

# 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

TABLE OF CONTENTS	i
ABSTRACT	iv
STATEMENT OF ORIGINALITY	vi
	vı
ACKNOWLEDGMENTS	VII
LIST OF PUBLICATIONS	viii
Conference Presentations Peer Reviewed Papers	viii viii
LIST OF FIGURES	ix
LIST OF TABLES	xiii
1. INTRODUCTION	1
Contributions of the study	3
2. LITERATURE REVIEW	6
2.1 Modelling of ecological systems and processes 2.1.1 Ecological Informatics	6 7
2.2 Artificial Neural Networks (ANNs)	8
2.2.1 Supervised Artificial Neural Networks (SNN)	
2.2.1.1 Supervised Feedforward Artificial Neural Networks (NSNN)	۱۱ ۱۸
2.3 Evolutionary Algorithms (FA)	10
2.3.1 Hybrid Evolutionary Algorithms (HEA)	
2.4 Summary of past modelling efforts at study sites	29
2.4.1 Differential equation based modelling of the Myponga Reservoir	29
2.4.2 Artificial neural network based modelling of the Myponga Reservoir	
2.4.3 Empirical based modelling of the Happy Valley Reservoir	32 22
2.5 Management history of the study sites	33 22
2.5.2 Happy Valley Reservoir	
2.5.3 Key water management issues	
2.5.4 Limitations of current management regimes	38
2.5.5 Potential improvements to management regime	40
<ul><li>2.6 Online, real-time monitoring and forecasting</li><li>2.7 Summary</li></ul>	42 46
3. MATERIALS, MODEL DESIGN AND APPLICATION	
3.1 Myponga reservoir site and data summaries	47
3.1.1 Myponga reservoir site description	

3.1.2 Historical data from Myponga reservoir	48
3.2 Happy Valley reservoir site and data summaries	49
3.2.1 Happy Valley reservoir site description	49
3.2.2 Historical data from Happy Valley reservoir	51
3.3 A comparison of study sites	
3.3.1 Water Quality Time-series Graphs of Myponga and Happy Valley reservoirs	
3.4 Hope Valley reservoir site and data summaries	
3.4.1 Hope valley reservoir sile description	01 21
3.4.2 Real-unite uata more valley reservoir	01
3.5 1 Data Pre-processing	
3.5.2 Recurrent Artificial Neural Network (RANN)	
3.5.3 Hybrid Evolutionary Algorithm (HEA)	
3.5.4 Kohonen Artificial Neural Networks (KANN)	
3.6 Method integration	72
4. ORDINATION AND CLUSTERING OF WATER QUALITY VARIABLES	74
4.1 Introduction	74
4.2 Aims and Hypotheses	
4.3 Methods and Materials	
4.3.1 Data	0/ ۲۲
4.5.2 MOUEL DESIGN	70 20
4.4 Results	78 78
4.4.2 Long-term dynamics and management related patterns	
4.4.3 Habitat preferences established by clusters according to ranges of physical/che	emical
conditions using merged data from both reservoirs	
4.5 Discussion	94
5. FORECASTING OF CHLOROPHYLL-A AND ANABAENA DYNAMICS	97
F 1 Introduction	07
5.1 INTroduction	97
5.2 Aims and Materials	
5.3 Methods and Materials	98
5.3.2 Model Desian	
5.4 Results	105
5.4.1 7-days ahead forecasting of Chl-a	105
5.4.2 7-days ahead forecasting of Anabaena abundance	114
5.4.3 Research progression	121
5.5 Discussion	124
6. RELATIONSHIPS OF CHLOROPHYLL-A AND ANABAENA DYNAMICS WITH PHYSI	CAL
AND CHEMICAL INPUT VARIABLES	128
6.1 Introduction	128
6.2 Aims and Hypotheses	128
6.3 Methods and Materials	129
6.3.1 Data	129

6.3.2 Model Design	129
6.4 Results	
6.4.1 Water temperature and algal abundance	
6.4.2 PO <sub>4</sub> concentrations and algal abundance	
6.5 Discussion	
0.5 Discussion	
7. DEVELOPMENT OF RULE-BASED AGENTS FORECASTING ALGAL DYNA	MICS USING
HEA	
7.1 Introduction	
7.2 Aims and Hypotheses	
7.3 Materials and Methods	
7.3.1 Data	137
7.3.1 Model Design	
7.4 Results	
7.4.1 Rule-based Uni-a agent	
7.5 Discussion	
8. CONCLUSION	156
8.1 Summary of findings	
8.2 Contributions of the study	
8.3 Recommendations	
8.4 The Future	162
ΑΡΡΕΝΠΙΧ Α	164
	1/4
I rophic state classifications	
APPENDIX B	166
The relationship between reinfall and turbidity	1//
The relationship between rainfail and turbidity	
APPENDIX C	
The relationship between colour and dissolved organic carbon (DOC)	
APPENDIX D	169
Chl-a as an input to HEA	
APPENDIX E	170
Most Influencing Parameter graphs	170
DEFEDENCES	4=0
KEFEKENUES	

## 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:..... Amber Lee Welk Date:....

# 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, 5<sup>th</sup> 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) 2<sup>nd</sup> 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

Figure 1. Basic conceptual structure of the two types of supervised artificial neural network (SNN): a) supervised feedforward; b) supervised feedback (from Recknagel and Cao, 2007)
Figure 2. Architecture of a recurrent artificial neural network (RANN) for predicting abundance of blue green algae, using typical inputs13
Figure 3. Results of RANN forecasting of Chl- <i>a</i> concentration 3 days in advance (Jeong et al, 2001)
Figure 4. Results of sensitivity analysis with disturbance of +/- 1 standard deviation to the input data (Jeong et al, 2001)
Figure 5. Results of sensitivity analysis with wide-ranged disturbance of +/- 2 standard deviations to the input data: a) evaporation and irradiance; b) pH and DO (Jeong et al. 2001)
Figure 6 - Kohonen Artficial Neural Network for non-linear cluster analysis of ecological data. (From Chon et al. 1996)
Figure 7. Summary of the process of evolutionary computing (from Morral, 2006)20
Figure 8. Results of Chl- <i>a</i> forecasting using a predictive rule-set obtained using the SASME framework (Bobbin, 2002)
Figure 9. An evolved rule set for predicting algae (Bobbin, 2002)
Figure 10. Flowchart of the hybrid evolutionary algorithm (from Cao et al. 2006)27
Figure 11. 7 day ahead forecasting of <i>Microcystis</i> and <i>Cyclotella</i> in Lake Kasumigaura (left column), and of <i>Anabaena</i> and <i>Asterionella</i> in Lake Soyang (right column) using a predictive rule set obtained by HEA (from Cao et al, 2006)
Figure 12. Sensitivity analysis with disturbance +/- standard deviation of input data for THEN (left) and ELSE (right) branches of predictive rule set for <i>Microcystis</i> (from (Cao et al, 2006)
Figure 13. Generic supervised feedforward artificial neural network for 14-days ahead forecasts of algal abundance (from Wilson, 2004)
Figure 14. Results of chlorophyll fluorescence forecasts, with 1, 2 and 3 days lead time, using a predictive equation (from Muttil and Lee (2005)45
Figure 15 - Locations of Myponga and Happy Valley reservoirs in South Australian (from SA Water Drinking Water Quality Report 2003-2004)

Figure 16. Time-series graph of preprocessed water quality data from Myponga and Happy Valley reservoirs, specifically a) Chl-a b) water temperature c) turbidity d) colour
Figure 17. Time-series graph of preprocessed water quality data from Myponga and Happy Valley reservoirs, specifically a) total phosphorus b) phosphate c) nitrate d) iron
Figure 18. Time-series graph of preprocessed water quality data from Myponga and Happy Valley reservoirs, specifically a) manganese b) dissolved oxygen c) conductivity
Figure 19. Time-series graph of preprocessed water quality data from Myponga and Happy Valley reservoirs, specifically a) Anabaena b) green algae c) diatoms60
Figure 20. Time-series graphs of available real-time water quality data from Hope Valley reservoir
Figure 21. Parameter settings of HEA for rule set discovery (from Cao et al, 2006)68
Figure 22 . Seasonal patterns visualised as a U-matrix, a k-means map and a component plane71
Figure 23. Framework for the integrated approach using KANN, RANN and HEA for explanation and forecasting of algal dynamics in Myponga and Happy Valley reservoirs
Figure 24. Ordination and clustering of main water quality variables by KANN with regard to seasonality
Figure 25. Ordination and clustering of dominant algal groups by KANN regarding seasonality
Figure 26. KANN, using k-means map, with corresponding component planes for major water quality variables clustered seasonally and separated into periods of different management at Myponga reservoir
Figure 27. KANN, using k-means map showing water temperature ranges, and corresponding component planes for dominant algal functional groups in Myponga and Happy Valley reservoirs
Figure 28. KANN, using k-means map showing PO <sub>4</sub> concentration ranges, and corresponding component planes for dominant algal functional groups in Myponga and Happy Valley reservoirs
Figure 29. KANN, using k-means map showing NO <sub>3</sub> concentration ranges, and corresponding component planes for the dominant algal functional groups in Myponga and Happy Valley reservoirs
Figure 30. Experimental progression of Chapter 5101
Figure 31. a) Chl- <i>a</i> forecasting results for RANN (left) and HEA (right) models tested on one year of data from Myponga reservoir, b) sensitivity analyses results from RANN (left) and HEA (right)

Figure 32. Chl- <i>a</i> forecasting results for RANN (left) and HEA (right) models tested on two years of data from Myponga reservoir
Figure 33. a) Chl- <i>a</i> forecasting results for RANN (left) and HEA (right) models tested on one year of data from Happy Valley reservoir, b) sensitivity analyses results from RANN (left) and HEA (right)
Figure 34. Chl- <i>a</i> forecasting results for RANN (left) and HEA (right) models tested on two years of data from Myponga reservoir
Figure 35. Chl- <i>a</i> forecasting results for merged RANN (left) and HEA (right) models tested on two years of data, one from each reservoir
Figure 36. Chl- <i>a</i> forecasting results for merged RANN (left) and HEA (right) models developed using only electronically measurable inputs, tested on two years of data, one from each reservoir
Figure 37. a) <i>Anabaena</i> forecasting results for RANN (left) and HEA (right) models tested on one year of data from Myponga reservoir, b) sensitivity analyses results from RANN (left) and HEA (right)
Figure 38. <i>Anabaena</i> forecasting results for RANN (left) and HEA (right) models tested on two years of data from Myponga reservoir
Figure 39. a) <i>Anabaena</i> forecasting results for RANN (left) and HEA (right) models tested on one year of data from Happy Valley reservoir, b) sensitivity analyses results from RANN (left) and HEA (right)
Figure 40. <i>Anabaena</i> forecasting results for RANN (left) and HEA (right) models tested on two years of data from Happy Valley reservoir119
Figure 41. <i>Anabaena</i> forecasting results for merged RANN (left) and HEA (right) models tested on two years of data, one from each reservoir
Figure 42. K-means map for temperature ranges (top left) with corresponding component planes for <i>Anabaena</i> and Chl- <i>a</i> (top middle and right) in Myponga Reservoir; sensitivity curve for <i>Anabaena</i> and Chl- <i>a</i> in response to temperature change (bottom)131
Figure 43. K-means map for PO <sub>4</sub> ranges (top left) with corresponding component plane for <i>Anabaena</i> and Chl- <i>a</i> (top right) in Myponga Reservoir; sensitivity curve for <i>Anabaena</i> and Chl- <i>a</i> in response to PO <sub>4</sub> change
Figure 44. K-means map for NO <sub>3</sub> ranges (top left) with corresponding component plane for <i>Anabaena</i> and Chl- <i>a</i> (top right) in Myponga Reservoir; sensitivity curve for <i>Anabaena</i> and Chl- <i>a</i> in response to NO <sub>3</sub> change
Figure 45. Development of rule-based agents by means of HEA and k-fold cross validation

Figure 46. Structure of the rule-based Chl- <i>a</i> agent for Myponga and Happy Valley reservoirs. a) input sensitivity of the THEN branch, b) input sensitivity of the ELSE branch
Figure 47. Validation results of the rule-based Chl-a agent for Myponga and Happy Valley reservoirs for all data (1999-2003)144
Figure 48. Validation results of the rule-based Chl- <i>a</i> agent applied 160 days of real-time data from Hope Valley144
Figure 49. Comparison of real-time and interpolated data for the same period of the year
Figure 50. Structure of the rule-based <i>Anabaena</i> agent (inc. Chl-a) for Myponga and Happy Valley reservoirs a) input sensitivity of the THEN branch, b) input sensitivity of the ELSE branch
Figure 51. Validation results of the rule-based Anabaena agent (inc. Chl-a) for Myponga and Happy Valley reservoirs (1996-2003)
Figure 52. Structure of the rule-based <i>Anabaena</i> agent (not inc. Chl- <i>a</i> ) for Myponga and Happy Valley reservoirs a) input sensitivity of the THEN branch, b) input sensitivity of the ELSE branch
Figure 53. Validation results of the Anabaena rule-based agent (not inc. Chl-a) for Myponga and Happy Valley reservoirs (1996-2003)
Figure 54. Rainfall and turbidity levels in Myponga reservoir 1993166
Figure 55. The relationship between colour and DOC in Myponga reservoir167
Figure 56. The relationship between colour and DOC in Happy Valley reservoir
Figure 57. Forecasting results from HEA rule using past Chl-a values as input to predict current Chl-a levels

# LIST OF TABLES

Table 1. Water quality data from Myponga reservoir Sampling Location 1	49
Table 2. Water quality data from Happy Valley reservoir Sampling Location 1	51
Table 3. Comparison of reservoir attributes	53
Table 4. Water quality data from Hope Valley reservoir	61
Table 5. Data used for each experiment in this chapter	76
Table 6. Classification criterion used in 4.4.1.1 and 4.4.2.1	77
Table 7. Classification criterion used in 4.4.3.1, 4.4.3.2 and 4.4.3.3	77
Table 8. Data used for each experiment in this chapter	99
Table 10. Information table for RANN and HEA models developed to forecast Chl-a	
concentration in Myponga reservoir (2test years)	108
Table 11. Information table for RANN and HEA models developed to forecast Chl-a	
concentration in Happy Valley reservoir (1test year)	109
Table 12. Information table for RANN and HEA models developed to forecast Chl-a concentr	ration
in Happy Valley reservoir (2 test years)	111
Table 13. Information table for RANN and HEA models developed using merged data to fore	ecast
Chl-a concentration in both reservoirs (2 test years)	112
Table 14. Information table for RANN and HEA models developed using merged data, with c	only
electronically measurable input variables, to forecast Chl-a concentration in both reserv	voirs
(2 test years)	113
Table 15. Information table for RANN and HEA models developed to forecast Anabaena	
abundance in Myponga reservoir (1test year)	114
Table 16. Information table for RANN and HEA models developed to forecast Anabaena	
abundance in Myponga reservoir (2 test years)	117
Table 17. Information table for RANN and HEA models developed to forecast <i>Anabaena</i>	
abundance in Happy Valley reservoir (1test year)	118
Table 18. Information table for RANN and HEA models developed to forecast Anabaena	
abundance in Happy Valley reservoir (2 test years)	119
Table 19. Information table for RANN and HEA models developed using merged data to fore	cast
Anabaena abundance in both reservoirs (2 test years)	120
Table 21. Myponga Reservoir database details	129
Table 22. Summary of experiment specific data used throughout the chapter	138
Table 23. Carlson's trophic state index (TSI) (according to Carlson (1977))	164
Table 24. UECD lake classification standard (according to Vellenweider and Kerekes (1982)	) 164
Table 25. German lake classification standard (according to Ryding and Rast (1989))	164
Table 26. Observed water quality data used for reservoir trophic state classification	165
Table 27. Reservoir trophic state classifications	165