

Current Scope, Case Studies and Future Directions of Ecological Informatics

F. Recknagel*

Ecology and Evolutionary Biology, University of Adelaide, Adelaide, 5005, Australia

Received 7 January 2012; revised 27 January 2013; accepted 17 March 2013; published online 31 March 2013

ABSTRACT. Ecological data management, analysis and synthesis as well as information processing and transfer in ecology are considered feature areas of ecological informatics. The paper presents case studies for analysis and synthesis of ecological data by means of unsupervised artificial neural networks and hybrid evolutionary algorithms to assist in computational bioindication of wetland water quality and early warning of cyanobacteria blooms in lakes and reservoirs. Integrated analysis and synthesis of ecological and genomics data, hybrid model libraries generic for ecosystem categories as well as internet-based data and model sharing are identified as key research areas of ecological informatics.

Keywords: ecological informatics, unsupervised artificial neural networks, evolutionary algorithms, sensitivity analysis, computational bioindication, benthic diatoms, macroinvertebrates, early warning, cyanobacteria blooms

1. Introduction

Ecological informatics (ecoinformatics) is an interdisciplinary framework for the understanding of information processing and transfer in ecology (Hogeweg, 2007) as well as management, analysis and synthesis of ecological data by advanced computational technology (Recknagel, 2006).

Ecological informatics currently undergoes the process of consolidation as a discipline. It corresponds and partially overlaps with the well-established disciplines bioinformatics and ecological modeling but is taking its distinct shape and scope. A comparison between ecological informatics and bioinformatics in Recknagel (2008) concluded that even though both are based on the same computational technology their focus is different. Bioinformatics focuses very much on determining gene function and interaction, protein structure and function as well as phenotypes of organisms utilizing DNA microarray, genomic, physiological and metabolic data. By contrast ecological informatics focuses to determine genotypes of populations by utilizing genomic, phenotypic and environmental data as well as to determine structure and functioning of ecosystems by utilizing community, environmental and climate data.

Comparing between ecological modeling and ecological informatics Recknagel (2008) stated that even though both rely on similar ecological data they adopt different approaches in utilizing the data. Whilst ecological modeling typically processes ecological data top down by ad hoc designed statistical or

mathematical concepts, ecological informatics infers ecological relationships from ecological data patterns bottom up by computational techniques. The cross-sectional area between ecological modeling and ecological informatics reflects a new generation of hybrid models that may gain the potential to predict emergent ecosystem structures and behaviors, as well as ecosystem evolution (e.g. Recknagel, 2003; Grimm et al., 2005). Typically hybrid models embody biologically-inspired computation in deterministic ecological models.

1.1. Ecological Data Management, Analysis and Synthesis

The Figure 1 describes the general flow of ecological information originating from different sources, undergoing different stages of integration and informing decisions in on- or off-line mode. The three sources of information are typically ecological data, heuristic and causal ecological knowledge that are integrated during the data analysis and synthesis. Before ecological data enter data analysis and synthesis they undergo data acquisition (e.g. Foody, 2007; Chang, 2010) and processing (e.g. Jones and Gries, 2010). Data visualisation by graphics, maps and animations becomes increasingly important in order to communicate ecological information with a broader audience of decision makers, stakeholders, citizens in an easy manner that is also understandable to non-experts. The information integrated by data analysis and synthesis supports either on-line or off-line decisions. On-line decisions require on-line monitoring by in situ sensors and/or remote sensing and allow early warning. Off-line decisions are typically supported by data warehouse queries, scenario and sensitivity analysis by computer simulation and forecasting.

It becomes obvious from Figure 1 that ecological data management, analysis and synthesis relies heavily on well-coordi-

* Corresponding author. Tel.: + 61 8 83133953; fax: + 61 8 83136222.
E-mail address: friedrich.recknagel@adelaide.edu.au (F. Recknagel).

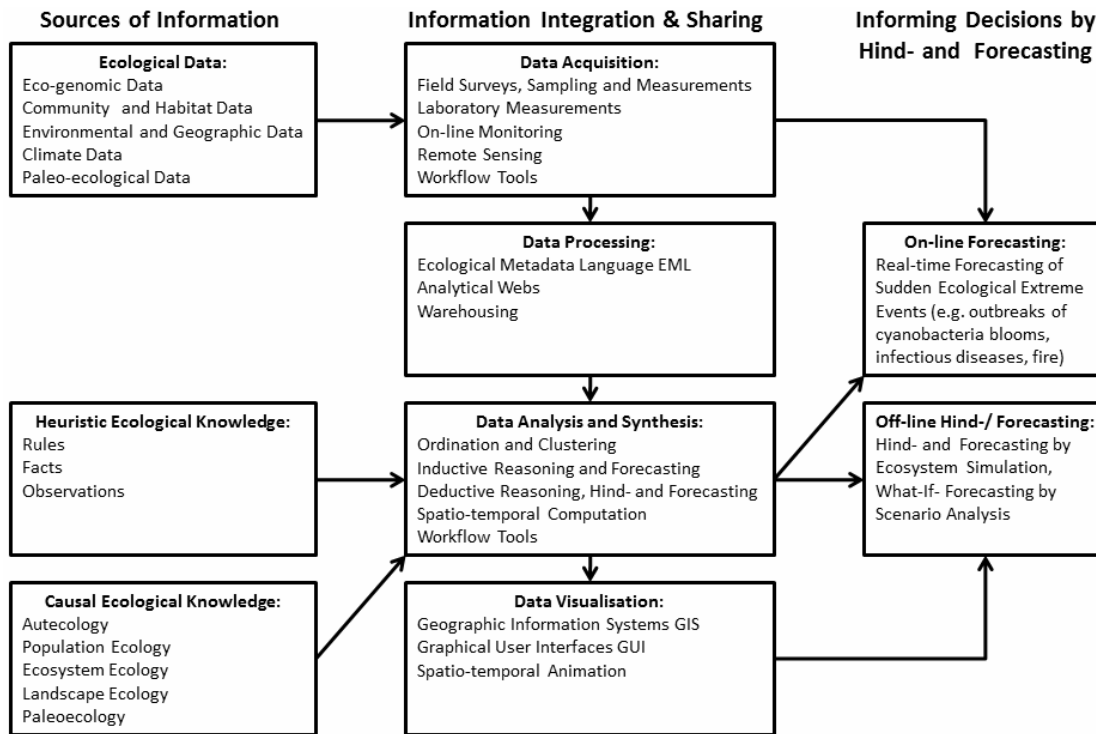


Figure 1. Flow of ecological information from the sources to the end-users.

nated inter-disciplinary research (Chon and Park, 2006) basically between natural scientists and computer scientists. Only the combination of highly reliable ecological data, competent ecological knowledge and novel computation and sensor technology will result in informed decisions to be trusted.

1.2. Information Processing and Transfer in Ecology

The Figure 2 has been designed to highlight the importance of information processing and transfer for determining ecosystem behaviour in response to environmental, habitat and climate changes. It indicates that inter- and intra-specific communication of organisms e.g. by info-chemicals (e.g. Voss et al., 2006; van Donk, 2007) and genomes is decisively controlling extent and direction of food web dynamics and nutrient cycles by taking into account the structure of the food web, nutrient stoichiometry, habitat and climate conditions. Insights into the structure and functioning of information carriers in ecosystems such as info-chemicals, genomes and paleo-ecological data will improve our understanding how ecosystems evolve through consecutive stable states.

Inter-disciplinary research is required between ecologists, molecular biologists, genomists, chemists and information scientists to improve our knowledge on ecological information processing and transfer.

2. Case Studies for Analysis and Synthesis of Ecological Data Exemplary for Ecoinformatics

Novel sensor, information and computing technology is generating an avalanche of physical, chemical, biological and

molecular data in ecology which are highly complex, diverse, fragmented and only constraint accessible. Ecoinformatics provides tailored tools such as artificial neural networks and evolutionary algorithms to integrate, analyse and synthesise ecological data allowing ecologists to unravel complexity, reveal patterns and generate broad generalities.

Unsupervised artificial neural networks (ANN) are designed to identify unknown data patterns based on similarities between the data variables. So-called self-organising maps (SOM) (Kohonen, 1989) are the most popular unsupervised ANN, which can be applied for ordination and clustering of complex non-linear data.

The principal approach of non-supervised ANN according to Kohonen (1989) is represented in a simplified diagram in Figure 3. It shows that the neurons of the ANN learn to distinguish between similar and dissimilar features of the normalised input data, which are mapped as clusters. Those features can be expressed by Euclidean distances, which are calculated between the inputs and weights.

Evolutionary algorithms (EA) 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 Figure 4). Holland (1975) provided the theoretical framework for the development of genetic and evolutionary algorithms that are being widely used for pattern recognition, forecasting, knowledge discovery, optimum control and parallel processing. EA have been successfully implemented as tools for solving ecological problems, which exhibit highest complexity. They allow to induce predictive models from ecological data sets by assembling multiva-

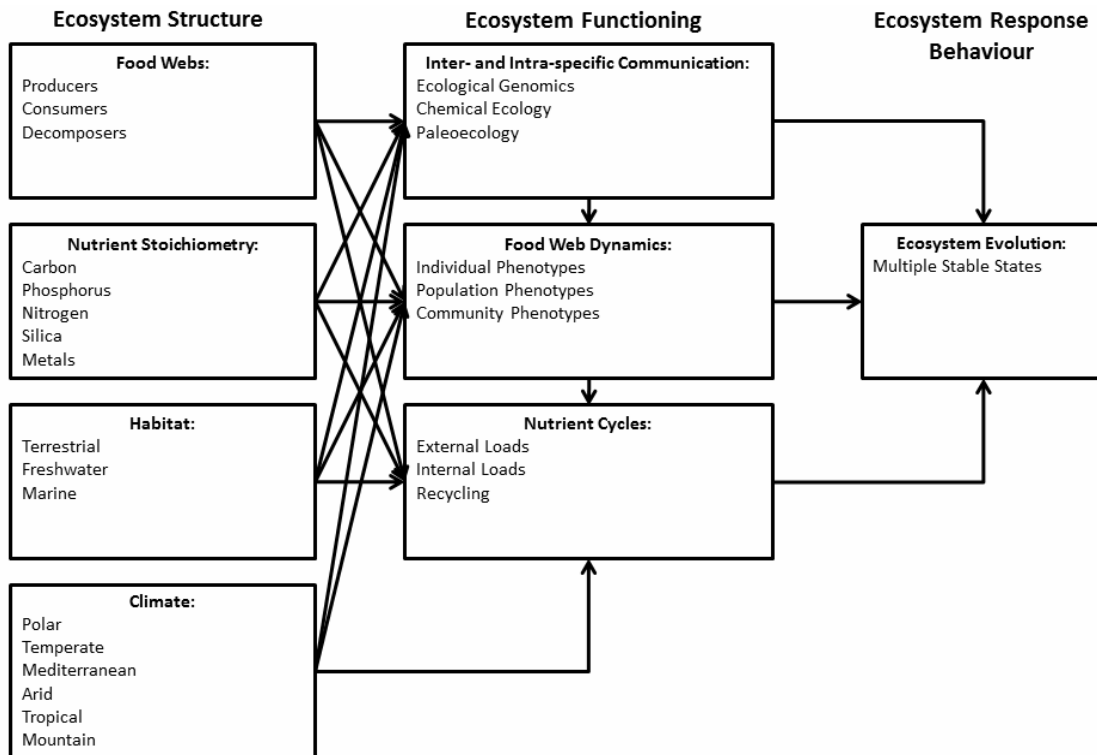


Figure 2. The role of inter- and intra-specific communication in determining ecosystem behaviour.

riate functions or rule sets which enable comprehensive sensitivity analyses as tools for knowledge discovery (e.g. Recknagel et al. 2013).

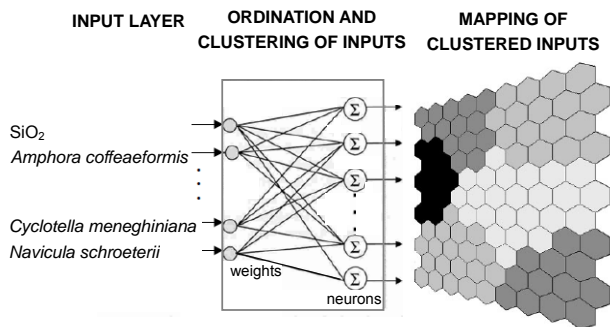


Figure 3. Conceptual diagram of the structure and functioning of non-supervised ANN.

2.1. Ordination, Clustering and Computational Bioindication of wetland habitats

A cross-sectional study of 47 floodplain wetlands along the River Murray in South Australia has been conducted by monitoring water quality conditions and surveying benthic diatoms and macroinvertebrates in April and May 2006. The floodplain conditions ranged from fully and partially inundated wetlands to paddles remaining after summer draught. As a result 15 species of diatoms as well as 15 taxa of macroinvertebrates were identified and quantified both occurring at distinctly different abundance across the 47 wetlands. The water quality

conditions varied significantly as well (see Figure 5).

Unsupervised ANN and EA were applied to identify relationships between the observed cross-sectional physical, chemical and biological data that would allow computational bioindication of wetland habitat conditions.

Figure 6a shows a rule set that has been synthesised by the hybrid evolutionary algorithm HEA (Cao et al., 2013) for the relationships between 4 species of diatoms, the mollusc Planorbidae and SiO₂ concentrations within the 47 floodplain wetlands. The IF-THEN rule (Figure 6a) appeared to be the fittest with a $r^2 = 0.94$ (see Figure 7) amongst 100 rules discovered by boot-strap training with randomly picked 25% testing data and 75% training data. The rule distinguishes between SiO₂ concentrations lower than 4 mg/l reflected by the sensitivity analysis for the THEN branch (Figure 6b) and concentrations higher than 4mg/l reflected by the sensitivity analysis for the ELSE branch (Figure 6c).The sensitivity analysis in Figures. 6b and 6c nicely reveals that in particular Cyclotella meneghiniana and Planorbidae show a preference for low SiO₂ concentrations whilst Staurosirella pinnata occurs abundantly at concentrations greater than 8 mg/l.

The evolutionary algorithm HEA has obviously selected suitable indicator species for the range of SiO₂ concentrations observed in the 47 floodplain wetlands under autumn conditions. In order to back up these findings an ordination and clustering of the data has been conducted by unsupervised ANN using the MATLAB Toolbox SOM (Vesanto et al., 2000). Figure 8 shows patterns of diatoms and Planorbidae abundances for three ranges of SiO₂ concentrations. The patterns clearly

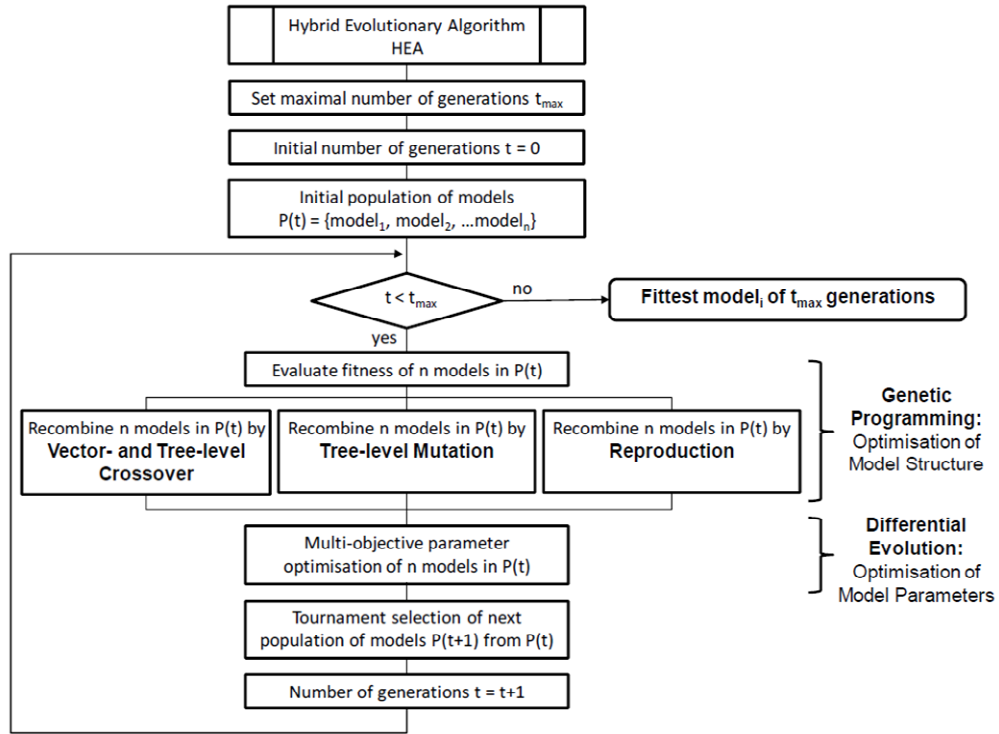


Figure 4. Conceptual diagram of the structure and functioning of the hybrid evolutionary algorithms HEA (Recknagel et al. 2013).

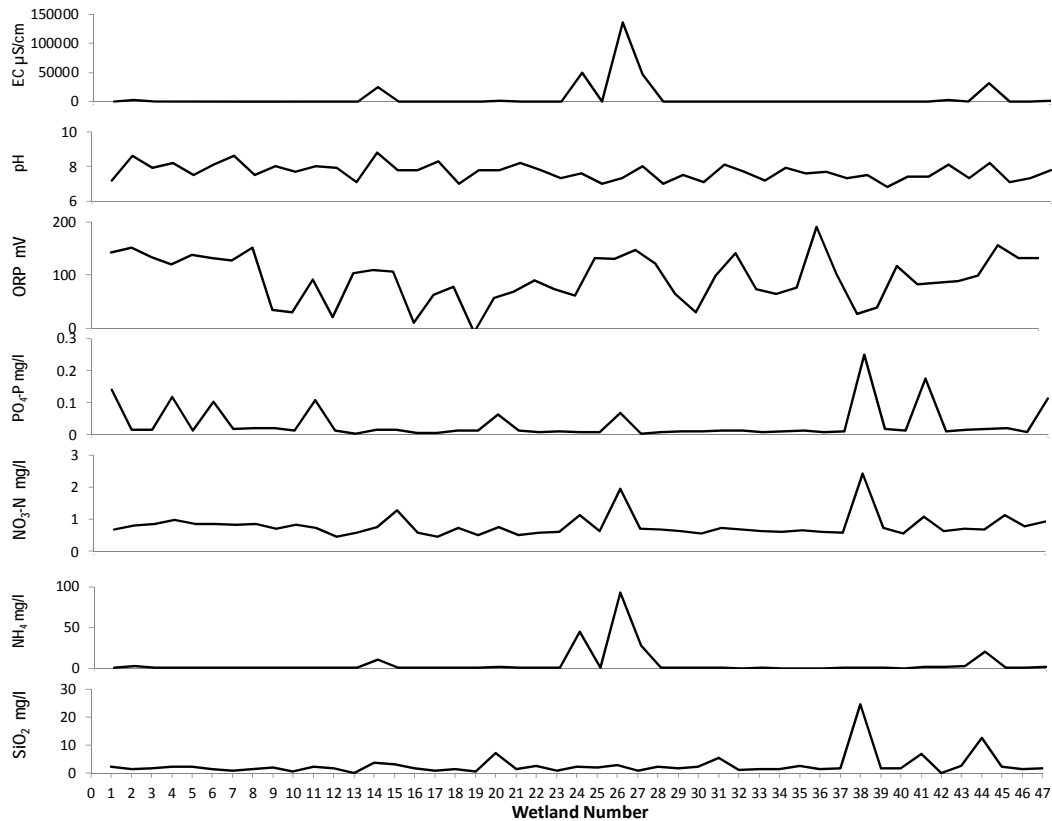


Figure 5. Range of water quality conditions across the 47 River Murray floodplain wetlands observed from April to May in 2006.

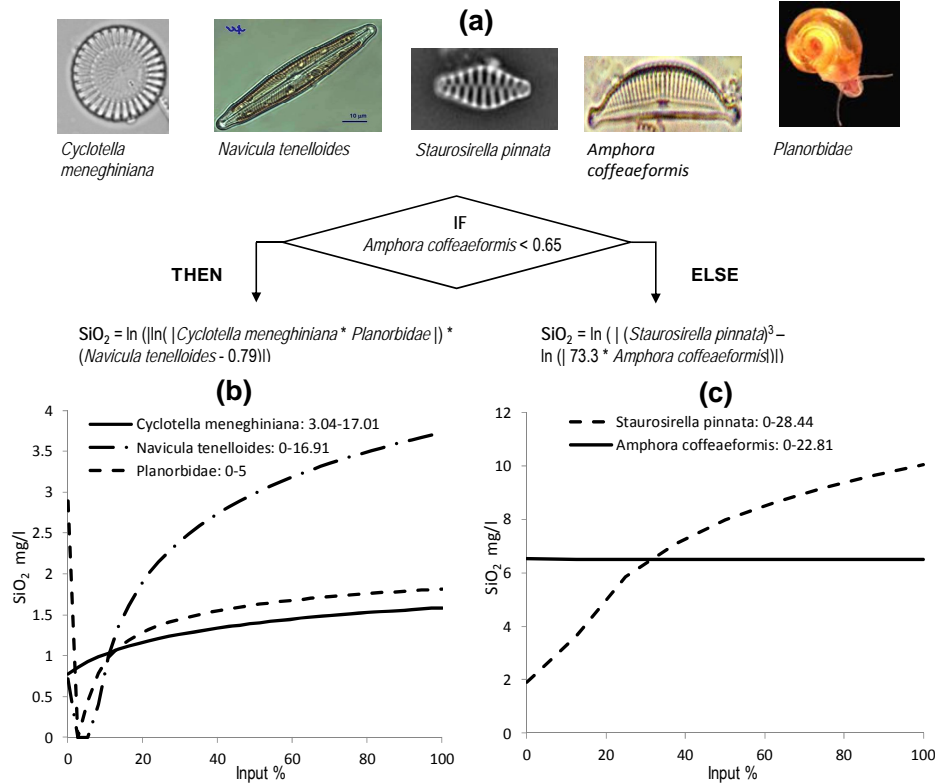


Figure 6. Predictive rule for indicating SiO₂ concentrations in 47 floodplain wetlands along the River Murray in South Australia by means of benthic diatoms and macroinvertebrates. a) rule structure, b) sensitivity analysis for the THEN branch of the rule, c) sensitivity analysis for the ELSE branch of the rule.

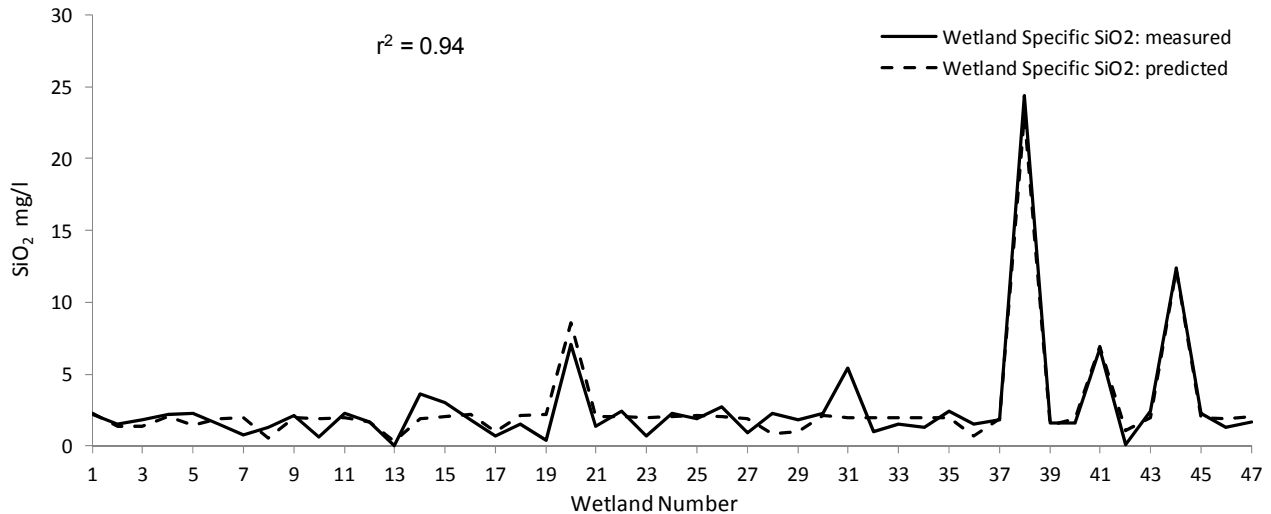


Figure 7. Validation of the predictive rule in Figure 6a for indicating SiO₂ concentrations in 47 floodplain wetlands along the River Murray in South Australia by means of benthic diatoms and macroinvertebrates.

support findings from the sensitivity analysis that *Cyclotella meneghiniana* and *Planorbidae* are potential indicators for low SiO₂ concentrations in these wetlands. Figure 8 also shows that the occurrence of *Stausirella pinnata* coincides with high and low SiO₂ concentrations and therefore appears not to be a ‘sharp’ indicator for high concentrations as concluded from the sensitivity analysis. By contrast it can be concluded from Fi-

gure 8 that *Amphora coffeaeformis* appears to be a strong indicator for SiO₂ at concentrations higher than 3 mg/l.

The validation result in Figure 7 suggests that the rule in Figure 6a can be used as predictive tool for identifying SiO₂ concentrations in these floodplain wetlands by using information on the abundances of the diatoms and *Planorbidae* represented in the rule. Predictive rules for the remaining 6 water

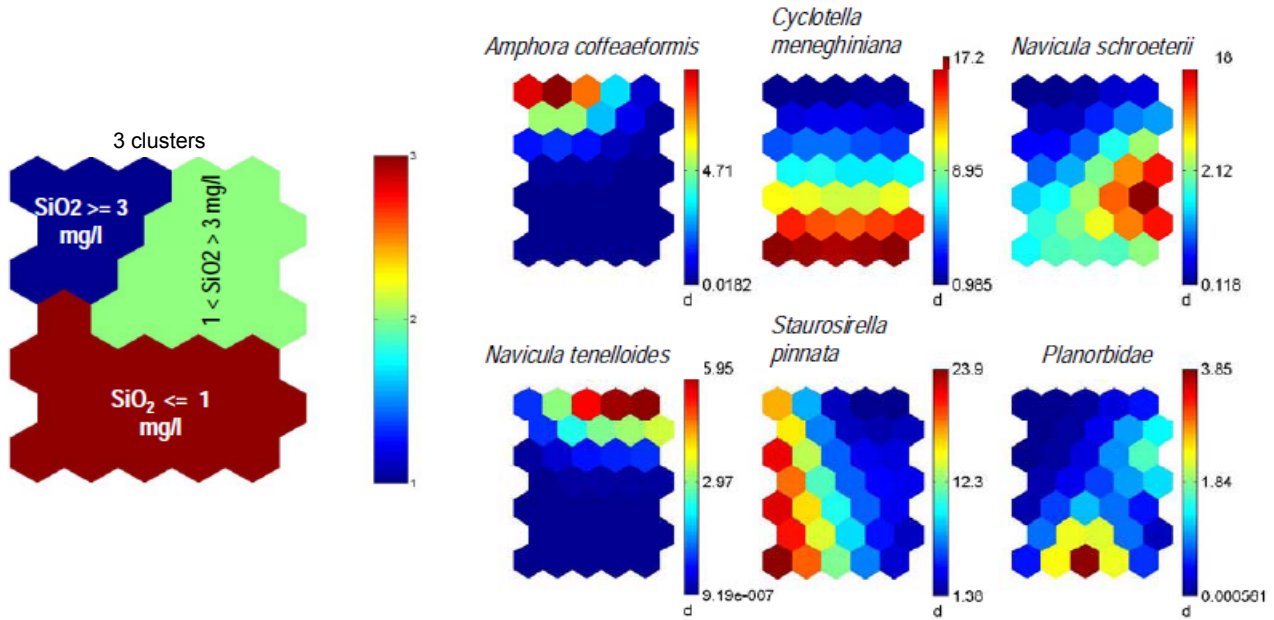


Figure 8. Ordination and clustering of relationships between ranges of SiO₂ concentrations and five diatom species as well as the snail Planorbidae in 47 floodplain wetlands along the River Murray in South Australia by means of unsupervised ANN.

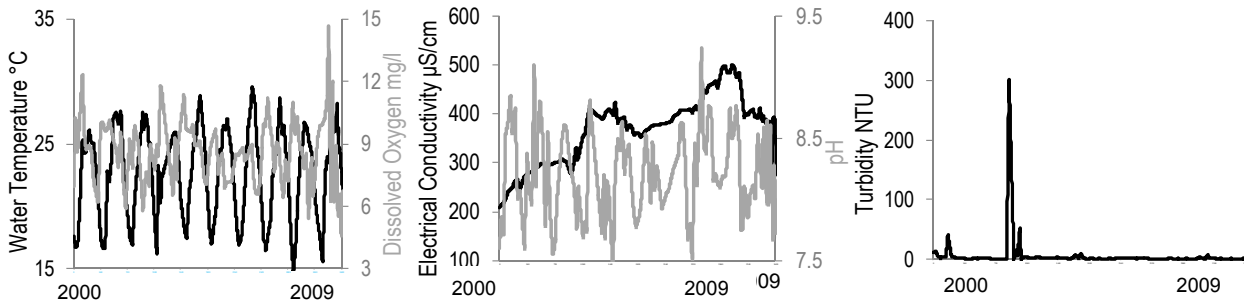


Figure 9. Time-series plots of water temperature, dissolved oxygen, electrical conductivity, pH and turbidity weekly measured from 2000 to 2009 in the Wivenhoe reservoir, Queensland.

quality variables shown in Figure 5 have also been developed solely driven by diatom and macroinvertebrate data with similar good accuracy.

2.2. Ordination, Clustering and Forecasting Cyanobacteria Bloom Data in Drinking Water Reservoirs

The tropical cyanobacteria *Cylindrospermopsis* is a major concern for water authorities in Queensland as it tends to recurrently bloom in drinking water reservoirs contaminating the raw water with high concentrations of the toxic cylindromopsin. Ten years of water quality time series of the Wivenhoe reservoir (Dunabin et al., 2009) provided by South-East Queensland Water have been analysed and synthesised by means of the hybrid evolutionary algorithms HEA and unsupervised ANN in order to develop an early warning system for outbreaks of *Cylindrospermopsis* in the Wivenhoe reservoir. Since the early warning system will be implemented for on-line decision support, only electronically measurable water quality variables were selected and processed as inputs for forecasting

models. The Figure 9 shows the daily-interpolated time-series plots of water temperature, dissolved oxygen, electrical conductivity, pH and turbidity weekly measured from 2000 to 2009 in the Wivenhoe reservoir. These variables are continuously monitored by in situ data loggers YSI 6820 that in future can be fed on-line into forecasting models for *Cylindrospermopsis*.

Figure 10a shows the IF-THEN rule for forecasting *Cylindrospermopsis* with a $r^2 = 0.64$ (see Figure 11) that appeared to be the fittest amongst 100 rules discovered during boot-strap training with randomly picked 25% testing data and 75% training data by the hybrid evolutionary algorithm HEA (Cao et al., 2013). The water temperature threshold of 25.5 °C of the rule distinguishes between *Cylindrospermopsis* abundances lower than 10,000 cells/ml reflected by the sensitivity analysis for the ELSE branch (Figure 10c) and abundances higher than 10,000 cells/ml reflected by the sensitivity analysis for the THEN branch (Figure 10b). The sensitivity analysis in Figures 10b and c reveals that lower alkaline values of pH coincide with high *Cylindrospermopsis* abundances and vice versa. By

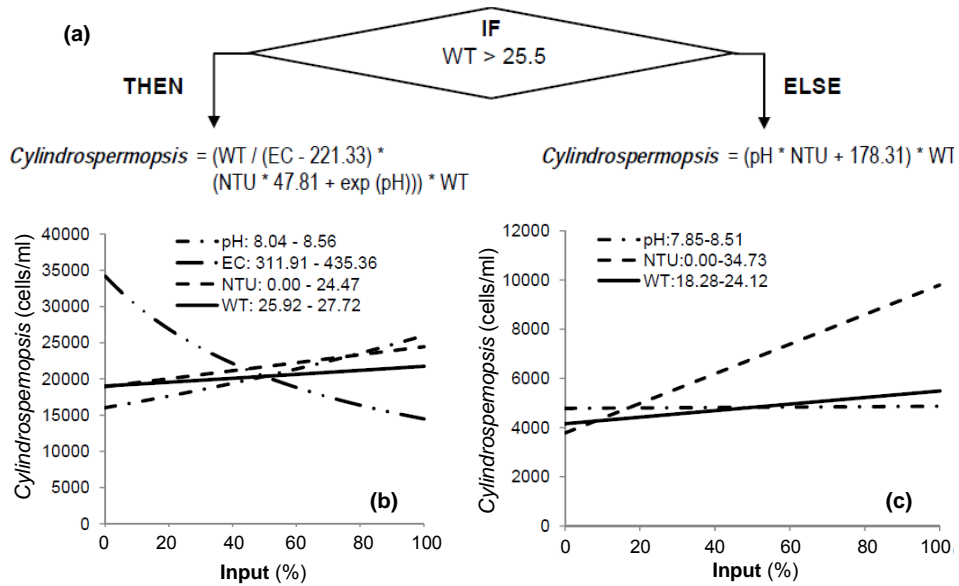


Figure 10. Predictive rule for forecasting 7-day ahead cell concentrations of *Cylandrospermopsis* in relationship with water temperature WT, pH, electrical conductivity (EC) and turbidity (NTU) in the Wivenhoe reservoir, Queensland. a) rule structure, b) sensitivity analysis for the THEN branch of the rule, c) sensitivity analysis for the ELSE branch of the rule.

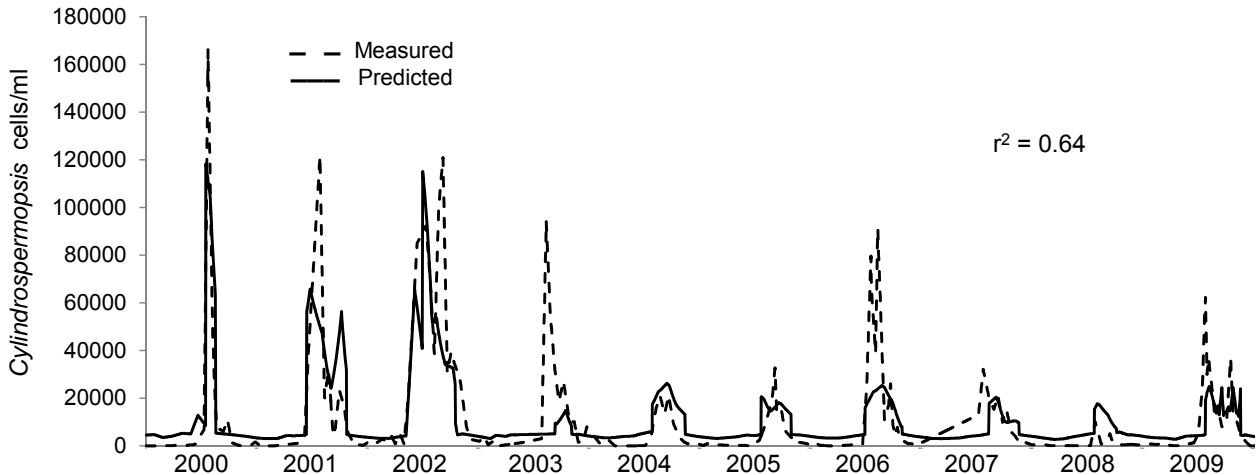


Figure 11. Cross-validation for 7-days ahead forecasting of *Cylandrospermopsis* in the Wivenhoe reservoir, Queensland.

contrast rising trends of conductivity, turbidity and temperature conform to growing trends of *Cylandrospermopsis* abundances.

The validation result in Figure 11 suggests that the simple rule in Figure 10a is performing the 7-days ahead forecasting quite accurately for the timing of fastest growth of *Cylandrospermopsis* in most of the testing years from 2000 to 2009. However it also shows that the magnitudes of *Cylandrospermopsis* abundance are underestimated for all years even though multiple peak events in 2001, 2002 and 2009 have been forecasted as well. Nevertheless these results are encouraging as they are solely based on electronically measurable predictor variables and will allow the on-line implementation of the model for early warning. Since in situ monitoring of phosphate, nitrate and silica is not yet supported by the YSI data logger the current forecasting results are not determined by nutrient concentrations.

In order to get an understanding of causal interrelationships between the chosen input variables and *Cylandrospermopsis* dynamics an ordination and clustering of the time series data has been conducted by unsupervised ANN using the MATLAB Toolbox SOM (Vesanto et al., 2000). The results in Figure 12 show that high oxygen concentrations coincide with either low cell numbers of *Cylandrospermopsis* indicating winter season at low water temperatures or with highest cell numbers suggesting photosynthesis at its peak. Low oxygen concentrations seem to be typical for medium high cell numbers that may indicate periods in the aftermath of bloom events with high microbial respiration by dying algal cells. The alkaline pH levels are highest at periods of highest cell numbers with photosynthesis at its peak and lowest in periods with low and medium high cell numbers. The pattern of electrical conductivity is neither distinctly related to abundances of *Cylandrospermopsis*.

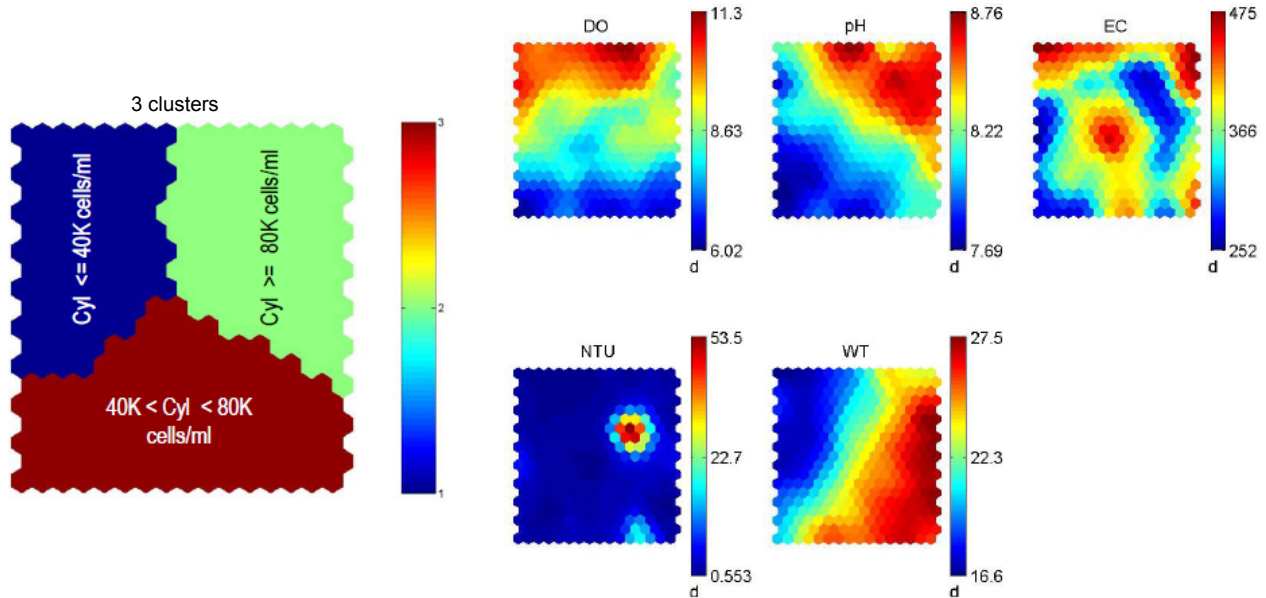


Figure 12. Ordination and clustering of relationships between 3 ranges of cell numbers of *Cylindrospermopsis* (Cyl) and water temperature (WT), pH, electrical conductivity (EC) and turbidity (NTU) in the Wivenhoe reservoir, Queensland by means of unsupervised ANN.

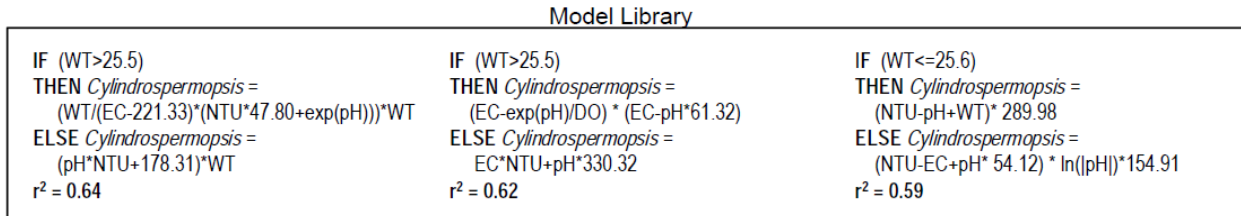


Figure 13. Model library for early warning of *Cylindrospermopsis* blooms in the Wivenhoe reservoir, Queensland.

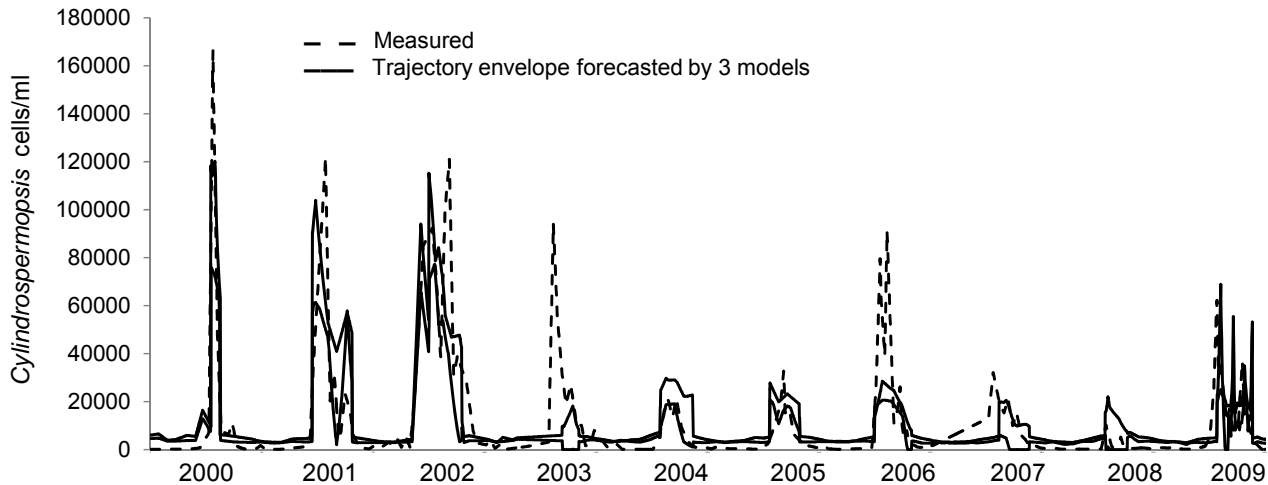


Figure 14. Envelope of 7-days-ahead forecasting of *Cylindrospermopsis* abundances in the Wivenhoe reservoir for 2000 to 2009.

mopsis nor to water temperature but indicates that it is not primarily driven by evaporation under subtropical climates but also by mineral loads through rain events. There seems to be a distinctive relationship between turbidity and *Cylindrosper-*

mopsis indicating that events with peak cell numbers cause highest turbidity levels. The pattern for water temperature backs up the 25.5 °C identified by the evolutionary algorithm HEA as threshold between medium and high cell numbers of *Cy-*

lindropermopsis (Figure 12).

To make early warning systems for cyanobacteria reliable and trustworthy tools for routine use it is desirable that envelopes are forecasted by a set of models rather than just point estimates by one model. The Figure 14 shows an envelope for Cylindropermopsis forecasted by the three models documented in Figure 13.

3. Future Directions of Ecological Informatics

Making informed decisions on the conservation of biodiversity and sustainable environments in spite of pollution and climate change is of vital importance for the habitat earth in the 21st century. Ecological informatics is committed to improve ecological understanding and contribute tools for integrating, analysing and synthesising the wealth of ecological knowledge and data for making informed decision at local, regional and global scale.

Ecological informatics is now focusing its efforts on: (1) integrated analysis of genomic, phenotypic and ecological data in order to better understand biodiversity and ecosystem behaviour in response to habitat and climate changes; (2) facilitating data sharing by www-based generic data warehouses tailored for ecosystem categories at global scale, and (3) implementing hybrid model libraries generic for ecosystem categories at global scale by object-oriented programming and internet access.

References

- Cao, H., Recknagel, F., and Orr, P.T. (2013). Enhanced functionality of the redesigned hybrid evolutionary algorithm HEA demonstrated by predictive modelling of algal growth in the Wivenhoe Reservoir, Queensland (Australia). *Ecol. Model.*, 252, 32-43. <http://dx.doi.org/10.1016/j.ecolmodel.2012.09.009>
- Chang, N.B. (2010). Advances of ecological remote sensing under global change. *Ecol. Inform.*, 5(5), 317-439. <http://dx.doi.org/10.1016/j.ecoinf.2010.07.005>
- Chon, T.S., and Park, Y.S. (2006). Ecological informatics as an advanced interdisciplinary interpretation of ecosystem. *Ecol. Inform.*, 1(3), 213-217. <http://dx.doi.org/10.1016/j.ecoinf.2006.02.007>
- Dunabin, M., Udy, J., Grinham, A., and Bruenig, M. (2009). Continuous monitoring of reservoir water quality: The Wivenhoe project. *Water*, 36(6), 74-77.
- Foody, G.M. (2007). Ecological applications of remote sensing and GIS. *Ecol. Inform.*, 2(2), 71-72. <http://dx.doi.org/10.1016/j.ecoinf.2007.06.001>
- Grimm, V., Revilla, E., Berger, U., Jeltsch, F., Mooij, W.M., Railsback, S.F., Thulke, H.H., Weiner, J., Wiegand, T., and DeAngelis, D.L. (2005). Pattern-oriented modeling of agent based complex systems: Lessons from ecology. *Science*, 310(5750), 987-991. <http://dx.doi.org/10.1126/science.1116681>
- Hogeweg, P. (2007). From population dynamics to ecoinformatics: Ecosystems as multilevel information processing systems. *Ecol. Inform.*, 2(2), 103-111. <http://dx.doi.org/10.1016/j.ecoinf.2007.01.002>
- Holland, J.H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor, MI.
- Jones, M.B., and Gries, C. (2010). Advances in environmental data management. *Ecol. Inform.*, 5(1), 1-2. <http://dx.doi.org/10.1016/j.ecoinf.2010.01.001>
- Kohonen, T. (1989). *Self-organization and associative memory*. Springer-Verlag, Berlin. <http://dx.doi.org/10.1007/978-3-642-88163-3>
- Recknagel, F. (2003). Simulation of food web and species interactions by adaptive agents embodied with evolutionary computation: A conceptual framework. *Ecol. Model.*, 170(2-3), 291-302. [http://dx.doi.org/10.1016/S0304-3800\(03\)00234-5](http://dx.doi.org/10.1016/S0304-3800(03)00234-5)
- Recknagel, F. (ed.) (2006). *Ecological Informatics. Scope, Techniques and Applications, 2nd Edition*, Springer-Verlag, Berlin, Heidelberg, New York. <http://dx.doi.org/10.1007/3-540-28426-5>
- Recknagel, F. (2008). Ecological Informatics: Overview. In: Jorgensen, S.E. and B Faith (Eds.), *Encyclopaedia of Ecology*, Oxford, Elsevier, Vol. 2, pp. 1041-1058, 2008.
- Recknagel, F., Ostrovsky, I., Cao, H., Zohary, T., and Zhang, X., (2013). Ecological relationships, thresholds and time-lags determining phytoplankton community dynamics of Lake Kinneret, Israel elucidated by evolutionary computation and wavelets. *Ecol. Model.* 255, 70-86. <http://dx.doi.org/10.1016/j.ecolmodel.2013.02.006>
- Van Donk, E. (2007). Chemical information transfer in freshwater plankton. *Ecol. Inform.*, 2(2), 112-120. <http://dx.doi.org/10.1016/j.ecoinf.2007.03.002>
- Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). *SOM Toolbox for Matlab 5*, Helsinki University of Technology, Finland.
- Voss, M., Vet, L.E.M., Wäckers, F.L., Middelburg, J.J., van der Putten, W.H., Mooij, W.M., Heip, C.H.R., and van Donk, E. (2006). Infochemicals structure marine, terrestrial and freshwater food webs: Implications for ecological informatics. *Ecol. Inform.*, 1(1), 23-32. <http://dx.doi.org/10.1016/j.ecoinf.2005.06.001>