A Contribution Towards Real-Time Forecasting of Algal Blooms in Drinking Water Reservoirs by means of Artificial Neural Networks and Evolutionary Algorithms

By Amber Lee Welk
Bachelor of Natural Resource Management (Hons.)

A Thesis Submitted for the Degree of
Doctor of Philosophy

Discipline of Ecology and Evolutionary Biology
School of Earth and Environmental Science
The University of Adelaide
December 2007

TABLE OF CONTENTS

ABSTRACT

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 models 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).

STATEMENT OF ORIGINALITY

This work contains no material that has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

Signed:………………………… Date:………………….
Amber Lee Welk

ACKNOWLEDGMENTS

I would like to thank my supervisors Associate Professor Friedrich Recknagel, Associate Professor Holger Maier and Mike Burch. Fred, my principal supervisor, has been invaluable with his enthusiasm and guidance – and his knack of pushing me when I need it!

This work was made possible by the financial support of the CRC for Water Quality and Treatment and I'd like to thank them for furthering my learning and development by allowing me to attend and present at numerous conferences throughout my candidature.

I must acknowledge Greg Ingleton, not just for the provision of the Happy Valley data, but also for patiently responding to all my pesky queries.

A big thank you goes to Dr Hongqing Cao, for developing the HEA that was used during the project.

Thanks to all in the Ecoinformatics and Watershed Ecology Lab for providing a friendly working environment. And special thanks must go to my fellow students Lydia Cetin, Anita Talib, Grace Chan and Bridget McDowell for sharing in my achievements and laughing at the disasters that have occurred throughout the project!

Most importantly, I'd like to express my appreciation for my family and friends. Firstly, thanks go to Dr Grewal for what can only be described politely as 'perseverance' in convincing me to go down this road. My friends have been a great help in ensuring I don't take things too seriously all the time, for their support and distractions I thank them very much. Thanks to the world's greatest grandparents for letting me retreat to Wilmington to hide from my project when it got too much. As in everything I do, my exceptionally patient mum has been my number one supporter, and her pride in me and what I accomplish makes me think I must be doing something right. My mum, my sister Naomi (who offered countless alternatives to work and a vast supply of chocolate) and partner Anthony have borne the brunt of my frustrations throughout the project but have been unwavering in their support and have offered endless encouragement. I hope you all know that you are part of anything and everything I achieve in life.

LIST OF PUBLICATIONS

Conference Presentations

Recknagel F, Welk A, Kim B, Takamura N. Explanation of processes determining phytoplankton abundances and succession in two lakes with different trophic states by means of long-term data patterns and non-supervised Artificial Neural Networks. *4th Conference of the International Society for Ecological Informatics (ISEI4), BEXCO, Busan, Korea, 24-28 October 2004.*

Welk A, Recknagel F, Burch M. Ordination, clustering and forecasting of phytoplankton dynamics in the Myponga drinking water reservoir by means of supervised and non-supervised artificial neural networks. *MODSIM05, International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, Melbourne, Australia, 12-15 December 2005.*

Welk A. Towards A Rule-based Agent for Forecasting Chlorophyll-a Concentrations in South Australian Drinking Water Reservoirs. *Cooperative Research Centre (CRC) for Water Quality and Treatment, 5th Postgraduate Student Conference, Melbourne, Victoria, 10-13 July 2006.*

Welk A, Recknagel F, Cao H, Chan W-S, Talib A. Rule-based agents for forecasting algal population dynamics in freshwater lakes discovered by hybrid evolutionary algorithms. *5th Conference of the International Society for Ecological Informatics (ISEI5), Santa Barbara, USA, 4-6 December 2006.*

Peer Reviewed Papers

Welk A, Recknagel F, Burch M. (2005) Ordination, clustering and forecasting of phytoplankton dynamics in the Myponga drinking water reservoir by means of supervised and non-supervised artificial neural networks. Proceedings of the International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, MODSIM05, Melbourne, Australia, 12-15 December 2005.

Recknagel F. Welk A, Kim B, Takamura N. (2006) Artificial Neural Network Approach to Unravel and Forecast Algal Population Dynamics of Two Lakes Different in Morphometry and Eutrophication. In: Ecological Informatics – Scope, Techniques and Applications (Ed: F. Recknagel) 2nd Edition. Springer-Verlag. Berlin, p325-345.

Recknagel F, Kim B, Welk A. (2006) 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.* 29 p 1497-1502

Recknagel F, Cao H, Kim B, Takamura N, Welk A. (2006) Unravelling and forecasting algal population dynamics in two lakes different in morphometry and eutrophication by neural and evolutionary computation. *Ecological Informatics,* 1 p 133-151.

Welk A, Recknagel F, Cao H, Chan W-S, Talib A. (2008) Rule-based agents for forecasting algal population dynamics in freshwater lakes discovered by hybrid evolutionary algorithms. *Ecological Informatics,* 3 p 46-54

LIST OF FIGURES

LIST OF TABLES