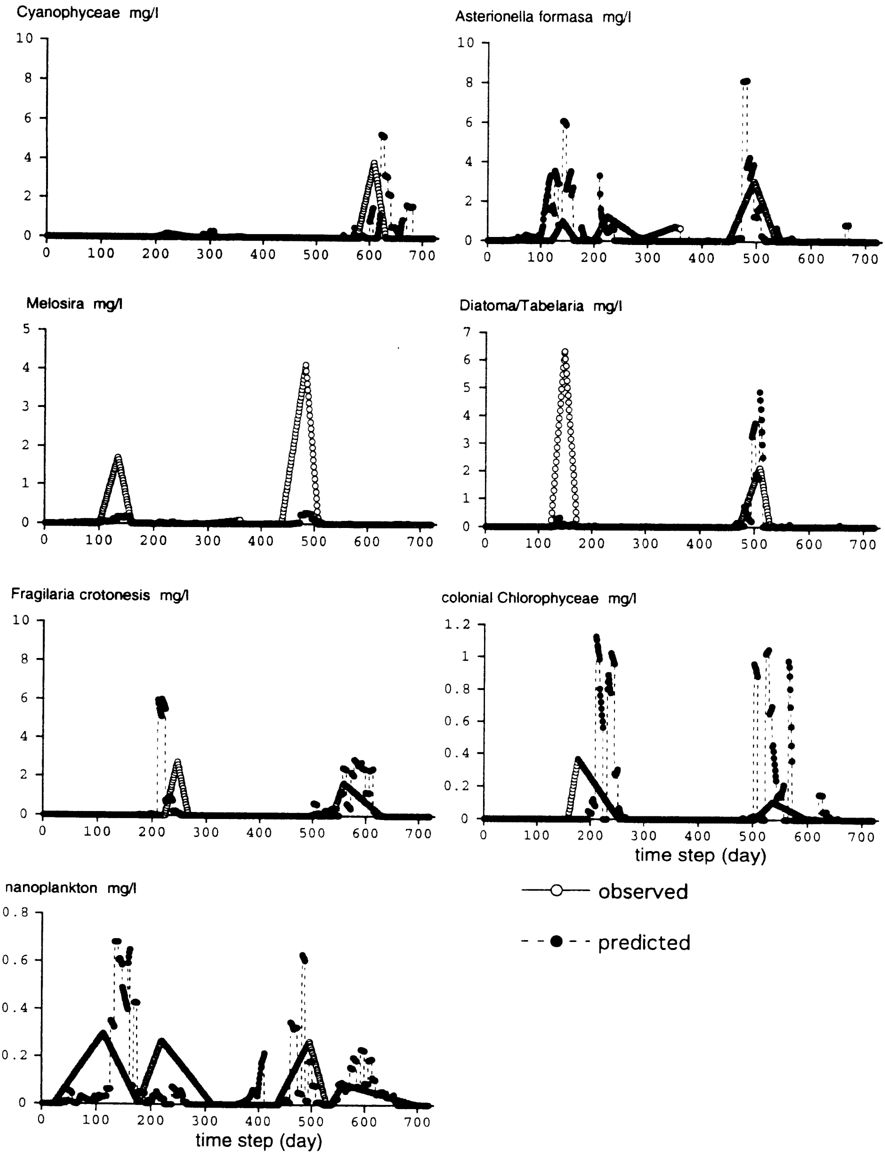

Figure 3: Validation of predictions using the trained neural network and input data from the Saidenbach Reservoir for 1980 and 1984

5 References

1. Vollennweider, R.A. and Kerekes, J. OECD cooperative programme for monitoring of inland waters. Eutrophication control. Synthesis Report. Paris 1980.
2. Jorgensen, S.E. A eutrophication model for a lake. *Ecological Modelling*, 2(1976), 147-162.
3. Straskraba, M. und Dvorakova, M. Probleme der Simulation von Talsperrenoekosystemen. *Acta hydrochim. hydrobiol.* 5(1977), 211-230.
4. Park, R.A., Collins, D.E., Leung, D.K., Boylen, C.W., Albanese, G., Caprariis, P. de, and Forstner, M. The aquatic ecosystem model MSCLEANER. In: Jorgensen, S.E.(ed.): *State-of-the-Art in ecological modelling. Vol. 2. ISEM*, Copenhagen 1979, 579-602.
5. Benndorf, J. and Recknagel, F. Problems of application of the ecological model SALMO to lakes and reservoirs having various trophic states. *Ecological Modelling*, 17(1982), 129-145.
6. Recknagel, F. and Benndorf, J. Validation of the ecological simulation model SALMO. *Int. Revue ges. Hydrobiol.* 67(1982)1, 113-125.
7. Reynolds, C.S. *The Ecology of Freshwater Phytoplankton.* Cambridge University Press, 1984.
8. Sommer, U., Gliwicz, Z.M., Lampert, W., and Duncan, A. The PEG-model of seasonal succession of planctonic events in fresh waters. *Arch. Hydrobiol.*, 106(1986), 433-471.
9. Recknagel, F., Petzoldt, T., Jaeke, O. and Krusche, F. Hybrid expert system DELAQUA - a toolkit for water quality control of lakes and reservoirs. *Ecological Modelling* 71(1994), 17-36.
10. Ruck, B.M., Walley, W.J., Reynoldson, T.B. and Day, K.E. A neural network predictor of benthic community structure in the Canadian waters of the Laurentian Great Lakes. *Water Pollution II, Modeling, Measuring and Prediction*, Wrobel and Brebbia, eds., pp. 287-294, Computational Mechanics Publications, Southampton, UK, 1993.
11. French, M.N., Krajewski, W.F., and Cuykendall, R.R. Rainfall forecasting in space and time using a neural network. *J. Hydrology*, 137(1992), 1-31.
12. Mirchandani, M.G. and Cao, W.C. On Hidden Nodes for Neural Nets. *IEEE Trans. on Circuits and Systems*, May 1989, 36, 661-664.
13. Simpson, P.K. *Artificial Neural Systems, Foundations, Paradigms, Applications, and Implementations*. Pergamon Press, Inc. Elmsford, New York, 210 pp., 1990.
14. Vemuri, V. (ed). *Artificial Neural Networks: Theoretical Concepts.* Computer Society Press Technology Series, Computer Society of IEEE, Washington, D.C., 145 pp., 1988.
15. Jones, W.P. and Hoskins, J. Back-Propagation, A generalized delta learning rule. BYTE, October, 1987, 155-162.
16. Horn, W. and Horn, H. Long-term relationships between phyto- and zooplankton in the meso-eutrophic reservoir Saidenbach. *Arch. Hydrobiol. Beih.*, 33(1990), 749-762.

6 Acknowledgments

The authors thank Heidemarie and Wolfgang Horn of the Hydrobiological Laboratory Neunzehnhain for providing data of the Saidenbach Reservoir.