



**ADELAIDE  
UNIVERSITY  
AUSTRALIA**



**PREDICTING FRESHWATER HABITAT CONDITIONS BY  
THE DISTRIBUTION OF MACROINVERTEBRATES  
USING ARTIFICIAL NEURAL NETWORK**

**THESIS**

**Submitted for the degree of Master of Applied Science**

**in**

**Faculty of Agriculture & Natural Resources Science**

**Department of Soil and Water**

**by Huong Thi Thu, Hoang**

**Adelaide, Aug, 2001**

# Contents

<b>Content</b> .....	<b>i</b>
<b>List of Figures</b> .....	<b>v</b>
<b>List of Tables</b> .....	<b>viii</b>
<b>Abstract</b> .....	<b>ix</b>
<b>Publications and Scientific Communications During Candidature</b> .....	<b>xi</b>
<b>Acknowledgements</b> .....	<b>xii</b>
<b>Declaration</b> .....	<b>xiv</b>
<b>1 Introduction</b> .....	<b>1</b>
<b>1.1. Habitat Condition Assessment of Freshwater Ecology</b> .....	<b>1</b>
<b>1.2. Computational Approach to Support Bioassessment</b> .....	<b>2</b>
<b>1.3 The Contribution of This Work</b> .....	<b>4</b>
<b>2 Background</b> .....	<b>6</b>
<b>2.1 Habitat Condition Assessment</b> .....	<b>7</b>
2.1.1 Stream Ecology .....	7
2.1.2 Biological Assessment of Stream Habitat .....	19
2.1.3 Biological Assessment Based on Macroinvertebrate Assemblages .....	24
2.1.4. Statistical Model to Support Bioassessment .....	28
2.1.5. Summary and Research Needs .....	33
<b>2.2 Artificial Neural Networks to support bioassessment</b> .....	<b>34</b>
2.2.1 Artificial Neural Networks - An Introduction .....	34
2.2.2 Application of ANN to modelling ecosystem .....	36
2.2.3 Current Achievement in Application of ANN to Assessment of Habitat Conditions using Macroinvertebrate Assemblages .....	39
2.2.3 Research Needs .....	41
<b>2.3 Summary and Thesis Aims</b> .....	<b>41</b>

<b>3</b>	<b>General Material and Methodology</b>	<b>43</b>
3.1	<i>Introduction</i>	43
3.2.	<i>Study Sites and Site Selection</i>	43
3.3	<i>Habitat Condition Database – A review</i>	46
3.3.1	<i>Physical Riparian and Other Predictor Variables</i>	46
3.3.2	<i>Environmental Water Quality Variables</i>	51
3.3.3	<i>Macroinvertebrate Distribution</i>	52
3.3.4	<i>Summary</i>	52
3.4	<i>Artificial Neural Networks</i>	54
3.4.1	<i>Fundamental Concept</i>	54
3.4.2	<i>Basic Models and Learning Rules</i>	58
3.4.3	<i>Feedforward Networks and Supervised Learning</i>	64
3.4.4	<i>Summary</i>	70
3.5	<i>Procedure for Network Development</i>	71
3.5.1	<i>Software</i>	71
3.5.2	<i>Data Preprocessing and Modelling</i>	73
<b>4</b>	<b>Adopting Clean Water Approach</b>	<b>78</b>
4.1	<i>Introduction</i>	78
4.1.2	<i>Clean Water Approach</i>	78
4.1.2	<i>Aims and Hypothesis</i>	79
4.2	<i>Materials and Methods</i>	81
4.2.1	<i>Data Analysis</i>	81
4.2.2	<i>Network Architecture</i>	81
4.2.3	<i>Methods of Training</i>	82
4.2.4	<i>Method of Validation</i>	83
4.2.5	<i>Method of Prediction</i>	84
4.3	<i>Sensitivity Analysis</i>	84
4.3.1	<i>Results of Sensitivity Analysis</i>	85
4.3.2	<i>Network Revision</i>	89
4.4	<i>Validation Results and Comparison with AusRivAS model</i>	90
4.4.1	<i>Validation Results</i>	90
4.4.2	<i>Comparison with AusRivAS model</i>	94
4.5	<i>Apply for Prediction Step – Prediction Results</i>	97
4.6	<i>Discussion</i>	99

4.6.1	<i>Performance of Artificial Neural Networks</i> .....	99
4.6.2	<i>Relationship between Predictor Variables and Macroinvertebrate Assemblages</i> .....	100
4.6.3	<i>Reference Condition Approach for Rapid Assessment</i> .....	104
4.7	<b>Chapter Summary and Conclusion</b> .....	106
<b>5</b>	<b>Developing Dirty Water Approach</b> .....	<b>108</b>
5.1	<b>Introduction</b> .....	108
5.1.1	<i>Dirty Water Approach</i> .....	108
5.1.2	<i>Aims and Hypothesis</i> .....	109
5.2	<b>Materials and Methods</b> .....	110
5.2.1	<i>Data Analysis</i> .....	110
5.2.2	<i>Network Architecture</i> .....	111
5.2.3	<i>Method of Training</i> .....	111
5.2.4	<i>Method of Validation</i> .....	112
5.3	<b>Sensitivity Analysis</b> .....	112
5.3.1	<i>Results of Sensitivity Analysis</i> .....	112
5.3.2	<i>Networks Revision</i> .....	116
5.4	<b>Validation results</b> .....	117
5.5	<b>Discussion</b> .....	121
5.5.1	<i>Performance of Artificial Neural Networks</i> .....	121
5.5.2	<i>Relationship between Macroinvertebrates and Habitat Conditions</i> .....	122
5.5.3	<i>Data Limitation</i> .....	132
5.6	<b>Chapter Summary and Conclusion</b> .....	133
<b>6</b>	<b>Elucidation of Freshwater Habitat Condition Discovered by a Sensitivity Analysis</b> ...	<b>134</b>
6.1	<b>Introduction</b> .....	134
6.2	<b>Design and Interpretation of Sensitivity Plots</b> .....	135
6.3	<b>Elucidation of Causal Relationship by Sensitivity Plots</b> .....	137
6.4	<b>Contradiction to Literature Findings</b> .....	153
6.5	<b>Limitation of the Method</b> .....	154
6.6	<b>Using Sensitivity Analysis for Management Purposes</b> .....	155
<b>7</b>	<b>General discussion</b> .....	<b>157</b>
7.1	<b>Performance of Artificial Neural Networks</b> .....	157
7.2	<b>Improvement of Input Data Representation</b> .....	160
7.3	<b>Temporal and Spatial Variations</b> .....	161
7.3.1	<i>Spatial Variance in Invertebrate Data</i> .....	161

7.3.1	<i>Temporal Variance in Macroinvertebrate Data</i> .....	162
7.3.3	<i>Summary</i> .....	163
7.4	<b><i>Representation of Data on Macroinvertebrate Taxa</i></b> .....	164
7.4.1	<i>Functional Feeding Groups</i> .....	164
7.4.2	<i>Abundance data</i> .....	166
8	<b>Thesis Conclusions and Recommendations for Further Research</b> .....	168
8.1	<b><i>Conclusions</i></b> .....	168
8.1.1	<i>Usefulness of Macroinvertebrates as Biological Indicators</i> .....	168
8.1.2	<i>Usefulness of Artificial Neural Network as Prediction Tools</i> .....	169
8.1.3	<i>Elucidation by Sensitivity Analyses</i> .....	170
8.1.4	<i>Limits of Using ANNs</i> .....	171
8.2	<b><i>Recommendations for Further Research</i></b> .....	172
8.2.1	<i>Further Research on ANN Performance</i> .....	172
8.2.2	<i>Further Research on Elucidation of Freshwater Habitat Conditions</i> .....	174
	<b>References</b> .....	175

# List of Figures

<i>Figure 3.1 Map of Queensland indicating location of reference sites of the Queensland stream database</i> .....	44
<i>Figure 3.2 Method used by QDNR to determine stream order (Strahler, 1957)</i> .....	47
<i>Figure 3.3 Example of wetted width, channel width and stream depth (Nichols et al., 2000)</i> .....	49
<i>Figure 3.4 Schematic diagram of a biological neuron (Kent, 1980)</i> .....	55
<i>Figure 3.5. Schematic diagram of a Mc. Culloch and Pitts neuron</i> .....	57
<i>Figure 3.6 Basic network connection geometries (Lin and Lee, 1996)</i> .....	60
<i>Figure 3.7 A simple perceptron</i> .....	65
<i>Figure 3.8 Three layer feedforward backpropagation network (Lin &amp; Lee, 1995)</i> .....	68
<i>Figure 3.9 Approach for data-preprocessing and ANN modelling ( Hoang et al., 2001)</i> .....	73
<i>Figure 4.1 ANN architecture of the stream habitat model for each macroinvertebrate taxon</i> .....	82
<i>Figure 4.2. Cross validation results of some training examples</i> .....	83
<i>Figure 4.3 Sensitivity of the taxa Cladocera to changes of inputs within their data range</i> .....	86
<i>Figure 4.4 Correct predictions of presence/absence of macroinvertebrate taxa before and after exclusions of redundant input</i> .....	92
<i>Figure 4.5 O/E criteria for training set</i> .....	93
<i>Figure 4.6 O/E criteria calculated for validation set</i> .....	94
<i>Figure 4.7 Distribution of Observed/Expected ratios for predicted stream sites using the AusRivAS and the ANN models. Plots are presented for the training data of reference sites.</i> .....	96
<i>Figure 4.8 Distribution of Observed/Expected ratios for predicted stream sites using the AusRivAS and the ANN models. Plots are presented for the independent validation data of reference sites.</i> .....	96
<i>Figure 4.9 Application of ANN to predict biological impairment of test sites</i> .....	98
<i>Figure 4.10 Relation between geographical inputs and distribution of macroinvertebrates</i> .....	101

Figure 4.11 Relation between seasons and macroinvertebrate distribution.....	102
Figure 4.12 Relation between water temperature and distribution of aquatic insects .....	103
Figure 5.1 Correct prediction of presence/absence of macroinvertebrates before and after exclusion redundant inputs explored by models of dirty water approach.....	119
Figure 5.2 Relationship between max H velocity and presence of black flies Simuliidae .....	123
Figure 5.3 Relation between Max H Velocity and presence of Dytiscidae.....	123
Figure 5.4 Relations between H velocity and Site velocity, and presence of Hydropsychidae .....	124
Figure 5.5 Blackflies Simuliidae and mayflies Leptophlebiidae favour low pH conditions .....	125
Figure 5.6 Relations between pH and mussel Corbiculidae, chironomid sf-Tanypodinae and mayfly Caenidae.....	126
Figure 5.7 Copepoda, Oligochaeta, Tanypodinae favour conditions of low conductivity .....	127
Figure 5.8 Dytiscidae and Hydropsychidae favour conditions of high conductivity .....	128
Figure 5.9 Relations between concentration of N and P and grazer Cladocera, blackflies Simuliidae and water tiger Dytiscidae .....	129
Figure 5.10 Relations between concentration of N and P and mayflies Leptophlebiidae, and dragon flies Gomphidae .....	130
Figure 5.11 Relations between concentration of N and P and grazer Cladocera, blackflies Simuliidae and water tiger Dytiscidae Caramoceratidae, and beetles Hydrophilidae.....	131
Figure 6.1 "Flat" responses .....	135
Figure 6.2 "Ramp" responses .....	136
Figure 6.3 "Threshold" responses.....	136
Figure 6.4 "Plateau" and "Valley" responses.....	137
Figure 6.5 Preference of Dugesiiidae to H velocity and pH .....	138
Figure 6.6 Preference of Oligochate to pH and site velocity.....	138
Figure 6.7 Preference of Atyidae to H velocity and slope .....	139
Figure 6.8 Preference of Palaemonidae to longitude and mean daily temperature.....	139
Figure 6.9 Preference of Copepoda to H velocity and Water Temperature.....	139
Figure 6.10 Preference of Cladocera to H Depth and water temperature .....	140
Figure 6.11 Physiological effects of water temp. to Daphnia (Lampert & Sommer, 1987)....	140
Figure 6.12 Preference of Ostracoda to H Depth and reach .....	141
Figure 6.13 Preference of Planobidae to slope and range in DSMM .....	141
Figure 6.14 Preference habitats of Thiaridae .....	142
Figure 6.15 Preference of Corbiculidae to stream order and pH .....	142
Figure 6.16 Preference of Acarina to Water temperature and N level .....	143

<i>Figure 6.17 Preference Baetidae to water temperature and H depth</i> .....	143
<i>Figure 6.18 Preference of Prosoptomatidae to Water temp and Latitude (H velocity)</i> .....	143
<i>Figure 6.19 Preference of Caenidae to Site max Velocity and Mean Channel Width</i> .....	144
<i>Figure 6. 20 Preference of Leptophlebiidae to Longitude and H velocity</i> .....	144
<i>Figure 6.21 Preference of Libellulidae to water temperature, Cordulidae to site max velocity, and Gomphidae to H Velocity and Silt-clay</i> .....	145
<i>Figure 6.22 Preference of Coenagrionidae to reach and mean daily temperature.</i> .....	145
<i>Figure 6.23 Preference of Gripopterygidae to Water temperature and (DSMR)</i> .....	146
<i>Figure 6.24 Preference of Corixidae to longitude and H velocity</i> .....	146
<i>Figure 6.25 Preference of Notonectidae to H depth(clean) and H velocity (dirty)</i> .....	146
<i>Figure 6.26 Preference of Pleidae to latitude</i> .....	147
<i>Figure 6.27 Preference of Veliidae to H velocity and slope (clean</i> .....	147
<i>Figure 6.28 Preference of Dytiscidae to H depth and H velocity</i> .....	147
<i>Figure 6.29 Preference of Elmidae to H velocity and alkalinity (clean)</i> .....	148
<i>Figure 6.30 Preference of Psephenidae to H velocity and % cobble</i> .....	148
<i>Figure 6.31 Preference of Hydrophilidae to H Depth (clean) and detrital cover</i> .....	148
<i>Figure 6.32 Preference of Orthoclaudiinae to reach and water temperature</i> .....	149
<i>Figure 6.33 Preference of Tanypodinae to stream order and water temperature</i> .....	149
<i>Figure 6.34 Preference of Simuliidae to H velocity and altitude</i> .....	149
<i>Figure 6.35 Preference of Ceratopogonidae and Tabanidae to Water temperature and Habitat depth</i> .....	150
<i>Figure 6.36 Preference of Leptoceridae to H velovity and water temperature</i> .....	150
<i>Figure 6.37 Preference of Hydropsychidae to longditude and H velocity</i> .....	150
<i>Figure 6.38 Preference of Ecnomidae to H velocity and Water temperature</i> .....	151
<i>Figure 6.39 Preference of Hydroptilidae to H depth and boulder content</i> .....	151
<i>Figure 6.40 Preference of Calamoceratidae to Water temperature H velocity</i> .....	151
<i>Figure 6.41 Preference of Philopotamidae to H velocity</i> .....	152
<i>Figure 6.42 Preference of Pyralidae to H Velocity and %Cobble</i> .....	152
<i>Figure 6.43 Examples of contradiction between sensitivity analyses and literature finding</i> .....	153



# List of Tables

<i>Table 2.1 Assessment of biological indicators of water quality (Cranston et al. 1996 [p.144]).....</i>	<i>23</i>
<i>Table 3.1 Selection criteria used to determine reference site condition (Conrick &amp; Cockayne, 2000).....</i>	<i>45</i>
<i>Table 3.2 Potential predictor variables .....</i>	<i>50</i>
<i>Table 3.3 40 most common macroinvertebrates taxa in Queensland stream system.....</i>	<i>53</i>
<i>Table 4.1 Division of O/E taxa into categories for reporting (Coysh et al, 2000).....</i>	<i>80</i>
<i>Table 4.2 Summary of the input sensitivity of 40 MI taxa by mean of percentages of output change for Clean Water Approach.....</i>	<i>87</i>
<i>Table 4.2 (continued) .....</i>	<i>88</i>
<i>Table 4.3 Number of inputs in the taxon-specific revised networks.....</i>	<i>89</i>
<i>Table 4.4 The correct predictions for presence/absence of 40 macroinvertebrate taxa of initial models with 39 inputs and revised models after sensitivity analysis with taxa-specific number of inputs.....</i>	<i>91</i>
<i>Table 4.5 Average numbers of taxa observed at sites in according environmental conditions assessed by the value of O/E criteria .....</i>	<i>98</i>
<i>Table 4.6 Numer of sites where taxa were present in total 180 validation sites (results from revised models).....</i>	<i>99</i>
<i>Table 5.1 Summary of input sensitivity of 40 MI taxa by mean of percentages of outputs change for dirty water approach.....</i>	<i>113</i>
<i>Table 5.1 (continued) .....</i>	<i>114</i>
<i>Table 5.1 (continued) .....</i>	<i>115</i>
<i>Table 5.2 Correct prediction of presence/absence of macroinvertebrate taxa in dirty water approach before and after sensitivity analyses.....</i>	<i>118</i>
<i>Table 5.3 Number of sites where presence of taxa was observed or predicted by revised models.....</i>	<i>120</i>

## Abstract

Stream and river ecosystems play a crucial role in the human existence even though they contribute only 0.00008% to the Global water budget. Rivers and streams are important sources for drinking water, industrial and agricultural water, and recreation. Hence, there is a lasting interest in the control of river health. The second half of Twentieth Century has been a period of intensive study of rivers. Many efforts had been spent for better understanding basic limnological processes including physical, chemical and biological processes. Researches realised that all these limnological processes could not be studied separately but they are in very intimate interrelationship with each other. Any change in any limnological process can upset the balance and lead to disturbance in the freshwater ecosystem. Assessment and prediction of river and stream health gain the great interest in management in order to maintain a sustainable balance in stream and river ecosystem for human activities now and for future generations.

River health had traditionally been assessed solely on the chemical analysis of water samples. In recent years there has been realisation that the structure of plant and animal communities of the river can give us more accurate and integrated information about conditions of river and stream health. Among these biological communities, macroinvertebrates are most widely used because they are abundant and diverse, and are sensitive to changes in water quality, flow regime and habitat conditions they inhabit. Impacts on these animals are relatively long lasting and can be detected for some time after the impact occurs.

The computational approach had been applied to analyse the relationship between habitat conditions and stream macroinvertebrate assemblages. Statistical models had gained some significant successes. However, they still have some constrains in dealing with complexity and highly non-linearity of the stream system. A new generation of computer program called Artificial Neural Network proves to be very efficient for the study complex and nonlinear processes. In the context of the given Master research project, Artificial Neural Networks were applied for modelling Queensland river and stream system.

Two approaches are developed by means of Artificial Neural Networks to study the Queensland river and stream network, which spreads over the territory of the federal state of Queensland (Australia) and covers the catchments of most major and many minor Queensland rivers.

The clean water approach was adopted to determine relationship between presence and absence of macroinvertebrate taxa and physical predictor variables, which are considered relatively stable under human activities. The model therefore studied data from reference sites in near pristine conditions. Validation results provided correct prediction of the presence/absence of these taxa with an average accuracy of 80 %. Trained models were applied to assess habitat conditions of impacted and test sites. The assessment of the health of specific sites was then based on the comparison between observed and predicted site data. Criteria O/E (Observed/Expected) was used to give rapid assessment of habitat at sites ranging from reference to badly degraded conditions.

The dirty water approach did not distinguish site into reference and degraded. Networks had been trained with data from both clean and degraded sites. This approach studied interrelationship among physical, chemical and biological processes. The input layer contained not only physical predictor variables but also chemical variables, which are altered under human impacts. Validation also was made by mean of correct prediction of macroinvertebrate taxa for both reference and impacted sites and provides average accuracy of 76%. Dirty water approach can be applied for quantitative prediction of habitat condition by mean of water quality.

Sensitivity analyses were carried out by manipulating the values of input parameters and assessing the resulting changes in outputs. This method identified the environmental predictor variables best able to predict the presence/absence of each family. The primary intention of this sensitivity analysis was to improve network performance by limiting input variables to those that were sensitive for each model. However, this process also provided new insights into relationships between environmental variation and the occurrence of Queensland stream fauna and enabled the identification of ecological traits of each taxon.

The two modelling approaches provided good results and can be applied for management purposes. Artificial Neural Networks proved to be an effective computational approach to support bioassessment. However, all models developed during this project studied only spatial variations of processes in stream and river ecosystem. Future research should focus on temporal variations of relationships between environmental variables and the distribution of macroinvertebrates as well. Model training and validation using databases from other Australian stream systems would further contribute to a generalisation of the ANN stream models.

# Publication and Communications in the Scientific Congress during the Candidature

## Paper

Hoang H., Recknagel F., Marshall J., Choy S. (*in press*) Predictive modelling of macroinvertebrate assemblages for stream habitat assessments in Queensland (Australia). *Ecological Modelling*

## Conferences attendance

1. 2nd Conference on Postgraduate Research on the Environment, Adelaide, September 13th 2000

Hoang T.H. "Habitat assessment of streams by Machine learning"

2. 2nd International Conference on Applications of Machine Learning to Ecological Modelling, Adelaide, November 27th - December 1st, 2000

Hoang H., Recknagel F., Marshall J. and Choy S. "Predictive modelling of macroinvertebrate assemblages for stream habitat assessments in Queensland (Australia)"

3. 2001 Societas Internationalis Limnologiae XXVIII Congress, Melbourne 4-10th February, 2001

Marshall J., Hoang H.\*, Choy S. and Recknagel F. "Relationships Between Habitat Properties and the Occurrence of Macroinvertebrates in Queensland Streams (Australia) Discovered by a Sensitivity Analysis with Artificial Neural Networks"

4. FENNER 2001 Biodiversity Conservation in Freshwater, Same Landscape, Different Perspective, Canberra 5-7th July, 2001

Hoang H\*, Marshall J., Recknagel F and Choy S "Elucidation of relationships between habitats and macro-invertebrates in Queensland streams by Artificial Neural Networks"

# Acknowledgements

First and foremost, I would like to thank God for not creating any unexpected difficulty in the last 18 months, and for giving me enough energy to overcome all the expected ones to finish this work.

Thanks to Dr Friedrich Recknagel for his supervision, valuable discussions and support in the research, as well as in all the administration work during my candidature. His encouragement in the hardest stage of the research was critical to my achievement.

I would like to extend my greatest thanks to the anonymous biologists and hydrographers from the Queensland Department of Natural Resources who worked very well with the collection and processing of the data used in this study. This study would never have been possible without their hard and invaluable works. Thanks also to Jon Marshall and Satish Choy; their great collaboration is a key to success of this study.

Special thanks to Jason Bobbin, Hugh Wilson, Mardi van der Wielen, Tumi Bjornsson and Li Wen in the Aquatic Research Group. Their continuous feedback and worthwhile suggestions improved the results, and later this manuscript in many ways.

Thanks are given to all those in the Department of Soil and Water, Adelaide University for their invaluable help during my candidature, especially John Davey and Tracey Parish, who are always very helpful whenever I needed a help, and Rob Murray, who was always ready whenever I came to bother him.

Thanks to Margaret Cargill for her patience and helpfulness. I am deeply indebted to her for improving my English, which is the greatest thing I have gained in my study time in Australia. Any new technology will be out-of-date one day, but English does not need to be updated regularly.

Thanks to my dearest Dũng for his infinite encouragement, enthusiasm and love. This work proves his ability to see the possibilities in me beyond my own belief. He was right; to believe that we can do means that we are already in half way there. However, there were a lot of efforts needed from me to go through the second half.

Thanks to Nga' for her support, care and friendship during our hard but cheerful time in Adelaide, which we began to get used to and love. The problem is we have too many hobbies. Nevertheless, I believe that all these hobbies are needed to support our biggest hobby – *love of knowledge*. Thanks to all our hobbies.

Thanks to my extended family, whom I cannot thank enough. Thanks to Mums, for their understanding and support during the time I could not fulfill my duty in the family.

Thanks to my friends in Adelaide, in Hanoi and all over the world for their support to keep me in the entire world. Special thanks are given to guys in *h1\_class group* and its Moderator for all their efforts; their E-mails were always the cheeriest things in my working days.

And finally, the financial support from AusAID for an 18-month scholarship is highly acknowledged.

Thank you all.



## Declaration

*I declare that this work contains no material, which has been accepted for the award of any other degree or diploma in any university or other tertiary institutions. To the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due references has been made in the text.*

*I consent to this copy of my thesis, when deposited in the University library, being made available for loan or photocopying.*

**Date**

18.04.01

**Signed**