

Early Warning for Algal Blooms in Drinking Water Reservoirs by Real-Time Forecasting

Amber Welk – PhD Thesis Abstract

Department of Environmental Biology, University of Adelaide

Historical water quality databases from two South Australian drinking water reservoirs were used, in conjunction with various computational modelling methods for the ordination, clustering and forecasting of complex ecological data. Techniques used throughout the study were: Kohonen artificial neural networks (KANN) for data categorisation and the discovery of patterns and relationships, recurrent supervised artificial neural networks (RANN) for knowledge discovery and forecasting of algal dynamics and hybrid evolutionary algorithms (HEA) for rule-set discovery and optimisation for forecasting algal dynamics. These methods were combined to provide an integrated approach to the analysis of algal populations including interactions within the algal community and with other water quality factors, which results in improved understanding and forecasting of algal dynamics.

The project initially focussed on KANN for the patternising and classification of the historical data to reveal links between the physical, chemical and biological components of the reservoirs. This offered some understanding of the system and relationships being considered for the construction of the forecasting models. Specific investigations were performed to examine past conditions and the impacts of different management regimes, as well as to discover sets of conditions that correspond with specific algal functional groups.

RANN was then used to build models for forecasting both Chl-*a* and the main nuisance species, *Anabaena*, up to 7 days in advance. This method also provided sensitivity analyses to demonstrate the relationship between input and output variables by plotting the reaction of the output to variations in the inputs. Initially one year from the data set was selected for the testing of a model, as per the split-sample technique. To further test the models, it was later decided to select several years for testing to ensure the models were useful under changed conditions, and that test results were not misleading regarding the models true capabilities. RANN were firstly used to create reservoir specific or ad-hoc models. Later, the models were trained with the merged data sets of both reservoirs to create one model that could be applied to either reservoir.

Another method of forecasting was trialled and compared to RANN. HEA was found to be equal or superior to RANN in predictive power, also allowed sensitivity analysis and provided an explicit, portable rule set. The HEA rule sets were initially tested on selected years of data, however to fully demonstrate the model's potential, a process for *k*-fold cross-validation was developed to test the rule-set on all years of data. To further extend the applicability of the HEA rule-set; the idea of rule-based agents for

specific lake ecosystem categories was examined. The generality of a rule-based agent means that, after successful validation on several lakes from one category, the agent could then be applied to other water bodies from within that category that had not been involved in the training process. The ultimate test of the rule-based agent for the warm monomictic and eutrophic lake ecosystem category was to be applied to a real-time monitoring and forecasting situation. The agent was fed with online, real-time data from a reservoir that belonged to the same ecosystem category but was not used in the training process. These preliminary experiments showed promising results. It can be concluded that the concept of rule-based agents will facilitate real-time forecasting of algal blooms in drinking water reservoirs provided on-line monitoring of relevant variables has been implemented.

Contributions of this research include: (1) to offer insight into the capabilities of 3 kinds of computational modelling techniques applied to complex water quality data, (2) novel applications of KANN including the division of data into separate management periods for comparison of management efficiency, (3) to both qualitatively and quantitatively elucidate relationships between water quality parameters, (4) research toward the development of a forecasting tool for algal abundance 7 days in advance that could be generic for a particular lake ecosystem category and implemented in real-time, and (5) to suggest a thorough testing method for such models (*k*-fold cross validation).