ADELAIDE
UNIVERSITY
AUSTRALIA

SUB CRUCE LUMEN

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

List of Figures

List of Tables

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 than 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

Marshall1 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 financially 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 **Signed**

18.04.01

What we have to learn to do

We learn by doing . . .

ARISTOTLE, *Ethics*

1 Introduction

1.1. Habitat Condition Assessment of Freshwater Ecology

Running water is the most important freshwater resource for man, being used for a variety of purposes. The maintenance of high quality running water has become an increasingly important issue in recent years, as greater demand has been placed on water resource. The quality of rivers and streams depends on their physical, chemical and biological properties. The latter are reflected by the types and density of living organisms present in the water.

Historically, water quality has been measured by physical, chemical and microbiological parameters such as biological oxygen demand, suspended sediments and bacterial counts. Chemical analyses determine concentrations of certain substances from sample taken at a specific point at a specific time. They therefore are often criticised because they only reveal the quality of the water at the time of sampling, and their further relevance has to be inferred by extrapolation from limited data (Hellawell, 1997). Biological monitoring, on the other hand, generally is considered to provide a more integrated appraisal of water and overall environmental quality (Hynes, 1960). Therefore, there is now widespread recognition that not only chemical analyses but biological techniques are required for an appropriate assessment of river quality (Wright, 1995). Moreover, biological surveillance of communities with special emphasis on characterising taxonomic richness and composition was claimed to be the most sensitive tool for quickly and adequately detecting alterations in aquatic ecosystem (Cairns & Pratt, 1993).

In aquatic ecosystems, such as streams and rivers, biological indicators have been proposed that use algae, fish, and macroinvertebrates (Hellawell, 1986). Amongst aquatic animals that can be used in bioassessment, macroinvertebrates proved to be an excellent indicator for the quality of freshwater streams (Rosenberg and Resh, 1993). Because of the crucial ecological functions of macroinvertebrates within stream ecosystems, great efforts are undertaken to preserve and restore their stream habitats. In terms of their use for biomonitoring, macroinvertebrates in streams have relatively long life cycle that allow exposing them to pollutants over a long period of time and integrating the effect of short-term pollution episodes.

Macroinvertebrate assemblages were an objective of many projects of bioassessment of habitat conditions, and river and stream quality. The British RIVPACS (River Invertebrate Prediction and Classification System) (Wright, 1995) was developed using macroinvertebrates for biological assessment of river quality in Great Britain. The RIVPACS approach was preferred to similar North American schemes, which had already been used successfully to assess river condition on national scale and also in regional framework. AusRivAS, the Australian River Assessment Scheme, is a national bioassessment program that uses aquatic macroinvertebrates to meet the first objective of the National River Health Program: assessment of 'health' or ecological condition of Australian rivers (Schofield & Davies, 1996).

1.2. Computational Approach to Support Bioassessment

While experienced biologist can make meaningful assessment of habitat condition from suitable biomonitoring data, it becomes difficult to comprehend extensive datasets when collections are made for long period of time and/or when large data sets need to be analysed and causal factors be identified. In ecological research, therefore, the processing and interpretation of data play an important role. The ecologist uses many methods, ranging from numerical, mathematical, and statistical methods to techniques originating from artificial intelligence such as expert systems (Recknagel, 1989), genetic algorithms (d'Angelo et al., 1995), and Artificial Neural networks (ANN) (Walley & Fontama, 1998) in order to study interrelations between biological communities and environmental parameters.

Stream modelling based on ecological knowledge and adequate stream monitoring data can substantially facilitate and improve assessment of stream habitats. Different modelling techniques have been developed and applied to freshwater streams. Moss et al. (1987) developed a statistical model for predicting macroinvertebrates occurring at some stream sites in Great Britain. The model worked with probability of the occurrence of macroinvertebrates and provided reasonable prediction results using environmental variables. Simpson et al. (1997) used a similar model for freshwater streams in Australia and New Zealand. Even though these models achieved some success to a variety of stream systems, they lack of ability to cope with non-linearity and high complexity of stream system.

Modelling freshwater quality is extremely difficult, as the interrelations between various influences are not well known. Hydrodynamic models are difficult to couple with chemical and biological models. The action of hydrological process on ecological processes has hardly been elucidated, as the requirements are different for both systems (Straskraba and Gnauk, 1985). The use of ANN may overcome many of these difficulties.

Walley and Fontama (1998) developed artificial neural network (ANN) to predict macroinvertebrate taxa in unpolluted river sites in the UK. Their results demonstrated the potential of ANN to model non-linear relationship between environmental variables and biotic indices. Schleiter et al. (1999) and Chon et al. (1996, 2000a, 2000b) went one step further to model the community dynamics of macroinvertebrates in German and Korean streams using ANN.

Artificial neural networks belong to a new generation of computer models based on machine learning techniques. ANNs were developed as models of biological neurons. They learn from experience in the database and can be able to solve real ecological problems in various areas (Lek et al., 2000). ANNs are universal function approximators, they are able to learn a complex non-linear mapping between independent and dependent variables from data. They do not require assumptions about mathematical relationship between state variables and the nature of the distribution of data. Machine learning models have the ability to extract temporal or spatial patterns and knowledge from highly nonlinear and complex data. Based on such patterns and knowledge they can predict future conditions. Machine learning models have successfully been applied to freshwater lakes (e.g. Recknagel 1997, Recknagel et al. 1998) and promise a new quality in stream modelling.

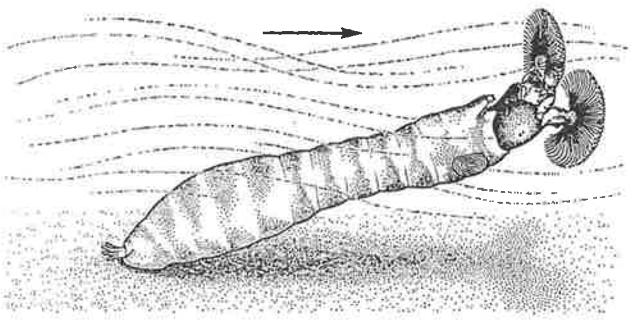
1.3 The Contribution of This Work

Even though much work has already been done in the field of applying Artificial Neural Networks in order to study interrelations between abiotic factors of freshwater stream ecology and different biotic community member around the world, no such research had been carried out yet in Australia. The aim of this work is to assess the suitability of the ANN models to determine the biological and environmental conditions of freshwater stream in Australia. The present case study was conducted by means of a comprehensive database of the Queensland stream system.

The work adopted the referential approach from the Australian River Assessment System (AusRivAS) and applied the Artificial Neural Networks as new computational tools to analyse and generalise the patterns within the database of the Queensland river and stream system. So-called “clean water approach” was applied to study the distribution patterns of macroinvertebrates in clean water. These patterns were then applied for predicting macroinvertebrate assemblage in clean water. The assessment criteria O/E from AusRivAS was also adopted in order to evaluate the performance of the newly developed models. Evaluation was made by comparing performances of ANNs and statistical models used in the AusRivAS protocol to demonstrate the potential of ANNs as alternative analytical tools.

In the second step, another approach was developed in an attempt to extend the capability of ANN models in dealing with abiotic factors. The so-called “dirty water approach” extended the capability of ANN models not only to work with physical predictor variables but also to work with water chemistry. More complex iteractions among physical predictors, water chemistry and macroinvertebrates themselves that determine the distribution of macroinvertebrates were studied and direction for applying this approach for management purposes were investigated.

The elucidation capability of ANN models was explored by sensitivity analyses. Sensitivity analyses drew the effects of single abiotic factor on presence of individual macroinvertebrate taxon. 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. This work demonstrated that the ANN technique applied for sensitivity analyses has the potential to enhance our understanding of how natural and anthropogenic impacts affect components of aquatic ecosystems.



A black-fly Simuliidae *larva in the typical filter-feeding posture (Gullan& Cranston,2000)*

2 Background

Stream and river ecosystems experience dramatic changes in response to human activities such as population growth and economic development. Predicting stream habitat condition is increasing of interest to water resources planners, policy makers, ecological researchers and especially limnologists. Traditional physical and chemical methods to examine freshwater habitat conditions have limited ability to deal with spatial and temporal variations. So-called bioassessment of stream habitat condition may be a suitable alternative (Hynes, 1970; Hellawell, 1986; Rosenberg & Resh, 1993; Loeb & Spacie, 1994). Fundamental to assessment of river health and biotic integrity is an understanding of the links between the habitat in which organisms live and factors shaping it (Norris & Thoms, 1999). Amongst aquatic organisms that can be used in bioassessment, macroinvertebrates have proved to be an excellent indicator for the quality of freshwater stream habitats (Rosenberg and Resh, 1993, Davis and Simon, 1995; Hawkers, 1997).

Stream modelling based on ecological knowledge and adequate stream monitoring data can substantially facilitate and improve assessment of stream habitats. Different modelling techniques have been developed and successfully applied to freshwater streams. Among the approaches to support bioassessment, Artificial Neural Networks (ANN) have been recognized as a potential tool in ecological modelling (Recknagel *et al.,* 1997, 1998; Schleiter *et al.*, 1999; Lek & Guegan, 2000). ANN models have flexibility to cope with temporal and spatial dynamics and are able to deal with the distinct non-linearity and high complexity of freshwater streams.

The main areas of discussion in this chapter are bioassessment of river habitat condition and the potential of machine learning in supporting bioassessment. Habitat condition monitoring and bioassessment of streams is discussed before bioassessment of streams by means of Artificial Neural Network techniques is introduced. Specific examples demonstrate the potential of ANN models in the areas of freshwater bioassessment. Among from recent research, a direction for the current studies is set which applied ANN techniques to the bioassessment of freshwater stream habitat conditions.

2.1 Habitat Condition Assessment

2.1.1 Stream Ecology

Streams and rivers are fundamental to human existence as well as to global diversity. Streams and rivers do not only affect the landscape over very long time periods, but are also in turn directly affected by the catchments where they originated and through which they flow. The understanding of stream ecosystem structure and function has progressed rapidly and continues to be one of the most active areas of research in aquatic ecology (Hauer & Lamberti, 1996).

Rivers are complex systems of flowing water draining specific land surfaces and are very important freshwater resources. Rivers are characterised by uni-directional current with a relatively high average flow velocity (0.1 to 1 m/s) in comparison with lakes and other water bodies. The river flow is highly variable over time. Prevailing current and turbulence cause thorough and continuous vertical mixing in rivers (Meybeck *et al.* 1996). Streams and rivers have a complex nature, which can be explained as a consequence of the three-dimensional geometry of channels with a long profile, a cross-section and mutual adjustment over a time scale (Allan, 1995). In river ecosystems, the physics, chemistry and biology of the water body are interrelated. Any substances introduced to a river are transported and transformed by physical, biological and biochemical processes. Consequently, the habitat condition of river water is changed spatially and temporally (Allan, 1995; Meybeck *et al.*, 1996c; Townsend *et al.*, 1997; Mason, 1996). Spatial and temporal variations in river ecosystem are crucial to the abundance and activities of freshwater organisms and to ecological processes in aquatic ecosystems, because they are main features of different types of water bodies and habitat conditions within them. Variation in the distribution and activities of aquatic organisms is evident at all spatial and temporal scales but especially in streams, where biotic differences are often obvious within and across a watershed. The distribution, abundance and activities of aquatic biota vary clearly with time, over temporal scales ranging from seconds or minutes to years (Stewart and Loar, 1994).

Hydrological processes, food resources, nutrient dynamics, riparian vegetation and many other factors intimately affect the structure and function of stream ecosystems (Hynes, 1970; Cummins, 1984; Allan, 1995). The following section discusses fundamental processes occurring in streams in their spatial and temporal variation as well as in their interrelations with each other.

Physical Processes

Hydrological processes

Hydrological processes strongly affect all processes occurring in streams. The most fundamental hydrological measurement that characterise all river and stream ecosystems is *discharge*, the volume of water flowing through a cross section of a stream channel per unit time (Gore, 1996). The amount of water flowing past a given point combined with the slope of the stream channel produces an indication of stream power. The potential energy of the stream is dissipated as friction heat loss on the streambed and when the stream picks up and moves material. The work performed by the stream influences the distribution of suspended sediment, bed material, particulate organic matter and other nutrients. The distribution of these materials has substantial influence on the distribution of riverine biota (Vannote et al. 1980)

Rivers and streams are integrated flowing systems that create and maintain aquatic habitat within the turbulent structure of the flow, as well as on and below the channel bed. At the catchment scale, the hydraulic condition of the flow may be generalised as uniform or gradually varying above and below interruptions in the longitudinal profile of the stream. At the stream reach scale, non-uniform flow conditions that occur in pools, riffles, and meanders can be distinguished. To distinguish the pattern of non-uniform flows, the mean depth, velocity and direction of the flow may be mapped on sketches or survey plans of a reach. The channel configuration and flow conditions are major components used to characterize the preferred habitats for different aquatic biota. At the habitat scale, individual stream flowlines and different states of flow can be outlined and analysed as rapidly varying, non-uniform flow (Newbury, 1996).

Stream morphometry and longitudinal patterns

Running water systems consist of tributary streams that erode the landscape following the weaker strata of bedrock, which then gradually coalesce to form the main river as it flows downhill. Streams do not flow far in straight lines, but tend to meander with gentle or sharper bends. Channels may also divide into a series of branches in response to variation in discharge, the nature of sediment, and the presence of erodable banks. In segments, water velocity varies longitudinally, and sediment on the stream bottom is eroded continuously from some areas and deposited in others (Giller & Malmqvist, 1998). This leads to alternating sequences of shallower, fast flowing, rifle areas with coarse substrates, and deeper pools with slow flow and fine substrate. Each of these areas represents a type of habitat with specific habitat conditions and stream biota.

Stream characteristics change longitudinally when upland streams turn into downland from the headwater streams. On the one hand, the stream sizes increase with the distance from source. On the other hand, the direct influence of the surrounding landscape on the functioning of the running water ecosystem decreases. The slope of the channel decreases, discharge increases, variability and nature of flow change and so does water chemistry. These longitudinal changes in physical and chemical characteristics impose significant consequential changes on ecosystem processes (such as decomposition, community respiration, primary production) and patterns (such as standing stock of organic biomass, species richness of invertebrates and fish, and community structure) (Statzner & Borchardt, 1994).

Temperature

Temperature is one of the most important variables in the stream ecosystem. Temperature affects movement of molecules, fluid dynamics, saturation constants of dissolved gases in water, metabolic rates of organisms, and a vast array of other factors that directly and indirectly affect life in the stream ecosystem (Hauer & Hill, 1996). Typically, the greatest source of heat in freshwater is solar radiation. However, in very heavily shaded streams, transfer of heat from air and flow from ground water are more important than direct solar radiation (Stanford et al., 1988).

Annual fluctuations in stream temperature are very important to stream organisms. Critical life history variables (e.g. reproduction, growth) of lotic plant and animals are regulated by temperature (Ward & Stanford, 1982). Many stream animals use temperature or temperature change as an environmental cue for emergence or spawning (Hauer & Hill, 1996). Temperature sets limits to the environment that species can live in, and species are generally adapted to certain temperature regime. The effect of temperature on the biota may be indirect through its influence on metabolic rates and oxygen concentration (Giller & Malmqvist, 1998).

A common misconception is that stream temperature is uniform among habitats within a stream reach. In reality, stream temperature may be highly variable between habitats only a few meters apart. Streams frequently experience significant changes in temperature from small shaded headwaters to broad, open canopied river reaches (Stanford et al., 1988).

Light

Light is a critical variable in most ecosystems. In streams, solar radiation is necessary for photosynthesis by attached algae. It is also the medium through which all-visual behaviors (e.g. predation by fish, macroinvertebrates) is expressed (Hauer & Hill, 1996). There is evidence to suggest that light can influence benthic invertebrate distribution. Some animal taxa show highest abundance in unshaded areas while other taxa prefer shaded areas (Giller & Malmqvist, 1998). The longitudinal downstream change in light regime and its consequences for stream bio-energetics is an integral part of stream ecosystems (Vannote et al., 1980).

Seasonal variation in lotic light regime is caused by changes in sun angle and day length and by phenological changes in streamside vegetation. Spatial variability in lotic light regimes also is high. Variation in the amount of shade cast by streamside vegetation is responsible for much of the spatial variability of light in streams. Streamside vegetation also plays a crucial role in the longitudinal gradient of light regimes in stream systems. As stream size enlarges progressively downstream, riparian trees and bushes shade proportionally less of the stream, allowing more diffuse sunlight to reach the streambed (Hauer & Hill, 1996).

Suspended sediment and bedload

Sediment concentration and bedload provide important information about stream systems that has direct significance for aquatic biota. Sediments are important for maintaining spawning gravels and the channel morphology of stream and river that form habitat for benthic organisms. Large amount of bedload transport may scour benthic plants and organisms, bury spawning gravels, or cause relatively rapid channel adjustments. The movement of sediment into stream systems generally occurs by two major processes: surface erosion and mass wasting or landslides. Furthermore, where soil, rock or previously deposited alluvium are being eroded by a stream or river system, these materials can represent an important source of sediment to aquatic systems. Numerous factors are involved in erosional processes. These factors include climate (precipitation, temperature), topography, vegetation (type and density of vegetation), soil (particle sizes, erodibility), and geology (characteristics of parent materials and bedrock). In addition, human perturbation and management practices that affect watersheds and stream systems can greatly increase natural rates of erosion and sediment yield. Inorganic sediments are typically characterised by two primary modes of transport: suspended sediment or bedload sediment. Each of these categories delineate relatively different groups of particle sizes with different implication for the morphology and ecology of a stream system (Beschata, 1996).

Sediment particles transported in suspension by a stream are typically < 0.1mm in diameter and consist mostly of silt and clay sized particles. Suspended sediment particles are transported downstream at essentially the same velocity as the flowing water. Bedload sediment consists of relatively large inorganic particles that are transported by water along the bed of the stream. They are relatively large (>1mm in diameter) and consist mostly of coarse sands, gravels, cobbles or larger stones. These sediments have important implications for aquatic plants and organisms because of their influence on the character of the stream substrate and channel morphology (Beschta, 1996). Moreover, light attenuation by suspended sediment also can reduce light penetration in streams, which is a significant factor within freshwater ecosystem as discussed in the previous section (Hauer & Hill, 1996)

Substrate

The substrate itself comprises a wide variety of inorganic and organic materials. The inorganic material (ranging in size from silt to sand, gravel, pebbles, cobbles, boulders and bedrock) is usually eroded from the river basin slopes, river channel and banks, and modified by the current. The organic materials vary from organic fragments and leaves, to fallen trees, derived ultimately from the surrounding catchment and upstream habitats, as well as aquatic plants such as filamentous algae, moss and macrophytes (Giller & Malmqvist, 1998).

The stronger the current velocity the larger the particle size that can be moved thus current velocity and substrate type are related, and mean substrate particle size generally declines downstream. However larger particles can protect smaller ones from being entrained in the current and carried away, and so in coarser substrates, finer sands and gravels will accumulate in between or behind the larger particles and increase the heterogeneity of the substrate. Temporal variability in substrate will occur naturally. The "stability" of the substrate refers to its resistance to movement and is proportional to particle size. Redistribution of the substrate and movement of particles will occur during periods of increased discharge following rainstorm (Giller & Malmqvist, 1998).

The nature of substrate is of prime importance for lotic invertebrates. It provides habitat space for a variety of activities such as resting and movement, reproduction, rooting or fixing to, and for refuge from predators and flow. It also provides food directly (organic particles) or surface on which food aggregates (e.g. algae, coarse and fine detrital particles) (Giller & Malmqvist, 1998). Diversity and abundance tend to increase with substrate stability and with the presence of organic detritus. Sandy substrates are thought to be poorest, due to instability. Stony riffles normally have a greater range of invertebrates than pools rich in silt (Allan, 1995). Heterogeneity is also important in controlling abundance and diversity, as mixed substrates provide a greater range of surfaces to colonize and microflow patterns (Giller & Malmqvist, 1998). Because of the above finding, the relationships among substrates and fauna diversity, biomass and abundance are not linear.

Water Chemistry and Chemical processes

Water Chemistry

Oxygen is required by all aerobic respiration. Oxygen enters water largely via diffusion from the air at the water surface. Oxygen solubility in water is negatively correlated with water temperature. Oxygen levels also vary with current speed and turbulence and are affected by the presence of macrophyte vegetation, as oxygen is a by-product of photosynthesis (Giller & Malmqvist, 1998). Species do differ in their respiratory ability and oxygen requirement, as evidenced by different responses to organic pollution that reduces oxygen, and these differences in turn contribute to differences in species distribution (Hynes, 1970).

A typical river is essentially a dilute calcium bicarbonate solution dominated by a few cations and anions (Wetzel, 1983). Other important variables to consider are pH (which measures the acidity of water), hardness (which measures concentration of Ca^{2+} and Mg^{2+}), conductivity (which measures the total ionic content) and alkalinity (which measures concentration of carbonates). These variables have direct and indirect effects on habitat conditions within streams and stream biota (Wetzel, 1983; Giller & Malmqvist, 1998).

Regular monitoring of water chemistry at a sampling location will show patterns of variation over time. Normally, short-term reversible changes in chemistry follow the rise and fall of water levels associated with rainfall events, or with long-term seasonal changes. If heavy rains follow a period of drought, accumulated solutes in soil water, which have increased in concentration through evaporation, undergo flushes of mineralisation or nitrification. The post-drought runoff water will contain large amounts of nitrates and other solutes (Hornung and Reynolds, 1995). Monitoring over long period of time can indicate directed changes in water chemistry that may fundamentally change the nature of the system, as in the case of acidification. Directed, long-term changes in nutrients, salinity, suspended solid load, and oxygen accompany gradual eutrophication of rivers caused by pollution. Long-term changes in water chemistry also follow changes to land use in the surrounding catchment such as afforestation or clearcutting (Giller & Malmqvist, 1998).

Within-river variation in water chemistry in space is a relatively well-known phenomenon. The concentration of most dissolved salts, nutrient levels and pH tend to increase from the river's source to its mouth. Changes in geology, soils, climate, vegetation, and in anthropogenic influence as one move from uplands to lowlands, also play a part. At a regional scale between rivers, geology and soils are the major factors influencing water chemistry, but local climate (especially rainfall patterns) and surrounding vegetation are also important (Allan, 1995; Giller & Malmqvist, 1998).

Solute Dynamics

The term "solute" refers to materials that are chemically dissolved in water. This includes materials such as calcium, chloride, sodium, potassium, magnesium, silica carbonate and more biological important solutes such as phosphate and nitrate. These solutes enter streams from three natural sources: the atmosphere (chloride, sodium, and sulfate from rainwater); soil and rock weathering (calcium, phosphate, silica and magnesium); and biological processes (nitrate from biological fixation by blue green algae) (Webster & Ehrman, 1996).

'Solute dynamics' refers to spatial and temporal patterns of solute transport and transfer. These processes are tightly coupled to the physical movement of water in streams. As materials cycle between the biotic and abiotic components of stream ecosystems, they are continuously or periodically transported downstream (Stream Solute Workshop, 1990). Primarily biochemical and hydrologic interactions occurring in whole watersheds as well as in-stream dynamics determine the dynamics of many solutes. The dynamics of a conservative solute are primarily driven by two processes: advection (down stream transport at the water velocity) and dispersion (molecular diffusion or turbulence). Dynamics of non-conservative solutes are more complicated because of the exchanges between solutes in the water column and on the stream substrate. These exchanges include abiotic processes (adsorption, desorption, precipitation and dissolution) and biotic exchanges (microbial uptake, plant uptake, leaching and mineralisation) (Webster & Ehrman, 1996).

Studies of solute dynamics in streams provide information on the rates of transport and transformation of the solutes themselves and quantification of various hydrological properties in streams.

Transport and storage of FPOM and CPOM

Fine particulate organic matter (FPOM) includes particles in the size range of $>0.45\mu m$ to $<1000\mu m$ that are suspended in the water column or deposited within lotic habitats. Suspended fine particulate materials include all living (e.g. bacteria, algae, protozoan, invertebrates) and non-living materials (amorphous organic matter, detritus, suspended organic sediment). FPOM can originate from many sources, including breakdown of larger particles by physical forces, animal consumption, microbial processes, flocculation of dissolved substances and terrestrial inputs

(Wallace & Grubaugh, 1996). FPOM functions as an important food resource for many filter-feeding invertebrates as well as for some vertebrates in river and streams (Wallace et al, 1991). The downstream transport of FPOM is also important to the theme of conceptualising streams as longitudinally linked systems (Vannote et al, 1980; Minshall, et al., 1985). Therefore, FPOM is important to many ecosystem processes as it represent a major pathway of organic matter transport and export.

Coarse particulate organic matter (CPOM) in streams is defined as any organic particle larger than 1mm in size. CPOM include components ranging from branches to entire trees that fall into stream channels and non-woody components donated by riparian vegetation (leaves, needles, fruits, flowers, seeds, frass) and materials produced within streams (fragmented aquatic plants, dead aquatic animals) (Lamberti & Gregory, 1996). CPOM is a major energetic resource for stream ecosystems. CPOM provides a large proportion of the fixed carbon in small streams and is important in larger streams (Cummins et al., 1983). CPOM that enters streams is transported downstream by the unidirectional flow of the lotic ecosystem. Trapping of this material is essential for the subsequent microbial colonisation that precedes consumption by shredding macroinvertebrates. These processes (retention) provide a critical link between input and the long-term storage and processing of CPOM (Vannote et al., 1980). The retentive capacity of streams for CPOM is a function of hydrologic, substrate related and riparian features (Lamberti & Gregory, 1996).

Stream Biota

Heterotrophic microorganisms (bacteria, protists, fungi) are important components of microbial communities, which function primarily as decomposers of dissolved (DOM) and particulate organic matter (POM) and are also consumed by higher trophic levels. An importance role of benthic bacterial communities is the assimilation of dissolved materials from overlying water. The ecological importance of these processes is that they result in the transfer of organic carbon associated with DOM, which is an important source of organic matter (Ward & Johnson, 1996).

Benthic stream algae are a ubiquitous group of photosynthetic organisms responsible for the majority of photosynthesis. Benthic algae are of fundamental importance to stream ecosystems. As organisms at the base of the food web, they are at the interface of the habitat conditions and biological communities. Photosynthesis by benthic algae provides oxygen for aerobic organisms in the ecosystem, and the fixed carbon provides food for algivores. Benthic algae enters to the food web through direct consumption from the substrata by macroinvertebrates such as snails or insects or through capture of drifting of benthic algae by filter feeders (Lowe & Laliberte, 1996). Many environmental factors interact to regulate spatial and seasonal growth and succession of phytoplankton populations, such as light, temperature, availability of phosphorous, nitrogen and silica, and dissolved organic compounds, which can influence phytoplankton metabolism by interacting with macro and micro-nutrients and influencing their availability (Wetzel, 1983).

Freshwater macroinvertebrates are ubiquitous; even the most polluted or environmentally extreme lotic environments usually contain some representatives of this diverse and ecologically important group of organisms. Most stream macroinvertebrate species are associated with surfaces of the channel bottom and other stable surfaces (fallen trees, roots, aquatic vegetation) rather than being routinely free swimming (Hauer & Resh, 1996). Macroinvertebrates play important roles within the stream community as a fundamental link in the food web between organic matter resources (leaf litter, algae, detritus) and fish (Hynes, 1970; Allan, 1995). Macroinvertebrate species composition changes between headwaters, middle reaches, and broader rivers, in response to changes in stream environment (Ward & Stanford, 1983).

The fish community is an assemblage of species inhabiting a prescribed area, that has the following properties: (1) richness, (2) diversity, (3) morphological and physiological attributes and (4) trophic structure (Li & Li, 1996). The number and kinds of species found can be ascribed to several ecological mechanisms. Physical tolerance to habitat quality (temperature, pH, dissolved oxygen, current, availability of substrata or cover) in the particular stream strongly affect membership in an assemblage of fishes (Matthews, 1987).

Biotic Interactions

One of the many advances in stream ecology in recent years is the increasing awareness of the importance of biotic interactions in the ecology of lotic organisms.

Stream Food Web

Stream food webs are essential for integrating studies of organic matter processing and community interactions. Food webs differ in structure and function among stream types, although they will all have some common elements. Most streams have approximately three or four trophic levels, but occasionally fewer or more may be present. Primary producers (including algae, bryophytes, macrophytes) and also detritus, occupy the lowest trophic level. There are groups of macroinvertebrates and some vertebrates (grazers and detritivores) apparently forming a primary consumer trophic level. However, macro consumers that feed on aquatic plants or plant detritus also ingest stream microbes and function as both primary and secondary consumers. Primary predators belong between level 3 and 4. Finally, predators that feed on other predators nearly always have mixed diets to include algivores and detrivores as well as other predators (Hersley & Peterson, 1996).

Species comprising stream food webs are constrained by many factors which then determine the structure and function of the food web of a particular stream, such as biogeography, geomorphology, substratum characteristics, gradient, riparian characteristics, temperature and inter-specific interactions (Cummins, et al., 1989; Ward & Stanford, 1982). The food web in any particular stream reflects all these factors, and among streams, a wide variation in food webs can be found.

Plant – Herbivore Interaction

Plants and animals interact in streams as they do in all ecosystems. Primary producers in streams consist of autotrophic bacteria, algae, bryophytes and vascular plants. The organic matter synthesised by primary producers in streams is a major energy source for benthic food webs. Herbivory (or grazing) is the consumption of living plants or their parts by animals. Herbivores have a major impact on plant assemblages in many streams, thus many structural and functional attributes of benthic algae are altered by grazers (Lamberti & Feminella, 1996). However, the strength and outcome of the algal-grazer interaction is also dependent on many abiotic factors such as light, nutrient, substratum, flow, season and disturbance (many works cited by Lamberti & Feminella, 1996).

Predator-Prey Interactions

Predator-prey interactions can have many types of effects on both predators and prey communities. In streams, the predominant predators are fish and some carnivorous macroinvertebrate species. Prey items include representatives of many orders of benthic macroinvertebrates. The effects of predators on prey populations depend on their predation rate compared to prey exchange rate (the rate at which prey moves in and out of areas where predation occurs) (Cooper et al., 1990). The impacts of predation on prey populations can be studied from two general perspectives: effects of predators-induced mortality in prey population and community, and consequences of anti-predatory behavior on prey fitness (Peckarsky, 1996).

Habitat Use and Competition

Species with similar morphology, life histories, and ecological requirements may coexist in many river and stream systems. The result is that there is potential for competition among these taxa. One of the important ways that such species coexist is through habitat partitioning. By exploiting different habitat or microhabitat patches at different times, potentially competing species can find opportunity to avoid competitive exclusion and thereby coexist (Connell, 1980). The differential use of habitats by closely related or similar species is also an important component of riverine biodiversity in that it promotes spatial complexity of biotic assemblages. Stream ecosystems are spatially heterogeneous, such as in the habitat diversity offered by polls, riffles and morphological features, or the convergence of flow velocity, depth, substratum, and temperature conditions that define different microhabitat patches within a single pool (Frissell & Lonzarich, 1996). Temporal heterogeneity, such as variations in flow or temperature over time, can also afford time-variant niches among which species are differentiated, therefore reducing or avoiding biotic interactions. Temporal variability in the environment can reduce competitive interactions between species and promote their continued coexistence regardless of overlap in their ecological niches (Connell, 1980).

Summary

Processes discussed in this section prove the high complexity and non-linearity of stream ecology. Streams and rivers are dynamic physical, chemical and biological entities, which interrelate with each other. Effects on biota are usually the final points of environmental degradation and pollution of streams (Noris & Thoms, 1999). Consequently, change in the health of a stream ecosystem will be reflected in the aquatic biological community. The biological communities that are exposed to pollutants act as integrators of the multiple past and present environmental effects (Cranston et al., 1996). Therefore using measurements of aquatic biota, to identify structural or functional integrity of ecosystems has recently gained acceptance for river assessment (Noris & Thoms, 1999). The concept of biological assessment of stream habitat is discussed in the next section.

2.1.2 Biological Assessment of Stream Habitat

As the community structure of an aquatic system is determined by conditions within a habitat (e.g. temperature, flow and salinity) and resources available (all things consumed by an organism), it is very sensitive to changes in these factors (Loeb & Spacie, 1994). The organisms that live in aquatic ecosystems are fundamental sensors that respond to any stress on that system, and only biological material could be used for adequate indication of spatial and temporal effects of chemical stressors in a river ecosystem (Cairns et al., 1993). Therefore, biological assessment is essential to assess the environmental health of aquatic ecosystems.

Advantages of Using Biological Assessment

Biological assessments are less time consuming than traditional chemical assessment as a single series of samples represents the sum effects of the prevailing conditions. In addition, animal and plant communities are not affected by temporary amelioration or usually by a transient deterioration of the effluent (Mason, 1996). Hynes (1960) also emphasized the advantages noted above of using biological assessment in polluted sites. Bioassessment can reveal long-term effects on ecosystems after the cause of the impact has passed and is itself undetectable (Ghetti & Ravera, 1994). Systematic biological monitoring and assessment is considered the most practical and cost effective approach to determine if human actions are degrading biological integrity (Davis and Simon, 1995). Such assessment provides both numeric and narrative descriptions of resource condition (Karr, 1998). Cairns et al. (1993) considered the role of the bioassay as a diagnostic tool for the restoration of desirable ecosystem conditions and as a predictive tool for preventing environmental impact.

For effective biological assessment, it is important to select organisms that will accumulate continuously over time. In addition, the group of organisms selected as a biological indicator should be widely distributed so that it is possible to compare the findings from one body of water to another (Patrick, 1994).

Biological Indicators

Indicators are environmental parameters selected and used in judging the degree to which specific environmental conditions have been changed or maintained (Cairns et al., 1993). Cairns et al. (1993) defined ***"indicator"*** after Hunsaker and Carpenter (1990) as "a characteristic of the environment that, when measured, quantifies the magnitude of stress to habitats, degree of exposure to stressor or degree of ecological response to the exposure".

Indicators are a shorthand description of aspects of an environment. Indicators are selected from a wide range of possible attributes and can be used singly or in combination to assess the conditions of the environment. They are key attributes, which give an impression of major trends and conditions (Walker et al., 1996).

Indicators may be used to understand the responses, adaptation and recovery of ecosystems and their inhabitants to both natural and anthropogenic disturbances (Johnson, 1995). Cairns et al. (1993) summarised the criteria for indicator selection. Ideal indicators should be:

- biologically and socially relevant;
- sensitive and broadly applicable to stressors;
- diagnostic of the particular stressors causing the problem;
- measurable, interpretable and integrative; and
- timely and cost effective.

To avoid confusion and errors in monitoring, biological indicators must also meet several requirements. Reliable biological indicators are taxa with narrow and specific tolerances. These indicators should be chosen considering the magnitude of the effects to be measured. They should not be applied in geographic locations for which they have not been designed (Cranston et al., 1996).

The biological indicator concept is well founded, shown by the commonplace observation that organisms reflect their environment. The concept of indicator species is of central importance in biological assessment. Richard *et al.* (1993) have defined "indicator species" as organisms which accumulate substances in their tissues in such a way so to reflect environmental levels of those substances or the extent to which the organism has been exposed to them. Good indicator species for freshwater quality management should have specific characteristics such as: be readily identified and easily sampled, have cosmopolitan distribution, readily accumulate pollutants, be easily cultured in the laboratory and have low variability (Hellawell, 1986).

Stream biota used as biological indicator

Benthic algae have a position at the interface of abiotic and biotic stream components and have many attributes that make them good organisms to employ in habitat condition assessment (Carbiener et al., 1990). Benthic algae are sessile and cannot swim away from potential disturbances. They must either tolerate their surrounding abiotic environment or die. Algal communities are usually species rich, and each species has its own set of environmental tolerances and preferences. Therefore the entire assemblage represents an information-rich system for environmental assessment. Algal identification is not difficult. Excellent taxonomic keys exist for the identification of benthic algae in most parts of the world (Lowe & Laliberte, 1996). The short life cycles of most stream algal species result in a rapid response to shifts in environmental conditions. Extant benthic algal communities are typically very representative of current environmental conditions, and indeed there is no better alternative group for studies of nutrient enrichment in open water (Hellawell, 1986). However, this advantage of algae for use as bio-indicators is a disadvantage in long-term assessment, when integration of present and past disturbances are required to assess habitat condition. Moreover, assessment methods based on aquatic plant communities are usually limited by the constraints on aquatic plant growth. The deepest parts of the water bodies, the areas shaded by trees, or those where flow velocity is too high, cannot be considered for ecological diagnosis, as vegetation growth is impeded (Amoros et al., 2000).

Fish communities are a highly visible and sensitive component of freshwater ecosystems. Fish provide several attributes that make them useful indicators of biological integrity and ecosystem health such as:

- Communities are persistent and recover rapidly from natural disturbance. Fish continually inhabit the receiving water and integrate the chemical, physical and biological histories of the water.

- They have large ranges and are less affected by natural microhabitat differences than smaller organisms. This makes fish useful for assessing regional and macro-habitat differences.

- Most fishes have long life span (from 2 - 10 years) and can reflect both long-term and current water resource qualities. The sampling frequency for trend assessment is less than for short-lived organisms.

- The taxonomy of fish is well-established (Simon & Lyons, 1995).

Fish communities respond predictably to changes in both biotic and abiotic factors (Karr 1981; Yoder & Rankin, 1995). Their characteristics have been used to measure relative aquatic habitat conditions (Simon & Lyons, 1995). However there are several difficulties in using fish to assess water habitat condition, especially in Australia:

- They are highly mobile and may often be migratory and therefore may be able to avoid exposure to adverse environmental conditions

- Water quality tolerances are poorly known for most Australian species

- The low diversity of fish in Australian waters means that few species are expected in any given river, reach or habitat (Cranston et al., 1996).

In addition, the use of the fish community in routine environmental surveillance is hampered by the necessity for extensive manpower and the difficulty in obtaining samples in deep, fast flowing rivers (Hellawell, 1986).

Cranston et al. (1996) discussed the possibility of using biotic communities as biological indicators of water quality. They assessed all possible biological indicators by 11 criteria and compared then with an ideal rating (Table 2.1). The results showed that macroinvertebrates offer the rating closest to the ideal, and they have been widely accepted to be included in the set of key indicators to assess stream habitat condition.

Table 2.1 Assessment of biological indicators of water quality (Cranston et al. 1996 [p.144])

Indicator	**Criterion**										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
Ideal rating	***H***	***L***	***H***	***H***	***H***	***H***	***H***	***H***	***H***	**-**	***H***
Mammals	M	M	H	L	H	M	H	L	M	G	M
Reptiles	M	M	L	L	M	M	H	M	M	G	M
Amphibians	H	L	H	L	M	M	H	M	M	G	M
Waterbirds	M	M	H	M	L	L	L	L	L	G	M
Fish	M	M	M	M	M	M	M	M	M	GD	L
Plants	H	L	M	M	M	H	M	H	H	G	M
Macroinvertebrates	M	M	H	H	H	H	H	M	H	GD	M
Biomarkers deformity	H	M	H	M	H	M	H	H	M	G	L
Biomarkers asymmetry	H	M	H	M	H	M	H	H	M	G	L
Bioassays	H	H	H	M	M	H	M	M	M	D	L

Selection criteria:
1. Ease to capture (High, Medium, Low)
2. Total cost/ha (H,M,L)
3. Standard methods available (H,M,L)
4. Interpretation criteria available (H,M,L)
5. Significant at catchment scale (H,M,L)
6. Low error associated with measurement (H,M,L)
7. Response to disturbance (H,M,L)
8. Stable over period of measurement (H,M,L)
9. Mappable (H,M,L)
10. Generic (G) / Diagnostic (D) application
11. Context data available (H,M,L)

Although there are some difficulties in selecting reliable indicators, the use of the biological indicator approach in aquatic ecosystems, in particular the use of macroinvertebrates, has received endorsement in biomonitoring programs (Rosenberg and Resh, 1993). The next section discusses the potential of macroinvertebrates as biological indicators for freshwater habitat condition and current biological assessment based on the distribution of macroinvertebrate assemblages.

2.1.3 Biological Assessment Based on Macroinvertebrate Assemblages

Macroinvertebrates and their significant role in stream ecosystem

Macroinvertebrates used in bioassessment are defined as those invertebrates that are retained by a 500-μm mesh sieve. Amongst the macroinvertebrates that fall below this size range many organisms are known or suggested to be valuable in aquatic biomonitoring. However, these invertebrates are not used in bioassessment because it is too difficult to identify them without a microscope (Cranston et al., 1996).

Macroinvertebrates are found in all streams and play crucial roles in organic matter dynamics and trophic energy transfer in stream ecosystems. As a group, they are the primary food source for most stream fishes. They also conduct important work of decomposing leaf litter and small particles of organic debris, and of grazing stream algae, fungi, and bacteria Functional importance of macroinvertebrates in aquatic ecosystems ranges from secondary producers to top predators. Therefore, their high diversity in streams reflects a variety of ecological and evolutionary processes (Hershey and Lamberti, 1998).

The study of macroinvertebrates is a central part of stream ecology. Earlier sections have focused on the multitude of interactive physical, chemical and biological variables that constitute the stream ecology. Geology, climate and other landscape features directly affect hydrologic patterns, and the movement and storage of inorganic and organic materials. Nutrients and the downstream transport of solutes are affected by channels and substratum complexity, the interactions of ground and surface waters, and by stream biota itself. Interactions between the stream channel, hyporheic zone, and riparian floodplains are important features in structure and function of the entire stream corridors (Stanford & Ward, 1993). These and many other factor affect the microhabitat structure of the stream and the distribution and abundance of stream macroinvertebrates.

Macroinvertebrates offer many advantages in bioassessment. They are ubiquitous and thus can be affected by environmental perturbation in different types of aquatic systems. The large numbers of species involved offer a spectrum of responses to environmental stresses. Their sedentary nature allows effective spatial analyses of pollutants or disturbance effects. Macroinvertebrates have variation in life cycles ranging from multivoltine (several generations per year) to merovoltine (two or three-year life cycle), which allow explanation of temporal changes caused by perturbation (Rosenberg and Resh, 1993; Hellawell, 1986). Because of that, macroinvertebrates act as continuous monitors of water they inhabit.

Relationship between Diversity and Environmental Disturbance

Variation in distribution and abundance of benthic macroinvertebrates may be caused by differences in flow rate among sites (Newbury, 1984), stream size and distance to the source (Minshall et al., 1985 cited by Linke et al., 1999), riffles and microhabitats (Robson & Chester, 1999), temperature and stream discharge (Boulton & Lake, 1992; Dinsmore et al., 1999), food resources and physicochemistry of the habitat (Townsend et al., 1997; Paradise & Kuhn, 1999). Altitude and slope have been found to be in correlation with invertebrate communities (Faith & Norris, 1989). Seasonal variability of such factors at a site is one of the prominent causes of temporal variation in the benthic macroinvertebrate community (Wade et al., 1989). Therefore, season should be explicitly taken into account in bioassessment (Linke et al., 1999). Many other aspects of the stream habitat condition affect the composition and abundance of stream macroinvertebrates. These factors include substratum, current velocity, dissolved oxygen and water chemistry (Hershey & Lamberti, 1998; Paradise & Kuhn, 1999; Dinsmore et al., 1999).

A ***disturbance*** is defined as a discrete event that disrupts the population, community, or ecosystem structure, often by changing resource abundance or physical environment (Resh et al., 1988). Effects of various types of disturbance on stream macroinvertebrates communities have been studied from many perspectives, including toxicants entering the stream, anthropogenic modifications of the channel, scour due to high discharge, drought, overexploitation of native fish species and introduction of exotic species (Hershey & Lamberti, 1998).

The responses of aquatic macroinvertebrate communities to environmental disturbances have therefore been incorporated into methods of bioassessment and biotic indices for the bioassessment of aquatic ecosystems. Typical observed responses to disturbance include increase abundance of certain species but general loss of diversity, especially with pesticide load or elevated nutrient level (organic enrichment) (Cranton et al. 1996). However, the intermediate disturbance hypothesis, as modified for streams, predicts the biotic diversity will be greatest in communities subjected to intermediate levels of disturbance. At low levels of disturbance, competitive interactions will result in lower diversity because of exclusion of species. High disturbance also will result in lower diversity because of exclusion of poor colonists or long –lived species. (Ward & Stanford, 1983)

Bioassessment of stream habitat using macroinvertebrates assemblage

Resh and McEltravy (1993) examined quantitative approaches used to study the effects of actual or potential disturbances on populations and communities of benthic macroinvertebrates and provided some suggestions to improve the role of biomonitoring in environmental assessment processes. Rapid Assessment is a very important application of macroinvertebrates in bioassessment. Rapid Assessment Biomonitoring is applied to identify water quality problems associated with both point and non-point source pollution and to document long term regional changes in water quality (Resh and Jackson, 1993). Using benthic macroinvertebrate community structures is a very fast and cost effective method in water quality monitoring by Rapid Assessment (Lenat & Barbour, 1994; Resh & Jackson, 1993; Resh, 1995).

Chessman (1999) presented a method that predicts macroinvertebrate community composition in flowing water from environmental data that has allowed pollution assessment from natural variability. The method uses a ***reference condition approach*** and predicts abundance of macroinvertebrates. The central idea of the referential approach is study biological relationship of sites in near pristine condition (reference condition) and then apply this relationship to predict the fauna at impacted sites if they were unimpacted *(sensu* Reynoldson et al., 1997*)*. It is based on the hypothesis that in the absence of pollution, river sites with similar natural environmental features will have similar macroinvertebrate faunas. The method showed a great distinction between human disturbed and undisturbed sites and high degree of correlation with physical and chemical indicators of human disturbance. However, the method worked with abundance, which can cause many difficulties in data collection, as abundance sampling is very subjective and thus data may have low reliability (Choy & Marshall, *pers comm*). Marchant et al (1995) found that patterns in macroinvertebrate communities were still evident when the taxonomic resolution was reduced from species to family level. Family level studies have been used successfully to describe biogeographical patterns across large areas (Corkum,

1989) and also to detect anthropogenic impacts on aquatic systems (Furse et al., 1984). Because of its simple conceptual basic and effective application, the taxon richness (the number of macroinvertebrates families) is concluded to be the most effective descriptor to use as the basis of a biocriterion for bioassessment (Bailey et al., 1998).

Assessment of river health involves comparisons. Indicators thought to represent river health are generally compared between sites that are thought to be similar in the absence of degradation (Norris & Thoms, 1999). A recent development in stream assessment has been use of the ***reference conditions***. These reference conditions serve as the control against which test site conditions are compared. The notion of a reference condition is really one of best available condition that could be expected at similar sites, and it is represented by several sites (Reynoldson et al., 1997). The reference condition is central to currently accepted ideas of "biocriteria" being developed by the US EPA (Davis & Simon, 1995). This approach is being used in Canada (Reynoldson et al., 1997; Bailey et al., 1998), in the UK (Wright, 1995) and Australia (Parsons & Norris, 1996) for stream assessment using macroinvertebrates assemblages. It involves testing an ecosystem exposed to a potential stress against a reference condition that is unexposed to such a stress. Several reference sites are sampled, and the variation among their communities represents the range of acceptable conditions. A test community falling outside of this range "fails", while a community that is within this range passes. The degree to which a test community falls outside the reference range is a measure of the magnitude of degradation at the site (Bailey et al., 1998). This reference condition approach is useful for estimating attainable conditions for evaluating temporal and spatial changes in ecological integrity, and for setting biological and environmental criteria (Hughes et al, 1986)

Predictive modelling based on habitat characteristics is central to many applications of the reference condition approach. Bioassessment uses predictive modelling to explain variation in reference communities considering the environmental conditions at these sites as predictors. The next sections discuss a computational approach to support bioassessment based on information about environmental and biological condition of sites.

2.1.4. Statistical Model to Support Bioassessment

Stream modelling based on stream monitoring data is a next step in studying and predicting the characteristics of stream habitats. Successful modelling based on biological factors of streams can improve bioassessment results. The suitability of predictive models for assessment programs is dependent upon their ease of application and practicality in providing management information at minimal cost and effort (Parsons & Norris, 1996).

Many studies require statistical analysis or even more complex numerical analysis to draw generalities and to detect and highlight patterns or trends in complex data set comprised of many variables in order to apply the data in further studies. Norris and George (1993) evaluated statistical analytical approaches used in data processing. Major methods of statistical modelling described in their study include Analysis of Variances (ANOVA), multiple regression (MR), Discriminant Function Analysis (DFA) and time series analysis. They are all very powerful tools for developing predictive models and associating physical, chemical and biological data together. ANOVA is applied to compare and partition total variability into components of the study. It depends on replicated sampling. Multiple regression is a continuous statistical approach used to examine relationships between biological measure and various environmental factors. DFA is an appropriate method that can be used to investigate the relationship between groups established from macroinvertebrate fauna and environmental variables. Time series analysis may be used to develop a predictive model based on variation in past time series.

However, these methods of statistical analysis often have stringent requirements of data, such as replicated collection of data, normally data distribution or high frequency of data collection. Some requirements are difficult to meet so simplified assumptions must be used in working with these methods. These assumptions and data requirements restrict the capability of statistical methods to cope with the non-linearity and complexity of water ecosystems. Statistical methods tend to minimize non-linearity in the processes. They are simple to implement if the relationships with variables are linear. If they are non-linear, transformation into linear becomes a major limitation of statistical methods in working with non-linear relationship of variables in the aquatic system (Lek *et al.*, 1996 and Paruelo *et al.,* 1997). In addition, each of the statistical methods mentioned above could be applied to certain problems in data processing, but integrated application of all the methods to work with a complex database may cause problems in implementation. Therefore, statistical methods appear to have restricted capacity for modelling complex and non-linear river quality data.

In order to overcome the limitation of statistical models, efforts have been undertaken resulting in some significant successes in this field. Welsh et al. (1996) used a statistical model but provided a method for identifying important environmental variables and constructing appropriate intervals to predict mean value of animal abundance. Reckhow (1993) applied a random coefficient to the same cross sectional data set to produce water body specific parameters rather than a single set of global parameters. It improved the results of the statistical models for chlorophyll *a*- nutrient relations in working with spatial variation. Yang *et al.* (2000) applied two-dimensional spatially distributed water quality data derived from the SPOT satellite to support one-dimensional water quality models (QAL2E) in estimating algal growth rate and respiration rate in a water ecosystem. However these supporting methods often make statistical models become too complicated and difficult to be applied in operational water quality assessment and management.

Recknagel (1989) and Recknagel *et al.*(1994) applied an expert system to water quality management. An expert system consists of two parts: a software product or expert system shell which contains the code handling the knowledge base and the knowledge base covering a set of rules for a specific problem (Straskraba, 1994). The Lake Ecosystem Model SALMO (Recknagel, 1989) is an example of a non-autonomous deterministic model. Although SALMO had achieved significant results applying some simplifying assumptions, it is unlikely that it would be effective with river systems, which are different in nature compared to lake ecosystems.

A widely used model in supporting bioassessment is simulation modelling, which can be done using STELLA software. Fischer (1994) successfully applied this model to study prey-predator relationships in order to control overcrowding in water bodies. His model consisted of prey submodels representing the population dynamics and growth of prey and predator submodels representing the predation process, population dynamics and growth of predators. Nevertheless, this was a mechanistic simulation model based on theoretical ecological and biological knowledge, which does not respond to all processes occurring in the studied systems. The main problem in working with structural dynamic models is that it is very difficult to obtain sufficient data to develop the models. Another problem is that the models do not reflect the real properties of ecosystems, particularly their adaptability and ability to meet changes in forcing functions with changes in species composition (Jorgensen, 1999).

Wright et al. (1984) first presented the results of a the project on British rivers that had two major objectives: to develop a biological classification of unpolluted running water sites and to assess if the type of macroinvertebrate community at a site may be predicted using physical variables. In this project the sites were classified by two-way indicator species analysis (TWINSPAN), and a multiple discriminant analysis (MDA) was employed to predict the group membership at sites using 28 environmental variables. The approach was found to be useful to the classification of running water sites by their macro-invertebrate fauna and the prediction of community type using environmental variables.

Based on this approach, Moss et al. (1987) developed a statistical model for predicting macroinvertebrates occurring at some stream sites in Great Britain. The model worked with probabilities of the occurrence of macroinvertebrates and provided reasonable results in prediction using environmental variables. The procedure is of practical value in the detection and assessment of pollution. However, they also acknowledged that the proposed applications did not provide an explanation for the macroinvertebrates response to environmental conditions. This is caused by the limitation of the applied discriminant analyses. If these techniques are used for explanatory purposes, a number of assumptions needs to be met including a jointly normal distribution of explanatory variables, equal covariance matrices amongst the groups being discriminated between, and accurate estimates of the prior probabilities of group membership.

Based on works of Wright et al., 1984; Moss et al. 1987; Wright, 1995; Wright et al., 1998, a software package have been developed in the British Institute of Freshwater Ecology's called RIVPACS (River Invertebrate Prediction and Classification System). RIVPACS has been applied on a nation wide scale to assess the biological quality of rivers and streams in the United Kingdom.

The system predicts the site specific macroinvertebrate fauna to be expected in the absence of major environmental stress (Wright, 1995; Wright et al., 1998; Moss et al., 1999). The statistical techniques used for RIVPACS are TWINSPAN classification of the reference sites based on their macroinvertebrate assemblages, followed by multiple discriminant analysis (MDA) of the resulting groups of sites using a limited number of environmental variables. Prediction of the fauna at a test site was achieved through MDA, leading to calculation of probabilities of capture of individual taxa based on the prediction of group membership for the test site (Moss et al., 1987).

In Australia, a similar predictive model called Australian River Assessment Scheme - AusRivAS was developed to use aquatic macroinvertebrates to assess the habitat condition of Australian rivers and streams (Schofield & Davies, 1996). AusRivAS models are based on RIVPACS, which also assess habitat condition in a river by predicting the macroinvertebrates families expected in the absence of environmental stress, such as pollution or habitat degradation (Coysh et al., 2000). Predictions are derived from a set of environmental measurements used to characterise the site. A predicted macroinvertebrates assemblage is compared with the actual assemblages and the ratio of observed/expected (O/E) families is used as a measure of ecological condition (Wright et al., 1984; Parsons & Norris, 1996; Marchant et al., 1997; Smith et al., 1999). There are two major differences between AusRivAS and RIVPACS. Firstly, macroinvertebrates are only identified to family level in AusRivAS. Second, major aquatic habitats (channel, riffle etc) are sampled and processed separately in AusRivAS (Smith, et al., 1999). The rationale behind habitat – specific sampling is that each habitat has a distinct macroinvertebrate community and within a given region, differences among habitats are greater than differences between sites. Unless comparisons between sites are based on the same habitats, they may be confounded by the occurrence of different habitats at each site (Parson & Norris, 1996).

The modelling approach for AusRivAS was similar to that of RIVPACS. Model building occurred in five steps. First, reference sites were classified into groups with similar macroinvertebrate communities using an agglomerative hierarchical fusion technique, Unweighted Pair-Group arithMetic Averaging (UPGMA). Second, once the optimal classification was chosen, stepwise discriminant function analysis (DFA) was used to identify which environmental variables best discriminated between groups in the classification. Third, the DISCRIM procedure in SAS statistical package was used to incorporate predictor variables into a discriminant function and assign sites to groups identified in the classification. Fourth, the probability of each family occurring at each site was calculated by multiplying the probability the probability of site belonging to a classification group by the probability of family occurring in that group and then summing the products to give the number of families expected (E). Fifth, using a preliminary model, O/E ratios of reference sites were calculated. The O/E score itself was used as a measure of impact at disturbed sites, with lower scores indicating greater disturbance. (Simpson et al., 1997; Smith et al., 1999; Coysh et al., 2000).

The AusRivAS model had been applied to study the effect of habitat- specific sampling on biological stream assessment for Australian Capital Territory (Parson & Norris, 1996), to classify macroinvertebrate communities across drainage basins in Victoria (Marchant et al., 1999), and to assess ecological condition of rivers in Western Australia (Smith et al., 1999). Even though the applications achieved some valuable success, some constraints appeared to have caused confounded assessment of biological impairment. Although statistics can be used to validate metric choices and predictions while building multimetric index, excessive dependence on the outcome of statistical tests can obscure meaningful biological patterns. A narrow focus on probability values (P-value) rather than on biological consequences limits the value of biological assessment. Dependence on narrow statistical approaches overlooks the fact that a statistically significant result (small P-value) may not equate with a large important effect, as researchers often assume; similarly, a statistically insignificant effect (large P- value) may well be biologically important (Karr, 1999).

Investigation of the RIVPACS classification based on statistical methods revealed that the composition of a few of the classification groups was less than optimal and could adversely affect the performance of parts of the prediction system (Wright et al, 1991). Moreover, the RIVPACS and AusRivAS statistical approach may be more difficult to apply to sites where environmental conditions are extreme or highly unpredictable and in consequence the biota are more difficult to document or show substantial year by year variation (Wright, 1995).

These constraints are caused by the assumptions and limited implementation of statistical methods in dealing with high non-linearity of stream data. As new computational techniques are becoming widely available, a number of alternative ordination and classification procedures are now being examined to determine whether a new procedure can deliver more reliable predictions (Wright, 1995).

2.1.5. Summary and Research Needs

Freshwater streams have a highly complex nature including distinct nonlinear processes over time and space. This nature makes habitat conditions in streams extremely difficult to assess and predict. Traditional methods often fail to cope with all these variations. Bioassessment with its conceptual advantages, reliable elucidation and practical implementation have proved to be a suitable alternative. Among indicator species living in streams macroinvertebrates provide many valuable characteristics and they are widely accepted to be a key indicator to assess stream habitat conditions.

Success in applying bioassessment in habitat condition management can be enhanced with the support of predictive modelling. Statistical and other mathematical and numerical models have to some extent been used successfully to support bioassessment in water quality. However, these models are still constrained by not being able to deal with the non-linearity and high complexity of stream ecology. New computational techniques are needed to find the way to overcome these difficulties. A new generation of models for bioassessment of freshwater streams may arise from application of machine learning techniques to be a potential alternative tool.

2.2 Artificial Neural Networks to support bioassessment

2.2.1 Artificial Neural Networks - An Introduction

Artificial Neural Networks are one of the most important applications of Machine Learning techniques. Machine learning is a subset of Artificial Intelligence, a branch of computer science that is concerned with the automation of intelligent behaviour. Machine learning focuses on knowledge acquisition by various automated induction techniques. Machine learning has proved to be a fruitful area of research, spawning a number of different problems and algorithms for their solution. These algorithms vary in their goals; in the available training data and in the learning strategies and knowledge representation languages they employ. However, all of these algorithms learn by searching through a space of possible concepts to find an acceptable generalisation (Luger and Stubblefield, 1992).

Kompare et al. (1994) showed by that using advanced machine learning techniques and general basic knowledge on ecosystems, it is possible to automatically generate better models and in less time than is the case by traditional model construction. Machine learning models have the ability to extract temporal or spatial patterns and knowledge from highly nonlinear and complex data. Based on such patterns and knowledge they can predict future conditions. Machine learning reduces to a great extent the need to query the expert in the way that computer extracts knowledge from the given data. It is able to identify and model a real world system that we do not fully understand yet.

As the significant application in this field, Artificial Neural Networks offer inductive approaches to model building. They are highly connective and simulate principles of natural evolution and knowledge discovery in large databases.

ANNs are non-linear mapping structures based on the function of the human brain. They are considered universal and highly flexible approximators for any data and are powerful tools for ecological modelling, especially with high non-linearity occasions when the data relationship are unknown (Lek & Guegan, 2000). They do not require assumptions about mathematical relationship between state variables and the nature of the distribution of data. All neural networks have in common the ability to learn from data. ANNs can identify and learn correlation between input data and corresponding target values. After training, ANNs are able to predict the output of new independent input data. ANNs may be broadly classified according to whether they learn in a supervised or unsupervised way (Waley and Fontama, 1998; Bishop, 1995).

A neural network learning model consists of two primary components: the topological structure of neural networks and an associated learning rule (Adeli and Hung, 1995). The backpropagation learning is one of the supervised learning methods.

Backpropagation (BP) networks based on the supervised procedure are preferred in ecological modelling, especially in water quality modelling. The architecture of the BP network is a layered feed forward neural network, in which the non-linear elements (neurons) are located in the hidden layer. The neurons feed a non-linear function by the sum of their inputs coming either from input nodes by feed forward or from output nodes by feedback. Neural networks determine the weighted connectance between the input and output nodes by these neurons (Recknagel et al., 1997; Recknagel et al. 1998, Lek et al., 1999).

Backpropagation is an algorithm for apportioning the error responsibility through a multilayered network. The neurons in a backpropagation network are connected in layers, with units in layer k passing their activations only to neurons in the layer k+1. In solving a problem, activation passes from the input units, through one or more internal layers of neurons (hidden layer) and ultimately passes to the output layer and the environment (Luger and Stubblefield, 1992).

Given the correct results, the network may calculate the error in the output units just as it did for a single-layer network. The error for a neuron in the layer directly below the output layer is a function of the errors on all the units that use its output. In general, the error for a neuron at layer n is a function of the errors of all neurons at layer n+1 that use its outputs. In a BP network, activation moves backward in a similar fashion (Luger and Stubblefield, 1992). Once BP has computed the error for each neuron in the network, the individual units may learn by applying the delta rule, the amount of learning is represented as the difference (delta) between the desired and computed outputs (Adeli and Hung, 1995).

These multi-layer artificial neural networks can learn in more complicated learning domains than those lacking hidden units. The feedforward net with BP of error has been found to be an effective learning procedure for classification problems (Rumelhart et al., 1986)

Maier and Dandy (1996) compared ANNs to statistical ARMA (Auto Regressive Moving Average) class of models widely used for modelling water resources time series in terms of advantages and disadvantages. They found that ANNs are more flexible in working with complex non-linear system and in providing long term forecasting. Similar comparisons between ANNs and other classes of statistical modelling provided by Lek et al. (1996) and Paruelo and Tomasel (1997) also emphasized the flexibility of ANNs.

Artificial neural networks thus bring an excellent alternative tool for analysing ecological data and for modelling thanks to their specific features of non-linearity, adaptivity through learning from samples, generalization and model independence (Schleiter et al., 1995).

2.2.2 Application of ANN to modelling ecosystem

ANNs have been applied to various fields of aquatic sciences and engineering, especially in modelling habitat condition. Modelling freshwater habitat condition is extremely difficult, as the interrelations between various influences are not known. Hydrodynamic models are difficult to couple with chemical and biological models. The action of hydrological process on ecological processes has hardly been elucidated, as the requirements are different for both systems (Straskraba and Gnauk, 1985). The use of ANN may overcome many of these difficulties. Unlike deterministic modelling, which is based on known theories and equations, ANN uses the measured data to determine relationships. Therefore, the problem of producing models that can address unidentified interactions and combine hydrodynamics with ecological processes is clearly possible using ANN.

Maier and Dandy (1996) used ANN as a viable means of forecasting salinity in the River Murray (South Australia) 14 days in advance. The results obtained had less than 7% average absolute percentage error. It was concluded that, ANN models appear to be useful tool for forecasting salinity in rivers.

Recknagel et al. (1997) applied ANNs to the task of modelling and prediction of algal blooms and to identification of the variables that play a major role in algal growth. In their study, major ecological factors of all chemical physical and biological categories, which could clearly define the environmental conditions of the aquatic system, were included as input variables and five dominating phytoplankton species were used as output variables. The resulting predictions on succession indicate the ability of ANNs to fit the complexity and non-linearity of complicated ecological phenomena. If an expanded database is available, not only a specific aim can be investigated but also cost-benefit strategies for management can be addressed applying ANN to scenario and sensitivity analysis (Recknagel, 1997; Recknagel et al., 1998).

ANN had been applied very successfully to eutrophication processes. Research has been done in Italy (Scardi, 1996), Japan (Yabunaka et al., 1997), and Turkey (Karul et al., 2000). Models used physical and chemical parameters and also biological variables as inputs to predict the behaviour of chlorophyll – a and other typical eutrophication indicator. The studies showed that nonlinear relationships in the eutrophication phenomenon could be modeled reasonably well. The ANN model can also estimate an extreme value that lies outside the boundaries of the training set. Conclusions were made that ANN models can be used to estimate the densities of certain species as functions of environmental parameters.

Wen and Lee (1998) applied ANN to the problem of optimising water quality management in a river basin. Their study focused on the objectives of environmental quality, treatment cost of wastewater and the assimilative capacity of a river to provide a solution to water quality management problems. The results of their work show that using the backpropagation algorithm and feed forward neural network, a multi objective programming model can simulate the decision makers' preferences and successfully overcome the disadvantages of unknown preferences of decision makers.

Recknagel and Wilson (2000) discussed the potential of ANN models in working with aquatic ecosystems. They compared presentations of 6 prototypes of inductive and deductive models for phytoplankton including a regression model; time series model; deterministic models for functional algal group succession and algal population; heuristic model; and ANN. The result of comparisons showed that only ANN provides an ability to predict both timing and magnitudes of species dynamics and species succession in the lake. ANN models can support both prediction and elucidation of ecosystem behavior with the potential to provide new insight into mechanisms of systems from the results.

Maier *et al.*(1998) used ANNs for modelling the incidence of cyanobacteria in rivers by forecasting the occurrence of a species group of *Anabaena* in the River Murray, Australia. ANNs provided a good forecast of both the incidence and magnitude of a growth peak of cyanobacteria within the limit required for water quality monitoring. The models also defined predominant variables in determining the onset and duration of cyanobacteria growth.

Lek-Ang *et al.*(1999) developed predictive modelling of *Collembolan* diversity and abundance on a riparian microhabitat scale. Biological variables that were retained to describe its structure in this model included abundance of dominant species, species richness and biological indices. In the input layer, the main environmental variables were considered. 80% samples were chosen randomly for the training process and the remaining 20% were used for model validation. The resulting habitat profiles illustrated the complex influence of each variable on the biological parameters of the assemblage and also the non-linear relationship between dependent and independent variables. The study gave satisfactory results over practically the whole range of values of dependent variables, which showed ANNs potential to predict biodiversity and structural characteristics of species assemblages.

Gozlan et al (1999) applied ANN with the aim to predict the abundance of six fish species in the river Garonne with back propagation as learning algorithm. The ANN was successful in predicting the abundance of 0+ fishes on a microhabitat scale, indicating that technique merits more frequent use in ecology and biodiversity studies. The explanatory part of the analysis, coupled with the predictive power of ANN, should facilitate the ecologically oriented management of aquatic ecosystems, providing that the duration of the study is extended.

In summary, ANNs based on machine learning techniques have proved to be powerful tool in bioassessment of aquatic ecosystems and stream habitat condition and have been applied worldwide in this field.

2.2.3 Current Achievement in Application of ANN to Assessment of Habitat Conditions using Macroinvertebrate Assemblages

Walley and Fontama (1998) firstly reported a successful application of ANN in prediction of macroinvertebrate taxa in unpolluted river sites and compared with the performance of RIVPACS. The objectives of predictions were average score per taxon (ASPT) and number of families presents (NFAM). Models were based on the standard backpropagation networks. The results showed that the ASPT model achieved a significantly higher level of performance in independent test data than the NFAM model. Results of their study demonstrated the ability of ANN in training with values of biological indices and understanding the relationship between environmental variables and biotic indices that is often a very complicated and non-linear problem. It was concluded from study that the neural networks performed marginally better than RIVPACS. They also discussed further improvement to the performance of neural network by extending the environmental data to include relevant catchment characteristics.

Schleiter et al. (1999) went one step further to model the population dynamics of macroinvertebrates in German streams using ANN. They tested the suitability of ANN for system analysis and impact assessment: (1) in temporal dynamics of water quality; (2) in bioindication of chemical and hydromorphological properties using benthic macroinvertebrates; (3) and long-term population dynamics of aquatic insects. The satisfactory results of the study showed that ANN can meaningfully be used in the analysis of effect-relation of species, including the identification and assessment of complex impact factors, and also for forecasting system behaviour which have specific, very complex and non-linear features. However, they admitted that as ANNs learn from examples, their quality depends heavily on the representativeness and compatibility of the database.

Chon et al (2000a) applied Artificial Neural network to classify and predict multivariate stream data even in a short period using benthic communities. This study demonstrated that temporal ANNs could be utilised to forecast and analyse short-period changes in multivariate data sets. The recurrent neural network appeared to be effective in patterning development of benthic communities in streams responding in a diverse manner to a wide range of pollution. The study also showed the advantage of specific forecasting for an individual taxon is that it could assist to characterise community changes.

Obach et al (*in press*) modelled the total amount of individuals of selected water insects based on a 30 years data set of population dynamics and environmental variables in stream in Central Germany using Kohonen self-organising maps in combination with some other types of neural networks. Results were interpreted on the basic of known species traits. The conclusion was made from the study that it is possible to predict the abundance of aquatic insects based on relevant environmental factors using Artificial Neural Networks.

Spatial analysis of stream invertebrate distribution in the drainage basin had been studied (Cereghino et al., 2000). The study provided a stream classification based on characteristic EPTC *(Ephemeroptera, Plecoptera, Trichoptera, Coleoptera)* insect assemblages at species level. The main interest of their results is that the stability of these theoretical assemblages may be used to refine representative and/or reference sites for biological surveillance, as a change in species composition within a given region can be considered as a biological indicator of environmental changes.

Pudmenzky *et al.*(1998) developed preliminary ANN models for predicting the distribution of macroinvertebrates in the Queensland stream system based on environmental variables. The network was trained with both categorical and continuous attribute input data. The ANN proved promising in predicting the taxa, which had the most even equal distribution of presence/absence (probability of occurrence around 0.5). As work had been done with a shareware version of the software package, only a subset of the data could be investigated. However, this is the first work done in applying ANN to biological assessment of habitat condition in Australia. Further research is highly recommended to investigate the possibility of ANN as computational alternative to AusRivAS in supporting bioassessment of habitat condition.

2.2.3 Research Needs

ANN hasproved to be a very effective approach to support bioassessment of habitat condition and had been applied all over the world with remarkable success. In Australia, anthropogenic effects on streams and rivers have resulted in considerable physical, hydrological and morphological changes in aquatic ecosystems (Smith et al., 1999). In response to declining conditions in Australian rivers, National River Health Program (NRHP) was established in 1992 with the aims to monitor and assess the ecological condition of Australian river and stream; to assess the effectiveness of current management practices; and to provide better ecological and hydrological data on which to base management decisions (Schofield & Davis, 1996). AuRivAS is a national bioassessment program that uses aquatic macroinvertebrates to assess the health of river and stream systems. AusRivAS uses statistical models as a prediction tools and exhibits some constraints in dealing with non-linearity and complexity of freshwater ecosystem. Artificial Neural Networks, which had been studied and applied in many areas all over the world with promising success, can be applied as alternative computational tool to the AusRivAS model.

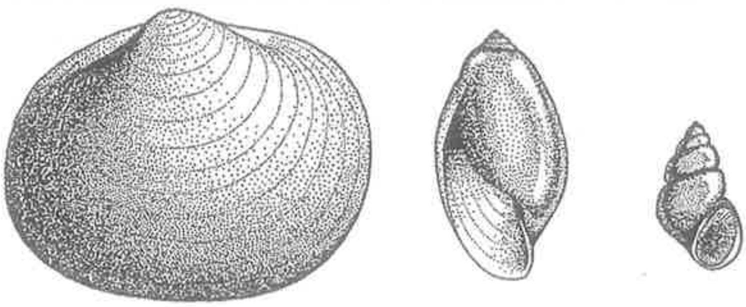
2.3 Summary and Thesis Aims

In conclusion, streams and rivers are very complex ecosystem with many processes in close interrelations. River health assessment is a way of examining waterways using tools such as water quality, habitat description, biological monitoring and flow characteristics to create an overall picture of the ecological health of that waterway.

Bioassessment of freshwater stream habitat is an effective method to obtain an accurate picture of condition or health of a waterway. Among indicator species used for bioassessment, macroinvertebrates prove to be very appropriate for use in studying stream and river habitat condition. Success in applying bioassessment in freshwater management can be enhanced with the support of computer modelling, especially using artificial neural network techniques.

Artificial Neural Networks had been applied in the field of habitat condition assessment using macroinvertebrate assemblages in many countries for a long time but no such of project had been done for Australian stream systems.

The goal of my research is to follow the preliminary study of Pudmenzky et al (1998) to apply Artificial Neural Networks for the bioassessment of stream ecosystem. The aim of this study is the development of an ANN model to predict habitat conditions in Queensland river systems based on environmental variables and colonisation patterns of 40 most common macroinvertebrate taxa. The predictions are based on a comprehensive database, which was previously subject to a preliminary case study by Pudmenzky et al. (1998). Beside the prediction, I will also test the elucidation capacity of ANN to explain the processes in freshwater ecosystems and to find out dominant factors affecting distribution of macroinvertebrates within the system.



Molluscs of solid objects and weed beds of depositing substrata in running water (Hynes, 1960).

3 General Material and Methodology

3.1 Introduction

This work investigated the application of an Artificial Neural Network model to predict the habitat conditions in the Queensland stream system. The project was aimed at determining if ANNs could be used as basic for a Queensland model. The work had been done based on a comprehensive database from Queensland Department of Natural Resources (QDNR) containing information taken from wet and dry seasons on water quality, habitat characteristics and occurrence patterns of macroinvertebrates for over 500 stream sites. This database was previously subjected to a preliminary case study by Pudmenzky et al. (1998). Different combinations of data had been studied and used for model development. This chapter discusses the structure and characteristics of data used for network development.

Computational approach had been applied to analyse the relationship between environmental variables and stream assemblages. The models had been developed by mean of Artificial Neural networks. Fundamental concept and method of modelling also are discussed in the chapter.

3.2. Study Sites and Site Selection

The Queensland river and stream network spreads over the territory of the federal state of Queensland (Australia). A diverse range of climatic conditions occur over the state, ranging from high rainfall area (1600 mm /annual) in the tropical Northeast to low rainfall area (200 mm/ annual) in the Southeast. Study sites are spread throughout the catchments of most major and many minor Queensland rivers.

The majority of sites in the stream database are situated in relatively high order streams of coastal lowland areas (see Fig. 4.1).

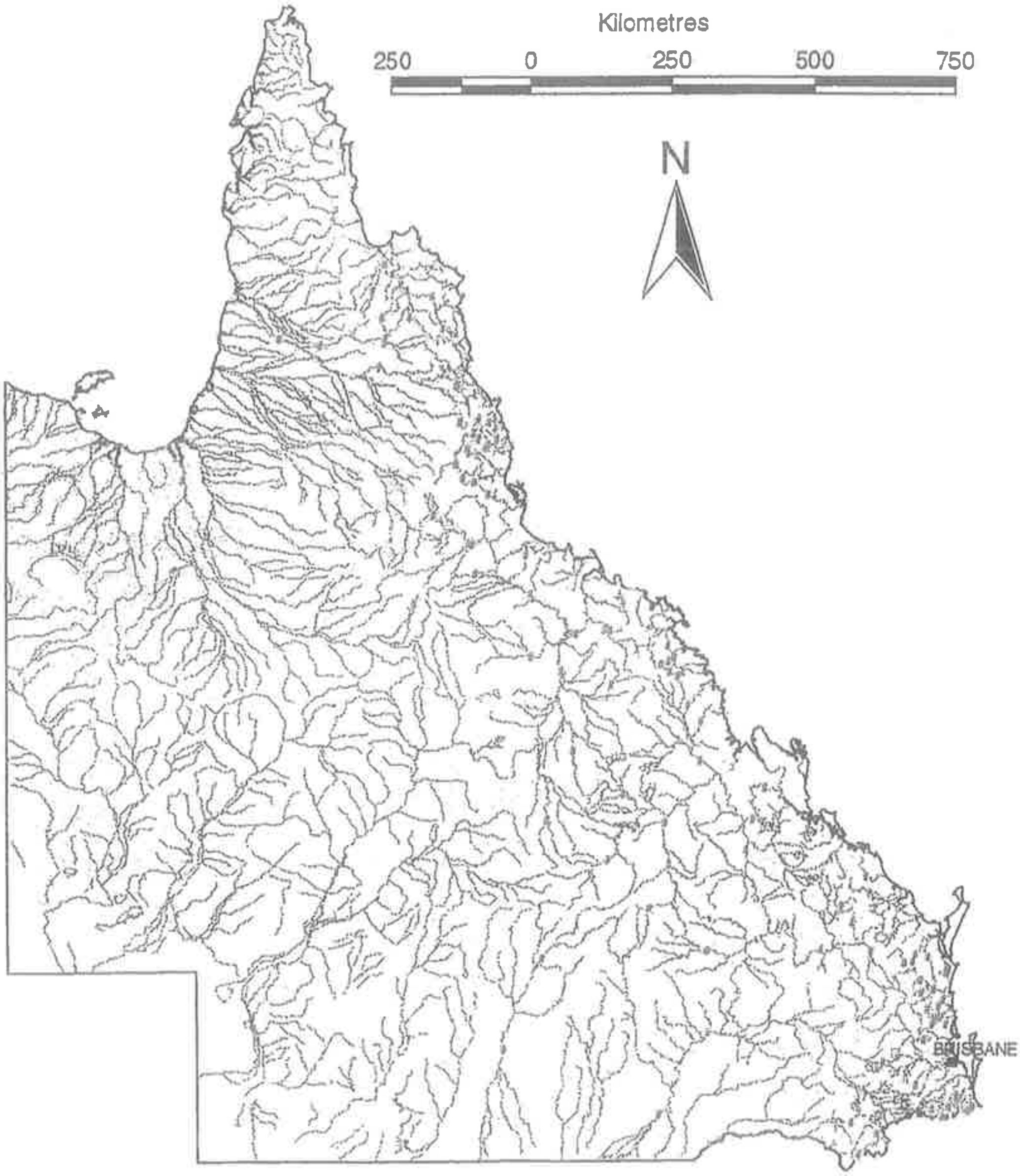


Figure 3.1 Map of Queensland indicating location of reference sites of the Queensland stream database

In this work the concept of reference site and test site had been used. Reference and test sites in Queensland were initially selected for the MRHI (Monitoring River Health Initiative) program using protocols outlined in the River Bioassessment Manual. Reference sites were those in near pristine condition. Test sites were those experiencing an impact from water quality or habitat degradation. QDNR currently uses a list of the 10 selection criteria to determine whether or not sites are in reference condition (Conrick & Cockayne, 2000).

Table 3.1 Selection criteria used to determine reference site condition (Conrick & Cockayne, 2000)

No.	Reference condition Selection Criteria	Level of Impact
1	No intensive agriculture within 20km upstream	
2.	No major extractive industry (current or historical) within 20km upstream	
3.	No major urban area (>5000 population) within 20km upstream	
4.	No significant point source waste water discharge within 20km upstream	
5.	No dam or major weir within 20km upstream	
6.	Seasonal flow regime not greatly altered	
7.	Riparian Zone of natural appearance	
8.	Riparian Zone and banks not excessively eroded beyond natural levels or significantly damaged by stock	
9.	Stream Channel not affected by major geomorphological change	
10.	Instream conditions and habitat not altered	
	SITE ASSESSMENT	**/30**

For all sites each criterion was assessed with the level of impact given from 1 meaning highly impacted to 3 with no/little impact. These levels were then summed to give a total site assessment out of a maximum of 30. If the site assessment score was <26 the site was considered as a test site: assessments ≥ 26 were considered as a reference site. Criterion 5 is crucial. Site failing to have criterion 5 score of 3 automatically fails the overall assessment.

Database contains information about habitat characteristics of 896 samples taken from reference sites and 1159 samples from test sites. Different combinations of data are used for training, validating networks and for testing network in prediction step.

3.3 Habitat Condition Database – A review

Habitat condition database contains information about habitat characteristics, water quality and colonisation pattern of macroinvertebrate in each site. For each site data set contains three sub-sets, which are discussed below.

3.3.1 Physical Riparian and Other Predictor Variables

Physical habitat data were collected from longitudinal profile. Potential predictor variables that are environmental variables that are relatively stable under the influences of human impacts. They are used for developing network based on reference condition approaches (Parsons and Noris, 1996; Simpson et al, 1997). Chemical variables such as dissolved oxygen, pH, nutrient concentration are often affected by anthropogenic impacts and they would provide spurious prediction if used to predict the membership of test sites to the reference site groups.

Data of habitat characteristics included 39 potential predictor variables consisting of discrete categorical or continuous data. Only some discrete categorical variables were formed by classification schemes such as stream order, most of them were represented just as empirical criteria for habitat characteristics such as soil types and vegetation type. These 39 potential predictor variables were used as input variables of the ANN models. Predictor variables for network development mainly belong to 3 categories: geographical, topographical and meteorological.

(1) Geographical

Latitude (S), Longitude (E), Altitude (m): Geographical information about location if site. Obtained by using GPS (Global Positioning System) and confirming readings on a 1:100 000 topographic maps.

Stream Order: Hierarchical-ordering system based upon the degree of branching (Strahler, 1957). Stream orders were determined using 1:100 000 scale maps. A second order stream is formed by the joining of two first order stream; the junction of two second- order stream form a 3^{rd} order stream etc. (Figure 3.2.)

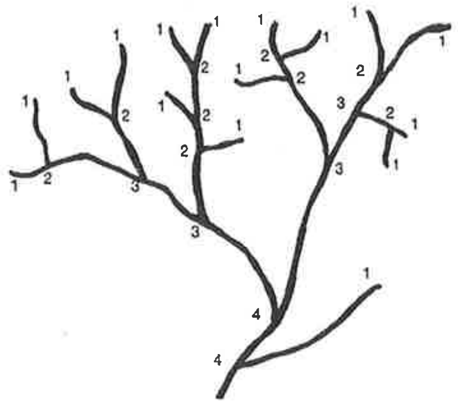


Figure 3.2 Method used by QDNR to determine stream order (Strahler, 1957)

Distance from Source (km): Distance from the site to longest thread of stream source.

0-2 Reach: An assessment of where in the catchment a site lies with relation to the watershed. This is a categorical variable. 0: lower, 1: middle and 2: upper.

(2) Topographical

Habitat 1-5: Categorical variable describes a predominant habitat type at the study site. Habitat types are prescribed because each habitat has a potentially distinct fauna. The performance of the predictive models will therefore not be confounded by differences in habitat availability between sites and time. In Queensland, five habitats most likely to be encountered are:

1. Riffle: This is a reach of relatively steep, shallow (<0.3m), fast flowing (>0.2m/s) and broken water over stony beds.

2. Edge/back water: edges are along the bank where there is little or no current and extend to approximately 0.5m from the bank. There may be some terrestrial vegetation, tree roots or the area may be bare. A backwater is a zone where the bank indents and a pool of water forms away from the main channel. The backwater may have a circular or back flow, and a silty bed with accumulated plant litter.

3. Run: This is a reach of relatively deep and fast flowing, unbroken water over a sandy, stony or rocky bed. The are features of stream during a flood events, below dams, where riffles have been 'drowned ' or in steep gradient streams flowing through gorge.

4. Pool bed: Pool bed habitats are zones of relatively deep, stationary or very slow flowing water over silty, sandy, stony or rocky beds. This habitat occurs in the main channel and should not be confused with backwaters. The velocity will indicate whether it is a pool or run. The classification factor is the bed type. Two main types are sandy/silty beds and rocky/gravel beds.

5. Macrophytes: Macrophyte habitats are areas where emergent, submergent and floating macrophytes or aquatic plants are present and can occur in slow to fast flowing areas.

Slope (m/m) calculated by dividing contour distance (m) to distance of stream between contour lines (m).

Substrate description: Visually estimates the composition of river substratum (to a depth of 10cm) into the following substrate categories. The sum of all substrate categories must total 100%.

- Bedrock (%)
- Boulder (%): >256mm
- Cobble (%): 64 – 256 mm
- Pebble (%): 16 – 64 mm
- Gravel (%): 4 – 16 mm
- Sand (%): 1 – 4 mm
- Silt/Clay (%): < 1mm

Substrate1-8: Categorical variable describes number of substrate types at the study site. Category 1 describes a site dominated by 100% of one substrate type while category 8 indicates site covering all types of substrate in different layers.

Soil Class Number (1–11): categorical variable attained from GIS map over lay

H Width (m) and H Depth (m): Width and Depth of habitat of the study sites.

0-4. Habitats: Categorical variable gives the assessment of the site. Nine criteria are numerically assessed from excellent to poor. These criteria are bottom substrate/available cover, embeddedness, velocity/depth category, channel alteration, bottom scouring and deposition, pool/riffle or run/bend ratio, bank stability, bank vegetative stability, and streamside cover. Habitat assessment sheet with full detail of assessment can be found in Conrick & Cockayne (2000). Final habitat assessment is the numerical score from 0 indicating poor condition to 4 indicating excellent habitat condition.

Site-mean phi: Visual estimates of the percentage cover of seven particle classes at a site were made: -9.5, -6.5, -4.5, -2, 2, 6.5, 9.5. These estimates were averaged to give a mean phi value for a site as a whole.

Mean Wetted Width, Mean Channel Width are measured or visually estimated if measurement can not be made (Figure 4.3)

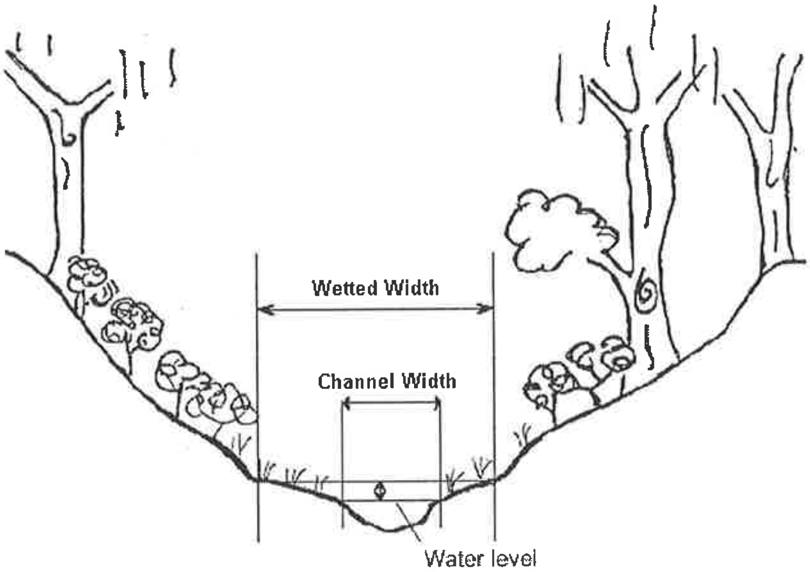


Figure 3.3 Example of wetted width, channel width and stream depth (Nichols et al., 2000)

Measurement applied method of Resh et al. (1996) for mean stream width and depth.

Mean stream width: Measure the width of the stream in meters, from water's edge to water's edge and perpendicular to the flow, for three different transects across the stream.

Mean stream depth: Along the same transects as above, measure the depth (in cm) at ¼ the distance from the water's edge, again at ½ the distance (midstream), and at ¾ of the way across. Add the three values and divide by 4 to account for the shallow water from the bank edge to the ¼ distance mark. Average depth in meters was recorded for each transects.

(3) *Meteorological variables*: Information can be extracted from the Bureau of Meteorology. These variables include mean wet season monthly rainfall (a), mean

dry season monthly rainfall (b), annual range in mean monthly rainfall, range in wet season monthly means, range in dry season monthly means, percentage rainfall in wet season, mean annual rainfall, mean daily max temp, mean daily min temp, mean daily temp range. *Season 1-2*: categorical variable contains 2 value: 1 means sample was taken in spring and 2 for sample taken in autumn.

(4) Some other physical variables used as predictors

Vegetation Type number (2-22): categorical variable describes number of both native and exotic vegetation types present at study site.

Soil Type Number (2 – 38): categorical variables

Water Temp (°C) measured at the site before sampling macroinvertebrates and disturbing the streambed.

Alkalinity: an expression of the buffering capacity of water, measured as the milliequivalents of hydrogen ions neutralised by a litre of water (expressed as $CaCO_3$ in mgL-1)

Table 3.2 summarised input variables used for network development. Totally, there are 39 predictors used to study interrelationship between physical and biological conditions of stream ecosystem

Table 3.2 Potential predictor variables

No	Predictor variables	Data Type	No	Predictor variables	Data Type
1	Season 1-2	categorical	21	0-8. substrate categories	categorical
2	Habitat 1-5	categorical	22	Site-mean phi	continuous
3	Latitude (S) Decimal	continuous	23	0-2 Reach	categorical
4	Longitude (E) Decimal	continuous	24	Mean Wetted Width	continuous
5	Altitude (m)	continuous	25	Mean Channel Width	continuous
6	Stream Order	categorical	26	Mean Depth	continuous
7	Slope	continuous	27	Mean wet season monthly rainfall	continuous
8	Distance From Source (km)	continuous	28	Mean dry season monthly rainfall	continuous
9	H Width (m)	continuous	29	Annual range in mean monthly rainfall	continuous
10	H Depth (m)	continuous	30	Range in wet season monthly means	continuous
11	Bedrock (%)	continuous	31	Range in dry season monthly means	continuous
12	Boulder (%)	continuous	32	Percentage rainfall in wet season	continuous
13	Cobble (%)	continuous	33	Mean annual rainfall	continuous
14	Pebble (%)	continuous	34	Mean daily max temp	continuous
15	Gravel (%)	continuous	35	Mean daily min temp	continuous
16	Sand (%)	continuous	36	Mean daily temp range	continuous
17	Silt/Clay (%)	continuous	37	Soil Type Number	categorical
18	Water Temp (°C)	continuous	38	Soil Class Number	categorical
19	Alkalinity (mgL-1 CaCO3)	continuous	39	Vegetation Type Number	categorical
20	0-4. Habitats	categorical			

3.3.2 Environmental Water Quality Variables

This sub- set contains environmental variables, which are altered by human impacts. They are used as inputs in dirty water approach. All water quality measurements and water samples are collected upstream of the biological sampling area. They are taken from a representative section of the streams.

Electrical Conductivity (µs/cm): is a measure of the total concentration of inorganic ions (salts) in the water.

pH: is a measure of the acidity or alkalinity of water and has scale from 0 (extremely acid) to 14 (extremely alkaline), with 7 being neural.

Turbidity (NTU): is a measure of the water "muddiness" and is caused by the presence of suspended particulate and colloidal matter consisting of suspended clay, silt, phytoplankton and detritus.

Chemical variables: Following chemical variables are analysed at laboratories.

- Total Hardness (mg/L $CaCO_3$)
- Total N (mgL-1 as N)
- Total P (mgL-1 as P)
- Na^+ (mg/L)
- K^+ (mg/L)
- Ca^{++} (mg/L)
- Mg^{++} (mg/L)
- HCO_3^- (mg/L)
- CO_3^{--} (mg/L)
- Cl^- (mg/L)

Following variables are also considered changeable under disturbance and belonged to this group

- Habitat Velocity - max (m/s)
- Detrital cover (%)
- Site Max Velocity
- Instantaneous Discharge

3.3.3 *Macroinvertebrate Distribution*

Sampling is not conducted when streams are in flood. If, during the scheduled sampling period, sites were consistently in flood, sampling resumed 4 – 6 weeks after floods have subsided to ensure that sampling only cover macroinvertebrates normally habit in the site. All macroinvertebrate samples were collected with a standard 250 - μm mesh dip net. Sample a total distance of 10m, covering a variety of velocities and different samples of habitat.

Colonisation pattern (presence/absence of macroinvertebrates) is preferred to use in this project. Abundance data was collected by many different people with different skill and experience, therefor quality control has shown that abundance data is not reliable enough to be used (Choy & Marshall, personal communication). All macroinvertebrate are identified to family level except for Oligochaeta (class), Copepoda, Osracoda (sub-class), Acarina (order), Cladocera (sub-order) and Chironomidae (sub-family). Adults and larvae for each family are combined for the purposes of data entry and analysis. 40 most common macroinvertebrates taxa are used as output for network development (listed in table3.3).

Outputs in the database receive only two values 0 and 1. 1 represents presence while 0 represents absence of this taxon at the study site.

3.3.4 *Summary*

Material used for neural network model development is a comprehensive database of the Queensland stream system having two parts.

- Part 1 containing 897 data set of reference sites is used for neural network model training and internal validation.

- Part 2 containing 1159 data set of test sites is used for neural network model external validation.

Each data set contains 3 subsets: 39 predictors describing physical condition of this site; 17 potential impacted environmental variables; and colonisation pattern of 40 macroinvertebrates taxa at the site. Different combinations of data are used for development of different neural network model approaches.

Table 3.3 40 most common macroinvertebrates taxa in Queensland stream system

No	Taxa	Upper classification	Common Name
1	Dugesiidae	Order Tricladida	Flat worms
2	Oligochaeta (Class)	Class	Segmented worms
3	Planorbidae	Class Gastropoda	Snails
4	Thiaridae	Class Gastropoda	Snails
5	Corbiculidae	Class Bivalvia	Mussels
6	Acarina (order)	Class Arachnida	Water mite
7	Copepoda (Sub-class)	Class Crustacea	Crustaceans
8	Cladocera (Sub-order)	Class Crustacea	Water fleas
9	Ostracoda (Sub-class)	Class Crustacea	Seed Shrimps
10	Atyidae	Class Crustacea	Freshwater Shrimps
11	Palaemonidae	Class Crustacea	Freshwater Prawns
12	Leptophlebiidae	Order Ephemeroptera	Mayflies
13	Baetidae	Order Ephemeroptera	Mayflies
14	Caenidae	Order Ephemeroptera	Mayflies
15	Prosopistomatidae	Order Ephemeroptera	Mayflies
16	Gomphidae	Order Odonata	Dragon flies
17	Corduliidae	Order Odonata	Dragon flies
18	Libellulidae	Order Odonata	Dragon flies
19	Coenagrionidae	Order Odonata	Damsel flies
20	Gripopterygidae	Order Plecoptera	Stone flies
21	Corixidae	Order Hemiptera	Bugs
22	Notonectidae	Order Hemiptera	Bugs
23	Pleidae	Order Hemiptera	Bugs
24	Veliidae	Order Hemiptera	Bugs
25	Dytiscidae	Order Coleoptera	Beetles
26	Elmidae	Order Coleoptera	Beetles
27	Psephenidae	Order Coleoptera	Beetles
28	Hydrophilidae	Order Coleoptera	Beetles
29	Tanypodinae (sub-family)	Order Diptera	True flies
30	Orthocladiinae (sub-family)	Order Diptera	True flies
31	Simuliidae	Order Diptera	True flies
32	Ceratopogonidae	Order Diptera	True flies
33	Tabanidae	Order Diptera	True flies
34	Leptoceridae	Order Trichoptera	Caddis flies
35	Hydropsychidae	Order Trichoptera	Caddis flies
36	Ecnomidae	Order Trichoptera	Caddis flies
37	Hydroptilidae	Order Trichoptera	Caddis flies
38	Calamoceratidae	Order Trichoptera	Caddis flies
39	Philopotamidae	Order Trichoptera	Caddis flies
40	Pyralidae	Order Lepidoptera	Moths

Identification key (Hawking & Smith, 1997; CSIRO, 1999)

3.4 *Artificial Neural Networks*

3.4.1 *Fundamental Concept*

Machine learning is a broad discipline in computer science that focuses on knowledge acquisition by various automated induction techniques. Using advanced machine learning techniques, comprehensive database and general basic knowledge on ecosystems, it is possible to automatically generate better models and in less time than is the case by traditional model construction. Machine learning reduces to a great extent the need to query the expert in the way that computer extracts knowledge from the given data. It is able to identify and model a real world system that we do not fully understand yet. (Kompare et al., 1994).

Two applications in this field, artificial neural networks and genetic algorithms, offer inductive approaches to model building. They are highly connective and simulate principles of natural evolution and knowledge discovery in large databases. In this project, I applied Artificial Neural Networks (ANN) as a tool to study the problems. This section discusses fundamental concept and mathematical background of the ANN used in the project.

Biological Neural Networks

ANNs are non-linear mapping structures based on biological principles of the functioning of theof human brain. Hence, to understand their operations, it is useful to understand the basic characteristics and operational mechanism of brain structure.

Human brain consists of approximately 10^9 to10^{12} fundamental units called neurons of many different types. A typical neuron has three major parts: the *cell body* or *soma*, the *dendrites*, and the *axon*. The *cell body* or *soma* is a main body of the nerve cell. The *cell body is* connected with filamentary input paths called *dendrites*. Dendritic trees are bunched into highly complex "dendritic trees", which have an enormous total surface area. *Axon* is a filamentary output path. The axon ends in a tree of filamentary paths called the axonic endings that are connected with dendrites of other neurons. The connection or junction between a neuron's axon and another neuron's dendrite is called a synapse. A schematic diagram of a typical biological neuron is shown in Figure 3.4.

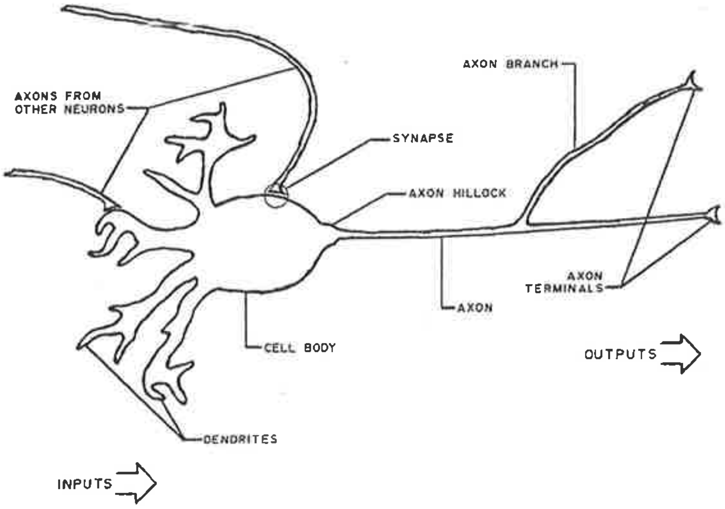


Figure 3.4 Schematic diagram of a biological neuron (Kent, 1980)

The neuron is the basic "gate" of the nervous system. It is a complex biochemical and electrical signal-processing factory. The neuron receives and combines signals from many other neurons through the *dendrites*. "Computations" (decisions) are performed in *the cell body* and the results are transmitted down the *axon* and its branches in pulse-codes digital form. The synapses are the inputs to the gate, where the pulse-coded information is reconverted to analog form. The inputs are subjected to weighted summation, and when a *threshold* is reached, the neuron fires, a new output pulse is placed on the axon. In this diagram, information flow is roughly from left to right through the neuron. (Kent, 1980)

A single neuron may have as many as 10,000 synapses and may be connected with some thousands neurons (Vemuri, 1992). However, not all synapses are excited at the same time. Because a received sensory pattern via the synapse probably stimulates a relatively small percentage of sites, an enormous number of patterns can be presented at the neuron without saturating the neuron's capacity. When the action potential reaches the axon ending, chemical messengers, *neurotransmitters,* are released. When a neurotransmitter is released, it drifts across the synaptic junction and initiates the depolarization of the postsynaptic membrane. The stronger the junction, the more neurotransmitters reach the postsynaptic membrane. Depending on the type of neurotransmitter, the effect on the postsynaptic potential is either *excitatory* (more positive) or *inhibitory* (more negative).

Decoding at the synapse is accomplished by temporal summation and spatial summation. In temporal summation each potential of an impulse adds to the sum of the potentials of the previous impulses. The total sum is the result of impulses and their amplitude. Spatial summation reflects the integration of excitations or inhibitions by all neurons at the target neuron. The total potential charge from temporal and spatial summations is encoded as a nerve impulse transmitted to other cells. The impulses received by the synapses of a neuron are further integrated over a short time as the charge is stored in the cell membrane. This membrane acts first as a capacitor and later as an internal messenger when complex biochemical mechanism take place (Kartalopoulos, 1996).

All integrated signals are combined at the soma, and if the amplitude of the combined signal reaches the *threshold* of the neuron, a "firing" process is activated and an output signal is produced. This signal, either a single pulse or a sequence of pulses at a particular rate, is transmitted along the cell's axonic endings.

In the real world of neural networks, the neurons do not all perform exactly the same function or in exactly the same way. The functions of sensory neurons and neural networks are quite diverse. This diversity adds to the complexity of the neural network. Whereas all neurons contain the same set of genes, individual neurons activate only a small subset of them. However, all neural networks exhibit certain properties such as:

- Many parallel connections exits between many neurons
- Many of the parallel connections provide feedback mechanism to other neurons and to themselves
- Some neurons may excite other neurons while inhibiting the operation of still others.
- Neural networks are asynchronous in operation
- Neural networks execute a program that is fully distributed and not sequentially executed
- Neural networks do not have a central processor. Instead processing is distributed

Biological neural networks are characterized by a hierarchical architecture. Lower-level networks preprocess raw information and pass their outcome to higher levels for higher-level processing.

Artificial Neural Networks

Artificial Neural Networks (ANNs) are systems that are purposely constructed to make use of some organizational principles resembling those of the human brain. They have following characteristics.

(1) ANNs have a large number of highly interconnected *processing elements* (also called nodes or units). These nodes usually operated in parallel and are configured in regular architectures. The processing elements in ANNs are called *artificial neurons.*

(2) The connections (weights) amongst neurons hold the knowledge.

(3) Artificial Neural Networks are neurally mathematical models.

Figure 3.5 below shows a simple mathematical model of biological neuron proposed by McCulloch and Pit (1943, cited by Lin & Lee, 1995), called an *M-P neuron.*

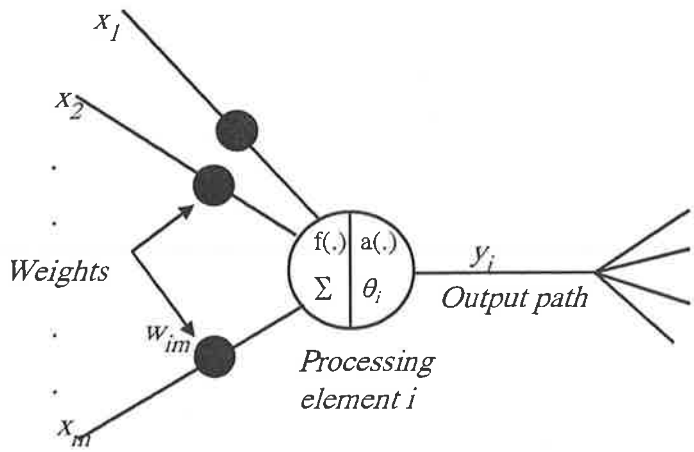


Figure 3.5. Schematic diagram of a Mc. Culloch and Pitts neuron

In this model, the i^{th} processing element computes a weighted sum of its inputs and outputs y_i according to whether this weighted input sum is above or below a certain *threshold* θ_i:

$$y_i(t+1) = a(f)\left(\sum_{j=1}^{m} w_{ij} x_j(t) - \theta_i\right), \tag{3.1}$$

where $a(f)$ is a the activation function or transfer function

The weight w_{ij} represents the strength of the synapse connecting neuron j (source) to neuron i (destination)

(4) A neuron can dynamically respond to its input stimulus, and the response completely depends on its local information; that is, the input signals arrive at the neuron via impinging connection and connection weights.

(5) Like a human brain, Artificial Neural Networks have collective behavior that demonstrates ability to learn, recall, and generalize information from training pattern. This collective behavior illustrates the computational power, and no single neuron carries specific information (Lin & Lee, 1996).

3.4.2 Basic Models and Learning Rules

There are three basic attributes that characterize the models of Artificial Neural Networks: models of the processing elements (neurons), models of synaptic interconnections, and the training or learning rules for updating the connecting weights. This section studies the basics of these three attributes.

Processing Elements

The function of an M-P neuron can be extended to a general model of a processing element (PE). The information processing of a PE consists of two parts: input and output. Associated with the input of a PE is an integration function f. The function combines information, activation, or evidence from an external source or other PEs into a *net input* to the PE. In the case of an M-P neuron, this is usually a linear function of the input x_j to the PE:

$$f_i = net_i = \sum_{j=1}^{m} w_{ij} x_j - \theta_i, \tag{3.2}$$

More-complex integration functions can also be considered as follows.

- *Quadratic function:*

$$f_i = \sum_{j=1}^{m} w_{ij} x_j^2 - \theta_i , \tag{3.3}$$

- *Spherical function:*

$$f_i = \rho^{-2} \sum_{j=1}^{m} (x_j - w_{ij})^2 - \theta_i , \tag{3.4}$$

where ρ and w_{ij} are the radius and the centre of the sphere, respectively.

- *Polynomial function:*

$$f_i = \sum_{j=1}^{m} \sum_{k=1}^{m} w_{ijk} x_j x_k + x_j^{\alpha_j} + x_j^{\alpha_k} - \theta_i , \tag{3.5}$$

where w_{ijk} is the weight on the conjunctive link connecting PE *j* and PE *k* to PE *i,* and α_j and α_k are real constants. This equation can be extended to include higher-order terms. A PE with a polynomial integration function is called a sigma-pi (ΣΠ) unit.

A second action of each PE is to output an activation value as a function of its net input through an *activation function* or *transfer function a (f)*. Some commonly used activation functions are as follows:

- *Step function:*

$$\text{a(f)} = \begin{cases} 1 & \textit{if} \quad f \geq 0 \\ 0 & \textit{otherwise} \end{cases} , \tag{3.6}$$

- *Hard limiter (threshold function):*

$$\text{a(f)} = \operatorname{sgn}(f) = \begin{cases} 1 & \textit{if} \quad f \geq 0 \\ -1 & \textit{if} \quad f < 0 \end{cases} \tag{3.7}$$

Where sgn(.) is the signum function

- *Ramp function:*

$$a(f) = \begin{cases} 1 & \textit{if} \quad f > 0 \\ f & \textit{if} \quad 0 \leq f \leq 1 \\ 0 & \textit{if} \quad f < 0, \end{cases} \tag{3.8}$$

- *Unipolar sigmoid function:*

$$a(f) = \frac{1}{1 + e^{-\lambda f}}, \tag{3.9}$$

- *Bipolar sigmoid function:*

$$a(f) = \frac{2}{1 + e^{-\lambda f}} - 1, \tag{3.10}$$

where $\lambda > 0$ determines the steepness of the continuous function $a(f)$ near $f=0$.

A PE with a linear integration function and a hard limited activation function is called *linear threshold unit* (LTU), and a PE with linear integration function and a graded activation function (Eq.(3.9) or (3.10) is called *linear graded unit* (LGU). The LTU and LGU are most frequently used models in ANNs (Lin & Lee, 1996). In my research, the LGU, unipolar sigmoid function had been used as transfer function for network performance.

Connections

Architecture defines the network structure, that is not only the number of processing elements but also their interconnectivity. Each PE is connected to other PEs or to itself; both delay and lag-free connections are allowed (Lin and Lee, 1996). There are five basic types of connection geometries.

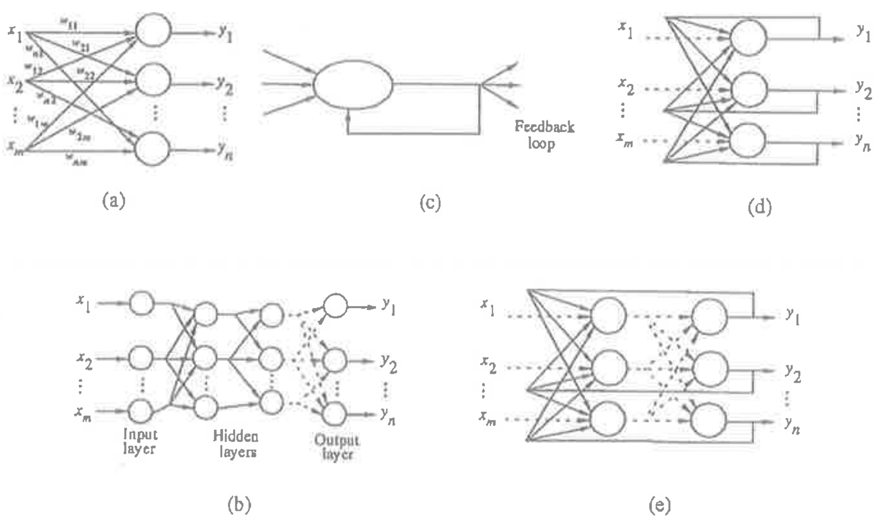


Figure 3.6 Basic network connection geometries (Lin and Lee, 1996)

(a) Single-layer feedforward network (b) Multilayer feedforward network (c) Single node with feedback to itself (d) Single-layer recurrent network (e) Multilayer recurrent network

In the *single-layer feedforward network* (Fig. 3.6a), a PE is combined with other PEs to make a layer of these nodes. Inputs can be connected to these nodes with various weights, resulting in a series of outputs, one per node.

Several layers can be interconnected to form *multilayer feedforward network* (Fig. 3.6b). Input layer receives inputs and typically performs no function other than buffering of the input signals. The outputs of the network are generated from the output layer. Any layer between the input and output layers is called a hidden layer because it is internal to the network and has no direct contact with the external environment. There may be no or several hidden layers in an ANN. The two mentioned types are *feedforward networks* because no PE output is an input to a node in the same layer or in a preceding layer.

The outputs can be directed back as inputs to same- or preceding-layer nodes, in this case, the network is a *feedback network.* If PE output is directed back as input to PEs in the same layer, the network is *lateral feedback.* Feedback networks that have closed loops are called *recurrent network.* A *single node with feedback to itself* is the simplest recurrent neural network (Fig. 3.6c)

In a *single-layer network with a feedback connection* (Fig. 3.6d) PE output can be directed back to the PE itself, to other PEs, or to both. In a multilayer recurrent network, a PE output can be directed back to the nodes in the preceding layer (Fig. 3.6e). A PE output can be also directed back to the PE itself and to the other PEs in the same layer.

More than one of the basic connection geometries can be used together in an ANN. Choice of neural network architecture define a priori probability distributions over non-linear functions. Feedforward neural networks such as multilayer perceptrons prove to be useful tools for nonlinear regression and classification problems (MacKay, 1997). This type of ANNs models have been applied to various fields of aquatic sciences such as modelling water quality and relating community characteristics with environmental variables (Schleiter et al., 1999).

Learning Rules

The neurodynamics of neural networks defines their properties, that is, how the neural network *learns, recalls, associates,* and continuously *compares* new information with existing knowledge, how it *classifies* new information, and how it develops new classifications if necessary. Learning is the process by which the neural network adapts itself to a stimulus, and produces a desired response. Learning is also a continuous classification process of input stimuli; when a stimulus appears at the network, the network either recognizes it or it develops a new classification. During the learning process, the network adjusts the synaptic weights in response to an input stimulus so that its actual output response meets the desired output response. When the actual output response is the same as the desired one, the network has completed the learning phase that means it has *"acquired knowledge"* (Kartalopoulos, 1996)

Assuming that there are *n* PEs in an ANN and each PE has an exactly m adaptive weight, then the *weight matrix* (or the *connection matrix*) W is defined by:

$$W = \begin{bmatrix} w^T{}_1 \\ w^T{}_2 \\ \vdots \\ w^T_n \end{bmatrix} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nm} \end{bmatrix}, \tag{3.11}$$

where $w_i = (w_{i1}, w_{i2},\ldots,w_{im})^T$, $i=1,2,\ldots,n$, is the weight vector of PE i and w_{ij} is the weight on the link from PE j (source node) to PE i (destination node).

Suppose that the weight matrix W contains all the adaptive elements of an ANN, then the set of all possible W matrices determines the set of all possible information processing configurations for this ANN. In other words, if the information processing performance is realised by this ANN, the ANN can be realized by finding an appropriate matrix W. Therefore, learning rules for ANNs need to be developed to efficiently guide the weight matrix W in order to approach a desired matrix that yields the desired network performance.

Learning rules are very important attributes to specify an ANN. In general, learning rules are classified into three categories: *supervised learning, reinforcement learning and unsupervised learning.* (Lin & Lee, 1996).

(a) Supervised learning

In the learning session of a neural network, an applied input stimulus results in an output response. This response is compared with a priori desired output signal, the *target* response. If the actual response differs from the target response, the neural network generates an *error signal*. This error signal is then used to calculate the adjustment that should be made to the network's synaptic weights so that the error is minimized closely to zero (the actual output matches the target output). The error minimization process requires a special circuit known as a *supervisor*.

The amount of calculation required to minimize the error depends on the algorithm used, this is purely a mathematical tool derived from optimization techniques. Some important parameters are time required per iteration, the number of iterations per input pattern for the error to reach a minimum during the training session, whether the network has reached the global or local minimum, and, if a local one, whether the network can escape from it or it remains trapped.

(b) Reinforcement learning

Reinforced leaning is an extreme case of supervised learning. In this case, a supervisor does not indicate how close the actual output is to the desired output but whether the actual output is the same with the target output or not. There is only a single bit of feedback information indication whether output is *right* or *wrong*.

If the supervisor's indication is "wrong", the network readjust its parameters and tries again and again until it get its output response "right". During this process there is no indication if the output response is moving in the right direction or how close to the correct response it is. Consequently, the process of correcting synaptic weights follows a different strategy than the supervised learning process.

Important parameters for the reinforcement learning are the same as of supervised learning. When reinforced learning rules are applied, certain boundaries should be established so that the trainee should not keep trying to get the correct response *ad infinitum.*

(c) Unsupervised learning

In unsupervised learning, there is no supervisor to provide any feedback information. There is no feedback from the environment to say what the output should be or whether they are correct. The network must discover for itself patterns, features, regularities, correlations or categories in the input data and code them in the output. While discovering these features, the network undergoes changes in its parameters; this process is called *self-organizing*.

A typical example is making an unsupervised classification of objects without providing information about the actual classes. The proper clusters are formed by discovering the similarities and dissimilarities among the objects. During the training session, the neural net receives at its inputs many excitations and it arbitrarily organizes the patterns into categories. When a stimulus is applied later, the neural net provides an output response indicating the class to which the stimulus belongs. If a class cannot be found for the input stimulus, a new class is generated. Even though unsupervised learning does not require a supervisor, it requires guideline to determine how it will form responses.

Supervised learning of rules is a very popular application of Artificial Neural Networks in pattern recognition work. The goal of process is to adapt the parameters of the network so that it performs well for patterns from outside the training set (Werbos, 1992). This goal meets the purpose of developing the ANNs to study bio-community of freshwater ecosystem. Feedforward Networks and Supervised Learning is discussed in the next session.

3.4.3 Feedforward Networks and Supervised Learning

Single - Layer Perceptron Networks

Single-layer feedforward networks, known as *simple perceptron* is shown in Figure 3.7.

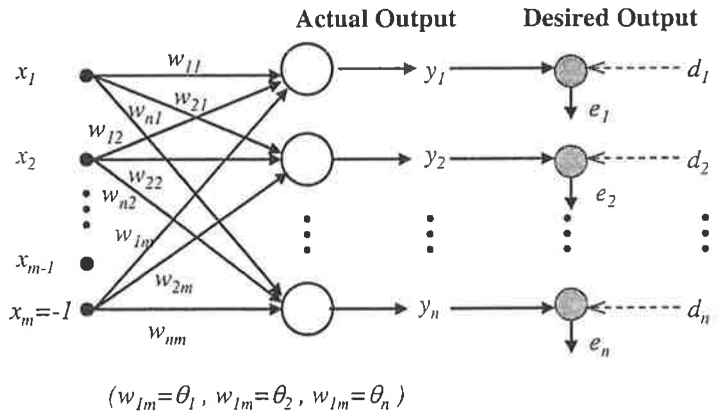


Figure 3.7 A simple perceptron

In which:

- Input pattern: $\mathbf{x}^{(k)}=[x_1^{(k)}, x_2^{(k)},\ldots, x_m^{(k)}]^T$; m: number of inputs
- Particular output pattern: $\mathbf{d}^{(k)}=[d_1^{(k)}, d_2^{(k)},\ldots, d_n^{(k)}]^T$; n: number of outputs
- Actual output pattern: $\mathbf{y}^{(k)}=[y_1^{(k)}, y_2^{(k)},\ldots, y_n^{(k)}]^T$; k=1,2,..,p, p: number of input-output pair in the training set

Desired performance of networks after training process is that the actual output pattern to be equal to the target pattern.

$$y_i^{(k)} = a(\mathbf{w}_i^T\mathbf{x}^{(k)}) = a\left(\sum_{j=1}^{m} w_{ij}x_j^{(k)}\right) = d_i^{(k)}, \tag{3.12}$$

i=1,2,...,n; k=1,2,...p,

where $\mathbf{w}_i^{\ T}=[w_{i1}, w_{i2},\ldots,w_{im}]^T$ is weight vector associated with PE i.

A simple learning rule determines the set of weights w_{ij} needed to achieve the desired performance for simple perceptron. A *Perceptron Learning Rule* is applied for simple perceptrons with linear threshold units (LTU) and *Widrow-Hoff learning rule* is applied for simple perceptrons with linear graded units (LGU) (Lin & Lee, 1995). In this section learning rule for PEs with continuous and differentiable activation functions - *Widrow-Hoff learning rule* - is discussed.

Widrow-Hoff Learning Rule

The learning problem that is of interest belongs to the class of supervised learning as indicated in Eq. (3.12). For a given set of p training patterns,$\{(\mathbf{x}^{(1)}, d^{(1)}), (\mathbf{x}^{(2)}, d^{(2)}),\ldots, (\mathbf{x}^{(p)}, d^{(p)})\}$, the goal is to find a correct set of weights w_i such that

$$\sum_{j=1}^{m} w_j x_j^{(k)} = d^{(k)}, \qquad k = 1,2,\ldots, p. \tag{3.13}$$

To find the weights from above equation, a cost function E (w), which measures the system's performance error, is defined by

$$E(\mathbf{w}) = \tfrac{1}{2}\sum_{k=1}^{p} (d^{(k)} - y^{(k)})^2 = \tfrac{1}{2}\sum_{k=1}^{p} (d^{(k)} - \mathbf{w}^T \mathbf{x}^{(k)})^2 = \tfrac{1}{2}\sum_{k=1}^{p} \left(d^{(k)} - \sum_{j=1}^{m} w_j x_j^{(k)} \right)^2 \tag{3.14}$$

The smaller E (**w**) is, the better w_j will be. E (**w**) is normally positive but approaches zero when $y^{(k)}$ approaches $d^{(k)}$ for k=1,2,...,p. The goal of the learning rule is to find the weights that will minimize the mean squared error E (**w**).

In general, learning rules start with a general initial guess at the weight values and than make successive adjustments based on the evaluation of an objective function. They eventually reach a near optimal or optimal solution in a finite number of steps. Given the cost function E (**w**) in Eq. (3.14), we can improve on a set of weights w_j by sliding downhill on the surface it defines in the weight space. The usual *gradient-descent algorithm* suggests adjusting each weight w_i by an amount Δw_i proportional to the negative of the gradient of E (**w**) at the current location

$$\Delta \mathbf{w} = \eta \nabla_{\mathbf{w}} E(\mathbf{w}) \tag{3.15}$$

That is,

$$\Delta w_j = \eta \frac{\partial E}{\partial w_j} = \eta \sum_{k=1}^{p} (d^{(k)} - \mathbf{w}^T \mathbf{x}^{(k)}) x_j^{(k)}, \qquad j = 1,2,\ldots, m. \tag{3.16}$$

If these changes are made individually for each input pattern x(k) in turn, then the change in response to pattern x(k) is simply:

$$\Delta w_j = \eta (d^{(k)} - \mathbf{w}^T \mathbf{x}^{(k)}) x_j^{(k)} \tag{3.17}$$

The learning rule in Eq. (3.17) is called *Widrow-hoff learning rule*. It is also refer to as the *least mean square* (LMS) rule. In this method, weights are initialized at any value. (Lin &Lee, 1995).

Multilayer Feedforward Networks

Single layer Perceptron Network is able to solve a problem with the condition that the input patterns of the problem be linearly separable or linearly independent. This limitation does not apply to feedforward networks with hidden layers between input and output layers. This section discusses the most popular learning algorithm applied for multi layered Artificial Neural Networks – the *Back- Propagation.*

Back - Propagation Learning Algorithm

The backpropagation (BP) was developed first by Rumelhart (1986) and since then, the back-propagation algorithm has been widely used as learning algorithm in feedforward multilayer neural networks. The BP is applied to feedforward ANNs with one or more hidden layers. Those networks associated with the back-propagation learning algorithm are called *backpropagation networks.* Based on this algorithm, the network learns a distributed associative map between the input and output layers. Given a training set of input-output pairs $\{(\mathbf{x}^{(k)}, \mathbf{d}^{(k)})\}$, the algorithm provides a procedure for changing the weights in a backpropagation network to classify the given input patterns correctly. The basic for this weight update algorithm is simply the gradient-descent method as used for simple perceptrons with differentiable units. The back-propagation algorithm performs two phases of data flow:

- The input pattern $\mathbf{x}^{(k)}$ is propagated from the input layer to the output layer and produces an actual output $\mathbf{y}^{(k)}$.
- The error signals resulting from the difference between $\mathbf{d}^{(k)}$ and $\mathbf{y}^{(k)}$ are *back-propagated* from the output layer to previous layer for them to update their weights.

Figure 3.8 shows three-layer networks with m PEs in the input layers, l PEs in the hidden layer and n PEs in the output layer. The solid lines show the forward propagation of signals and the dashed lines show the backward propagation of errors.

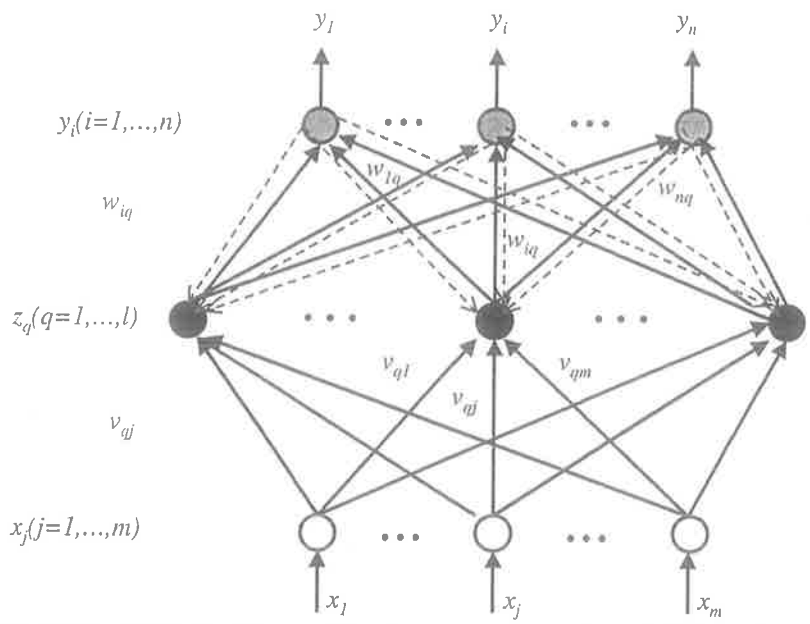


Figure 3.8 Three layer feedforward backpropagation network (Lin & Lee, 1995)

Given an input pattern **x**, a PE q in the hidden layer receive a net input of

$$net_q = \sum_{j=1}^{m} v_{qj} x_j \tag{3.18}$$

and produce an output of

$$z_q = a(net_q) = a\left(\sum_{j=1}^{m} v_{qj} x_j\right). \tag{3.19}$$

The net input for a PE i in the output layer is then

$$net_i = \sum_{q=1}^{l} w_{iq} z_q = \sum_{q=1}^{l} w_{iq}\, a\left(\sum_{j=1}^{m} v_{qj} x_j\right), \tag{3.20}$$

And it produces an output of

$$y_i = a(net_i) = a\left(\sum_{q=1}^{l} w_{iq} z_q\right) = a\left(\sum_{q=1}^{l} w_{iq}\, a\left(\sum_{j=1}^{m} v_{qj} x_j\right)\right), \tag{3.21}$$

The above equations indicate the forward propagation of input signals through the layers of neurons and their back propagation. The error signals and their back propagation will be considered next. Firstly, a cost function is defined:

$$E(\mathbf{w}) = \tfrac{1}{2}\sum_{i=1}^{n}(d_i - y_i)^2 = \tfrac{1}{2}\sum_{i=1}^{n}[d_i - a(net_i)]^2 = \tfrac{1}{2}\sum_{i=1}^{n}\left[d_i - a\left(\sum_{q=1}^{l} w_{iq} z_q\right)\right]^2, \tag{3.22}$$

Then according to the gradient-descent method, the weights in the hidden-to-output connections are upgraded by

$$\Delta w_{iq} = -\eta \frac{\partial E}{\partial w_{iq}}, \tag{3.23}$$

In which, η is learning ratio.

Using described equation and chain rule of $\partial E/\partial w_{qi}$,we have

$$\Delta w_{iq} = -\eta \left[\frac{\partial E}{\partial y_i}\right]\left[\frac{\partial y_i}{\partial net_i}\right]\left[\frac{\partial net_i}{\partial w_{iq}}\right] = \eta[d_i - y_i][a'(net_i)][z_q] \underline{\underline{\Delta}} \eta\delta_{oi} z_q, \tag{3.24}$$

where δ_{oi} is the error signal of the *i*th node in the output layer, that defined by:

$$\delta_{oi} \underline{\underline{\Delta}} -\eta \left[\frac{\partial E}{\partial y_i}\right]\left[\frac{\partial y_i}{\partial net_i}\right] = \eta[d_i - y_i][a'(net_i)], \tag{3.25}$$

where net_i is the net input to PE *i* of the output layer and $a' = \partial a(net_i)/\partial net_i$.

For the weight update on the input-to-hidden connections, we use the chain rule with the gradient-descent method and obtain the weight update on the link weight connecting PE *j* in the input layer to PE *q* in the hidden layer,

$$\Delta v_{qj} = -\eta \left[\frac{\partial E}{\partial v_{qj}}\right] = -\eta \left[\frac{\partial E}{\partial net_q}\right]\left[\frac{\partial net_q}{\partial v_{qj}}\right] = -\eta \left[\frac{\partial E}{\partial z_q}\right]\left[\frac{\partial z_q}{\partial net_q}\right]\left[\frac{\partial net_q}{\partial v_{qj}}\right], \tag{3.26}$$

Each error term $[d_i - y_i]$, $i=1,2,\ldots,n$, is a function of z_q. Evaluating chain rule, we have

$$\Delta v_{qj} = -\eta \sum_{i=1}^{n} \left[(d_i - y_i)a'(net_i)w_{iq}\right]a'(net_q)x_j, \tag{3.27}$$

$$\Delta v_{qj} = -\eta \sum_{i=1}^{n} \left[(d_{oi} w_{iq}\right]a'(net_q)x_j = \eta\delta_{hq} x_j, \tag{3.28}$$

where δ_{hq} is the error signal of PE *q* in the hidden layer and is defined as

$$\delta_{hq} = -\frac{\partial E}{\partial net_q} = -\left[\frac{\partial E}{\partial z_q}\right]\left[\frac{\partial z_q}{\partial net_q}\right] = a'(net_q)\sum_{i=1}^{n} \delta_{oi} w_{iq}, \tag{3.29}$$

The error signal of a PE in a hidden layer is different from the error signal of a PE in the output layer. Because of this differences, the above weight update procedure is called *generalized delta learning rule*. One important feature of the back-propagation algorithm is that the update rule is local. To compute the weight change for a given connection, we need only quantities available at both ends of that connection.

The learning procedure requires only that the change in weight be proportional to $\partial E/\partial w$. True gradient descent requires that infinitesimal steps s be taken. The constant of proportionality is the learning ratio η. The larger this constant, the larger the change in weights, the more rapid learning but it might results in oscillations. One way to increase the learning rate without leading to oscillation is to include a *momentum* term to the generalized delta rule (Rumelhart & McClelland, 1988). This scheme is implemented by giving a contribution from the previous time step to each weight change:

$$\Delta v(t+1) = \eta \delta_{hq} x_j + \alpha \Delta v(t) \tag{3.30}$$

In which, $\alpha \in [0,1]$ is a momentum parameter.

Constant α determines the effect of past weight changes on the current direction of movement in weight space. This provides a kind of momentum in weight space that effectively filters out high-frequency variations of the error-surface in the weight space. A value of $\alpha=0.9$ is often used (Rumelhart & McClelland, 1988; Lin & Lee, 1996).

3.4.4 Summary

In summary, the artificial neural network is an adaptive communication network that communicates a "cost function" for a desired output. Mathematically speaking, a neural network represents a dynamic system that can be modelled as a set of coupled differential equation.

The performance of the Artificial Neural Networks is described by the figure of merit, which expresses the number of recalled patterns when input patterns are applied that are complete, partially complete or noisy. In designing an artificial neural network, following parameters were considered to be very important

- Network topology: number of layers in the networks, number of neurons or nodes per layers, interconnections among neurons
- Learning algorithm
- Number of iterations per pattern during training
- Number of calculations per iteration

- Choice of transfer function and the range of operation of the neuron

The wide choice of architectural configurations, in conjunction with variety of learning rules, led to the development of over thirty types of neural network models. The choice of a type of network models depends on a number of factors. However, once a particular architecture and learning rule has been proposed, its properties can be analyzed and studied in detail.

Feedforward Back-propagation is now the most widely used tool in the field of supervised Neural Networks. It is a very powerful for application with pattern recognition. It is generally used with a very simple network design but the same approach can be used with any network of differentiable functions (Werbos, 1992). Recently, Multilayer Feedforward Back-propagation had been implemented for studies of stream hydrological and ecological responses to climate change (Poff et al., 1996), modelling water quality (Schleiter et al., 1999), and also for studying biological condition by mean of macroinvertebrates (Walley & Fontama, 1998). It proves to be a useful tool to study freshwater ecosystem.

In this research, I develop Artificial Neural Networks with multilayer feedforward connection characteristic with backpropagation algorithm in an attempt to study high non-linearity and high complexity of nature of freshwater ecosystem.

3.5 Procedure for Network Development

3.5.1 Software

The project had been performed with software package NeuroSolution ver. 3.0 – the Neural Network Simulation Environment, which is a product of NeuroDimension Incorporated company.

NeuroSolutions provides an object-oriented simulation environment for neural network design and application. It has quickly evolved into the software tool of choice for both the neural network beginner and expert alike. This software combines a modular, icon-based network design interface with an implementation of advanced learning procedures including backpropagation and backpropagation through time.

NeuroSolutions can be used to design neural networks to solve many different types of problems in a variety of fields. The result is a virtually unconstrained environment for designing neural networks to solve real-world problems such as forecasting, pattern recognition, process control, targeting marketing, and many more.

NeuroSolutions is based on the concept that neural networks can be broken down into a fundamental set of neural components. By allowing the user to arbitrarily interconnect these components, a virtually infinite number of neural models can be constructed. Neural components, such as axons, synapses, and gradient search engines, are laid out on a graphical breadboard and connected together to form a neural network. Input components are used to inject signals, and probe components are used to visualize the network's response.

Neural networks are often criticized as being a "black box" technology. With NeuroSolutions' extensive and versatile set of probing tools, this is no longer the case. Probes provide you with real-time access to all internal network variables, such as: inputs/outputs, weights, errors, hidden states, gradients, sensitivity analysis.

Networks are developed using NeuroSolutions for Excel, which is one powerful tool in the NeuroSolutions package. NeuroSolutions for Excel was designed to allow user to develop a complete solution to a problem in one easy to use package while also giving the flexibility to customize its operation using Visual Basic for Applications as a scripting language.

NeuroSolutions for Excel is a revolutionary product which benefits both the beginner and advanced neural network developer. For the beginner, NeuroSolutions for Excel offers visual data selection, one step training and testing, and automated report generation. For the advanced user, NeuroSolutions for Excel offers the ability to perform parameter optimization, run batch experiments, and create custom batch experiments programmatically. The best part is that all of these tasks can be performed without ever leaving Microsoft Excel.

3.5.2 Data Preprocessing and Modelling

Flow chat below shows the procedure of network development (Figure 3.9).

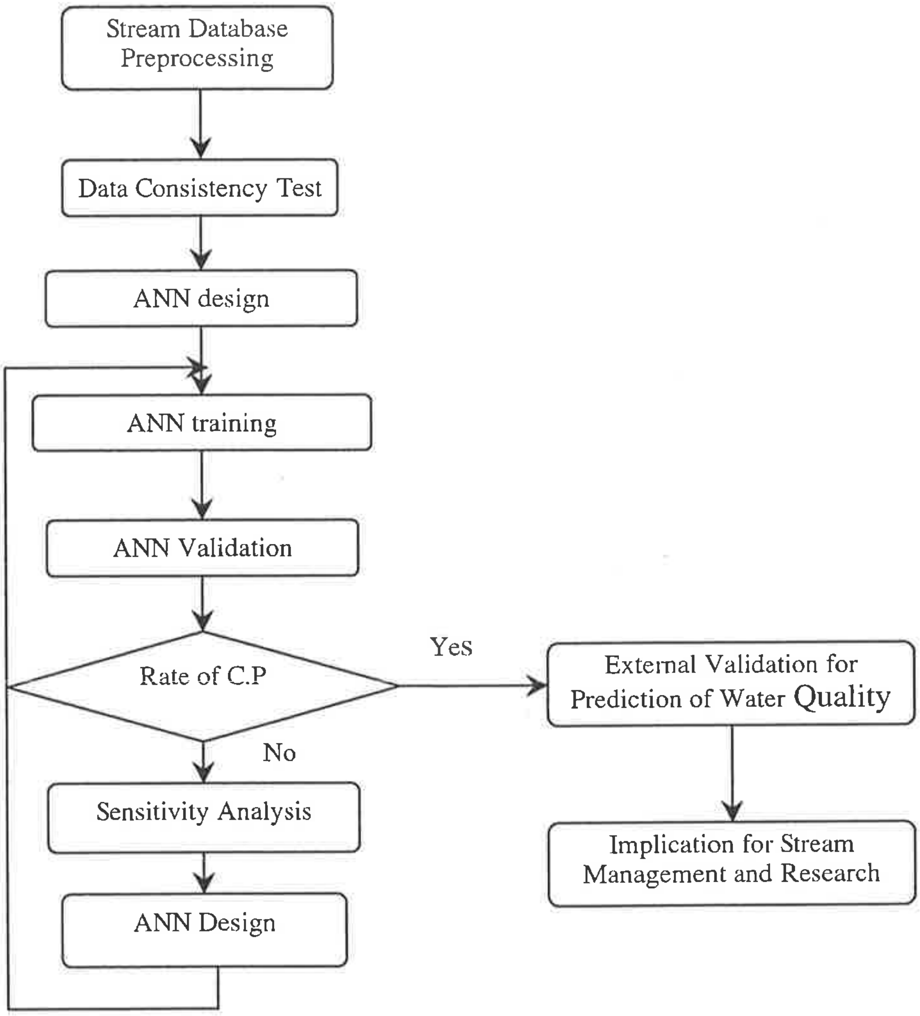


Figure 3.9 Approach for data-preprocessing and ANN modelling (Hoang et al., 2001)

Processes in each box are explained below

Data preprocessing and consistency test

Data used for training processes are designed dependent on method of modelling. Data set of sites containing incomplete or unreasonable data will be removed. The remaining data sets will be randomly divided into training and validation data sets. From the total number of samples in the data set, 80% is randomly taken for ANN the training processes and the remaining 20% are used for network internal validation.

ANN Design

Design step decides structure of network and training algorithm. Feedforward backpropagation is used as training algorithm.

Multilayer perceptrons (MLPs) are layered feedforward networks typically trained with static backpropagation. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are easy to use, and that they can approximate any input/output map. The key disadvantages are that they train slowly, and require lots of training data (typically three times more training samples than network weights).

Generalized feedforward networks are a generalization of the MLP such that connections can jump over one or more layers. In theory, a MLP can solve any problem that a generalized feedfoward network can solve. In practice, however, generalized feedforward networks often solve the problem much more efficiently. A classic example of this is the two-spiral problem. Without describing the problem, it suffices to say that a standard MLP requires hundreds of times more training epochs than the generalized feedforward network containing the same number of processing elements (NeuroDimension, 1999). Therefor this paradigm is chosen for network development in this research.

Design step also decides the architecture of networks including

- number of inputs in the input layer
- number of outputs in the output layer
- number of hidden layers as well as number of neurons (Processing Elements) contained in each hidden layer

NeuroSolutions simulations are vector based for efficiency. This implies that each layer contains a vector of PEs and that the parameters selected apply to the entire vector. The parameters are dependent on the neural model, but all require a non-linearity function to specify the behavior of the PEs. In addition, each layer has an associated learning rule and learning parameters. The number of PEs and learning parameters are entered in the corresponding fields. Parameters such as step site, momentum coefficient, number of iterations characterise the performance of the designed networks. Architecture and performance parameters will be changed experimentally during the training and internal validation time in order to optimise results for designed input layer.

ANN "training" performance

Neural network determines the weighted connectance between input and output nodes by the neurons (processing element). The neurons are located in the hidden layer and feed a non-linear sigmoid function. A learning process (training) forms the interconnection between the neurons and the nodes.

The aim of the training of a neural network is to minimize the output error with the respect to the known desired output. This error is defined to be the sum of the differences between the network outputs and the measured outputs they are supposed to predict. Once formed by training, the interconnections remain fixed in the hidden layer and the neural network can be used for predictions.

The cross validation set is used to determine the level of generalization produced by the training set. Cross validation is executed in concurrence with the training of the network. Every so often, the network weights are frozen, the cross validation data is fed through the network, and the results are reported. The stop criteria of the controller can be based on the error of the cross validation set instead of the training set to insure this generalization. This is an indication that the network has begun to overtrain. Overtraining is when the network simply memorizes the training set and is unable to generalize the problem.

Cross validation was estimated by means of mean square error (MSE) between calculated and targeted outputs. Chapter 4 will discussed more detal on how cross validation can be use to identify overtrained situation. Overtraining results in increasing values of MSE with increasing number of iteration. MSE of training set continues to decrease but in this case, network only memories the database from training set but does not generalise the patterns.

Training process will be carried out by data from training set.

ANN validation process

ANNs will be validated with independent data set to test the performance of networks. A comparison between the actual values collected from sites with the value predicted by the model will be made to evaluate performance of networks. The validation results are represented by percentage of correct predictions of colonisation pattern of each macroinvertebrate taxa.

Sensitivity Analysis

A sensitivity analysis will be conducted for all ANN models for specific macroinvertebrate taxa in order to improve models' validity. Method of sensitivity analysis as follows: The first input is varied between its mean +/- a defined number of standard deviations so that it covers whole range of this input in the database while all other inputs are fixed at their respective means. The output is calculated for a certain number of steps above and below the mean. The processes will be repeated for all inputs and for each of 40 outputs.

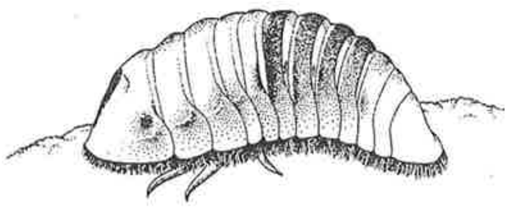
The sensitivity will be qualified in term of percentage output change over the range of input data. Input makes the output change below 40% will be considered redundant for this specific macroinvertebrate taxon. Refer to results of sensitivity analysis we can decide which variables are sensitive to distribution of specific macroinvertebrates and which variables are redundant.

Refining networks

Redundant inputs for each network will be taken out and new ANNs will be designed with taxa specific input layers. Training and internal validation processes will be repeated for refined networks in order to receive designed rate of correct prediction. The condition is designed in accordance with the purpose and requirement of network user. Processes are repeated until validation results satisfy the required condition of the rate of correct prediction. In this project, the condition of 70% correct prediction for all macroinvertebrate taxa is applied for fulfillment of network performance.

External validation and Stream Site Prediction

Trained networks are applied in prediction step. External validation and site prediction are made to test ability of trained networks to be applied in predicting habitat conditions at sites for management purposes. Details of these methods are addressed in chapters 4 and 5, where each model approaches are discussed.



Lateral view of the larva of a species of water penny Psephenidae *(Gulland & Cranston, 2000))*

4 Adopting Clean Water Approach

4.1 Introduction

4.1.2 Clean Water Approach

This model had been developed based on the ***referential approach*** (Reynolson et al., 1997) to predict the fauna at impacted sites as if they were unimpacted.

The concept of a reference condition is now being widely applied for biomonitoring and bioassessment of aquatic resources. The reference condition is central to the idea of "biocriteria" developed by the US Environmental Protection Agency (Davis & Simon, 1995). The same approach has been used in the UK for river classification and water quality assessment (Wright, 1995) and is fundamental for the National River Health Program in Australia (Parson and Norris, 1996).

Reference condition is defined as the condition that is representative of a group of minimally disturbed sites organised by selected physical, chemical and biological characteristics. Reference conditions are described based on pre-established criteria that exist at a wide range of sites. The reference conditions then serve as the control against which test conditions are compared. Reference condition represents the best available conditions and is made upby information from numerous sites. (Reynolson et al., 1997). Selection criteria used to determine reference site conditions in the Queensland stream system are described in chapter 3

In the current application, the reference condition is employed to compare the biological attributes of individual test sites with a group of reference sites. Reference condition uses an array of reference sites that characterise the potential biological conditions in a region for which assessments are to be made. A test site is subsequently compared to what is either the most appropriate subset of reference sites or to the entire reference site.

The main feature of this approach is that comparisons need to be made where site attributes are expected to yield similar invertebrate communities in the absence of disturbance. This analytical approach for comparisons with reference condition was adopted from RiVPACS, used for predicting the macroinvertebrate fauna in flowing water in UK (Wright, 1995) and AusRivAS in Australia (Parson & Noris, 1996). In both RiVPACS and AusRivAS, the number of taxa expected is calculated as the sum of the probabilities of those predicted (Moss et al., 1987). The number of those taxa actually collected is then compared with the number expected.

The severity of any environmental impact is assessed based on how much the number of taxa observed (O) deviates from the number expected (E), calculated as the O/E ratio. When the O/E ratio indicates impairment, the types of organism predicted to occur but not collected, or not predicted but collected, are used for interpretation.

4.1.2 Aims and Hypothesis

The aim of the stream modelling was to determine biological conditions of sites with respect to reference conditions based on the presence and absence of invertebrate taxa. The model was trained by means of reference data. Therefore the model outputs strictly reflect "reference condition". The assessment of the health of specific sites is than based on the comparison between observed and predicted site data.

A typically observed response of aquatic macroinvertebrate communities to environmental disturbance is general loss of diversity, especially with pesticide load or elevated nutrient enrichment (Cranton et al., 1996). The hypothesis of neural network model development by the clean water approach is that the number of taxa observed at the degraded sites should be less than the expected number, which reflects reference conditions. The value of the criterion O/E should range from a minimum of 0 (indicating that none of the families expected at a site were actually found at that site) to a theoretical maximum of 1, indicating a perfect match between the families expected and those that were found. In practice, this maximum can exceed 1, indicating an unusual diverse site, which should be a subject of further research to explain the cause and nature of such overpopulation (Coysh et al, 2000).

In order to simplify interpretation and aid tmanagement decisions, the O/E ratio can be divided into bands representing different levels of biological condition ranging from reference to severely degraded (Table 4.1).

Table 4.1 Division of O/E taxa into categories for reporting (Coysh et al, 2000)

Value of O/E	O/E taxa	Environmental condition
O/E> 1.2	More families found than expected	Richer invertebrate community than pristine – potential nutrient enrichment
1.2≥O/E≥0.8	Expected number of families within the range found at 80% reference site	Near pristine condition
0.8>O/E≥0.4	Fewer families than expected	Mildly to moderately impaired site
O/E<0.4	Very few of expected families remain	Moderately to severely degraded sites

Series of ANN models were developed with database from the Queensland stream system to study interrelations between macroinvertebrates assemblages and abiotic factors at reference habitat conditions. Developed neural network models then could be applied for predicting the conditions of freshwater environment. Habitat characteristics will be expressed by the criteria O/E range sites from reference to severely degraded condition.

4.2 *Materials and Methods*

4.2.1 Data Analysis

In the context of the modelling framework, site specific habitat features are used to predict the occurrence of invertebrate taxa at a site not affected by environmental stress. Habitat characteristics such as altitude, stream order and annual rainfall are suitable as such predictor variables. By contrast, chemical variables such as dissolved oxygen, pH and nutrient concentrations could easily be affected by anthropogenic impacts and would not be suitable as predictor variables. They would cause misleading predictions on the membership of test sites to the reference site groups (Smith et al., 1999). As a result, 39 physical riparian from the first subset of the database were used as predictor variables for neural network modelling, as listed in the Table 3.2 (Chapter 3). These predictor variables include both discrete categorical and continuous data. Only some discrete categorical variables are formed by classification schemes such as stream order; most are represented just as empirical criteria for habitat characteristics, such as soil types and vegetation type.

After removing incomplete or unreasonable data, 896 data sets from reference sites were used for neural network modelling, in which 716 data sets were taken randomly for training and the remaining 180 sets were used for internal validation. 1159 data sets from test sites were used for external validation or prediction steps.

4.2.2 Network Architecture

The design of the ANNs resulted in the selection of 39 environmental predictor variables considered as input nodes in the input layer and each of 40 the macroinvertebrate taxa considered as output nodes. Accordingly, 40 models were developed in this study.

The backpropagation algorithm was used for the training of the ANNs. One of the disadvantages of backpropagation algorithm is that it is difficult to determine in advance the number of hidden layers and number of nodes in each hidden layer. Many optimisation studies were carried out to select the best model configuration. For the initial ANNs, the best neural network was set up with a single hidden layer with 15 neurons. Figure 4.1 represents the general ANN architecture of the stream habitat model for each macroinvertebrate taxon.

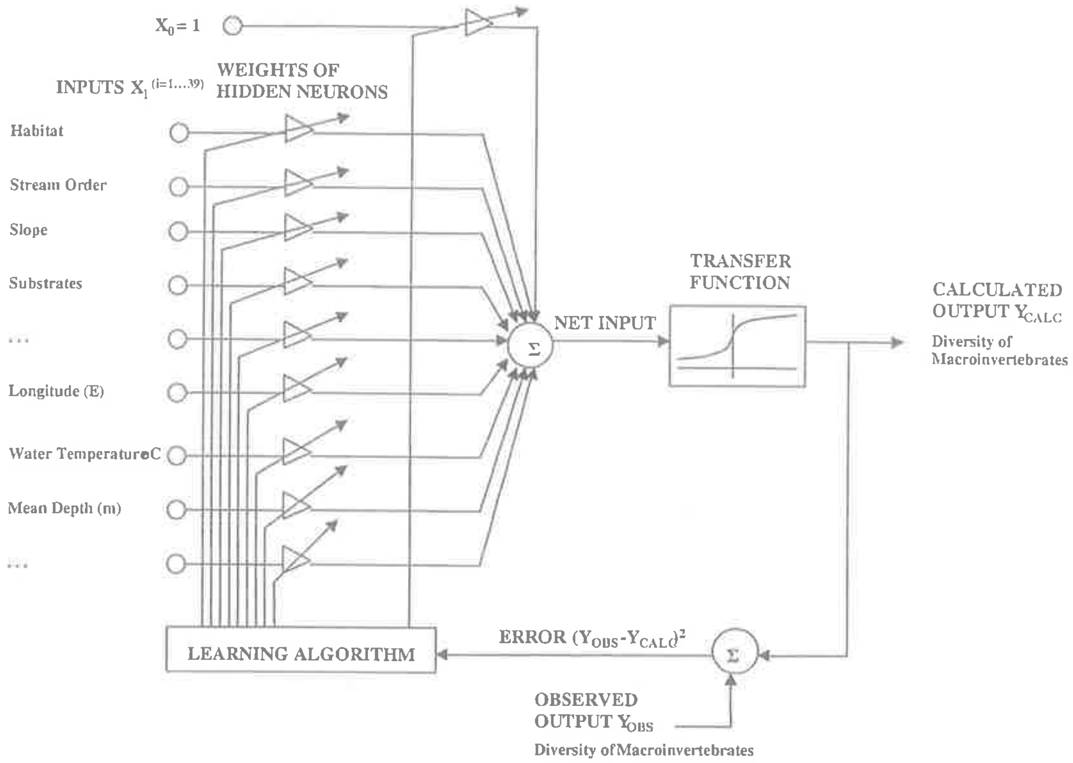


Figure 4.1 ANN architecture of the stream habitat model for each macroinvertebrate taxon

4.2.3 Method of Training

ANN training by the specific data is an important step of the modelling process. The feedforward backpropagation is applied as a learning algorithm to adjust the connectivity weights through convergence (Rumelhart et al., 1986), and the sigmoidal is used as the transfer function. The error term is the sum of the differences between the output and the targeted data, and the chosen criterion for the error term allowing convergence was 0.01.

Cross validation was applied to control overtraining. Cross validation was estimated by means of the mean square error (MSE) between calculated and targeted outputs. Figure 4.2 illustrates plot to show how cross validation was used to identify an overtrained situation. Overtraining results in increasing values of MSE with an increasing number of training iterations. The MSE of training sets continues to decline but, in this case, the neural network only memorises the database from the training set but does not generalise the patterns. The optimum for training is reached when a minimal MSE of both cross validation and training are observed.

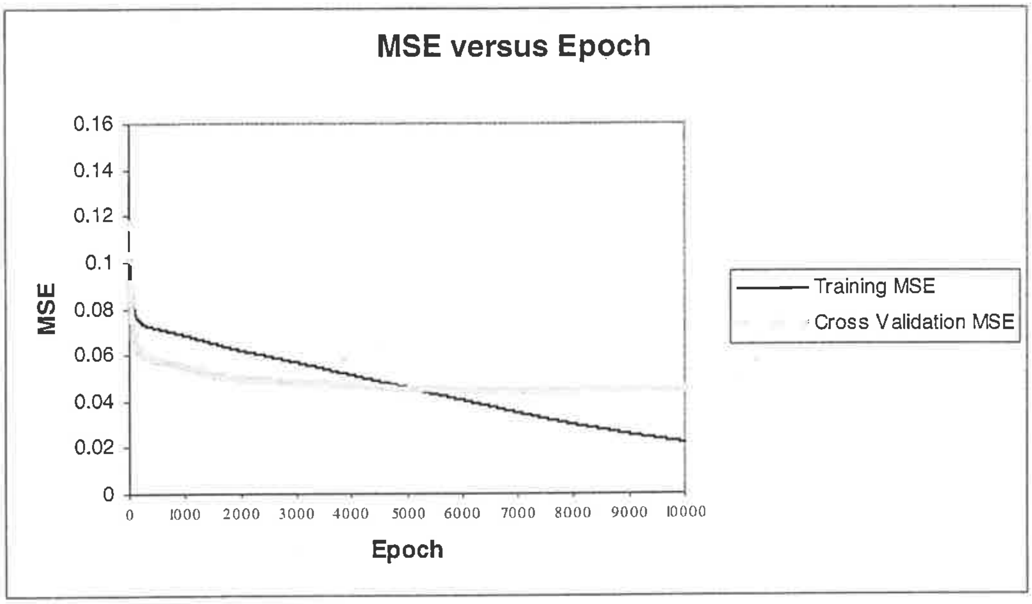


Figure 4.2. Cross validation results of some training examples

After a number of training trials, the network obtained the optimal technical training parameters as follows:

Number of iterations: 5000

Learning rate for hidden layer and for output layer: 0.1 and 1 respectively

Momentum coefficient: 0.9

4.2.4 Method of Validation

Validation is conducted in order to test the ability of networks to generalise relationships between habitat condition and macroinvertebrate assemblages in the reference condition. Neural networks may provide up to 100% correct answers when applied to training sets simply because they memorise the whole database but do not generalise the relationships between the inputs and targeted output. Trained networks were validated with data from an other 180 sites. Data of these sites had never been introduced to these neural networks before. They are therefore called *"independent sites"* and they also belong to reference conditions.

The validation is tested by means of correct predictions of the presence/absence of each macroinvertebrate taxa in 180 sites. If the validation gave appropriate results, and the conclusion could thus be drawn that the neural networks have already generalised relationships within the system, networks can be used in a prediction stage.

Output nodes in the database contained only the values 0 and 1 representing presence and absence of macroinvertebrate taxon respectively. Neural networks provided output values in the range from 0 to 1, describing the probability of this taxon occurring at the site. It was decided that, if the probability of occurrence ≥ 0.5, then the taxon was considered as present at the site; if the probability of occurrence was $<$ 0.5 then the taxon was considered as absent at the site.

4.2.5 Method of Prediction

Validated neural network models can be applied for predictions. In the prediction, studied sites are tested by the models for biological impairment. The networks predict the presence and absence of each taxon at these sites given that they are unimpacted. Consequently, the sum of taxa present at the site gives an expected number of macroinvertebrate taxa that should occur at the sites. This expected number would be compared with the observed number of macroinvertebrate taxa at the sites to determine a value of the criterion O/E.

The criterion O/E can be used as a biotic index to evaluate the habitat characteristics of stream sites ranging from reference conditions to severely degraded sites. The values of the O/E figures identify levels of degradation in the test sites.

4.3 Sensitivity Analysis

A comprehensive sensitivity analysis was conducted for the 40 ANN models for specific macroinvertebrate taxa. The method of sensitivity analysis is described in Chapter 3. For neural networks developed in this chapter, the inputs were varied between their mean +/- five levels of standard deviations, in order to cover the whole range of inputs in the database and the outputs were calculated for 150 steps above and below the mean.

Sensitivity analyses were only conducted for continuous inputs and those categorical inputs that were formed by classification schemes such as stream order, 0-4 habitat. Categorical variables represented just as empirical criteria for habitat characteristics, such as soil types and vegetation type, were not investigated by sensitivity analysis.

4.3.1 Results of Sensitivity Analysis

Plots were generated for each input variable for each taxon specific ANN model, illustrating the network output over the range of the varied input in the database. As an example, Figure 4.3 shows the sensitivity of *Cladocera* to 10 input variables.

The results of the sensitivity analysis are summarised in Table 4.2

The primary intention of this sensitivity analysis was to identify sensitive inputs for each model in order to improve network performance by removing insensitive inputs. However, this process also provided new insights into relationships between environmental variation and the occurrence of Queensland stream macroinvertebrates. Chapter 6 discusses detailed examples of such relationships and the potential of the technique to enhance our understanding of anthropogenic impacts on components of aquatic ecosystems.

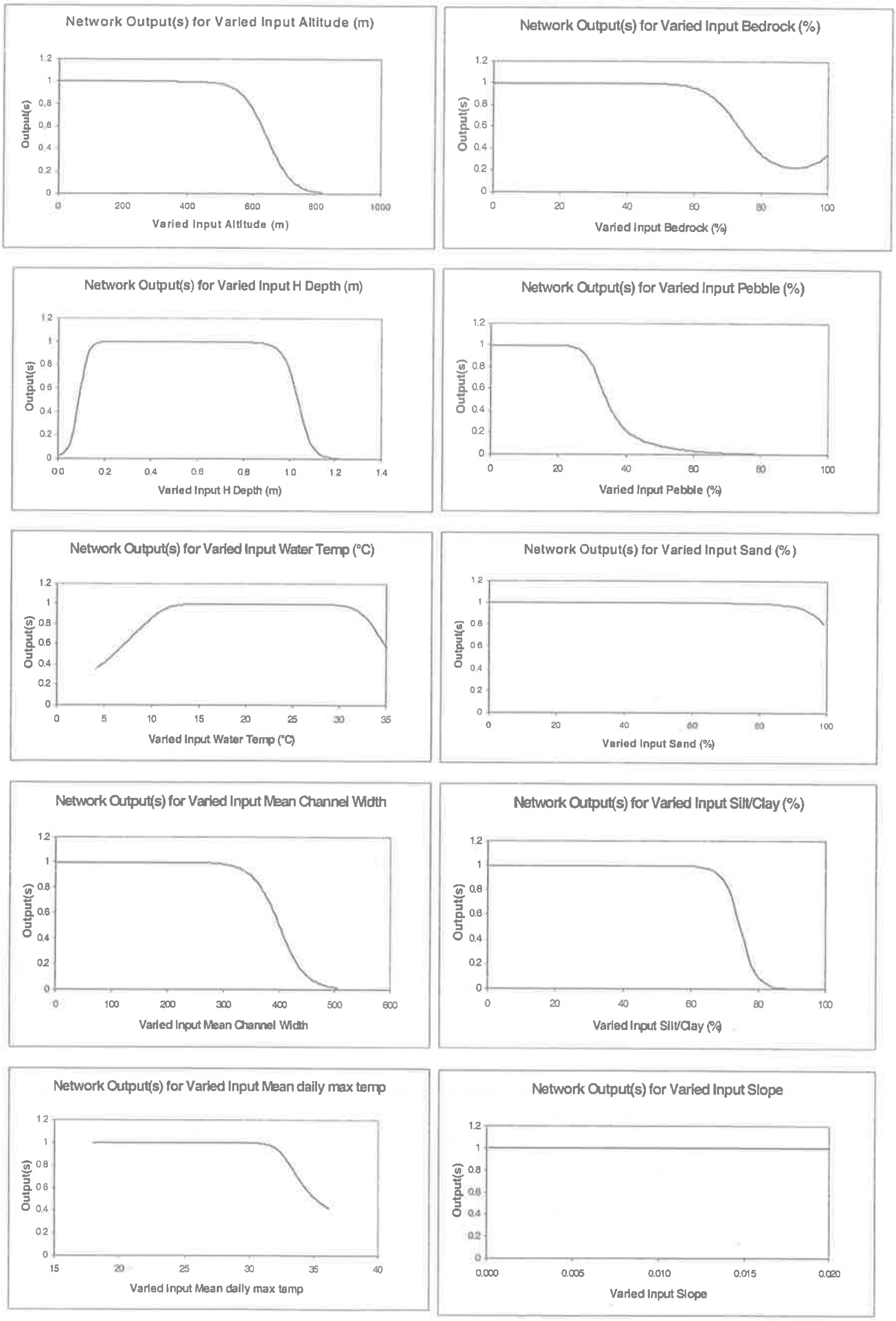


Figure 4.3 Sensitivity of the taxa Cladocera *to changes of inputs within their data range*

Table 4.2 Summary of the input sensitivity of 40 MI taxa by mean of percentages of output change for Clean Water Approach

(S)(Stt	Taxa	Latitude (S)	Longitude (E)	Altitude (m)	Stream Order	Slope	Distance FrSource	H Width (m)	H Depth (m)	Bedrock (%)	Boulder (%)	Cobble (%)	Pebble (%)	Gravel (%)	Sand (%)	Silt/Clay (%)
1	Oligochaeta	**98**	0	0.1	0.1	0.7	0.0	0.0	2.6	0.0	**96**	1.4	**63**	**100**	0.0	**69**
2	Acarina	**86**	**92**	**90**	**89**	**43**	**48**	**94**	**85**	**74**	**77**	**84**	36.0	**91**	**77**	**91**
3	Copepoda	**97**	3.1	**86**	**87**	**98**	10.0	2.8	2.9	2.7	**99**	**43.0**	**97**	**96**	**98**	31.0
4	Cladocera	12.0	21	**99**	0	0	1.7	0	**100**	**77**	**59.0**	**86**	**100**	5.3	18.7	**100**
5	Ostracoda	**100**	**82**	8.2	**98**	**40.5**	**77**	**95**	**98**	4.8	**95**	**72**	35.8	9.2	4.3	18.0
6	Atyidae	25	**45**	25	23	**47**	28	4.5	**40.6**	6.4	1.1	36.0	31.6	13.0	6.9	38.7
7	Palaemonidae	5.9	**90**	39	1.1	0.4	0.4	0.7	1.9	**100**	**99**	1.2	12.3	0.2	32.0	31.5
8	Leptophlebiidae	24	**41**	**49**	**97**	**47**	12.2	11.2	**89**	**97**	15.3	6.8	38.6	20.9	20.0	**80**
9	Baetidae	**96**	**96**	**90**	**87**	**64**	**87**	**65**	**93**	**93**	**81**	**95**	**97**	35	**96**	17.7
10	Caenidae	27	**83**	**98**	**76**	11.4	**66**	**95**	**90**	9.5	**97**	**99**	**59**	5.2	38	**99**
11	Gomphidae	**81**	**93**	**99**	**85**	**92**	**96**	30	**88**	17.2	**92**	**86**	**93**	36	**91**	**87**
12	Libellulidae	38	3.4	2.0	1.9	0.9	0.2	**100**	**100**	**83**	**90**	0.4	**95**	**97**	**96**	2.7
13	Coenagrionidae	0.1	5.2	25	0.7	0.1	0	0	**99**	0	0.7	35	**98**	0	1.0	13
14	Corixidae	**61**	**77**	22	9.9	34	28	**51**	**63**	32	**71**	**83**	**97**	**97**	**99**	**85**
15	Dytiscidae	**100**	**100**	**90**	**95**	**92**	**92**	**44**	**96**	**82**	**86**	**82**	**57**	**92**	**80**	**79**
16	Elmidae	**83**	2.6	**91**	**96**	**77**	**48**	**91**	8.2	11.1	14	**50**	32	30	**88**	**98**
17	Hydrophilidae	7.6	8.4	7.5	20.9	0.9	3.2	1.4	**75**	**83**	27	9.2	23	1.2	2.1	**99**
18	Tanypodinae	**64**	**96**	8.2	**74**	**97**	5.2	**98**	22	**84**	12.0	**46**	**82**	14.0	**89**	6
19	Orthocladiinae	**96**	**45**	**75**	**73**	**70**	**63**	**77**	**72**	**81**	**52**	**86**	**71**	**50**	**75**	**83**
20	Ceratopogonidae	**40.5**	**63**	**98**	35	7.7	**96**	3.8	**94**	**47**	12.1	11.7	**90**	22	20	10.4
21	Leptoceridae	1.8	**95**	1.2	11.3	6.1	**77**	2.7	4.3	**97**	**41**	13.3	**41**	5.0	0.8	**85**
22	Hydropsychidae	0.2	13	0	0	0	**97**	0.7	2.4	5.4	0	**98**	0.7	**45**	0	0.5
23	Hydroptilidae	16	1.9	**60**	7.9	25	**94**	**90**	**94**	13.8	**95**	**92**	9.8	**50**	**93**	**85**
24	Dugesiidae	**51**	2.1	3.8	2.2	2.6	2.4	**55**	27	17	2.3	5.0	13	6.9	34	10.0
25	Planorbidae	**88**	**97**	**84**	**73**	**95**	**97**	22	26	**93**	**82**	**96**	**98**	**96**	**80**	**98**
26	Thiaridae	**94**	**93**	**99**	**99**	7.1	48	0.5	47	**96**	**86**	**97**	**86**	**95**	**96**	0.6
27	Corbiculidae	2.1	16.5	**45**	32	29	**99**	14.0	**59**	4.8	3.0	6.2	6.8	14.5	7.6	8.6
28	Corduliidae	5.2	**66**	**65**	1.7	2.5	2.4	**85**	**96**	5.5	3.8	2.6	7.2	**51**	3.1	32
29	Notonectidae	0.2	12.0	0.1	0.2	0.2	0	2.0	**68**	**98**	0	0	0.1	0	5.2	10.6
30	Pleidae	6.3	0	0	0.5	0	0	15.8	14.8	0	0.4	0	0	0	1.1	**80**
31	Veliidae	**81**	**100**	**87**	36	**71**	**67**	28	**62**	36	**62**	26	33	**70**	**66**	**58**
32	Psephenidae	0	0.1	0.3	0	0	0	0	0.1	0	0	**99**	0	0	0	0.1
33	Simuliidae	14	**95**	**98**	12.0	13.0	11.7	**71**	**43**	**99**	4.5	**100**	**49**	9.5	20	10.0
34	Tabanidae	4.2	4.3	0	0	0	2.7	0	0	5.2	0	**98**	0	0	0	0
35	Ecnomidae	**65**	13.7	**62**	2.5	**97**	**99**	0.7	**69**	0.7	18.8	10.5	1.4	2.0	4.9	3.7
36	Calamoceratidae	22	5.9	**72**	10.0	16.7	**96**	36	**92**	15.9	39	**91**	33	**56**	37	**88**
37	Philopotamidae	0	0	0.2	0	0	0	0	0	0	0	0	**61**	0.2	0	0
38	Prosopistomatidae	0	0	0	0	0	0	0	0	0	0	0.2	0	2.1	0.4	0
39	Gripoterygidae	0.1	**81**	7.2	0.2	0	0	**53**	1	0	0	0.5	**60**	0	0	0
40	Pyralidae	**76**	6	**100**	2.9	0.8	0.7	1.1	2	20	**40**	4	0.7	15	0.5	1

Table 4.2 (continued)

(S)(Stt	Taxa	Water Temp (°C)	Alkalinity	Site-mean phi	Mean Wetted Width	Mean Channel	Mean Depth	Mean WSMR (a)	Mean DSMR (b)	Annual range in MMR	Range in WSMM	Range in DSMM	%rainfall inWS	Mean annual rainfall	Mean daily max temp	Mean daily min temp	Mean daily temp range
1	Oligochaeta	5.2	0.0	**64**	0.0	0.1	0.0	0.0	0.8	0.0	0	0	0	3.7	**96**	0.1	**95**
2	Acarina	**66**	**75**	**94**	**94**	**48**	**75**	**94**	**74**	**92**	**45**	37	**97**	**85**	**66**	**97**	**91**
3	Copepoda	**97**	4.0	**91**	6.1	20.5	1.3	8.9	**99**	5.3	24	2.7	**99**	2.3	2.8	37	**84**
4	Cladocera	**65**	0.0	7.0	0.0	**99**	0.0	**78**	40.0	27.0	0	0	0	**63**	**58**	5.2	10
5	Ostracoda	21.2	**98**	22.8	8.1	**85**	**75**	11.3	**97**	37.5	**94**	8.4	14.3	7.5	**96**	**95**	7.1
6	Atyidae	27.0	**42.0**	28.2	9.1	8.2	16.7	7.1	17.0	6.0	38	7.6	0	36	15.6	6.7	**41**
7	Palaemonidae	20.4	**52.0**	4.8	0.9	0.5	2.2	26.0	**93**	0.3	23	16.0	1.9	17.9	**99**	**97**	**100**
8	Leptophlebiidae	**98**	**96**	**47.5**	**94**	**100**	**81**	**71**	**57.4**	**52.6**	**99**	**77**	**87**	5.5	4.6	**60**	**98**
9	Baetidae	**100**	**97**	**92**	**47**	**45**	**77**	**58**	38	21	**98**	**83**	20	**48**	**96**	**97**	**95**
10	Caenidae	**94**	**85**	**45**	7.0	**88**	**99**	**66**	**84**	**94**	**100**	**46**	5.7	22	**93**	**100**	**90**
11	Gomphidae	**71**	19.0	16.0	**57**	**99**	8.2	14.5	11.7	**78**	**97**	**82**	6.3	**84**	**90**	**57**	**97**
12	Libellulidae	**100**	0.5	8.9	1.3	12.0	0.2	**100**	5.7	**93**	**76**	0.6	**67**	0.7	**100**	2.3	**45**
13	Coenagrionidae	0.2	0	0.3	0	0	0	**93**	0.8	**87**	**84**	40	33	0	30	3.2	**70**
14	Corixidae	18.1	**71**	35	**73**	**58**	**84**	14.4	24	29	**46**	**54**	**100**	13	22	16.7	**90**
15	Dytiscidae	**100**	**92**	**98**	**97**	34	26	**95**	**86**	**88**	**99**	**49**	**91**	36	**77**	**79**	**95**
16	Elmidae	8.1	**99**	37	**89**	77	9.2	**99**	**99**	**96**	**45**	9.0	**94**	**90**	**94**	**99**	**94**
17	Hydrophilidae	2.8	**94**	**69**	1.5	10.9	0.2	36	**96**	6.3	**88**	1.9	1.7	**87**	6.7	**97**	6.0
18	Tanypodinae	**100**	17.0	39	31	**99**	9.7	35	8.3	3.7	6.3	8.3	1.3	27	7.1	5.6	**83**
19	Orthocladiinae	**99**	**72**	**96**	**73**	**70**	**78**	**87**	**48**	**95**	**72**	**90**	**71**	**77**	38	**46**	**64**
20	Ceratopogonidae	**77**	**63**	2.7	2.8	6.2	**93**	**67**	2.9	23	**100**	**95**	3.8	3.7	**70**	**94**	**65**
21	Leptoceridae	23	8.0	10.0	**80**	18.6	9.7	2.9	34	**70**	9.6	**90**	**98**	0.9	9.7	1.0	**99**
22	Hydropsychidae	0.3	1.4	0.3	0	7.2	0	2.7	0	1.0	4.8	0.1	0	0	4.5	**83**	0
23	Hydroptilidae	**50**	21	**53**	6.7	**63**	3.1	**49**	**95**	**84**	**93**	**44**	**93**	**81**	**92**	**95**	**92**
24	Dugesiidae	**100**	2.3	14.7	2.5	22	19.5	14.2	**63**	3.5	2.7	2.6	8.5	11.7	**99**	**61**	20
25	Planorbidae	**100**	4.9	15.6	12.7	**66**	31	**93**	**77**	17.3	**65**	15.0	**82**	**71**	**95**	**97**	**98**
26	Thiaridae	**99**	4.7	9.0	2.7	4.6	0.2	2.8	**92**	4.5	**97**	**86**	**99**	0.7	**73**	**77**	**90**
27	Corbiculidae	**100**	**98**	2.3	**88**	3.2	19.8	**72**	**99**	2.0	**83**	**92**	12.3	15.8	**58**	9.3	5.7
28	Corduliidae	5.7	6.5	4.5	1.6	2.3	1.0	3.2	1.0	**93**	28	4.6	7.0	1.2	**41**	**94**	3.7
29	Notonectidae	0	2.0	0	0.3	0.5	0	2.7	0	0	0.5	0	**60**	1.7	**69**	0.8	0
30	Pleidae	0.2	0.2	0.3	0	**49**	0	0	0.2	0	0	0	3.4	0.1	**83**	2.7	0
31	Veliidae	**97**	39	**57**	**83**	**64**	15.9	**81**	**91**	**60**	**78**	24	**76**	**73**	**71**	34	**94**
32	Psephenidae	0.6	0	2.1	1.0	0.8	0.4	0	0	0	0	0	0	0	0.3	0	0
33	Simuliidae	31	4	**82**	10.8	10.6	2.5	10.6	19.5	6.2	**59**	4.1	0.8	38	8.5	6.2	**97**
34	Tabanidae	0	0	0	0	0	0	0	0	0	0.1	0	0.5	0	0	0	11.0
35	Ecnomidae	**100**	16.7	1.8	**79**	0.6	2.6	9.7	1.6	1.6	10.7	1.7	**63**	**74**	**44**	3	**47**
36	Calamoceratidae	**81**	**81**	13.0	18.0	24	4.6	**76**	**73**	**80**	**94**	17.5	**99**	20.6	20.1	12.6	15
37	Philopotamidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
38	Prosopistomatidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	Gripoterygidae	0	0	0	0	0	0	4	**100**	0	0	0	0	**40**	6	0.3	0
40	Pyralidae	**70**	1.5	15	11	**90**	**47**	4	11	0	**100**	15	1	5	0.6	6	0.8

4.3.2 Network Revision

An input showing a range of output changes less than 40% over its range was considered redundant and was excluded from the network architecture (Chapter 3). After excluding defined redundant inputs, the neural networks were revised for each taxon. Inputs for percentages of substrate compositions are closely linked to each other. Therefore, network refinements did not take into account those inputs, although sensitivity analyses had been conducted for them. These inputs together with empirical categorical variables, which had not considered as part of the sensitivity analysis, remained in the revised neural network models.

Some taxa appeared to be very sensitive against changes of most of the inputs, including *Acarina, Leptophlebiidae, Dytiscidae, Baetidae and Orthocladiinae.* Therefore, criteria to consider inputs as sensitive were extended for these cases to 50% of output change over the range of input change, in order to consider only the most sensitive inputs for improving these ANNs. The numbers of input nodes were different for different taxa. Table 4.3 summarises the number of input nodes for each taxa specific network after revision

Table 4.3 Number of inputs in the taxon-specific revised networks

N	Taxa	Number of inputs	Number of nodes in HD	N	Taxa	Number of inputs	Number of nodes in HD
1	Dugesiidae	19	10	21	Corixidae	25	15
2	Oligoçhaeta	17	10	22	Notonectidae	17	10
3	Planorbidae	29	15	23	Pleidae	15	10
4	Thiaridae	27	15	24	Veliidae	31	15
5	Corbiculidae	24	10	25	Dytiscidae	35	15
6	Acarina	35	15	26	Elmidae	33	15
7	Copepoda	22	10	27	Psephenidae	13	10
8	Cladocera	20	10	28	Hydrophilidae	20	10
9	Ostracoda	27	15	29	Tanypodinae	21	10
10	Atyidae	18	10	30	Orthocladiinae	36	15
11	Palaemonidae	19	10	31	Simuliidae	20	10
12	Leptophlebiidae	31	15	32	Ceratopogonidae	27	15
13	Baetidae	34	15	33	Tabanidae	13	10
14	Caenidae	34	15	34	Leptoceridae	20	10
15	Prosopistomatidae	13	10	35	Hydropsychidae	15	10
16	Gomphidae	32	15	36	Ecnomidae	24	10
17	Corduliidae	20	10	37	Hydroptilidae	29	15
18	Libellulidae	22	10	38	Calamoceratidae	23	10
19	Coenagrionidae	18	10	39	Philopotamidae	13	10
20	Gripopterygidae	17	10	40	Pyralidae	19	10

Taxon-specific models were developed for each macroinvertebrate taxon as a result of revision. Each taxon required a specific set of input variables. The numbers of nodes in the hidden layers were selected after a series of optimisation trials. It appeared to be optimal that models having more than 25 variables as inputs were developed with 15 nodes in the hidden layer, and models with less than 25 inputs were developed with only 10 nodes in the hidden layer.

Other technical parameters including learning rates for hidden and input layers, momentum remained the same as for initial models. Neural network models were trained with 5000 iterations.

4.4 Validation Results and Comparison with AusRivAS model

4.4.1 Validation Results

The settled target, to achieve correct predictions for 70% of the cases in the validation data set of 180 reference sites, was achieved. After refining each taxon - pecific ANN model based on sensitivity analysis and validation, the chosen condition for the network performance was achieved. The overall rate of correct predictions of taxa presence at stream sites in the validation data set by all models was better than 70 %. Table 4.4 and Figure 4.4 illustrate the development of correct predictions from the initial models prior to sensitivity analysis to the final refined models.

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

No	Taxa	Initial (39 inputs)	*Revised*	No	Taxa	Initial (39 inputs)	*Revised*
1	Prosopistomatidae	98.37	100	21	Cladocera	76.04	82.04
2	Gripoterygidae	92.39	96.20	22	Leptoceridae	75.45	81.44
3	Pleidae	91.01	92.81	23	Dugesiidae	74.25	80.24
4	Philopotamidae	90.41	92.81	24	Libellulidae	73.65	79.64
5	Planorbidae	89.82	91.62	25	Veliidae	71.86	75.45
6	Notonectidae	89.82	89.82	26	Oligochaeta	70.06	77.84
7	Tabanidae	87.43	89.22	27	Ostracoda	70.06	77.25
8	Hydropsychidae	86.83	91.62	28	Tanypodinae	68.86	76.65
9	Psephenidae	85.63	88.02	29	Atyidae	68.26	77.84
10	Coenagrionidae	83.23	83.83	30	Baetidae	68.26	72.65
11	Pyralidae	83.11	86.69	31	Caenidae	67.66	73.05
12	Simuliidae	82.63	86.23	32	Hydroptilidae	67.66	79.64
13	Calamoceratidae	82.04	85.03	33	Orthocladiinae	66.47	70.65
14	Corbiculidae	80.84	86.23	34	Dytiscidae	65.87	72.05
15	Ceratopogonidae	80.24	83.83	35	Elmidae	65.87	73.65
16	Thiaridae	80.24	86.23	36	Copepoda	65.27	79.64
17	Leptophlebiidae	79.04	83.23	37	Corixidae	65.26	76.05
18	Ecnomidae	79.04	84.43	38	Palaemonidae	64.07	73.65
19	Acarina	77.84	79.04	39	Hydrophilidae	64.07	74.85
20	Corduliidae	76.65	83.23	40	Gomphidae	63.69	73.05

Results demonstrate that all macroinvertebrate models had been improved, even though specific improvement rates differed greatly. The *Copepoda* model achieved the highest improvement with correct prediction increasing by 14.37%, followed by the *Gomphidae* model with 10% improvement in correct prediction. Performances of the model developed for *Notonectidae* remained unchanged. The mean value of correct predictions for the validation set was 76.74% for the initial models and 82.16% for taxa specific models after exclusion of redundant inputs.

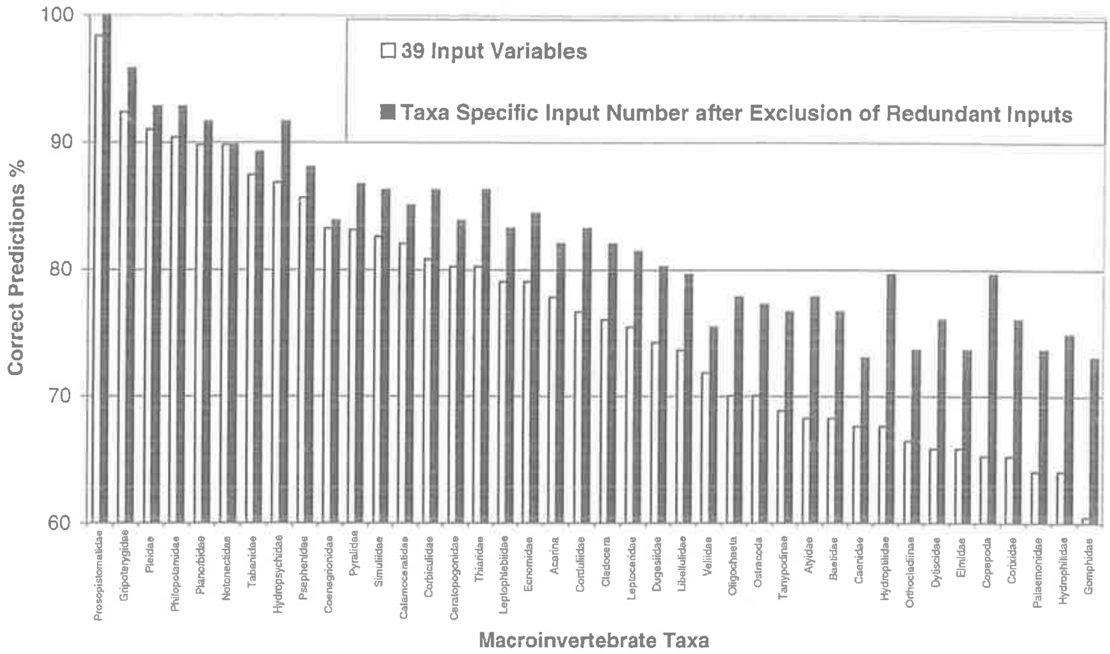


Figure 4.4 Correct predictions of presence/absence of macroinvertebrate taxa before and after exclusions of redundant input

To promote further validation of the ANN models, predictions of stream sites using the training and validation data sets were conducted and evaluated by the ratio of observed to predicted (O/E) data. An O/E value in the range from 0.8 to 1.2 was selected as corresponding with the central 80% of reference sites to indicate whether a specific site was biologically degraded or not. This criteria was previously suggested for applications of AusRivAS models (Coysh et al., 2000). Under optimal conditions, the O/E values of both models should meet this range for reference samples, as reference samples were used for developing and validating both models.

The results in Fig. 4.5 clearly indicate that there is a 95.6% correspondence with the reference sites for the training data. The 4.4% of sites outside the range were the sites with extreme conditions. Even though they belong to reference conditions, only 4 or fewer from a total of 40 taxa were observed at 2.71% sites having O/E< 0.8 (MRHI1/250, MRHI2/456 MRHI3/200, MRHI4/067, MRHI2/299, MRHI4/172, MRHI4/294).

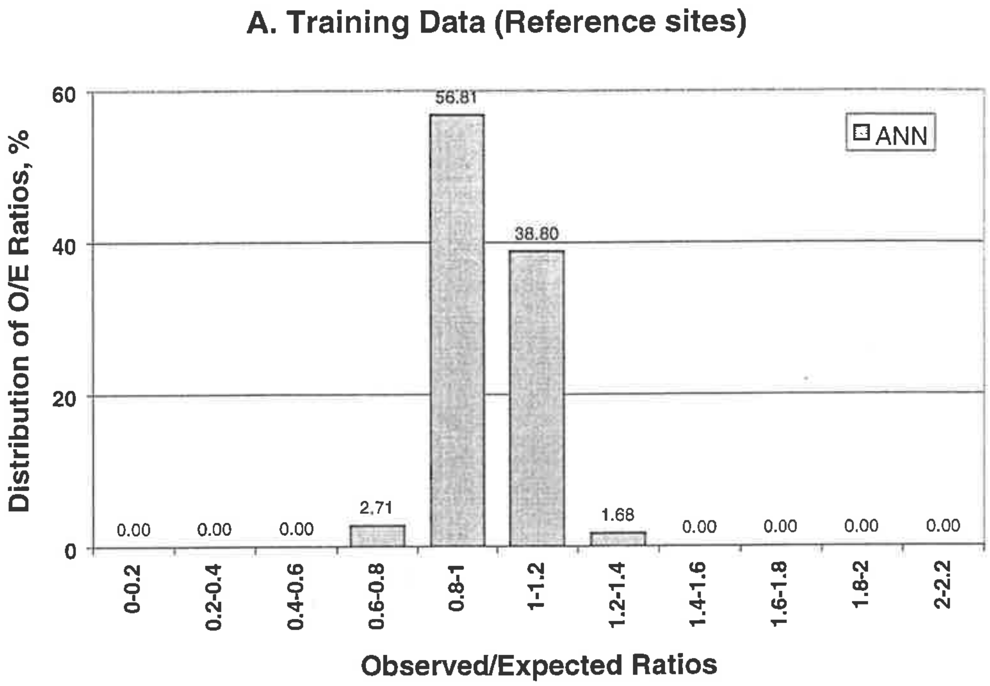


Figure 4.5 O/E criteria for training set

Figure 4.6 shows the validation results for 180 sites in the validation set. 85.03% correspondence with the reference sites for the validation data. Again, sites with given O/E criteria outside the range 0.8-1.2 appeared to have extreme number of taxa observed. The following 13 sites (MRHI1/311, MRHI3/161, MRHI1/300, MRHI2/484, MRHI1/012, MRHI2/488, MRHI4/319, MRHI1/051, MRHI1/388, MRHI2/345, MRHI3/160, MRHI4/141, MRHI4/259) having 7 or less taxa had been predicted with O/E <0.8. Other 11 sites (MRHI1/529, MRHI3/072, MRHI3/113, MRHI4/349, MRHI3/019, MRHI3/111, MRHI4/112, MRHI4/128, MRHI4/129, MRHI3/137) having 22 or more taxa had been predicted with O/E >1.2. However the value of O/E predicted for these sites were only slightly outside the range indicating the reference conditions.

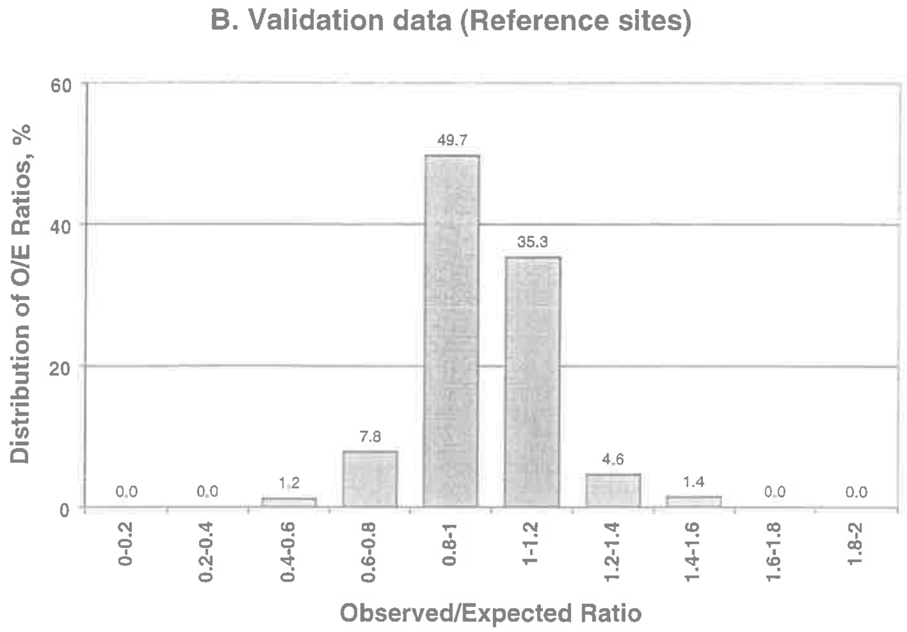


Figure 4.6 O/E criteria calculated for validation set

Even though some models had failed to predict O/E criteria for extreme sites, the good validation results for majority of the ANN models finally justified their application to the prediction and assessment of stream test sites.

4.4.2 Comparision with AusRivAS model

In order to evaluate the relative performance of the ANN models, a comparison with the model AusRivAS (Coysh et al., 2000) was carried out. It was applied to the Queensland stream system based on the same training and validation data of reference sites as used for ANN modelling.

Data restriction

There were no sites of macrophyte habitat considered in this comparison as no AusRivAS models have been developed for macrophyte habitats. To obtain a site assessment by AusRivAS, the appropriate biological and habitat data from the test site under investigation were entered and preliminary analyses are performed to determine whether the test site fell within the experience of that model. Any sites with no appropriate reference group for comparison were identified "outside the experience of the models" (Coysh et al., 2000). These sites were also excluded from the data set. Therefore, the data sets were reduced. The O/E ratios were calculated for the training data (data from 600 reference sites used to generate the model), and the validation data (data from 160 randomly withheld reference sites).

In both cases these calculations are restricted to the 40 taxa. Rare taxa had to be removed according to AusRivAS protocols prior to classification. Taxa occurring at less 10% of sites if there are less than 100 sites or taxa occurring at less than 10 sites if there are more than 100 sites considered rare taxa. *Prosopistomatidae* occurred at less than 10 sites from total 760 studied sites and they were removed from models. Two models were therefore comparable.

Comparison results

The O/E ratios number of taxa for samples were calculated from AusRivAS output. In accordance with the AusRivAS methodology (Simpson et al., 1997), the expected number of taxa for each sample was calculated as the sum of probabilities of occurrence of taxa with a 50% or greater probability of occurrence. Ratios O/E number of taxa was also calculated from ANN output. The expected number of taxa as sum of taxa was predicted by models after data preprocessing (Hoang et al., *in press*).

The comparison of O/E data calculated from outputs of the ANN and AusRivAS models are plotted in Figure 4.7 and 4.8.

The results of the model comparison in Fig. 4.7 clearly show that the ANN models identifying much better the reference sites as unimpaired than AusRivAS. While the ANN models identified the correspondence of training data to reference sites by 94.2%, AusRivAS did only by 60.10%.

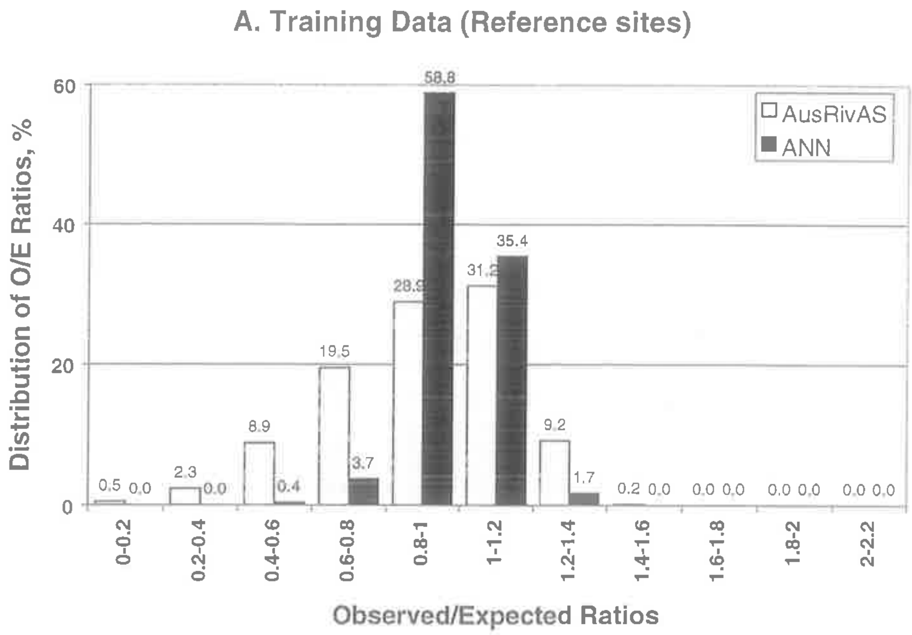


Figure 4.7 Distribution of Observed/Expected ratios for predicted stream sites using the AusRivAS and the ANN models.

The ANN models performed similarly well for the validation data (81.07%) compared to 54.71% by AusRivAS models (Figure 4.8).

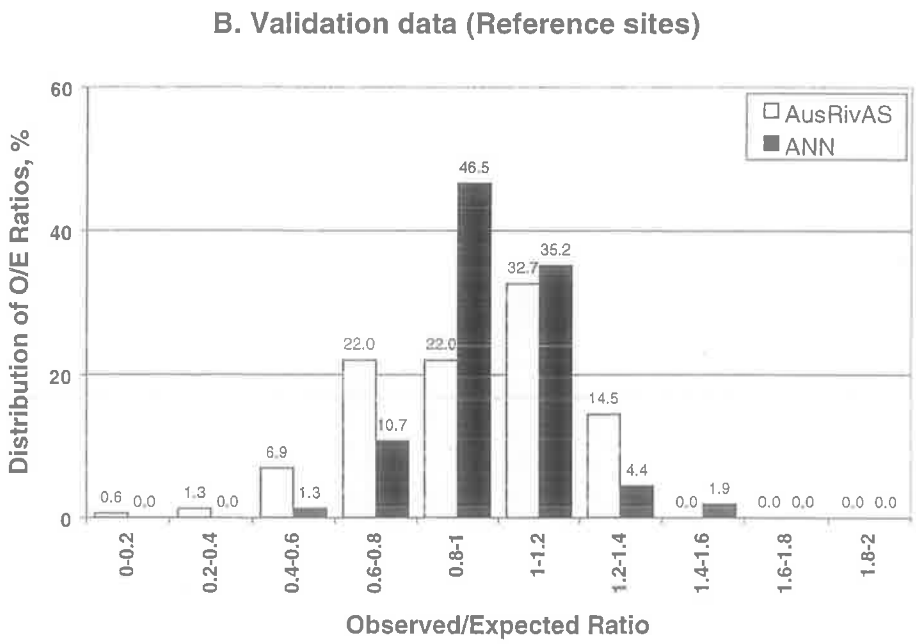


Figure 4.8 Distribution of Observed/Expected ratios for predicted stream sites using the AusRivAS and the ANN models.

AusRivAS failed to predict extreme sites as discussed earlier. Moreover, the values of O/E predicted by AusRivAS for extreme poor sites was <0.4. AusRivAS even failed to cope with some normal sites, which are defined to have number of taxa in the range of mean $\pm$ SD (calculated as 14.64 ± 4.36 for whole data set).

The results of this study clearly illustrate the excellence of the ANN models compared to the AusRivAS models in terms of validity.

4.5 Apply for Prediction Step – Prediction Results

Test sites refer to the site being tested by model regarding biological impairment. The test sites may be sites with unknown or suspected impacts, sites selected for regional assessment or reference sites resampled for periodic testing of the model. It is important that all stream and river types, which may be represented by the test sites, have been sampled at sites considered to be equivalent to reference conditions. That is to ensure that test sites will be compared against reference conditions that can be expected for in the absence of impact (Coysh et al., 2000).

The application of the validated ANN models to an independent test data set of stream sites resulted in a slightly positively skewed distribution of the O/E data (Fig. 4.9) that might be representative for a field data set.

Whilst the majority of data (36.15 %) were in the range of 0.8<O/E<1.2 and indicated no degradation effects of corresponding stream sites, 16.82 % of the data indicated mildly to moderately impaired sites. In 47.02 % of the data richer invertebrate communities were indicated than observed at reference sites that might be subject to research for causal clarification.

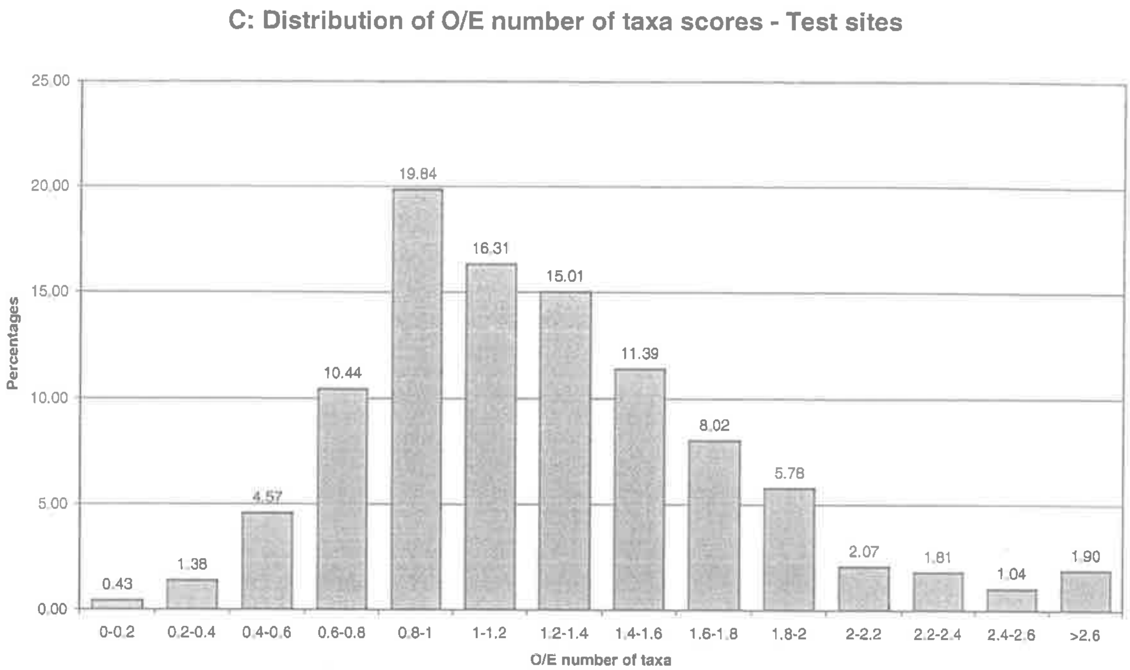


Figure 4.9 Application of ANN to predict biological impairment of test sites

Table 4.5 shows the average number of taxa present in each range of O/E. As hypothesised, the reference sites showed significantly higher taxon richness than possibly degraded sites

Table 4.5 Average numbers of taxa observed at sites in according environmental conditions assessed by the value of O/E criteria

Value of O/E	Average number of taxa at sites (average number of taxa at reference sites – 14.64 taxa)	Environmental condition classified by AusRivAS
O/E> 1.2	17.7 ± 3.63	Richer invertebrate community than pristine – potential nutrient enrichment
1.2≥O/E≥0.8	13.73 ± 3.23	Near pristine condition
0.8>O/E≥0.4	9.19 ± 2.65	Mildly to moderately impaired site
O/E<0.4	3.18 ± 1.51	Moderately to severely degraded sites

This implies that the reference sites were well chosen and showed minimal impairment relative to the test sites.

4.6 *Discussion*

4.6.1 Performance of Artificial Neural Networks

Results of the comparison of the ANN models with AusRivAS models showed that the performance of the ANN models is in general better than statistical predictions. Walley & Fontama (1998), Schleiter et al. (1999) and Gabriels et al. (2000) also came to similar conclusions for predicting macroinvertebrates in freshwater streams and rivers based on different sets of environmental characteristics.

ANNs provide similar good results for validations with both the complete and the restricted database used to compare with AusRivAS models, in which macrophyte habitat could not be assessed. The results showed that ANNs are able to work with all habitat types provided that samples from these habitats have been included in the learning procedure.

Reliability is not only obtained by the correct prediction of presence/absence of macroinvertebrate taxa at sites. For all taxa, numbers of sites where taxa presented were similar for predicted and observed data in the validation set.

Table 4.6 Numer of sites where taxa were present in total 180 validation sites (results from revised models)

No	Taxa	Observed	Predicted	No	Taxa	Observed	Predicted
1	Prosopistomatidae	2	2	21	Cladocera	45	47
2	Gripoterygidae	17	19	22	Leptoceridae	119	122
3	Pleidae	20	21	23	Dugesiidae	37	38
4	Philopotamidae	26	24	24	Libellulidae	64	70
5	Planorbidae	22	19	25	Veliidae	39	42
6	Notonectidae	29	26	26	Oligochaeta	72	70
7	Tabanidae	23	24	27	Ostracoda	58	62
8	Hydropsychidae	54	58	28	Tanypodinae	130	124
9	Psephenidae	33	35	29	Atyidae	93	94
10	Coenagrionidae	62	57	30	Baetidae	123	128
11	Pyralidae	38	37	31	Caenidae	119	123
12	Simuliidae	41	40	32	Hydroptilidae	59	61
13	Calamoceratidae	24	26	33	Orthocladiinae	85	91
14	Corbiculidae	26	26	34	Dytiscidae	67	72
15	Ceratopogonidae	62	57	35	Elmidae	75	82
16	Thiaridae	45	44	36	Copepoda	98	98
17	Leptophlebiidae	93	96	37	Corixidae	80	76
18	Ecnomidae	32	37	38	Palaemonidae	88	90
19	Acarina	118	117	39	Hydrophilidae	48	50
20	Corduliidae	30	29	40	Gomphidae	82	82

This fact demonstrates that correct prediction provides the meaningful information for the reliability of network performance. Predictions of rare taxa like *Prosopistomatidae, Gripoterygidae* and *Philopotamidae* are examples. If networks could not generalise presence patterns for these taxa from database but only statistically predicted that these taxa never occur in any site, this information had already achieved more than 90% correct prediction, because these taxa occured at less than 10% of sites observed. The high percentages of correct prediction of taxa present at around 50% of sites such as *Leptophlebiidae, Ceratopogonidae, Leptoceridae, Oligochaeta, Atyidae, Copepoda* are really a significant achievement.

The validation results proved that the ANN models had worked successfully not only with very common but also with very rare species such as *Prosopistomatidae, Gripoterygidae, Pleidae,* and *Planorbidae.* These results showed that ANN models could work with all taxa with nonzero probabilities of presence at sites.

4.6.2 Relationship between Predictor Variables and Macroinvertebrate Assemblages

The results of the neural network models developed indicated that physical predictor variables have close relationships with presence/absence of macroinvertebrates. Geographical predictors easily measurable from maps, such as latitude, longitude, altitude and distance from sources explain significant variation in the benthic macroinvertebrate community. These variables appeared to be highly sensitive for 32 out of the 40 taxa in this study (see examples in Figure 4.10). Corkum (1989) and Bailey et al. (1998) also found geographical predictors to be useful in studing benthic invertebrate communities in various streams in North America. This information makes the test of a community's deviation from reference conditions much more sensitive. Therefore, taking advantages of the predictable component of variation in benthic invertebrate communities with these easily measured geographical variables can improve biological assessments of the community expected at test sites.

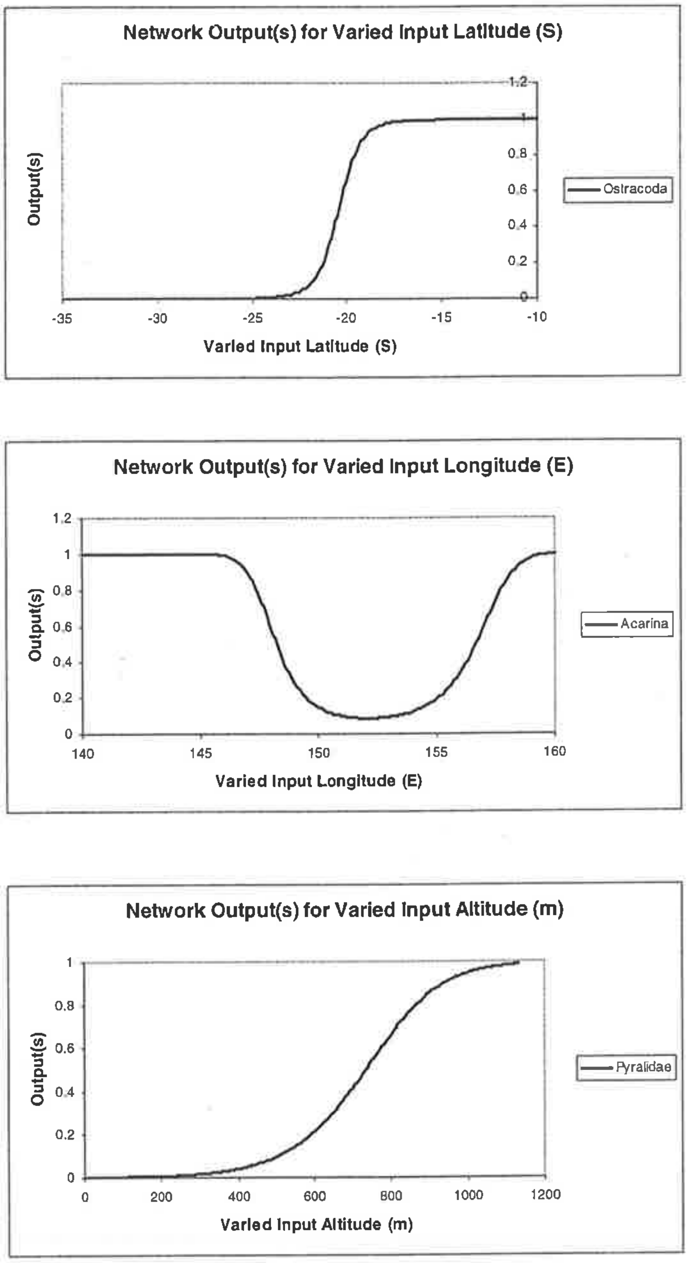


Figure 4.10 Relation between geographical inputs and distribution of macroinvertebrates

Sensitivity analysis showed relatively consistent changes between seasons for 16 taxa. Figure 4.11 illustrates some of these changes. Other taxa showed no clear change according to seasons.

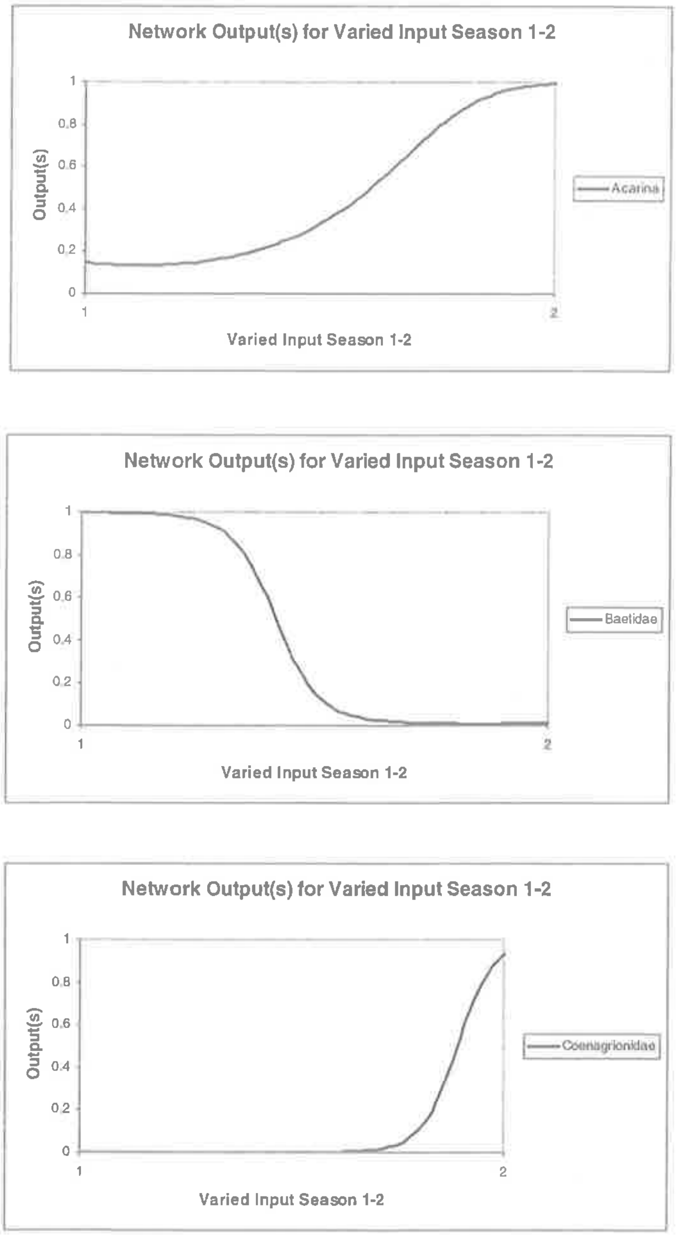
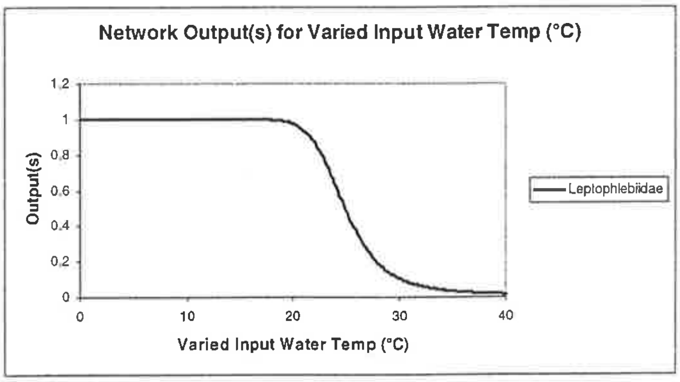


Figure 4.11 Relation between seasons and macroinvertebrate distribution

This seasonal variability is also reflected in water temperature, which is highly sensitive for 21 taxa. Dodelec (1989) found that many macroinvertebrate taxa, which are typical for high quality water, favoured cold water. Cool habitats are also favorite by aquatic insects as cool water contains more oxygen at saturation than does warm water (Ward, 1992). These earlier findings are supported by results of current project (Figure 4.12).



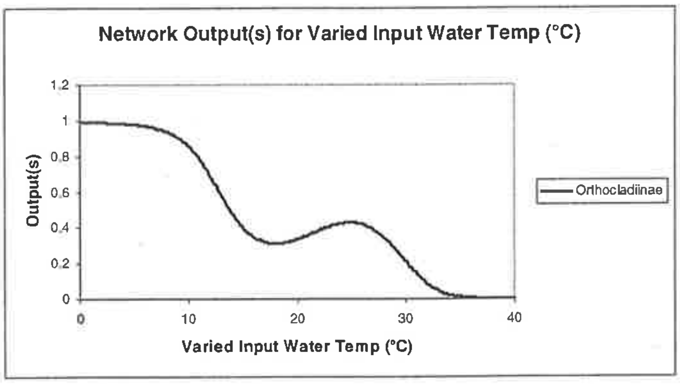


Figure 4.12 Relation between water temperature and distribution of aquatic insects

Substrate compositions were found to be very sensitive for most of the taxa even for rare taxa such as *Philopotamidae* and *Gripopterygidae*. For the taxa *Psephenidae, Philopotamidae* and *Tabanidae*, substrate compositions are the only driving variables for their distributions in the streams. The presence of only a few taxa (*Atyidae, Dugesiidae, Corbiculidae, Ecnomidae* and *Prosopistomatidae*) does not rely on substrate compositions.

Climatic and meteorological variables also provided significant impacts on macroinvertebrate distribution. The presence of all taxa, except for four rare species, relies on weather conditions. Among these variables, *range in wet season monthly means*, *mean daily max temperature* and *mean daily temperature range* are the most important variables. Sensitivity analyses showed that they are highly sensitive for most taxa. The high sensitivity of these easily obtained hydrological and meteorological variables demonstrated the applicability of the proposed methods.

Topographical characteristics also play an important role in the presence of macroinvertebrates. *Habitat Width* and *Depth* were highly sensitive for 28 taxa. Slope was less sensitive, although it still was greatly sensitive for 10 taxa and moderately sensitive for 4 others. Stream order was very sensitive for 13 taxa, while Site-mean phi was significant for 12 taxa.

More details of the sensitivity analysis and relationships between predictor variables and distribution of macroinvertebrates are discussed in Chapter 6.

4.6.3 Reference Condition Approach for Rapid Assessment

The reference condition approach is considered one of the most effective ways of using the information available from biological communities to established "biocriteria" such as O/E (Bailey et al., 1998). This approach is widely used for the rapid assessment of river health (Resh & Jackson, 1993).

The first objective of using rapid assessment is to reduce efforts and costs in assessing environmental conditions at a site. This can be achieved by: (1) reducing the number of habitats sampled and replicate sample units taken per habitat, (2) considering only a fraction of the animals collected, which means fewer have to be identified.

Another objective of rapid assessment approaches is to summarise the results of site surveys in a way that can be understood by non-specialists such as managers, decision-makers and the concerned public. This is done by using analysis measures that express results as single scores, as well as by placing the scores obtained in categories of environmental quality based on regional data (Resh & Jackson, 1993).

The success of the rapid assessment approach ultimately depends on the ability to detect impacted and unimpacted conditions. A test community falling outside of the range "fails", while a community that is within this range "passes". Communities either above or below criteria values may fail because they are unusual relative to reference communities (Bailey et al., 1998).

At the assemblage level, presence–absence data appear sufficiently strong to allow detection of reasonably subtle differences among sites. Previous works had shown that O/E could distinguish four bands of degradation with acceptable errors (Wright, 1995; Norris, 1996, Clarke et al., 1996). These bands represent classes of biological impairment ranging from 'equivalent to reference' to 'poor'. The robustness of presence-absence data in allowing an assessment of biological condition has important implications for bioassessment.

However, the clean water approach applied in this work appeared to have some limitations. Reference conditions are classified as near pristine condition. In practice, a reference condition of nearly pristine can hardly be found at the present time; they only represent the least impaired sites within the area of interest (Hawkins et al., 2000). Even site, which meets all requirements for reference conditions, may experience many types of disturbance. The empirical foundation of the method itself therefore contains some information relating to pollution. Test sites can only be judged by the relative conditions to the reference. Moreover, the approach cannot be applied for areas, where near pristine or even least impaired sites no longer exist, and where data of study sites in pristine conditions has never been collected. Therefore, it is extremely difficult to obtain appropriate databases for training the networks.

Moreover, the O/E criteria showed some problems in application for management purposes. O/E takes into account only the richness, the number of taxa present at a site, as a single descriptor to use as a basic biocriterion for bioassessment work. It assumes that the importance of all taxa is the same, while in fact the responses of some taxa are different to different environmental stressors (Hellawell, 1986). The presence of some indicator families actually indicates conditions with some degree of pollution. Therefore, the predicted taxon list is also required to provide a "target" invertebrate community to rectify identified impacts. The type of predicted taxa may also provide clues as to the type of impact that a test site is experiencing. For example, the absence of predicted *Leptophlebiidae* might indicate an impact on a stream from trace metal input (Hellawell, 1986). Families in *Ephemerotera, Plecoptera* and *Tricoptera* orders are taxa identified in some Rapid Assessment Protocol (Plafkin et al, 1989; Resh & Jackson, 1993) because they provide more "indicator value" than others taxa, and distribution of these taxa can give elucidation for environmental stressors (Ward, 1992).

Interpretation of the O/E criterion is not very clearly understood. The hypothesis was that low O/E ratio (O/E<<1) implies that test sites are adversely affected by some environmental stressors. Practical examples showed that the assessment score of test sites did not vary in understandable ways. Interpretation for sites having O/E >>1 is poorly understood.

AusRivAS suggested that higher levels of the criterion O/E exceeding 1 are probably caused by nutrient enrichment (Coysh, 2000). However, it is argued that although intermediate levels of organic enrichment may favour certain suspension- or deposit-feeding macroinvertebrate groups, changes in substratum and low dissolved oxygen concentrations that often occur at high levels of organic enrichment usually result in the disappearance of intolerant taxa (Hynes, 1960; Hellawell, 1986).

Organic enrichment in the aquatic ecosystem is one of the oldest and most fully documented forms of pollution. Several taxa either increase or decrease in response to organic effluents in rivers. At intermediate levels of organic enrichment, numerical changes occurred mostly within the established, ambient taxa complex within each system. However, as the severity of organic pollution increased, system similarities increased as well. Ultimately this can be result in a predominance of some species such as *Chironomus riparius* pupae at high pollution levels (Johnson et al., 1993).

4.7 Chapter Summary and Conclusion

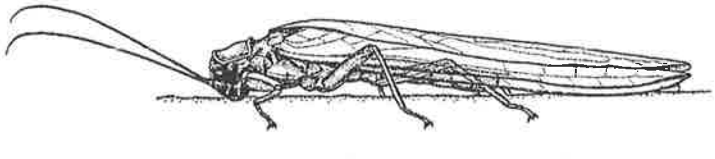
The Artificial Neural Network models developed for Queensland stream systems are producing promising results. The "clean water approach" models adopted the referential condition approach from the AusRivAS models to provide first-pass assessment of rivers over large areas and to help identify what additional information is required. Better results of the ANNs models compared to AusRivAS showed that ANNs can be suitable alternative tools to study distinct non-linear relationships within freshwater ecosystem. The clean water approach provides qualitative assessments and should be satisfactory for state-of-the-environment reporting (Smith et al., 1999).

The usefulness of the reference condition approach can be increased not only by modelling but also by explaining variation in the community descriptors among the reference sites. This elucidation possibility of predictive models can be achieved by means of sensitivity analyses with ANNs. Further details are discussed in Chapter 6.

The clean water approach can successfully be used in bioassessment if the community descriptor adequately summarise the nature of community present at a reference site, and ideally respond when degradation by the stressors is present. The predictor variables must describe, either directly or indirectly, the habitat of stream pertinent to the invertebrate community, be relatively easy to measure, and be unaffected by potential stressors.

The most important element of this approach is the explanation of community variation with variation in habitat conditions at the sites. Clear and fully reliable elucidation of these variations is unfortunately not available.

The alternative to this are "dirty water" models, which utilise a wider range of input variables including those that describe chemistry of water quality. The aim of a dirty water model is to identify input variables that exert some influence on outputs, and then run simulations of various scenarios through the model to predict the ecological consequences of altering input variables. Full details of these models are discussed in Chapter 5.



Adult Gripoterygidae, *presence of which is a good indicator for clean water condition (Gullan & Cranston, 2000)*

5 Developing Dirty Water Approach

5.1 *Introduction*

5.1.2. Dirty Water Approach

It has been widely demonstrated that interactions among chemical and physical processes create environmental conditions at a range of scales that strongly influence the distribution and abundance of lotic taxa, and thus the composition of lotic assemblages (e.g. Hynes, 1970). Many studies have identified substrate composition, complexity and heterogeneity as major determinants of in-stream biota (e.g. Ward, 1992). Turbidity, an optical property of water and a measure of light attenuation by suspended particular matter, plays an important role in the life of benthic macroinvertebrates. Suspended solid loads < 40 mg/l above normal levels resulted in a 25% reduction of aquatic insects of riffles. Densities were reduced 60% when suspended solid were 120 mg/l or more above normal (Gammon, 1970). Dissolved oxygen plays a major role in spatial temporal distribution patterns of aquatic insects (Ward, 1992). Responses include microspatial positioning, depth distribution, migratory behaviour, and predator-prey interaction. pH influences food availability and leads to differences in species richness, the distribution of species among order of insects in different conditions of acidity in watershed (Otto & Svensson, 1983). A survey of streams in the southeastern United States revealed that mollusks, mayflies, beetles and dipterans were better developed in the hard highly alkaline streams, while stoneflies and caddisflies were better developed in the soft water of low alkalinity (Neel, 1973). Other abiotic factors such as flow velocity (e.g. Statzner et al., 1988) and water chemistry (e.g. Bunn et al., 1986) have been found to also influence biotic composition.

The physical and chemical properties of running waters and their effects on the community are driven by numerous environmental variables such as climatic conditions, production–respiration ratio, urban storm water run-off and waste water effluents. The interactions and dependencies among these properties are only partly understood. Since knowledge of species-habitat interrelation remains insufficient, consequently, prognostic assessment of ecosystem properties is not presently available (Vannote et al., 1980; Townsend, 1989).

Some research has been done to present applicability of ANNs in bioindication of chemical and hydro-morphological habitat characteristics with benthic macroinvertebrates (Schleiter et al., 1999; Chon et al., 1996; Chon et al., 2000) and demonstrated potential of ANNs in this field.

Main feature of this approach is that neural nets study the relations between distribution of macroinvertebrates and habitat characteristics from both reference and potentially impacted sites. Data on water polluted sites were used for modelling, the approach therefore is called *"Dirty Water Approach"*.

5.1.2 Aims and Hypothesis

The aim of the model was to study functional interrelation between water quality, habitat characteristics and colonisation patterns of benthic macroinvertebrates within stream and river ecosystems.

'Dirty water' models utilised a wider range of input variables, including those that were altered by anthropogenic impacts. The aim of a dirty water model was to identify input variables that exerted some influence on outputs, and then run simulations of various scenarios through the model to predict the ecological consequences of altering input variables.

Investigation of sensitivity curves derived from dirty water ANN models using the methods outlined in chapter 3 should greatly enhance our understanding of the effects of impacts of various types on individual macroinvertebrate taxa. This would enable impact specific indicator taxa to be readily identified and should enhance our capacity to determine and mitigate the effects of human activities on stream ecosystems.

The hypothesis was that the distributions of macroinvertebrates at family level were determined not only by physical variables but also by the chemical water quality and by the distributions of other macroinvertebrate taxa. Using both chemical and physical predictor variables, Artificial Neural networks were able to predict combinations of macroinvertebrate taxa present at study sites.

Furthermore, a hierarchy of factors determining the community structure of invertebrates could be identified from number of impact variables (Schleiter et al., 1999).

5.2 *Materials and Methods*

5.2.1 Data Analysis

Physical characteristics are always the driving variables to define the typical taxa, which should occur at each specific habitat site and condition. All 39 predictor variables used in *clean water approach* were also used as inputs in this approach.

Chemical variables such as ions, pH and nutrient concentrations could easily be affected by anthropogenic impacts and would not be suitable as predictor variables for clean water approach. In the dirty water approach, models intended to study effects of water chemistry in impacted sites, these chemical variables therefore were also considered as driving variables to predict presence/absence of macroinvertebrate taxa. There were 17 variables belonging to this group (chapter 3). They were all continuous data.

Dirty water approach also considered interrelations among macroinvertebrate taxa. To serve this purpose, neural network models determined a combination of taxa presence instead of a distribution of each individual taxon as developed in clean water approach. Therefore, 40 macroinvertebrate taxa were used as outputs in only one model.

The database contained information of 716 reference sites and 1159 sites potentially having impacted conditions. From this database, 80% of sites were considered as training set. The remaining 20% were used to validate models' performance.

The training set contained data from 1193 sites of both reference and potentially impacted conditions, from which 100 sites had been taken randomly for cross validation

The validation set contained data from 300 sites, 115 of which were potentially impacted, 185 others belonged to reference condition.

5.2.2 Network Architecture

One ANN model was developed to study functional interrelation. The structure of the network was as follows:

- Input layer contained 56 nodes
- Single hidden layer with 33 nodes
- Outputs layer with 40 nodes

5.2.3 Method of Training

The ANN model had been trained with data from the training set where cross validation was applied to control overtraining. The training parameters that appeared to be optimal read as follows:

- Number of iterations: 10000.
- Step sizes: 1 and 0.1 for hidden layer and output layer respectively.
- Learning rule: Momentum
- Momentum: 0.7

5.2.4 Method of Validation

The validation was conducted with data from validation set. Validation results were estimated by means of correct prediction of presence/absence of all macroinvertebrate taxa in the validation set. A target of 70% correct predictions for all taxa in the validation set was settled to manage reliability of network performance.

5.3 Sensitivity Analysis

5.3.1 Results of Sensitivity Analysis

Method of sensitivity analysis was the same as applied for clean water approach. Inputs were varied between their mean +/- only three standard deviations that proved to be sufficient to cover whole set of inputs in the database.

The ANN sensitivity was also estimated by means of output change over the range of inputs in the database as discussed in the chapter 3. There were 56 plots calculated for each taxon. The results are summarised in table 5.1

Table 5.1 Summary of input sensitivity of 40 MI taxa by mean of percentages of outputs change for Dirty WaterAapproach

Taxa	H Velocity - max	Detrital cover (%)	Conductivity (µscm-1)	pH	Turbidity (NTU)	Total Hardness	Total N(mg/l)	Total P(mg/l)	Na+(mg/l)	K+(mg/l)	Ca++(mg/l)	Mg++(mg/l)	HCO3-(mg/l)	CO3--(mg/l)	Cl-(mg/l)	Site Max Velocity
Dugesiidae	15	5	9	**82**	5	1	12	10	1	2	23	9	12	6	1	28
Oligochaeta	10	8	**62**	**57**	11	15	8	11	2	5	21	9	17	5	0	**85**
Planorbidae	16	**49**	3	21	3	4	7	13	2	29	4	2	4	10	0	12
Thiaridae	4	5	2	**36**	2	8	**42**	6	0	5	9	1	3	9	1	2
Corbiculidae	15	16	4	**72**	24	18	28	18	2	11	**31**	8	27	20	4	28
Acarina	28	26	26	**49**	23	17	**33**	**34**	2	**34**	26	**40**	**67**	8	3	20
Copepoda	**99**	15	**88**	7	21	20	13	**38**	0	**51**	4	12	**62**	18	1	**55**
Cladocera	**66**	18	**35**	20	9	28	**83**	12	2	10	2	7	8	4	1	**36**
Ostracoda	4	7	1	3	1	0	0	2	0	2	1	0	1	2	0	5
Atyidae	**97**	10	11	12	10	2	25	3	2	23	26	12	**42**	11	1	13
Palaemonidae	16	18	13	**30**	15	10	28	25	9	28	24	6	13	29	0	23
Leptophlebiidae	14	16	6	**93**	1	**14**	**81**	27	1	**45**	**45**	1	8	2	1	15
Baetidae	**67**	**42**	**43**	14	22	20	12	**30**	17	16	**35**	**41**	**30**	17	6	**84**
Caenidae	**41**	**55**	20	5	26	11	6	**44**	8	17	**37**	**34**	**33**	4	5	**30**
Prosopistomatidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Gomphidae	11	15	21	**55**	8	16	**44**	15	3	3	15	1	**38**	15	1	20
Corduliidae	**52**	1	1	4	1	0	1	1	1	6	1	0	2	0	0	**53**
Libellulidae	2	5	1	7	2	0	2	1	0	6	2	1	6	0	0	2
Coenagrionidae	31	4	6	3	4	9	4	6	1	35	6	5	31	2	1	25
Gripopterygidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Corixidae	**100**	15	**34**	10	26	11	**33**	**34**	8	16	15	1	17	13	2	14
Notonectidae	**38**	9	28	27	3	12	**69**	11	7	15	8	1	19	7	1	26
Pleidae	9	1	2	4	1	5	6	1	0	1	4	0	5	2	0	3
Veliidae	**100**	9	8	13	4	10	23	12	2	3	4	6	10	5	3	**40**
Dytiscidae	**97**	15	**60**	**58**	15	21	**88**	16	3	11	23	3	18	9	0	18
Elmidae	**92**	28	9	21	23	23	**53**	8	5	**43**	**55**	16	**56**	29	3	**38**
Psephenidae	**93**	25	18	**67**	19	25	29	13	0	5	**33**	1	**35**	5	2	24
Hydrophilidae	25	**97**	4	19	15	9	**56**	31	6	18	3	6	**68**	19	7	31
Tanypodinae	**44**	10	**96**	**50**	16	5	4	13	1	12	4	2	14	0	0	29
Orthocladiinae	**70**	**52**	20	**45**	28	3	**31**	24	2	11	**33**	26	37	6	1	**44**
Simuliidae	**99**	4	3	**79**	3	0	**76**	14	1	4	3	0	1	2	1	3
Ceratopogonidae	9	2	2	3	3	3	2	4	0	6	2	3	7	15	1	8
Tabanidae	16	0	0	0	0	0	0	1	0	1	1	0	11	1	0	2
Leptoceridae	**99**	1	4	11	2	2	24	5	1	5	25	3	28	0	0	15
Hydropsychidae	**99**	**93**	**32**	**42**	15	9	25	28	2	13	**34**	15	33	6	0	**50**
Ecnomidae	**81**	**36**	25	**42**	3	21	18	22	5	**47**	**33**	3	25	7	5	24
Hydroptilidae	2	3	2	15	4	0	7	6	2	1	6	5	8	4	1	4
Calamoceratidae	**83**	**95**	**42**	23	3	**33**	**32**	8	8	13	**39**	8	29	13	6	15
Philopotamidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pyralidae	**90**	2	0	15	5	2	6	3	0	3	7	2	2	1	1	24

Table 5.1 (continued)

Taxa	Instant. Discharge	Latitude	Longitude	Altitude (m)	Stream Order	Slope	Distance FrSource	H Width (m)	H Depth (m)	Bedrock (%)	Boulder (%)	Cobble (%)	Pebble (%)	Gravel (%)	Sand (%)	Silt/Clay (%)
Dugesiidae	22	**76**	**35**	17	8	5.3	9.6	1.4	3.8	20	15	28	1.6	2	13	11
Oligochaeta	11	**41**	**50**	**57**	**48**	5.3	**45**	8.5	**32**	**45**	4.9	4	14	6.8	15	21
Planorbidae	0.5	**35**	6.5	4.7	15	5.4	5.2	3.4	3.6	2.6	9.4	3	2	0.8	7.4	1.3
Thiaridae	4	7.8	**78**	**96**	19	8.5	**58**	0.1	24	**46**	2.9	12	3.2	7.1	15	13
Corbiculidae	25	**78**	**79**	**84**	**87**	**36**	**68**	2.3	10	**64**	**47**	16	1.9	24	21	**60**
Acarina	30	**61**	**75**	**70**	**48**	**61**	28	18	**51**	23	**55**	**43**	17	**33**	**56**	**61**
Copepoda	31	**51**	**98**	27	**87**	**57**	**68**	15	**65**	9	**41**	29	**33**	23	21	**59**
Cladocera	24	**57**	**48**	**62**	18	25	**82**	16	6	10	8	10	8	16	14	7
Ostracoda	1	3	1	**32**	6	2	7	0	3	1	4	2	1	1	1	2
Atyidae	31	25	**93**	**74**	10	15	10	1	28	**48**	25	**32**	**46**	1	**60**	29
Palaemonidae	23	**40**	26	**72**	**44**	20	21	4	**49**	14	8	2	22	24	23	25
Leptophlebiidae	**82**	24	**69**	**72**	19	9	**40**	9	16	**84**	21	15	11	26	**69**	15
Baetidae	14	**56**	**61**	**30**	**42**	25	**45**	8	24	18	38	27	27	5	28	5
Caenidae	**83**	**57**	**97**	**93**	**43**	**61**	**54**	21	**39**	27	**43**	**59**	33	**48**	**73**	**60**
Prosopistomatidae	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
Gomphidae	29	**85**	19	30	**54**	29	14	16	4	**32**	**30**	27	12	25	**42**	**36**
Corduliidae	3	2.7	2.1	36	0.6	1.7	4.2	0.3	11	1.8	2.4	2	1.8	1.9	5.6	1.8
Libellulidae	4.7	**30**	8.2	8.7	15	19	1.3	0.1	2.7	2.6	4.3	**36**	2.4	2.2	5.4	5.8
Coenagrionidae	6	**52**	15	9	**42**	9.4	**66**	5.7	**45**	5.3	**36**	**41**	4	2.8	16	**40**
Gripopterygidae	0.1	0.1	0.8	0.1	0.1	0.6	0.1	0	0.1	0	0.1	0.6	0.1	0.1	0.1	0.1
Corixidae	**46**	**84**	**60**	**50**	28	18	**52**	1	31	19	18	22	10	12	24	22
Notonectidae	17	**49**	23	**55**	**55**	12	**88**	20	4	36	15	32	15	19	8	5
Pleidae	1	19	1	6	6	2	34	2	2	2	2	5	1	1	4	5
Veliidae	2	10	15	9	6	37	36	1	6	11	9	4	11	3	13	**45**
Dytiscidae	**31**	26	13	**59**	**32**	15	**46**	5	38	7	3	**50**	3	15	26	14
Elmidae	11	**35**	**51**	**53**	**30**	16	**47**	11	**42**	**62**	5	25	17	5	26	10
Psephenidae	**60**	24	26	22	26	29	**39**	19	14	25	25	24	13	25	**45**	18
Hydrophilidae	20	**31**	20	**55**	20	**63**	**39**	7	**32**	26	12	**49**	25	20	**37**	7
Tanypodinae	**45**	23	**44**	7	10	11	8	13	3	27	17	15	13	25	**36**	15
Orthocladiinae	40	**60**	**71**	26	85	21	7	15	9	29	16	**37**	16	4	**49**	18
Simuliidae	20	**39**	10	2	12	2	3	1	3	2	6	18	2	4	1	6
Ceratopogonidae	11	25	15	13	25	3	5	2	17	3	3	4	8	9	7	26
Tabanidae	1.6	4.7	7.7	1.4	0.7	2.1	2	0.6	1.7	1.3	0.4	0.1	0.6	0.8	0.2	0.3
Leptoceridae	6	26	**84**	**45**	5	8	4	1	2	7	2	10	3	2	4	5
Hydropsychidae	**48**	**54**	23	**47**	**32**	24	26	3	21	24	19	21	6	24	27	**41**
Ecnomidae	**61**	**55**	**63**	**44**	**50**	22	43	16	28	**48**	28	**32**	**43**	21	**47**	**31**
Hydroptilidae	9	**45**	11	6	18	3	5	1	23	1	3	3	3	1	7	2
Calamoceratidae	**56**	**51**	47	**67**	**52**	**54**	47	9	27	**49**	5	27	**30**	12	**57**	13
Philopotamidae	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Pyralidae	9	6	**49**	5	5	3	7	3	2	6	6	**32**	2	6	7	7

Table 5.1 (continued)

Taxa	Water Temp (°C)	Alkalinity	Site-mean phi	Mean Wetted Width	Mean Channel Width	Mean Depth	Mean WSMR (a)	Mean DSMR (b)	Annual range in MMR	Range in WSMM	Range in DSMM	%rainfall in WS	Mean annual rainfall	Mean daily max temp	Mean daily min temp	Mean daily temp range
Dugesiidae	**55**	13	25	1	20	3	12	3	16	**49**	3	**70**	9	**30**	28	11
Oligochaeta	28	**31**	8	2	36	15	25	5	16	24	**43**	**44**	26	29	26	21
Planorbidae	17	16	16	4	8	2	6	1	26	6	15	13	5	3	4	6
Thiaridae	23	11	3	2	5	5	9	**59**	11	18	8	**72**	**26**	**78**	25	4
Corbiculidae	**76**	21	28	8	17	1	24	**82**	**58**	25	15	**82**	26	**63**	**72**	2
Acarina	**75**	16	**69**	2	**40**	28	**43**	**41**	**54**	**56**	**43**	72	**47**	**51**	**56**	30
Copepoda	**36**	12	**46**	11	**65**	20	8	21	**85**	**66**	**37**	**31**	**39**	28	22	37
Cladocera	**74**	13	15	15	26	13	10	18	**33**	**92**	14	28	24	19	22	33
Ostracoda	16	4	1	0	5	1	1	1	2	3	1	4	5	5	9	6
Atyidae	27	**31**	26	2	9	3	12	15	**39**	**94**	25	**58**	21	**50**	8	26
Palaemonidae	**33**	**32**	5	6	19	24	**72**	**32**	**37**	8	25	**77**	**87**	**68**	**49**	**32**
Leptophlebiidae	17	26	18	3	9	2	15	**53**	22	20	**32**	**95**	**33**	18	**86**	**50**
Baetidae	**49**	**56**	9	6	35	7	**72**	35	**50**	**51**	**59**	**58**	**54**	**48**	**40**	**54**
Caenidae	**48**	**70**	**50**	26	**42**	**36**	**57**	15	19	**68**	**53**	**51**	**46**	**57**	**34**	**33**
Prosopistomatidae	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Gomphidae	**30**	15	13	14	5	3	25	20	**50**	**61**	**69**	19	19	22	9	27
Corduliidae	4	0	1	1	1	1	2	12	8	4	1	8	1	6	1	2
Libellulidae	15	1	2	0	7	1	2	2	**87**	12	11	**42**	2	2	1	5
Coenagrionidae	10	3	25	2	6	3	6	13	**88**	24	26	**46**	11	19	**30**	**37**
Gripopterygidae	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Corixidae	**77**	**48**	16	6	13	25	19	**33**	28	**37**	13	**79**	23	**63**	**78**	**52**
Notonectidae	**38**	19	16	15	37	10	45	23	8	**36**	9	**83**	**79**	10	**58**	39
Pleidae	2	1	0	1	2	1	4	12	4	1	0	4	2	6	2	5
Veliidae	21	19	22	3	16	4	**45**	10	4	34	5	**68**	**71**	14	20	16
Dytiscidae	28	12	11	3	8	15	11	**30**	21	**43**	19	**63**	**54**	**35**	**55**	**51**
Elmidae	**63**	**38**	**34**	8	**41**	**31**	**45**	13	**36**	**43**	**30**	**34**	**45**	**34**	**80**	**83**
Psephenidae	15	7	20	5	**43**	22	18	15	7	**62**	4	23	22	**48**	**38**	10
Hydrophilidae	**72**	**68**	12	15	17	**36**	22	**39**	**51**	**37**	15	**86**	9	**65**	16	18
Tanypodinae	5	9	6	8	4	10	16	12	21	**49**	7	**61**	16	15	14	20
Orthocladiinae	**56**	27	26	3	18	28	**53**	**74**	27	**61**	15	**37**	**43**	**51**	**47**	**79**
Simuliidae	31	2	5	1	2	4	4	18	10	4	1	**36**	10	3	2	2
Ceratopogonidae	11	2	2	1	5	1	14	5	10	6	1	20	28	21	9	18
Tabanidae	12	2	1	1	1	1	1	2	1	3	1	5	1	3	1	0
Leptoceridae	16	4	6	1	12	1	12	4	16	**43**	10	**92**	8	1	5	4
Hydropsychidae	30	16	**44**	0	9	20	**64**	**58**	**54**	14	13	34	**59**	**41**	**79**	**60**
Ecnomidae	**51**	17	**49**	4	19	11	**39**	**41**	**55**	**46**	**53**	**53**	**44**	19	**53**	**54**
Hydroptilidae	15	**53**	5	3	1	9	22	37	1	12	**36**	15	13	29	17	19
Calamoceratidae	15	12	14	3	**34**	4	23	**35**	**66**	**46**	20	**37**	**30**	77	**63**	**54**
Philopotamidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pyralidae	15	2	20	0	**37**	1	3	3	**32**	7	5	4	4	4	13	**41**

5.3.2 Networks Revision

Redundant inputs needed to be removed in order to improve transparency and validity of the ANN model. ANNs with simpler architectures required less time to learn and generalise pattern from database.

Each input had a certain effect on output nodes and as the number of inputs used in this approach was much higher than in clean water approach, the individual sensitivity level of each input was lower compared to its sensitivity in the clean water approach. For this reason, the criteria for declaring input as redundant needed to be revised as less strict. Those inputs affecting the range of output changes by less than 30% over their range were considered as redundant

Some inputs appeared to be highly sensitive for certain taxa but less sensitive for others. However this approach considered combinations of macroinvertebrate families as ANN outputs, therefore, effects of single input need to be assessed for the whole set of outputs. Inputs, which caused small change in presence/absence of most macroinvertebrate taxa, were considered redundant in overall. Turbidity, CO_3^{2-}, Cl^- concentrations, *H Width* were redundant for all outputs. *0-4 Habitat* were also redundant for all outputs except for *Copepoda*. However, many other inputs provided sensitivities at higher level than *0-4 Habitat* for this taxon, thus, *0-4 Habitat* was removed from the data set. Similar situations occurred for the inputs *Mean depth* and *Hardness*.

As the results of sensitivity analyses, 10 inputs were considered as redundant and were removed from the revised ANN. They were from both physical and chemical variable:

- Physical variables: H Width, Mean Wetted Width, 0-4 Habitat and Mean Depth
- Chemical variables: Turbidity, Hardness, Na^+, Mg^{+2}, CO_3^{-2}, Cl^- concentrations.

The final ANN model was developed with 46 nodes in the input layer, same 40 nodes as outputs and 30 nodes in the single hidden layer.

It had been run with the same technical parameters as mentioned before for training with 5000 iterations.

5.4 Validation results

Validation results are summarised in table 5.2 and plots are shown in figure 5.1

Even though the settled condition for verification of network performance was not achieved for *Oligochaeta* (61% of overall correct prediction), and just nearly achieved for *Palaemonidae* (69.67%) and *Hydrophilidae* (69.33%), clear improvements were observed for all taxa at quite different rates by means of the revised ANN structure. Clear improvements were obtained for both reference sites and impacted sites Average value of correct predictions for all taxa improved from 75.75% to 79.01% for reference sites, from 74.7% to 79.07% for potential impacted sites, and from 75.23% to79.03% in overall.

Predictions for test sites achieved slightly better improvement. The biggest improvement was observed for the prediction of presence of *Oligochaeta* at test sites where correct predictions increased by nearly 14%. Prediction for *Tabanidae* at the test sites was the only case when correct predictions declined after revision (from 96.55% by initial network to 93.97% by revised network). Prediction remained the same for few cases (*Simulidae* at test sites, *Hydrophilidae* at reference sites)

Even though the average correct prediction rates at reference and test sites were similar, the rates of correct predictions were quite different at reference and test sites for different taxa.

Table 5.2 Correct prediction of presence/absence of macroinvertebrate taxa in dirty water approach before and after sensitivity analyses

		Initial Models			After Exclusion of Redundant Inputs		
	Taxa	**56 Input Ref. Sites**	**56 Input Test Sites**	**56 Input Overall**	**46 Input Ref. Sites**	**46 Input Test Sites**	**46 Input Overall**
1	Prosopistomatidae	98.37	98.28	98.33	98.91	99.14	99
2	Gripoterygidae	92.39	95.69	93.67	96.20	96.55	96.33
3	Philopotamidae	88.59	93.10	90.33	90.76	93.97	92
4	Tabanidae	85.87	96.55	90	88.59	93.97	90.67
5	Psephenidae	85.87	90.51	87.67	86.41	92.24	88.67
6	Hydropsychidae	86.96	85.34	86.33	89.67	88.79	89.33
7	Pleidae	85.87	82.76	84.67	88.59	87.07	88
8	Corbiculidae	87.50	80.17	84.67	89.13	85.34	87.67
9	Notonectidae	85.33	82.76	84.33	85.87	87.07	86.33
10	Calamoceratidae	79.89	91.37	84.33	82.61	92.24	86.33
11	Planorbidae	80.98	80.17	80.67	86.41	87.93	87
12	Pyralidae	78.26	81.9	79.67	85.33	81.03	83.67
13	Veliidae	76.09	80.17	77.67	77.72	81.03	79
14	Corduliidae	76.63	78.45	77.33	79.35	82.76	80.67
15	Simuliidae	78.26	75	77	82.07	75	79.33
16	Dugesiidae	77.17	75.86	76.67	75.54	75.86	75.67
17	Elmidae	71.2	81.9	75.33	73.91	83.62	77.67
18	Acarina	76.63	72.41	75	78.8	74.14	77
19	Dytiscidae	78.26	68.10	74.33	78.8	72.41	76.33
20	Ecnomidae	73.91	74.14	74	76.63	81.03	78.33
21	Thiaridae	77.17	68.97	74	83.7	77.59	81.33
22	Ceratopogonidae	69.56	79.31	73.33	70.11	813	74.33
23	Cladocera	75.54	67.24	72.33	81.52	73.28	78.33
24	Coenagrionidae	74.46	66.38	71.33	79.35	73.28	77
25	Copepoda	74.46	65.51	71	79.89	72.41	77
26	Caenidae	70.1	71.55	70.67	77.72	75.86	77
27	Baetidae	73.91	64.66	70.33	76.09	70.69	74
28	Libellulidae	71.2	68.1	70	75.54	75.86	75.67
29	Atyidae	70.11	67.24	69	70.65	74.14	72
30	Tanypodinae	70.11	65.52	68.33	72.28	75	73.33
31	Leptoceridae	66.85	70.69	68.33	71.2	75	72.67
32	Hydrophilidae	69.02	67.24	68.33	69.02	69.83	69.33
33	Leptophlebiidae	67.94	67.24	67.67	75	73.28	74.33
34	Corixidae	69.57	63.79	67.33	70.11	71.55	70.67
35	Orthocladiinae	66.31	66.37	66.33	71.74	72.41	72
36	Palaemonidae	66.85	63.79	65.67	69.57	69.83	69.67
37	Ostracoda	65.76	63.79	65	71.74	68.97	70.67
38	Hydroptilidae	62.5	65.52	63.67	70.11	71.55	70.67
39	Gomphidae	60.87	63.79	62	73.37	68.1	71.33
40	Oligochaeta	56.52	46.55	52.67	60.33	62.07	61
	Average	**75.57**	**74.7**	**75.23**	**79.01**	**79.07**	**79.03**

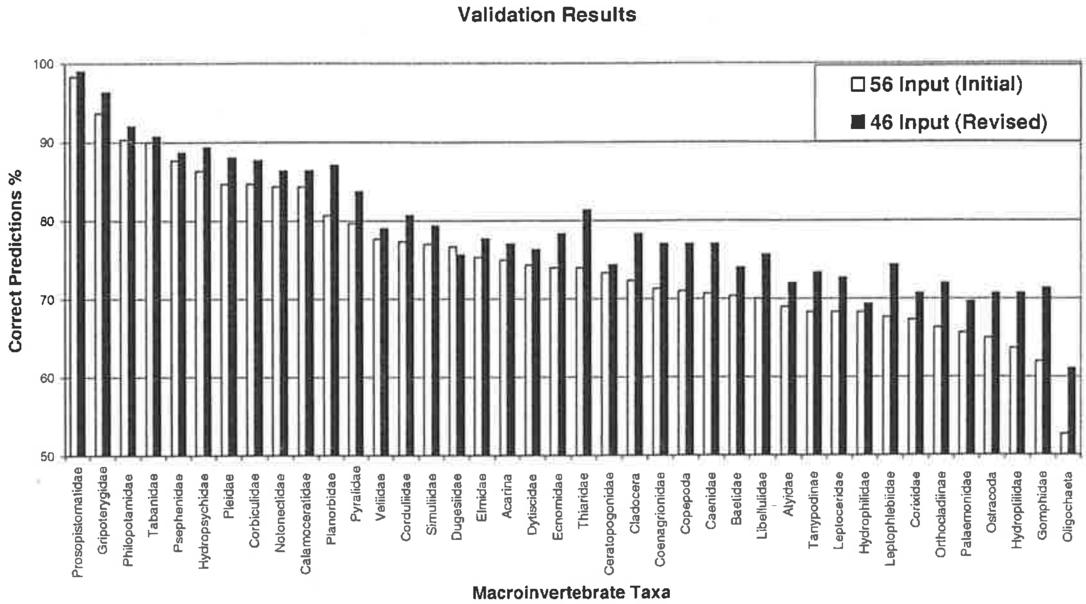


Figure 5.1 Correct prediction of presence/absence of macroinvertebrates before and after exclusion redundant inputs explored by models of dirty water approach.

Numbers of sites where taxa observed and predicted by models as present are comparing in table 5.3. Comparisons show that models can predict approximately the same number of sites where taxa are present. The results verify the reliability of correct predictions obtained by models shown in table 5.2.

Table 5.3 Number of sites where presence of taxa was observed or predicted by revised models

	Taxa	185 reference sites		115 test sites		300 overall validation sites	
		Observed	Predicted	Observed	Predicted	Observed	Predicted
1	Dugesiidae	35	31	25	23	60	54
2	Oligochaeta	80	76	70	64	150	140
3	Planorbidae	23	20	16	15	39	35
4	Thiaridae	52	48	57	53	109	101
5	Corbiculidae	25	22	28	25	53	47
6	Acarina	140	143	74	76	214	219
7	Copepoda	100	105	65	67	165	172
8	Cladocera	56	53	37	39	93	92
9	Ostracoda	71	70	58	55	129	125
10	Atyidae	94	97	56	60	150	157
11	Palaemonidae	103	100	50	48	153	148
12	Leptophlebiidae	116	113	48	50	164	163
13	Baetidae	139	149	82	89	221	238
14	Caenidae	127	128	80	87	207	215
15	Prosopistomatidae	4	4	3	3	7	7
16	Gomphidae	88	84	43	41	131	125
17	Corduliidae	36	35	19	15	55	50
18	Libellulidae	78	75	50	43	128	118
19	Coenagrionidae	57	56	47	50	104	106
20	Gripopterygidae	17	15	4	3	21	18
21	Corixidae	86	86	59	63	145	149
22	Notonectidae	24	21	12	15	36	36
23	Pleidae	28	25	17	15	45	40
24	Veliidae	53	49	22	23	75	72
25	Dytiscidae	74	68	40	41	114	109
26	Elmidae	81	79	37	36	118	115
27	Psephenidae	28	28	11	8	39	36
28	Hydrophilidae	58	53	41	36	99	89
29	Tanypodinae	139	146	84	88	223	234
30	Orthocladiinae	105	105	70	67	175	172
31	Simuliidae	45	47	31	30	76	77
32	Ceratopogonidae	55	52	26	25	81	77
33	Tabanidae	29	26	6	4	35	30
34	Leptoceridae	124	130	82	86	206	216
35	Hydropsychidae	62	62	39	32	101	94
36	Ecnomidae	45	40	29	25	74	65
37	Hydroptilidae	57	53	33	31	90	86
38	Calamoceratidae	37	34	13	12	50	46
39	Philopotamidae	30	25	10	10	40	35
40	Pyralidae	43	41	27	24	70	65

5.5 *Discussion*

5.5.1. Performance of Artificial Neural Networks

The ANN models provided similarly good validation results for both reference sites and probably impacted sites. A series of chemical and hydro-morphological properties could be modelled with reasonable low error. The results clearly indicated functional relationships between colonisation patterns of benthic macroinvertebrates and chemical and hydro-morphological habitat characteristics within river and stream ecosystems. Moreover, a hierarchy of factor determining the community structure of invertebrates may be identified from numerous impact variables. The rule induction algorithm of the ANN model correctly chooses the families that are indicative of habitat conditions, in conjunction with other families that refer to co-occurrence of specific taxa.

The ANN architecture used in this approach was extremely complex. The complexity of model could induce noise that may have negative impacts on the ANN performance. Evidence was given by the fact that sensitivity levels were significantly reduced for most of inputs compared to their sensitivity obtained by clean water models, where fewer input variables were applied to develop models. That means that the performance of ANNs can be improved by simplifying their structure. Specific procedures applied for selection of the most relevant variables can reduce input dimension and therefore simplify the complexity of network in a rational way.

In ANN models the best pre-selection method was a sensitivity analysis (Schleiter et al., 1999). Sensitivity analyses can be conducted several times until an optimal structure is obtained. Within the framework of this research, sensitivity was conducted only once and improvement was clearly observed. The procedure can be repeated to further remove redundant inputs and make models more transparent.

Good results were obtained for common taxa as well as rare and moderately frequent taxa. Results again revealed that ANN models are able to cope with taxa represented by probability of presence at sites greater than 0.

5.5.2 Relationship between Macroinvertebrates and Habitat Conditions

Sensitivity analyses showed that the effects of physical predictor variables to macroinvertebrates distribution were similar to relationship studied by the clean water approach. The trends of output changes over the ranges of inputs appeared to be the same for most taxa, even though the magnitudes of changes were largely different

Relationship between physical predictors and the distribution of macroinvertebrates was discussed for the clean water approach in chapter 4 and will be discussed further in next chapter. In this chapter, the discussion focuses on relationships between the distribution of macroinvertebrates and input variables, which are potentially affected by anthropogenic activities.

Current and discharge

So-called *H Velocity*, *site max velocity* and *instantaneous discharge* were identified as sensitive inputs for the ANN performance. Habitat velocity was highly sensitive for distribution of 16 taxa and also affected 5 others. Discharge also appeared to be significantly sensitive for 12 taxa and moderately for 5 others. Site max velocity also influenced to the presence of 19 taxa.

Much of the ecological information on water current-insect interactions deals with the relationship between distribution patterns and the corresponding spatial variations in water velocity (Ward, 1993). In addition to their direct effects, current and discharge play major roles in structuring habitat conditions for stream macroinvertebrates by influencing mineral and organic substrate, suspended and aquatic flora. Changes in discharge increase the number of aquatic insects drifting downstream

Simulidae, filter-feeding black flies have feeding mechanisms that depend on current (Chutter, 1969). Many species belonging to the *Simulidae* family loses their feeding capability at still water. The cephalic fans of *Simulium ornatum* var. *nitidifrons,* for example, close and lose their feeding function at velocity < 19cm/sec (Harrod, 1965). Figure 5. 2 illustrate result of sensitivity analysis shows that *Simulidae* prefers fast flowing habitat with H velocity exceeding 0.5 m/s.

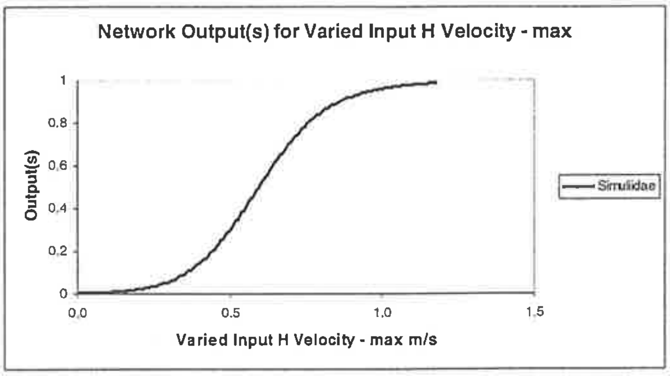


Figure 5.2 Relationship between max H velocity and presence of black flies Simulidae

Predacious diving beetles *Dytiscidae* favour habitat amongst weed in still water (Hawking & Smith, 1997). The relation was revealed by sensitivity analyses as shown in figure 5.3

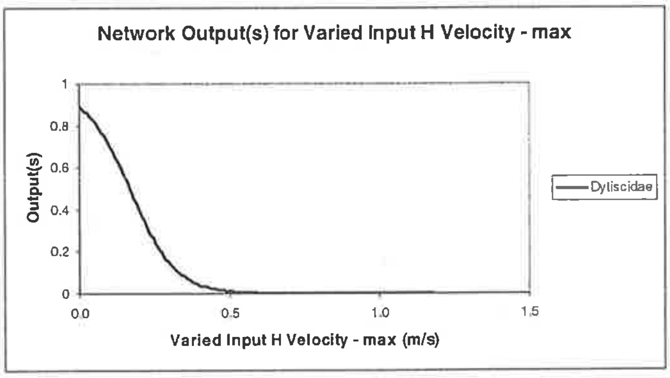


Figure 5.3 Relation between Max H Velocity and presence of **Dytiscidae**

Edington (1968) found that the nets of various species were concentrated in certain velocity ranges. *Hydropsycchidae* is a rapid family that constructs net in the fast water (15 – 100 cm/sec). Figure 5.4 shows that fast flowing streams are favourite habitats of *Hydropsycchidae.*

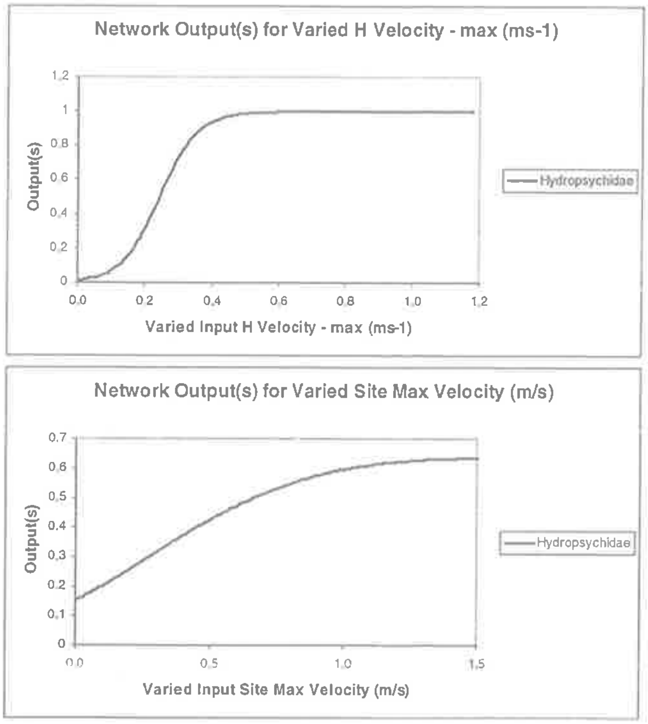


***Figure 5.4 Relations between H velocity and Site velocity, and presence of* Hydropsychidae**

pH

pH has many influences on the aquatic life. Aquatic organisms have ranges of tolerance and optima of pH themselves as the enzyme functions are controlled by specific pH values. Acidity levels influence the solubility and formation of metals ions in the stream water. Those ion concentrations have direct impacts on the life of macroinvertebrates (Lamberti & Sommer, 1997). *Cleanwater* rivers vary in acidity (Giller & Malmqvist, 1998). Conditions within the Queensland stream system confirm this fact. While reference sites have pH values in the range from 4.4 to 9.43, test sites have pH value in the range of 5.1 - 10. Low pH is observed at catchments with hard, igneous rock being low in dissolved salts and buffering capacity. On the other hand, catchments with sedimentary rock are rich in carbonates and the originating streams are usually well buffered, hard water systems with high pH.

pH was observed to be at medium sensitive level for colonisation patterns of macroinvertebrates. However results showed that it was really sensitive for 12 taxa. Several invertebrate families appeared to be absent from low pH sites (e.g. molluscs, mayflies) while others were usually well presented (e.g. stoneflies, blackflies).

Mayflies posses a well-known sensitivity to acid conditions, only few mayflies can tolerate low pH (Giller & Malmqvist, 1998). However, exceptions are some *Leptophlebiids*, which can be found in water of very low pH. Sensitivity analyses showed that mayflies *Leptophlebidae* and blackflies *Simulidae* were mostly present at low pH condition (figure 5.5).

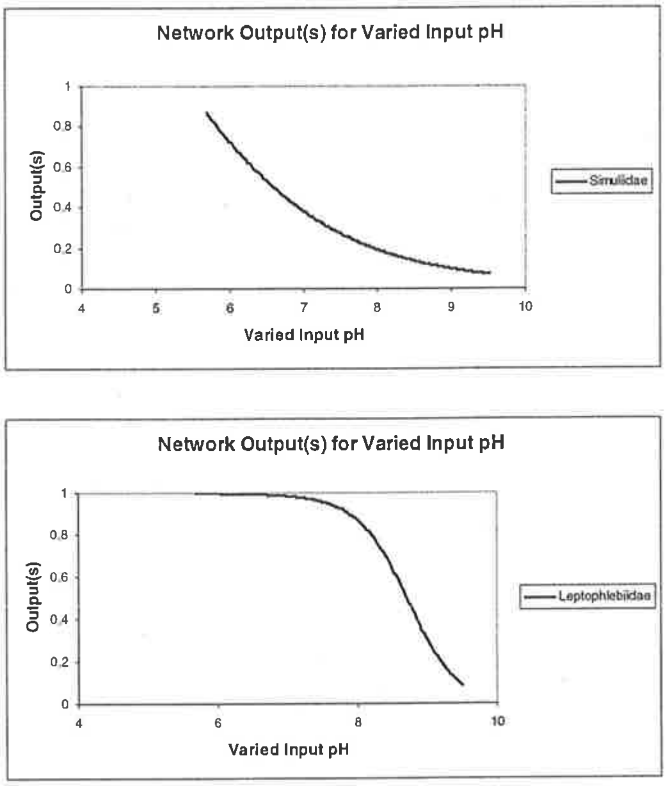


Figure 5.5 Blackflies* Simulidae *and mayflies* Leptophlebidae *favour low pH conditions

Physiological effects of low pH have also been demonstrated on chironomids, mayflies and crustaceans (Giller & Malmqvist, 1998). Sensitivity analyses (Figure 5.6) revealed such relationship. Examples are shown for *Corbiculidae* belonging to the class mussel Bivalvia and for the sub family *Tanypodinae* of chironomids, which can only be present at conditions of high pH. *Caenidae* favour high pH conditions but also tolerate lower level of pH down to 5 (Otto & Svensson, 1983)

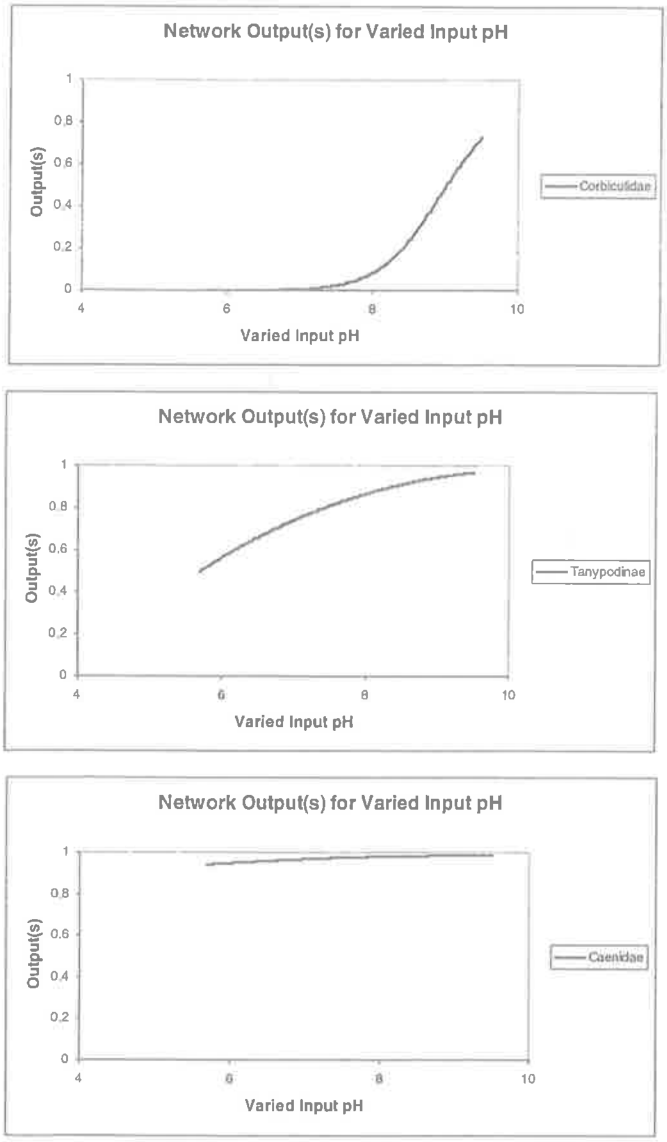


Figure 5.6 Relations between pH and mussel Corbiculidae, chironomid sf-Tanypodinae and mayfly Caenidae

Conductivity and ion concentrations

Conductivity takes into account the total concentration of inorganic ions in the water. Inorganic ions can significantly impacts on biology of macroinvertebrates in freshwater ecosystem. Low calcium levels, for example, can cause osmotic problems and affect shell cuticle secretion in invertebrates as observed for crustaceans, crayfish and snails (Giller & Malmqvist, 1998). The ANN models used several concentrations of inorganic ions as inputs, from which only concentrations of K^+, Ca^+ are considered relatively sensitive variables to the distribution of macroinvertebrates in the Queensland stream system. Conductivity caused significant impacts on presence of 9 taxa. Figure 5.7 and 5.8 illustrate some examples.

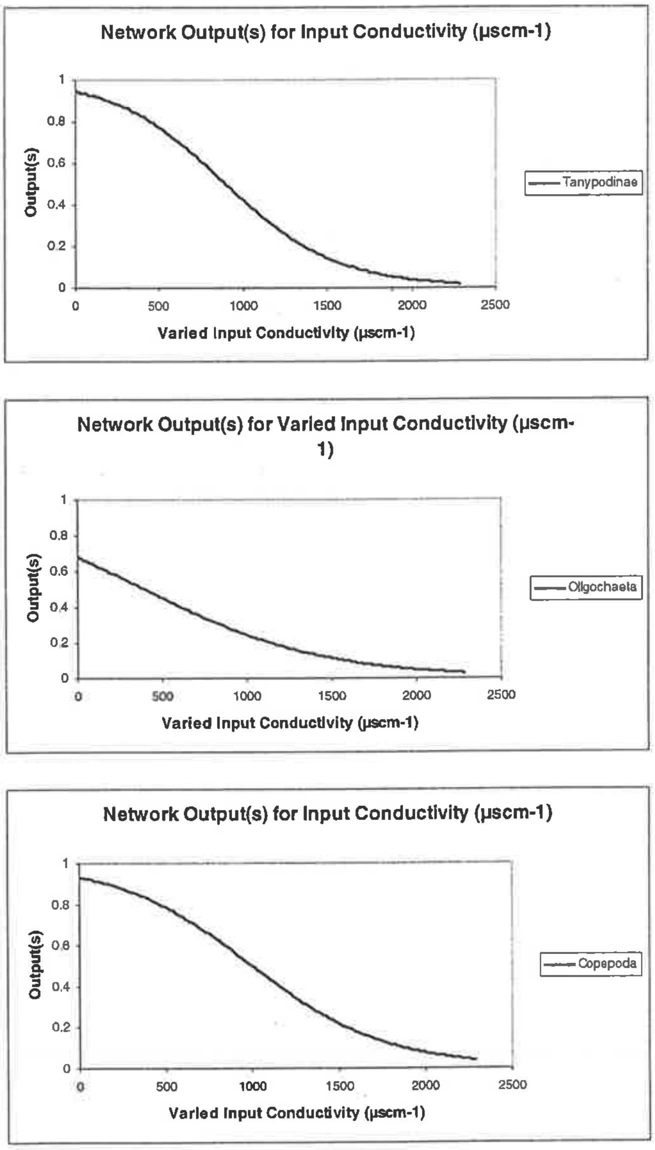
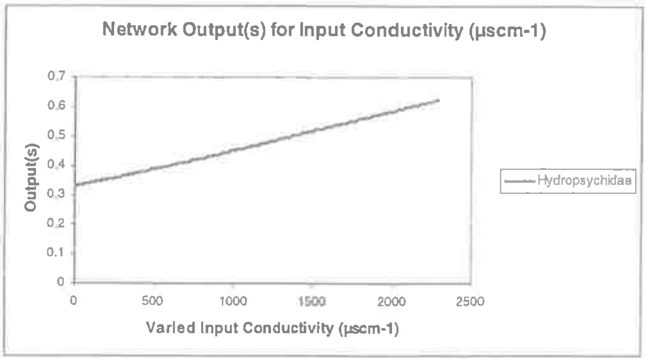


Figure 5.7 **Copepoda, Oligochaeta, Tanypodinae** ***favour conditions of low conductivity***



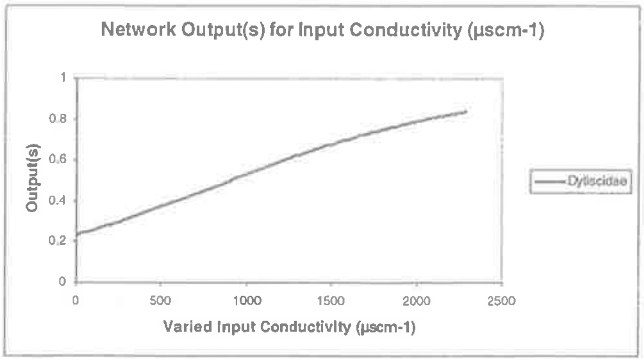


Figure 5.8 **Dyticsidae** *and* **Hydropsychidae** *favour conditions of high conductivity*

Mineral nutrient concentration

Mineral nutrients are usually divided into the macro element (N, P, S, K, Mg, Ca, Na, Cl) which usually make up >0.1% of the organic matter, and the trace elements (Fe, Mn, Zn, B, Si, Mo...) according to the amounts required. All these elements can be considered as mineral nutrients dissolved in water. Theoretically, any of these elements could become an essential, limiting resource. In most freshwater ecosystems, however, many of them are almost always in excess, so that the spectrum of limiting nutrients can be narrowed to N, P and some trace elements (Lamberti & Sommer, 1997). The productivity of autotrophic stream producers is influenced by nutrient concentrations, particular phosphorus and nitrogen (Giller & Malmqvist, 1998). Nutrient-rich rivers can be dominated by filterers, grazers and some predators (Peterson et al., 1993).

Figure 5. 9 illustrate favorite habitat conditions for grazer *Cladocera*, blackflies *Simuliidae* and carnivore water tiger *Dytiscidae* occur at high concentration of N and P.

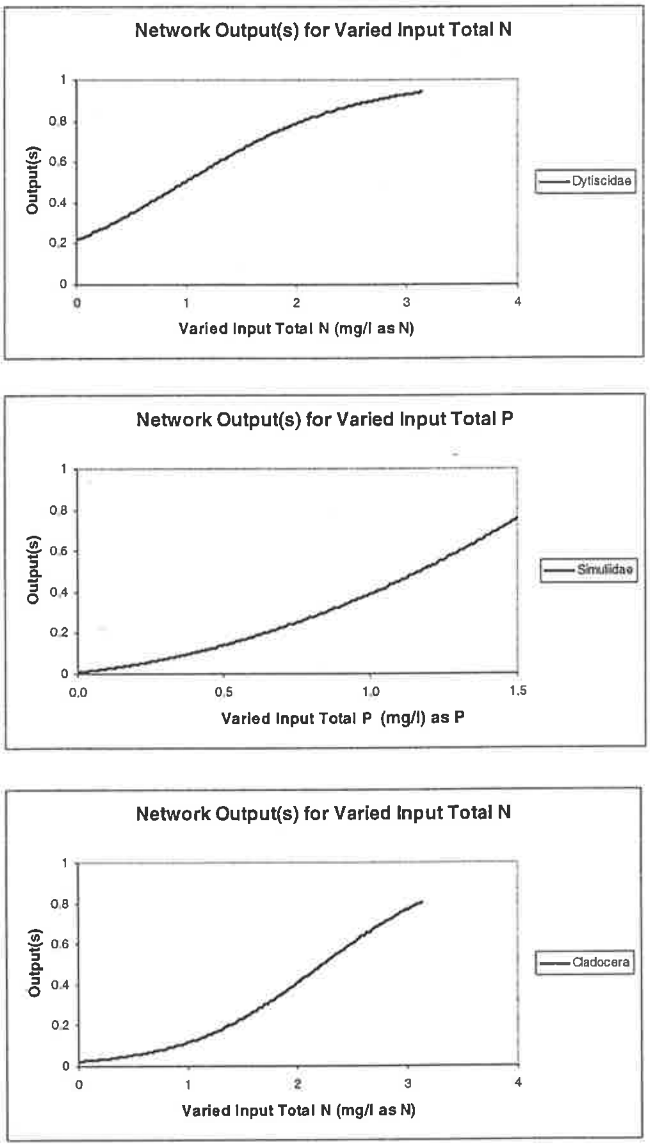
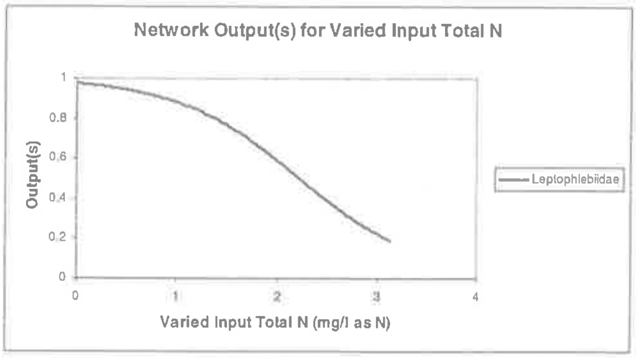


***Figure 5.9 Relations between concentration of N and P and grazer* Cladocera*, blackflies* Simuliidae *and water tiger* Dytiscidae**

Figure 5.10 shows that shredders such as *Leptophlebidae* and predators such as dragonflies *Gomphidae* favour low concentration of N and P.



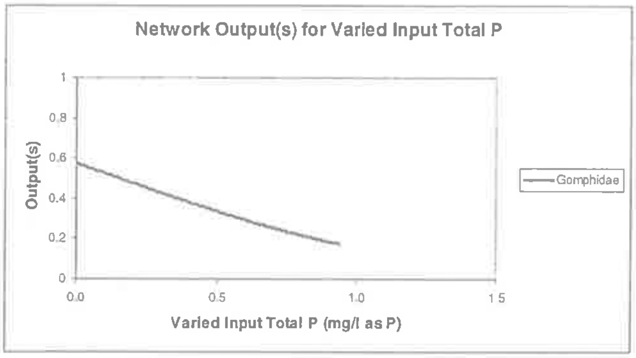


***Figure 5.10 Relations between concentration of N and P and mayflies* Leptophlebidae*, and dragon flies* Gomphidae**

Detrital cover

Detrital cover should significantly influence the presence of macroinvertebrates, especially those feeding on detritus. Wetzel (1983) defined detritus as " organic carbon lost by non-predatory means from any trophic level (includes egestion, excretion, secretion, and so forth) or inputs from sources external to the ecosystem that enter and cycle in the system". Detritus is all dead organic carbon, distinguishable from living organic and inorganic carbon. Detritus originating as ungrazed primary production support a "detritus food chain", which is defined as any route by which chemical energy contained within detrital organic carbon becomes available to the biota (Weltzel, 1983). Therefore, detritus, as a component of environment, can either directly or indirectly affect the distribution of macroinvertebrates in freshwater ecosystems.

In this study, detrital cover appeared to be sensitive for 12 taxa. Examples include models for *Libelullidae, Hydropsychidae, Hydrophilidae, Dugesiidae, Calamoceratidae.* Figure 5.11 illustrates relations between detrital cover and some taxa, where taxa presence of which appeared to be sensitive to detrital cover.

Relations between detrital cover and distribution of macroinvertebrates should be highly complex. Even though non of direct correlations between detrital cover and presence of certain macroinvertebrate taxa had been found in the literature so far, the relation plots discovered by sensitivity analyses provided new insight and hypotheses for further research in this field.

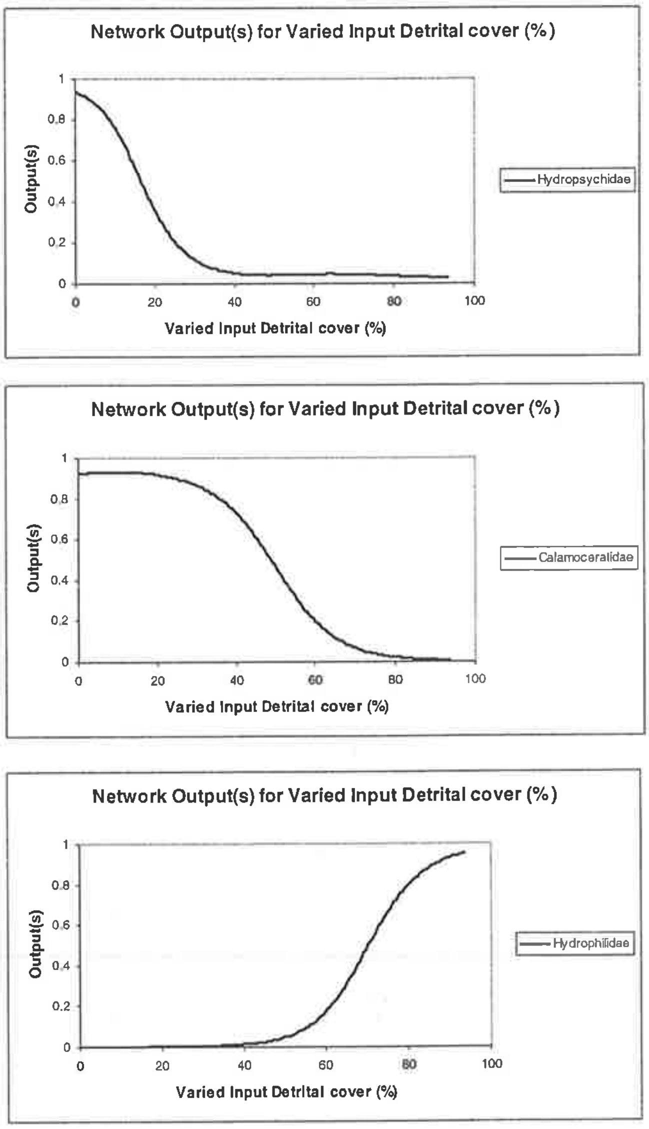


Figure 5.11Relations between concentration of N and P and grazer **Cladocera,** ***blackflies*** **Simuliidae** ***and water tiger*** **Dytiscidae Caramoceratidae,** ***and beetles*** **Hydrophilidae**

5.5.3. Data Limitation

Several chemical variables such as oxygen concentration, which can probably partly explain of the variation between different macroinvertebrate communities, are not available in the database. Oxygen is considered crucial to the life of aquatic fauna especially macroinvertebrates that depend on oxygen in solution to meet their respiratory needs. In addition, the interactions of oxygen with other variables such as current, substrate, or temperature are considered important in the context of aquatic ecology (Hynes, 1960; Hellawell, 1986, Ward, 1992). Oxygen condition in water can be expressed by dissolved oxygen (DO) or biological oxygen demand (BOD).

It has been known that quantities of certain trace elements exert a positive or negative influence on aquatic plant and animal life including the distribution of macroinvertebrates. Trace elements play important part on the enzyme functions. Bivalves and crustaceans are extremely sensitive to heavy metals concentrations. On the basis of their biology, they are excellently suited for use as heavy metals indicator organisms (Forstner & Wittmann, 1983).

Impacts of toxicity, biotic and abiotic degradation of insecticide and pesticide residues on aquatic organisms are very important issues that need to be considered. Aquatic organisms including macroinvertebrates may be contaminated by chemicals through several pathways: directly via uptake through gills or skin as well as indirectly via ingestion of food or contaminated sediment given. Insects are known to be highly sensitive to insecticide toxicity, crustaceans are at lower level as well. Macroinvertebrates represent the most sensitive to biological response range to DDT, malathion and endrin (Mason, 1996).

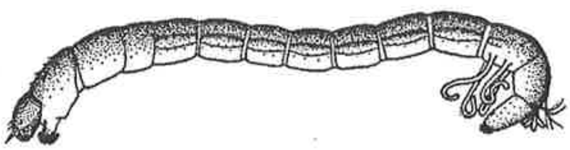
Unfortunately, data of the parameters discussed above is not available for all samples. Some data of oxygen level are still in preparation. We hope to be able with an extended database to develop more reliable models in order to improve prediction capability of ANN models based on dirty water approach.

5.6 Chapter Summary and Conclusion

The dirty water approach proved to be useful to assess stream habitat conditions. Good prediction results for both clean and potentially impacted sites prove the possibility of the approach to be applied for management purposes. Once functional interrelations between water chemistry and distribution of macroinvertebrates have been determined, the approach can be applied in the reverse way, using colonisation patterns of macroinvertebrates and easily-measured physical predictor to predict chemical variables, which are representative for the distributions of macroinvertebrates. This application provides a quantitative assessment of stream habitat conditions.

Investigation of sensitivity curves derived from dirty water ANN models using the methods described in this study greatly enhanced the understanding of the effects of impacts of various types on individual macroinvertebrate taxa. Results will enable to identify impact specific indicator taxa. The shape of the sensitivity curves of taxa would indicate how important it is to manage disturbances within certain bounds in order to maintain healthy aquatic ecosystems. More details about sensitivity analysis and how to apply results of sensitivity analyses as management tools are discussed in the next chapter.

However, a better database is required especially with data about water quality to get more meaningful interrelation such as oxygen and nutrient level, concentrations of trace metal and poisonous elements in stream habitats.



True flies Chironomid, are among the most pollution-tolerant macroinvertebrates (Hynes, 1960)

6 Elucidation of Freshwater Habitat Condition Discovered by a Sensitivity Analysis

6.1 Introduction

After training a neural network, sensitivity analysis is carried out in order to find out the effect that each of the network inputs is having on the network output. Sensitivity analysis is a method for extracting the causal relationships between the inputs and outputs of the network. The input variables that produce low sensitivity values can be considered insignificant and can usualy be removed from the network. This will reduce the size of the network, which in turn reduces the complexity and the training time. Furthermore, this may also improve the network performance.

Sensitivity analysis was conducted for each of ANN models. Each input variable was varied between its mean +/- a certain number of standard deviation while all other inputs were fixed at their respective means. The number of standard deviation is defined so that the computed range covers whole value of this input in the database. By theory, mean +/- 3 standard deviations can cover 99% of database, except for some extreme values (Bury, 1975). The network output was computed for a defined number steps above and below the mean. This process was repeated for each input variable. Plots were generated for each input variable and for each taxon specific ANN model illustrating the change of network output over the range of the varied inputs.

The primary intention of this sensitivity analysis was to improve the neural network model performance by limiting input variables to those that were sensitive for each model. Analysis of sensitivities allows to estimate the percentage change of the each output within the range of specific input. However, this process also identified important relationships between environmental variation and the occurrence of taxa.

6.2 *Design and Interpretation of Sensitivity Plots*

Various shapes of sensitivity plots can be distinguished which illustrate relationships between ANN input variables and outputs. Names indicate family-level identification of taxa. Outputs on the Y-axes are the predicted occurrence of a taxon ranging between 0 and 1, and inputs on the X-axes are environmental input variables over the ranges they were varied.

Some input variables had little influence on outputs. In these cases there was either no output response to input variation (Figure 6.1a) or the change in output response occurred over a very small range (Figure 6.1b). This type of response indicated that inputs had little or no influence on the occurrence of the macroinvertebrate family of concern.

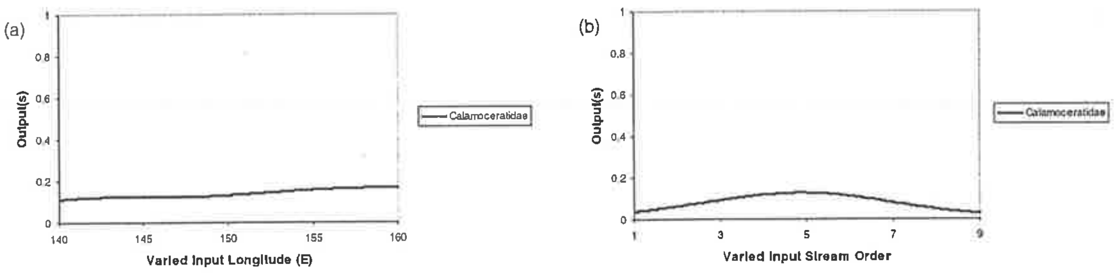


Figure 6.1 "Flat" responses

Of greater interest were the input parameters that resulted in a large range of output response as they vary (Figures 6.2, 6.3, 6.4). These were the variables that had a large impact on whether the ANN models predicted taxa to be present or absent in a sample. The nature of the relationships between input variables and the presence/absence of macroinvertebrate taxa was classified based on the shape of the sensitivity plots as follow.

As they varied, some input parameters produced a 'ramp' response in the predicted probability of a taxon being present, in which probability changed gradually from high to low. The ramp were either positive, meaning that the probability of taxon presence increased with increasing values of the input variable (Figure 6.2a) or negative, meaning the probability of taxon presence decreased with increasing values of the input variable (Figure 6.2b).

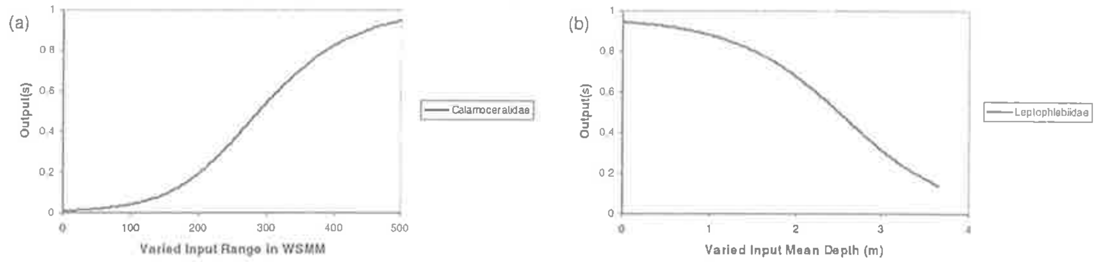


Figure 6.2 "Ramp" responses

The second type of response was a 'threshold' whereby the probability of occurrence changed from high to low over a narrow range of input variability. Once again the relationships were positive (Figure 6.3a) or negative (Figure 6.3b). The slope of the threshold can vary from almost a vertical drop (a 'cliff') to a much gentler 'incline'.

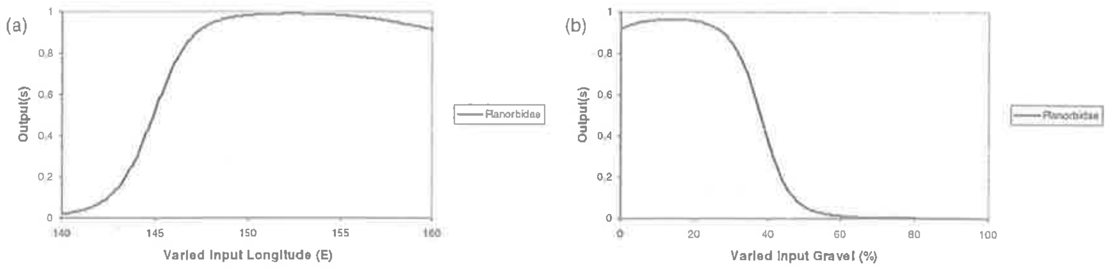


Figure 6.3 "Threshold" responses

The two remaining types of relationship between inputs and outputs identified from the sensitivity plots were a 'plateau' or a 'valley'. The plateau response had a continuous range of input values in which there was a high probability a taxon was present, with a drop off to low probability at either end of the range (Figure 6.4a). The valley response was the converse with low probability over a range of input values and high probability at either end (Figure 6.4b). In both cases there was variation between examples in the extent of the range of variable values at the top of the plateau or the bottom of the valley, producing either a broad or narrow plateau or valley. There was also variation in how steeply the output responses rose or fell outside the input ranges at the bottom of the valley or top of the plateau, and in some examples there was asymmetry between the slopes at the two ends of the range.

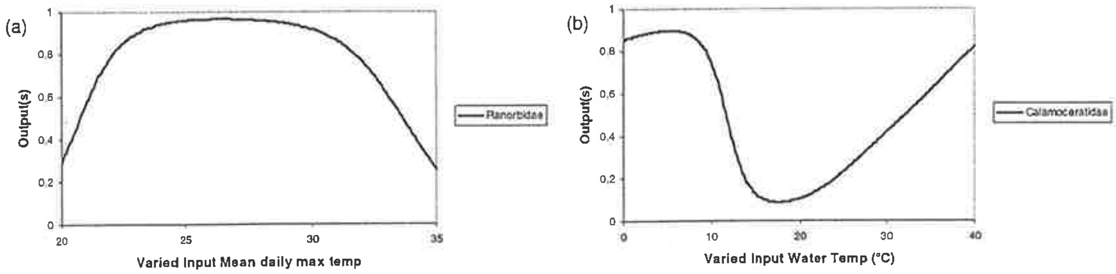


Figure 6.4 "Plateau" and "Valley" responses

Interpretation of these sensitivity curves provides new insights into relationships between the occurrence of macroinvertebrate families in Queensland streams, and variation in some physical properties as well as water chemistry of aquatic habitats.

6.3 *Elucidation of Causal Relationship by Sensitivity Plots*

The shapes of the sensitivity curves indicated the ecological response of taxa to an environmental variable. Ramp curves indicated a gradual change in preference over a range of an environmental parameter, whereas a threshold appeared more indicative of a more abrupt cut off in the tolerable range of a variable. A plateau response was really a modification of either the ramp or threshold curves with drop offs at either end rather than just one. The valley curve was more perplexing and may represent different preference and tolerance profiles of two or more species in a family. In principle, where more than two species with differing requirements are included in a family, sensitivity curves could contain multiple valleys or plateaus.

There is not the scope to present and discuss all sensitivity plots in this study, because each taxon may need comprehensive study for response to habitat conditions. I hereby only select some example to illustrate the types of relationships that were evident between input environmental variables and the predicted presence/absence of macroinvertebrate taxa. The objective of this is to demonstrate the usefulness of the technique, rather than to report specific relationships.

Worms

Dugesiidae are large, free-living flatworms. This is the most well-known of the freshwater flatworm families, and is widespread and common in streams in Australia. Worms are sensitive species to pH and are dominant in high-pH conditions (Hawking & Smith, 1997).

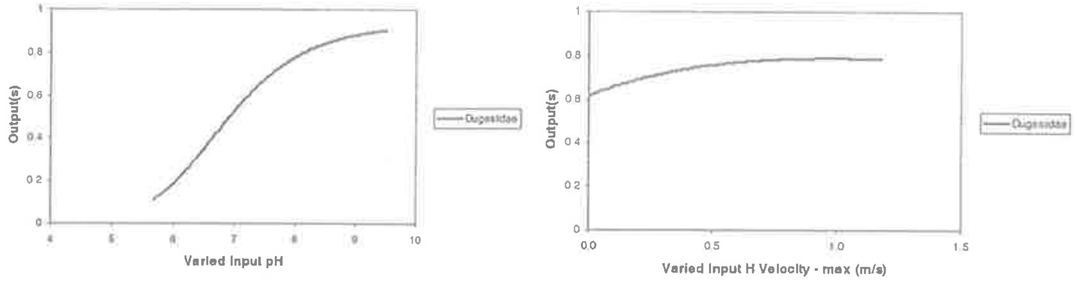


Figure 6.5 Preference of* Dugesiidae *to H velocity and pH

Oligochaeta (earthworms, segmented worms) are predominantly an aquatic class. Segmented worms inhabit on substrata of still and slow-flowing waters and also sensitive to acidic conditions. Figure 6.5 and 6.6 clearly support theoretically expected preference of *Dugesiidae* and *Oligochaeta.*

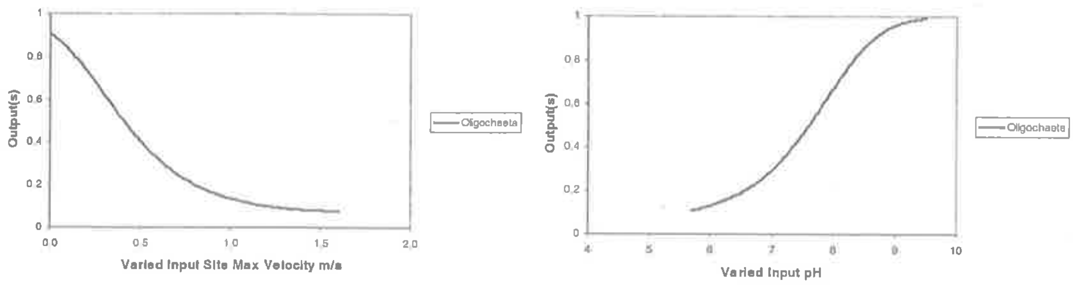


Figure 6.6 Preference of* Oligochate *to pH and site velocity

Crustaceans

Atyidae (freshwater shrimp) live in many types of water body but prefer comparatively still waters where they congregate under banks, large submerged boulders and aquatic vegetation. Most of them prefer surface-waters (Choy & Horwitz, 1995). Figure 6.7 illustrate the preference of *Atyidae* in flat habitat with slow current velocity.

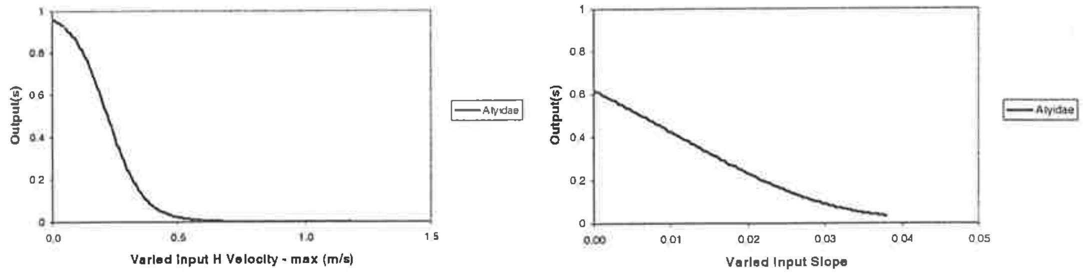


Figure 6.7 Preference of **Atyidae** *to H velocity and slope*

Palaemonidae (freshwater prawns) live in running or still permanent waters, and away from the coast (Horwitz, 1995). They are tropical species and therefore are mostly observed in high temperature waters (Figure 6.8)

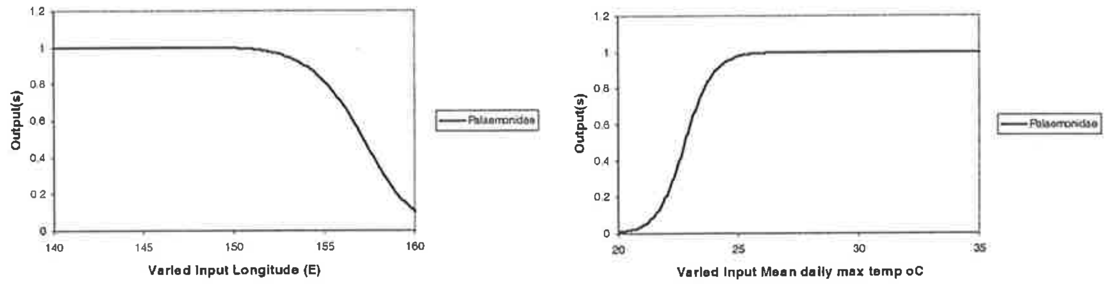


Figure 6.8 Preference of **Palaemonidae** *to longitude and mean daily temperature*

Copepoda appeared to occur at warm and still water (Figure 6.9)

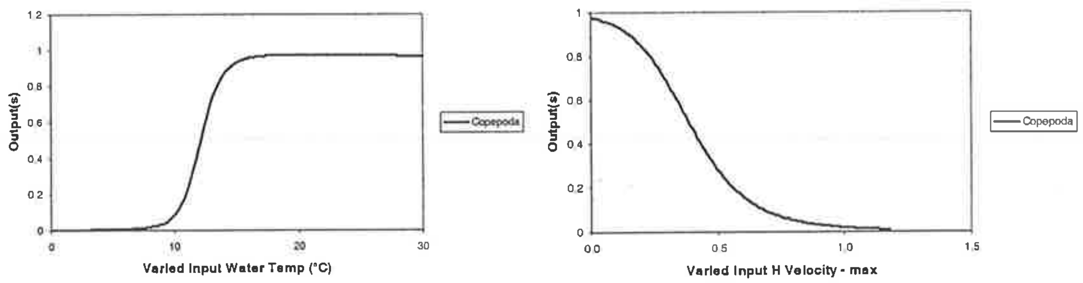


Figure 6.9 Preference of **Copepoda** *to H velocity and Water Temperature*

Cladocera (water fleas) is known to be freely swimming (nektonic) and requires slow flowing water at depth. While water depth is increasing flow velocity is slightly decreasing towards high order downland streams at low altitudes with small channel widths. The sensitivity curve in Figure 6.10 indicates optimum depths for *Cladocera* species from 0.2 to 1.2 m.

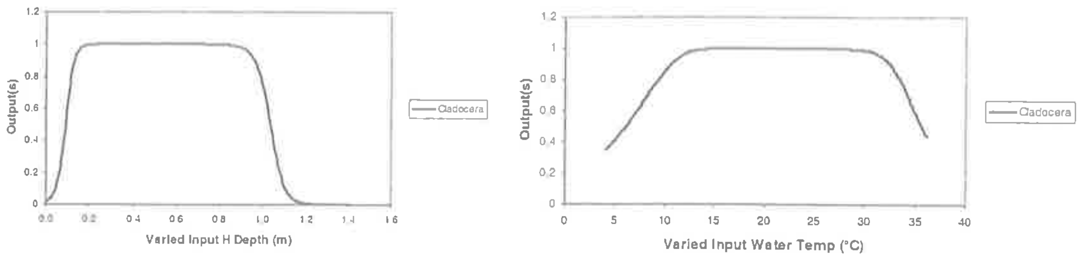


Figure 6.10 Preference of **Cladocera** *to H Depth and water temperature*

Figure 6.11 shows optimal temperature for several Daphnia species regarding physiological effects such as ingestion and reproduction rates maximum (Lampert and Sommer 1997). This relationship between temperature and biological activities is described as unimodal curves. A decrease in activity above the maximum is usually more rapid than the increase in the activity rate at sub-optimal temperature. A similar shaped relationship was discovered for water temperature (see Figure 6.10) that indicates optimum conditions for Cladocera in the range from 10 to 30°C.

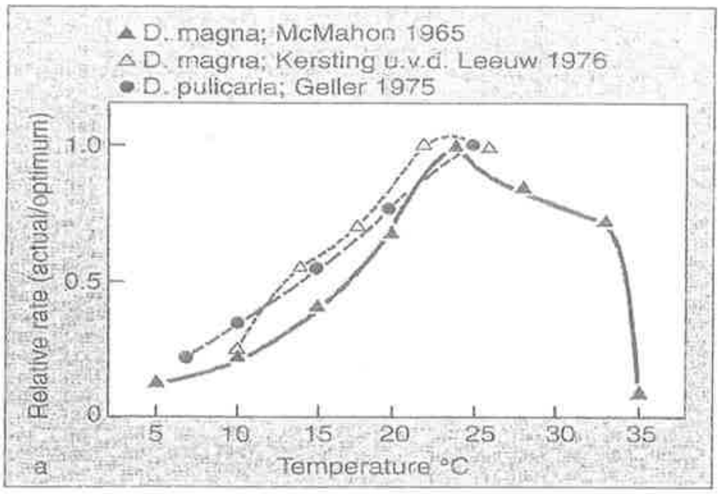


Figure 6.11 Physiological effects of water temp. to **Daphnia** *(Lampert & Sommer, 1987)*

Ostracoda (seed shrimps) are also good swimmers and often found to be abundant at water surface of fresh or saline water (Hawking & Smith, 1997). The sensitivity curve (Figure 6.12) shows preferred habitat depth for *Ostracoda* between 0.5 and 1m in the lower reaches.

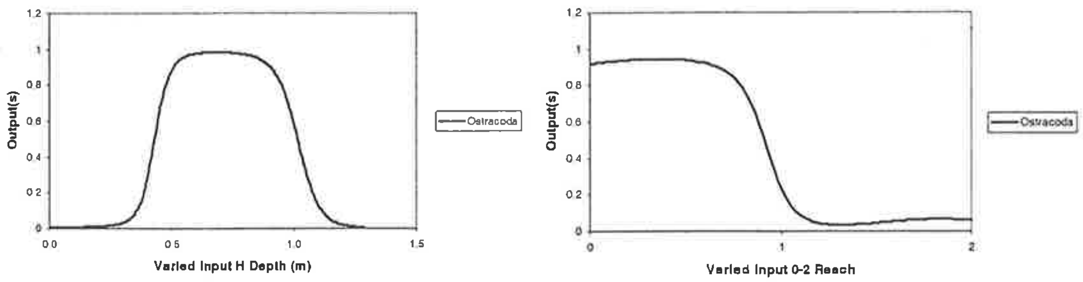


Figure 6.12 Preference of **Ostracoda** *to H Depth and reach*

Snails and mussels

Planorbidae is a cosmopolitan family of mainly left-coiled freshwater snails, which is confined to waters of low salinity, usually with algal growth or water weed on which the animals feed. Some species occur among dead leaves or other debris of slow-flowing rivers. They have been observed at habitat, which were highly polluted, oxygen depleted or very deep. They utilise haemoglobin or carry other respiratory modifications for coping with such conditions. Some exhibit considerable drought resistance. Planorbids often are the dominant molluscs at a site (Smith, 1996). Figure 6.13 shows that *Planobidae* can survive in drought condition with even extremely low dry-season-monthly-mean (DSMM) rainfall.

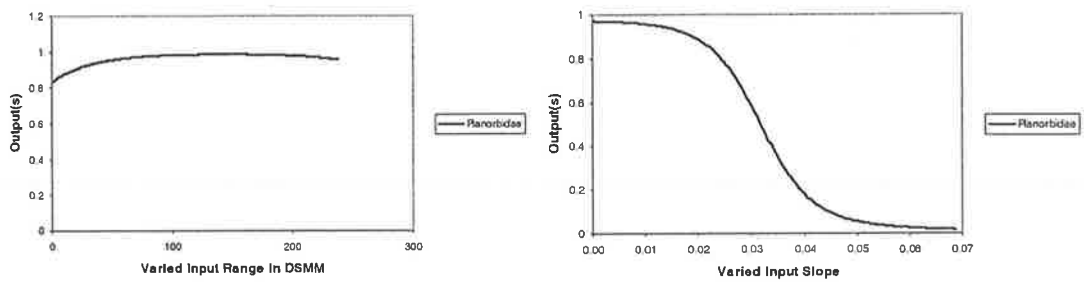


Figure 6.13 Preference of **Planobidae** *to slope and range in DSMM*

Thiaridae (marsh snails, black snails) is a worldwide-distributed family of almost exclusively freshwater snails. They are often found in great numbers, and in coastal lowland Queensland streams, where they are regularly the dominant molluscs, often in muddy habitats (Smith, 1996). The sensitivity plots in Figure 6.14 supports literature finding on *Thiaridae*.

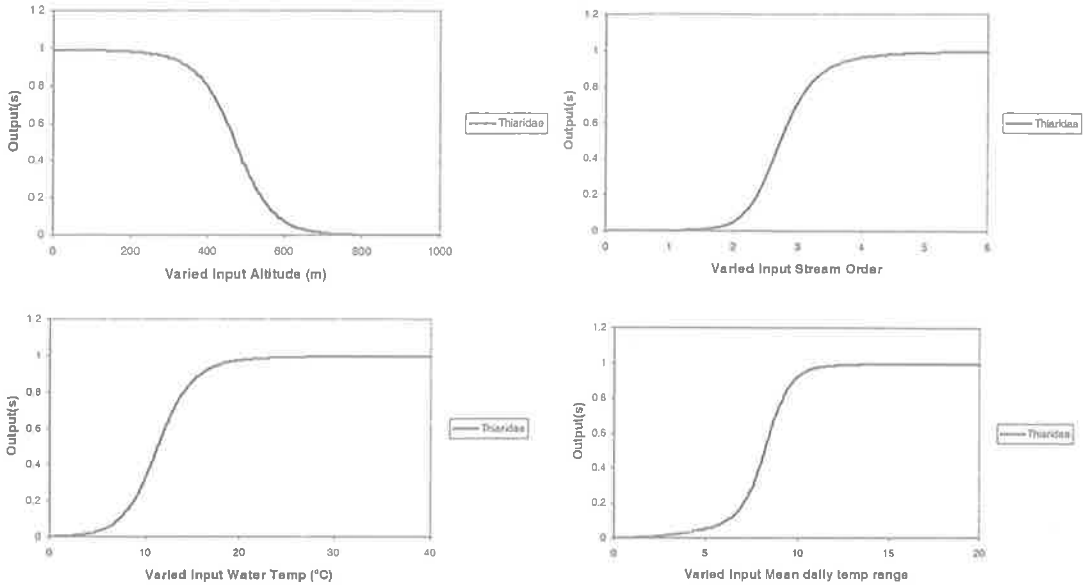


Figure 6.14 Preference habitats of **Thiaridae**

Corbiculidae are strong-shelled bivalves with a preference for flowing dowland river with sandy substrate (Smith, 1996). Corbiculidae are very sentivitive to low pH. They also accumulate toxic chemicals in the tissues. Corbiculids therefore have been used to monitor pH and various chemical contaminants. The sensitivity plots in Figure 6.15 show distinct sensitivity of *Corbiculidae* to pH and stream order.

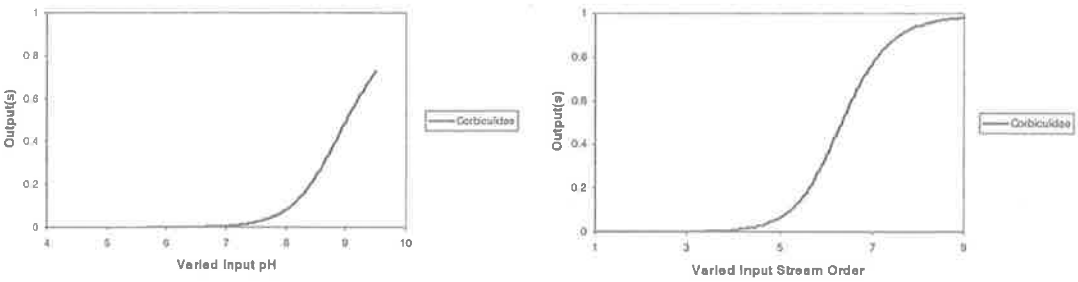


Figure 6.15 Preference of* Corbiculidae *to stream order and pH

Water mites

Norton et al. (1988) suggested that water mites (*Acarina*) are demographically as least as conservative as soil-dwelling relatives. They live in cold, oligotrophic waters and have multi-year generation time.

Models in both clean and dirty water approaches revealed that *Acarina* occurred in condition of low water temperature and low level of N and P, which characterise oligotrophic water (Figure 6.16).

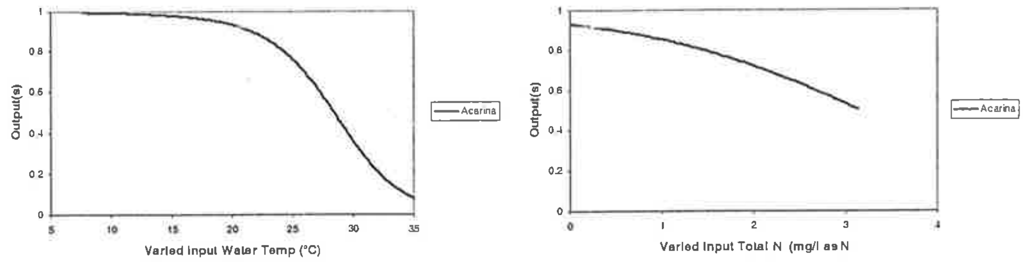


Figure 6.16 Preference of **Acarina** *to Water temperature and N level*

Mayflies

Baetidae are most common in clear, cold streams (Suter, 1996). They are amongst the earliest of mayflies to emerge where some appear on warm days at the end of winter. In Queensland, *Baetidae* was observed in southern part and never found in tropical areas. As fast swimmer (Hawking & Smith, 1997), *Baetidae* nymphs prefer deep habitat. This information was confirmed by the sensitivity analyses curves (Figure 6.17).

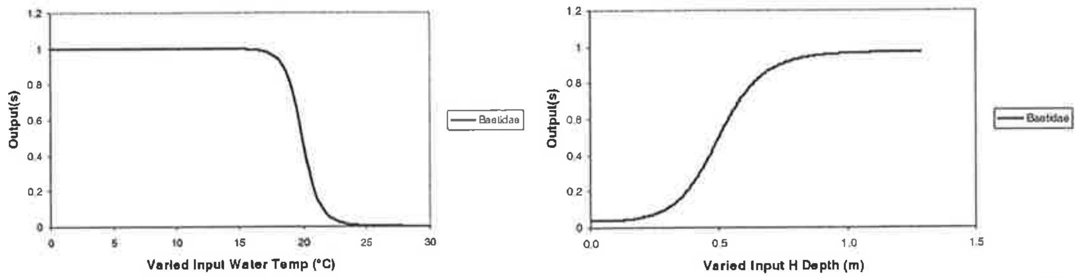


Figure 6.17 Preference **Baetidae** *to water temperature and H depth*

Prosopistomatidae is unusual mayfly from tropical rivers in north Queensland. They are common in riffles and fast-flowing, warm water (Dean, 1996) and are regarded as rare. Sensitivity curves show the trends of this taxon towards latitude and water temperature (Figure 6.18).

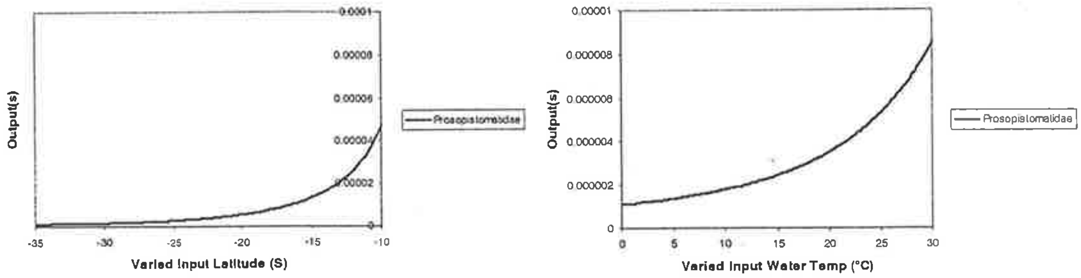


Figure 6.18 Preference of **Prosopistomatidae** *to Water temp and Latitude (H velocity)*

Caenidae prefers slow-moving stream. They are rare found in swift flowing waters. Nymphs burrow into the mud and sediment on the bottom of ponds and standing rock pools (Ward, 1992). Large channel also plays the role in current velocity. Narrow channels likely have higher velocity than wide channels. Figure 6.19 shows the preference of *Caenidae* in wide channels with slow-current streams.

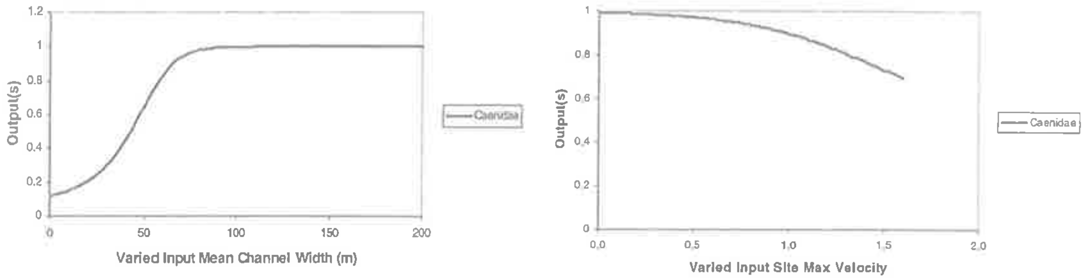


Figure 6.19 Preference of* Caenidae *to Site max Velocity and Mean Channel Width

Leptophlebiidae are adapted to various habitats from warm standing waters of coastal watersheds to melted snow of sub-alpine areas. In Queensland, nymphs were found in southeastern parts. As discussed in Chapter 5, these mayflies prefer low pH conditions. (Figure 6.20)

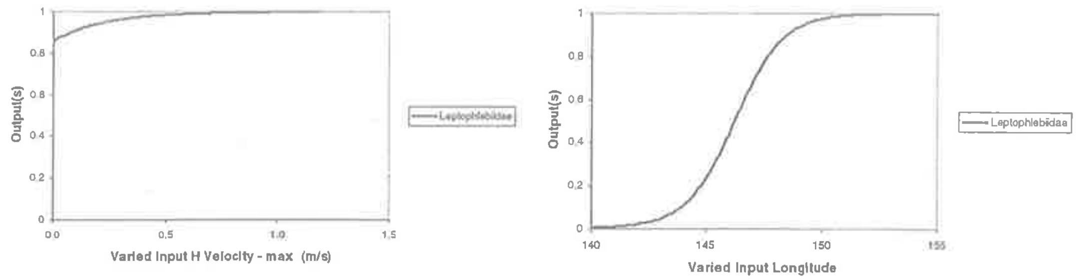


Figure 6. 20 Preference of* Leptophlebiidae *to Longitude and H velocity

Dragonflies, damseflies and stoneflies

Many species of *Odonata* burrow in fine sediment. Dragon fly nymphs typically lie buried in silt with only the eyes and respiratory aperture above the sediment. *Gomphidae* are cosmopolitan, swift, lender forms. They are frequent in running water. *Corduliidae* are found in eastern Australia. They are usually observed at clear fast streams. *Libellulidae* are tropical origin, mostly of still or slowly running water. The sensitivity plots in Figure 6.21 support these findings.

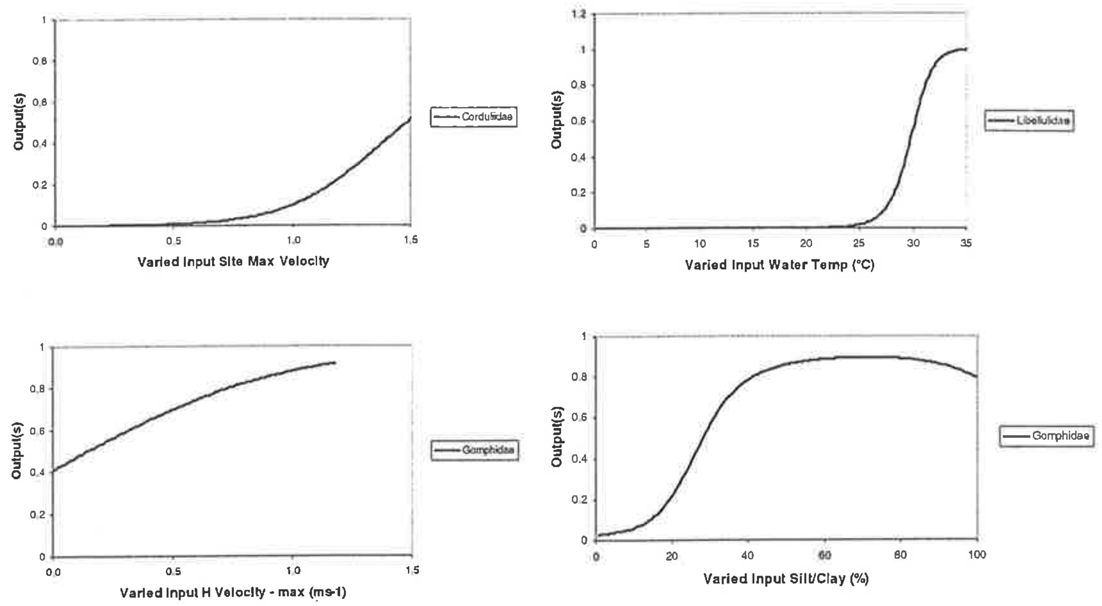


Figure 6.21 Preference of **Libellulidae** *to water temperature,* **Cordulidae** *to site max velocity, and* **Gomphidae** *to H Velocity and Silt-clay*

*Coenagrionida*e are mainly tropical as resulted in Figure 6.22. Adults can be found in low-ordered static water, while nymphs live among aquatic plant.

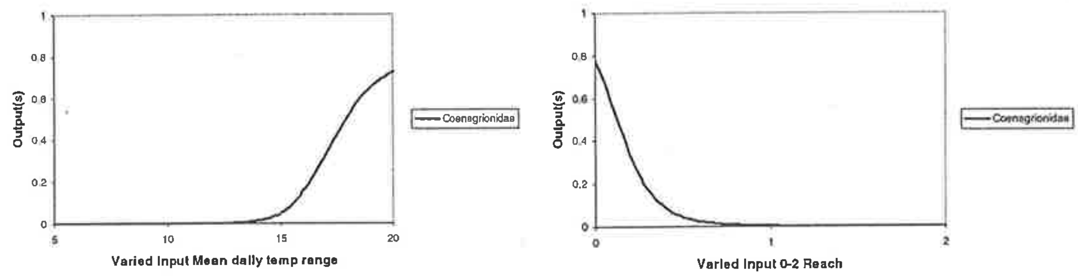


Figure 6.22 Preference of **Coenagrionidae** *to reach and mean daily temperature.*

Stoneflies *Gripopterygidae* nymphs live in the water with stony or gravel bottom. They require cool well-aerated water and can be found in shallow upper reaches of streams. Larvae occur on swift or on slow-flowing streams. Their distributions are observed in North Queensland. Because of the often highly specific environmental requirements of nymphs, stoneflies are particularly good water-quality indicators, especially where oxygen-demand pollutants are concerned (Ward, 1992). Dirty water models predicted presence of stoneflies at 10% of clean-water sites and only at 2% of dirty-water sites. Stoneflies tend not to occur where temperatures can exceed 25oC due to their oxygen requirement (Hynes, 1970). This is also demonstrated by the sensitivity plots in Figure 6.23.

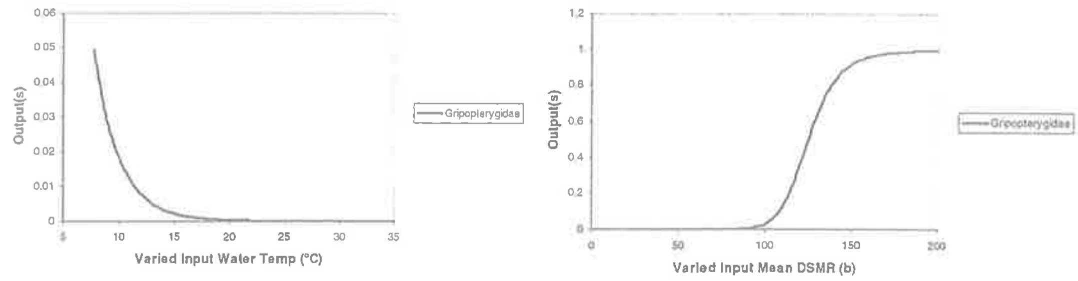


Figure 6.23 Preference of **Gripopterygidae** *to Water temperature and (DSMR)*

Bugs

Corixidae are predacious bugs that swim actively in the still and slow-flowing water and feed mainly on insect larvae in the bottom ooze in coastal areas (Carver et al., 1991). Sensitivity plots in Figure 6.24 confirm these findings.

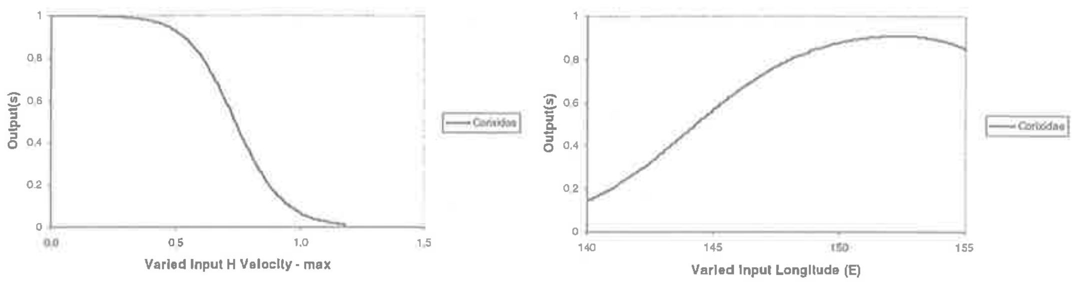


Figure 6.24 Preference of **Corixidae** *to longitude and H velocity*

Notonectidae, predatory bugs, swim upside-down, usually just under the water surface, and are common in still or slowly running waters as indicated by Figure 6.25. The family is cosmopolitan and occurs throughout Australia (Carver et al., 1991).

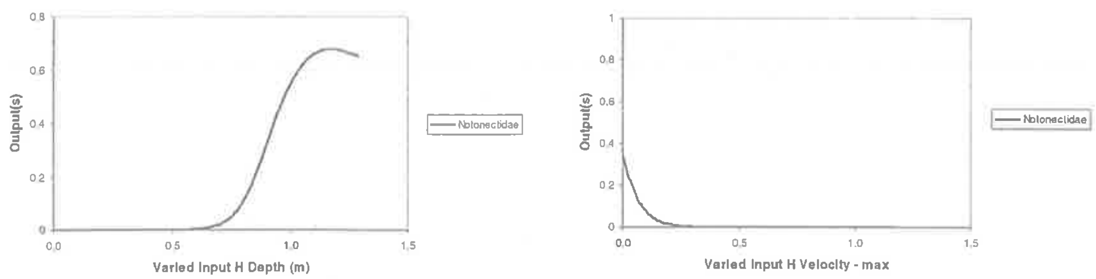


Figure 6.25 Preference of **Notonectidae** *to H depth(clean) and H velocity (dirty)*

Pleidae are frequently abundant in the tropical north (see Figure 6.26) in swamps and lakes (Carver et al., 1991).

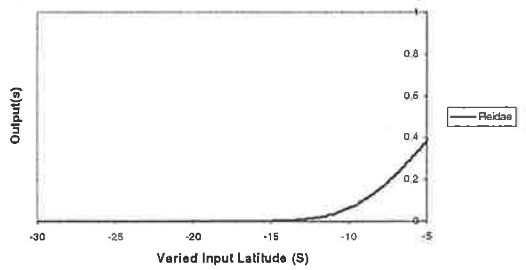


Figure 6.26 Preference of **Pleidae** *to latitude*

Veliidae are found swimming in large groups in sheltered side pools or along the edges of bodies water. They live amongst emergent vegetation and floating leaves at the surface of quiet areas of still and flowing water (Carver et al., 1991). Figure 6.27 shows the preference of *Veliidae* in flat habitat with habitat current velocity less than 0.5 m/s.

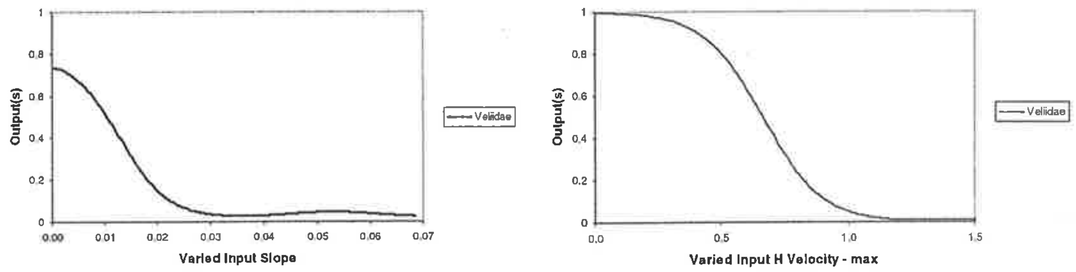
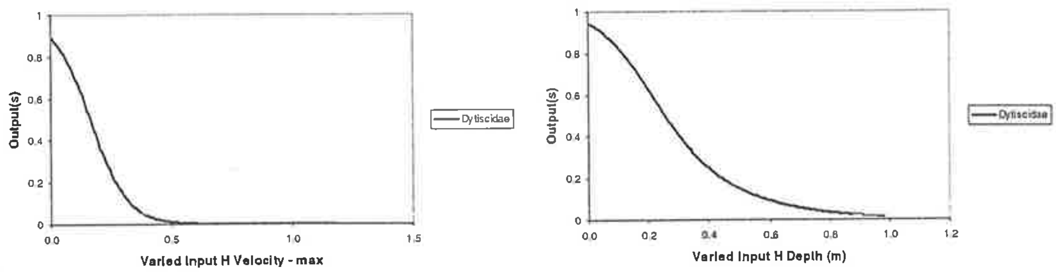


Figure 6.27 Preference of **Veliidae** *to H velocity and slope (clean*

Beetles

Dytiscidae (Figure 6.28) live in a variety of aquatic habitats but are most common in the littoral zone at the edges of lakes and ponds, and are found in running and still waters (Lawrence & Britton, 1991).

Figure 6.28 Preference of **Dytiscidae** *to H depth and H velocity*



Elmidae (Figure 6.29) were found in all kinds of streams. They are more common in shallow running water, rocky bottoms, clear water and high oxygen contain. They frequently the only *Coleopterans* present in torrential streams (Ward, 1992).

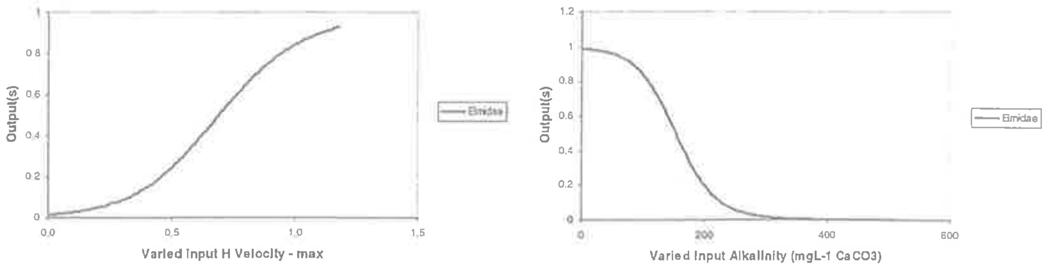


Figure 6.29 Preference of* Elmidae *to H velocity and alkalinity (clean)

Larval *psephenids* (Figure 6.30) are streamlined animals. They feed on algae on rocks in exposed positions with high-energy flows.

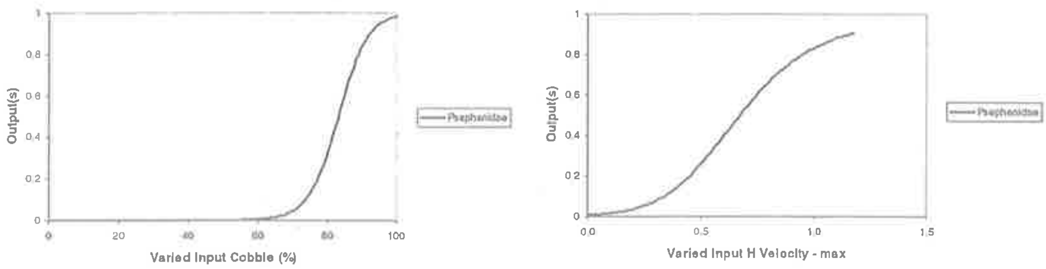


Figure 6.30 Preference of* Psephenidae *to H velocity and % cobble

Most larval *Hydrophilidae* (Figure 6.31) are fully aquatic, occur in a wide range of lotic habitats, though shallow well-vegetated margins of still waters, quiet areas of flowing water and stream banks are most favoured (Lawrence & Britton, 1991)

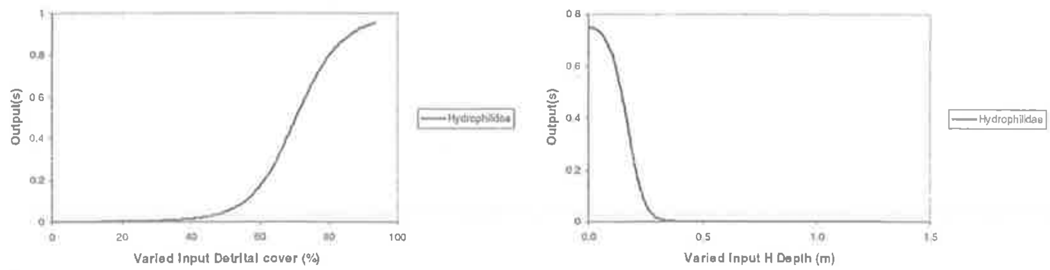


Figure 6.31 Preference of* Hydrophilidae *to H Depth (clean) and detrital cover

True flies

Midge larvae (*chrinomids)* are the most dominant in aquatic ecosystem. The neural network models considered two subfamilies *Tanypodinae* and *Orthocladiinae*. *Orthocladiinae* have cold-stenothermic nature and are dominant in subalpine and mountain streams, where the maximum water temperature in summer may reach 10oC. In middle and lowland stream, where temperature may exceed 20oC, *Orthocladiinae* significantly decrease (Lindegaard & Brodersen, 1995). Sensitivity analyses confirmed the presence of *Orthocladiinae* at upper reach and cold water (Figure 6. 32)

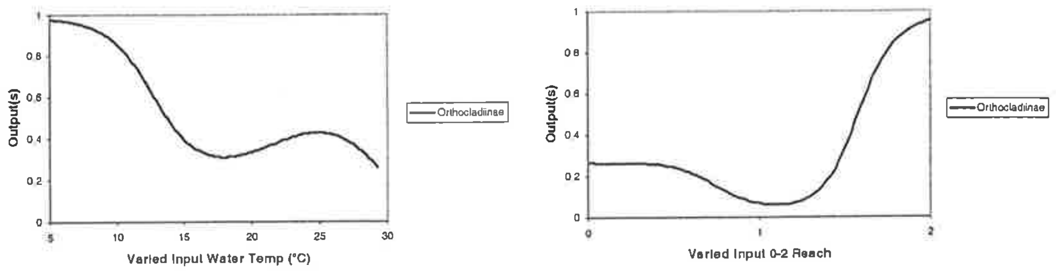


Figure 6.32 Preference of* Orthocladiinae *to reach and water temperature

Tanypodinae appear very few in montane and subalpine streams and increase furthers downstream the river continuum with higher water temperature (Lindegaard & Brodersen, 1995) (Figure 6. 33)

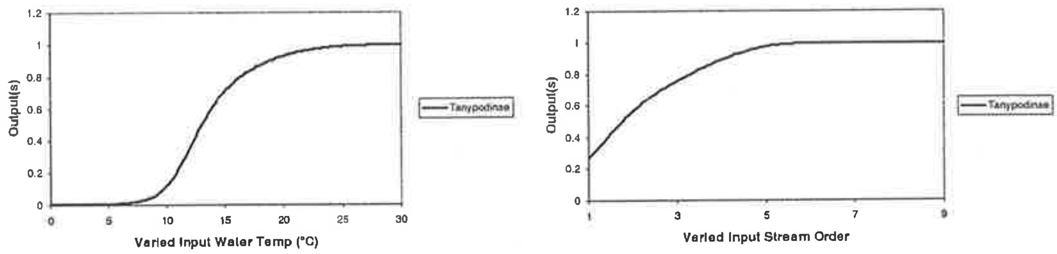


Figure 6.33 Preference of* Tanypodinae *to stream order and water temperature

The *Simuliidae* (blackflies) are amongst the most characteristic running water macro-invertebrates. The characteristic habitat of blackflies is attaching to substratum of flowing water in highland stream as indicated in Figure 6.34.

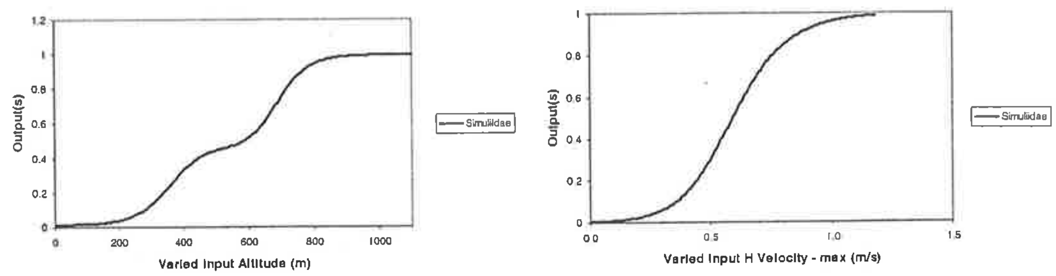


Figure 6.34 Preference of* Simuliidae *to H velocity and altitude

Ceratopogonidae and *Tabanidae* (horse flies, marsh flies) (figure 6.35) are amongst the commoner of the "higher" *Diptera* in aquatic habitats. They vary in habit, but prefer wet sand and mud at cool stream margins.

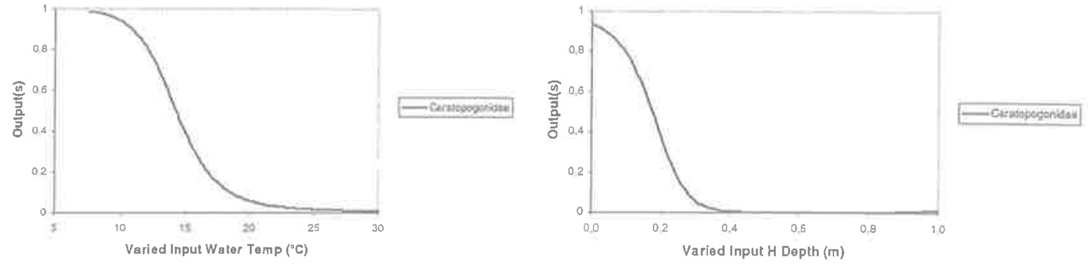


Figure 6.35 Preference of* Ceratopogonidae *and* Tabanidae *to Water temperature and Habitat depth

Caddis flies

Leptoceridae (Figure 6.36) are very common Trichoptera found in a wide range of habitats: from mountain streams to lakes, including temporary pools and saline waters. *Leptoceridae* was included in the list of aquatic insect families characteristic of potamal zones, where the annual range of water temperature exceeds 20oC; the current is slow and dissolved oxygen deficits occur at time (Ward, 1992).

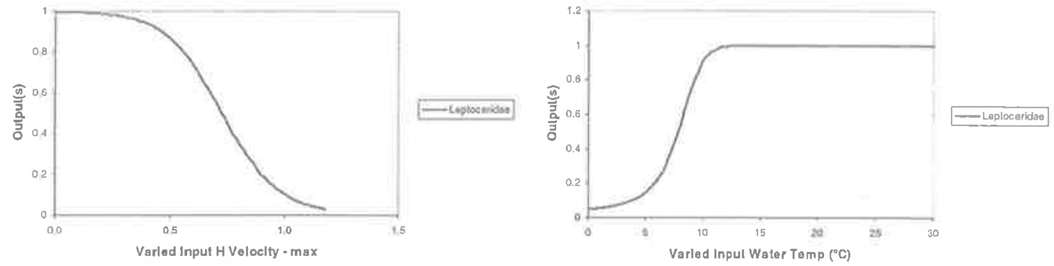


Figure 6.36 Preference of* Leptoceridae *to H velovity and water temperature

Hydropsychidae (figure 6.37) build fixed retreats of plant material or rock fragments in fast running water and constructs a filter net in the current to capture algae, organic debris and small invertebrates as food. Distribution is observed in eastern Australia (Neboiss, 1991, Hawking & Swmoth, 1997).

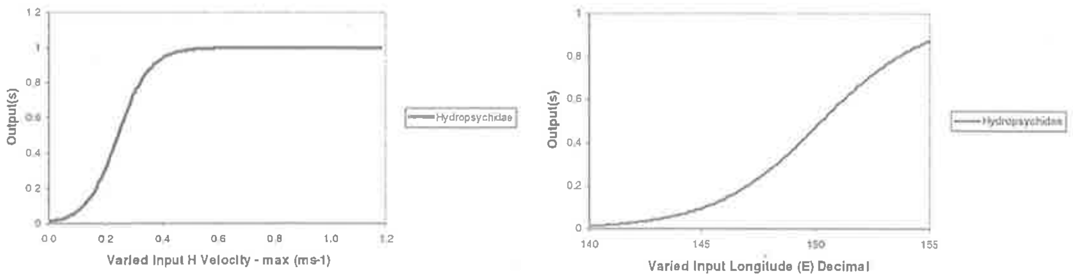


Figure 6.37 Preference of* Hydropsychidae *to longditude and H velocity

The larvae *Ecnomidae* construct fixed tubes of silk incorporating plant and mineral material, attached to logs and rocks (Neboiss, 1991). Their habitats include still and flowing water thus sensitivity analyses shows that they prefer fast flowing water, and as most of caddish flies they prefer cool water temperature (Figure 6.38) (Hawking & Smith, 1997; Hellawell, 1986).

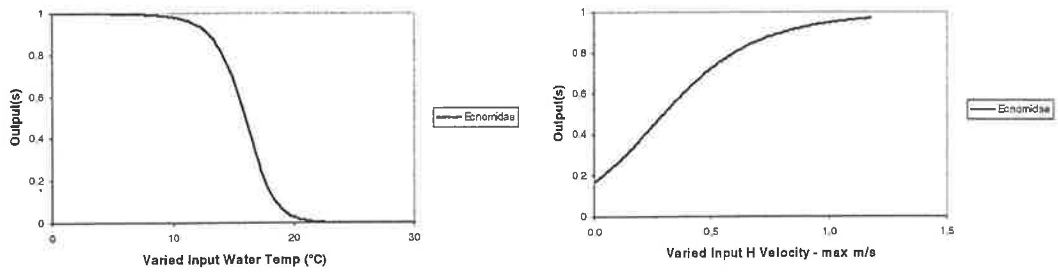


Figure 6.38 Preference of **Ecnomidae** *to H velocity and Water temperature*

Hydroptilidae are very small caddisflies (micro-caddis), which live in ponds, backwaters and areas of deep silt. The larvae are free-living. The case is attached to the substrate immediately prior to pupation. The sensitivity analyses shows the shallow habitat with few boulder substrates was favourite for them (Figure 6.39)

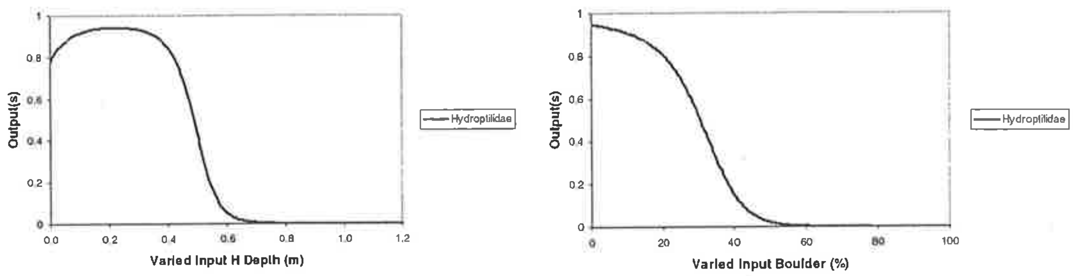


Figure 6.39 Preference of **Hydroptilidae** *to H depth and boulder content*

Calamoceratidae inhabit sluggish still or slow flowing rivers and feed on plant debris (Neboiss, 1991). The illustrated response to water temperature (Figure 6.40) may represent the different ecological requirements of *Calamoceratid* caddis fly species from north and south Queensland (Marshall et al., 2000)

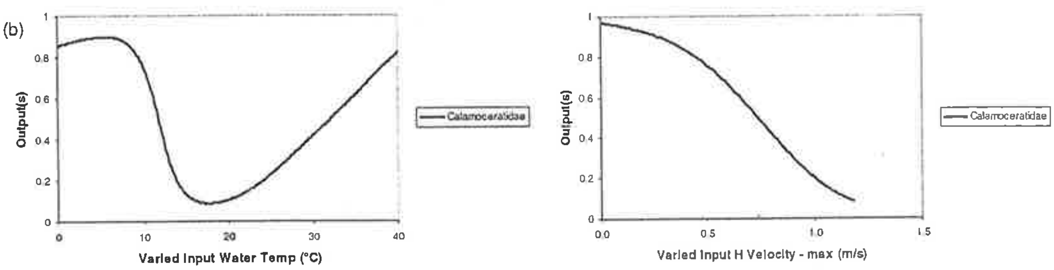


Figure 6.40 Preference of **Calamoceratidae** *to Water temperature H velocity*

Philopotamidae Small to medium sized *Trichoptera* found mostly amongst large stone in clear, rapid streams where they build fixed silken tubes or sack-like nets (Cartwright, 1997). Sensitivity analyses shows the trend of these caddis flies to present in rapid habitat velocity and highland streams (Figure 6.41)

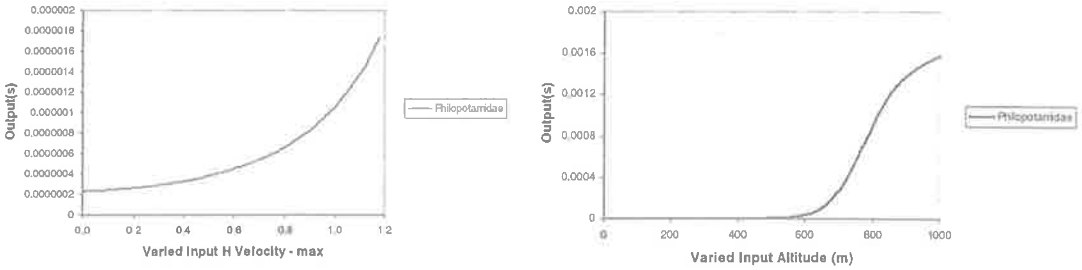


Figure 6.41 Preference of* Philopotamidae *to H velocity

Moths

Most moth species with aquatic larvae belong to the family *Pyralidae* (Figure 6.42). They are attached to bedrock in fast-flowing streams and rivers (Hawking & Smith, 1997).

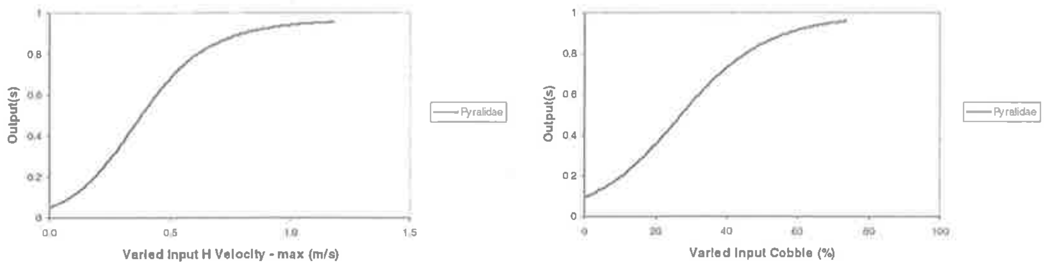


Figure 6.42 Preference of* Pyralidae *to H Velocity and %Cobble

Many other relations were detected and were in most cases confirmed with related ecological results, when information was available. The results indicate that developed neural network models in many cases work in ecological meaningful manner.

6.4 *Contradiction to Literature Findings*

Most of the relationships found by sensitivity analyses were confirmed by literature finding. However, there were few cases where contradiction was observed.

Plots in Figures 6.43 is discussed some cases as examples.

Giller & Malmqvist (1998) discussed that a majority of stream living *triclads* (only one family *Dugesiidae*) is cold-living species. The sensitivity curve (Figure 6.43a) however shows that *Dugesiidae* could present only in conditions of water temperature exceed 20oC.

Libellulidae are tropical origin but sensitivity curve (Figure 6.43b) shows their absence at latitude above –20(S), which is characterised for tropical zones of Queensland.

Dytiscidae have the greatest diversity occurring in the southeast part of Australia and are most common in the littoral areas (Lawrence & Britton, 1991). Information obtained by sensitivity analysis (Figure 6.43c and d) shows that they were found only in North Queensland at low-ordered streams.

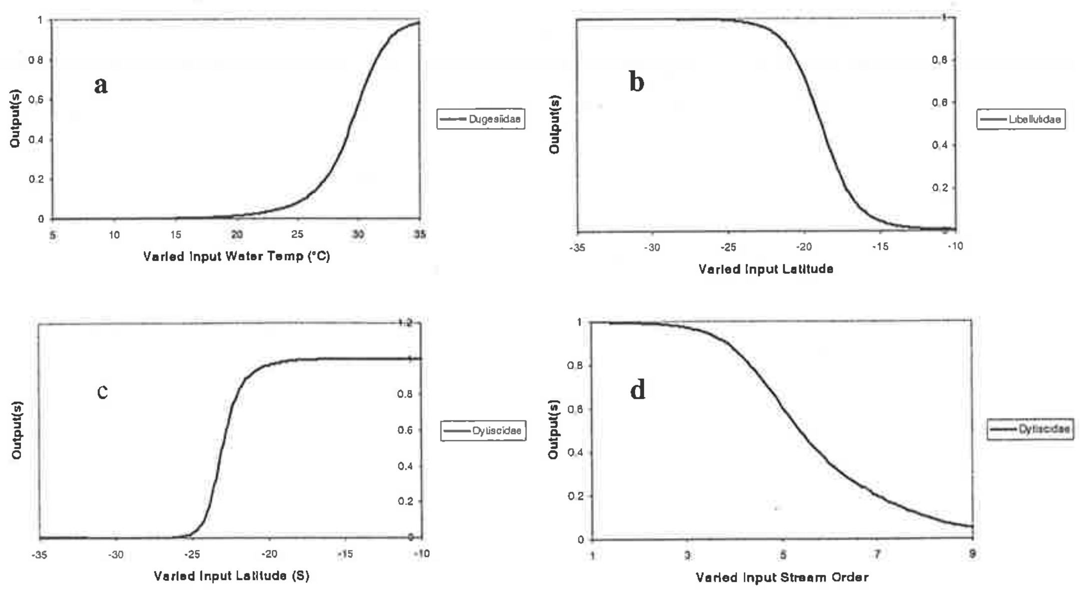


Figure 6.43 Examples of contradiction between sensitivity analyses and literature finding

These contradictions are very interesting parts of results as well. They require further research to find out whether the models generate wrong patterns or they provide new insight into those relationships.

6.5 *Limitation of the Method*

The method only considered the impact of single input on output response. In fact, effects of habitat conditions on distribution of macroinvertebrates are always multivariate patterns. Many factors are not independent but closely interrelated.

The interrelation among current, temperature and oxygen demands of macroinvertebrates is an example. The current continually replenishes water and hence also oxygen in the immediate vicinity of the respiratory surface of the animal and quite low levels can be tolerated in strong currents that renew oxygen at a high rate. Generally, metabolic rates and oxygen demand are higher in stream invertebrates than in still water forms at a given temperature. Respiration is temperature-related and rates can increase by10% or more per 1oC temperature rise. Thus increased temperature does not only reduce oxygen availability but it also increases oxygen demand that can add to the physiological stress of organisms (Giller & Malmqvist, 1998).

The most important hydraulic characteristic for individual organism is the prevailing current velocity striking the organism head-on (Statzner et al., 1988). Species do react differently to current velocity, show differential preferences and consequently different flow conditions lead to divergent assemblages of organisms. In a detailed survey, boundary layer Re (Renolds number) was the most strongly correlated individual variable with invertebrate distribution and taxon richness in two New Zealand streams but a combination of mean velocity, substrate size, and depth gave stronger correlation than any single variables (Quinn & Hickey, 1994). It appears that the interaction between current velocity and stream substrate size is important in determining invertebrate distributions.

Orth & Maughan (1983) determined optimum velocity, depth, and substrate for major taxa on benthic macroinvertebrates of warm-water woodland stream. The combination of current velocity of 60cm/sec, a depth of 34 cm and rubble-boulder substrate resulted in optimal diversity of benthic assemblages. Recognising that habitat selection by benthos may be based on factor combinations, the investigators derived "joint preference factors" using the product of the individual preference factors.

Ecological patterns that underlie these multivariate patterns are characterised, for example, by the fact that mean species richness and total species pools increase with pH (Hildrew & Townsend, 1987). Stream with a pH as high as 6.5 but low alkalinity (low Ca^{2+}) often show similar features to more acidic waters (pH< 5.5, Willoughby & Mappin, 1988). Effects of pH on aquatic fauna are different on different conditions of water temperature (Hynes, 1970). Food supply also depends on current speed, either to convey particles to filter feeding organisms or to deposit detritus (Hellawell, 1986). Toxicity of ammonia and hydrogen sulfide to aquatic organism is dependent on both temperature and pH conditions (Smith & Maasdam, 1994).

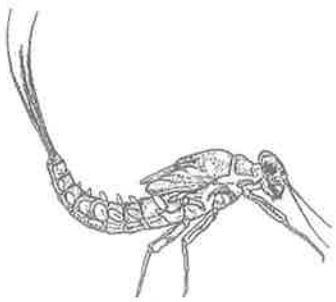
These and other examples illustrate the multivariate effects of different habitat condition variables on distributions of macroinvertebrates. To study the effect of individual variable while keeping all other variables at their respective means ignores the very important pattern–combination of factor preferring by macroinvertebrates. Further research needs to consider this fact and improve the practicability of sensitivity methods on elucidation of the freshwater habitat conditions.

6.6 Using Sensitivity Analysis for Management Purposes

The sensitivity curves obtained by sensitivity analyses can contribute information for better management and maintenance of good condition of streams and rivers. The shape of the sensitivity plots of taxa would indicate how important it is to manage disturbances within certain bounds in order to maintain healthy aquatic ecosystems. Taxa with a threshold response to a disturbance would be eliminated from a stream beyond a certain disturbance level, whereas taxa with ramp responses would gradually become rarer as disturbance intensified. The identification of such thresholds would provide water resource managers with a powerful tool.

Investigation of sensitivity curves derived from dirty water ANN models with more comprehensive database using the methods outlined here should greatly enhance our understanding of the effects of impacts of various types on individual macroinvertebrate taxa. This will enable impact specific indicator taxa to be readily identified and should enhance our capacity to monitor and mitigate the effects of human activities on stream ecosystems.

Family level identifications were used in this study because data were collected to develop rapid assessment techniques. Sensitivity curves could be more usefully applied to generic and preferably species level data to better study the ecological responses of macroinvertebrates to freshwater habitat conditions.



Mayflies Baetidae ar*e usually one of the most frequently encountered families in lotic habitats (Giller Malmqvist, 1998)*

7 General discussion

7.1 *Performance of Artificial Neural Networks*

The capabilities of Machine Learning technique in general, and Artificial Neural Networks in particular, do not just only come from the specific induction method, but from a proper and meaningful formulation of the problems, representative data, and from crafting the model representation to make learning tractable. For example, there is no point in developing an ANN to predict water quality variables in stream from hydrological and meteorological information. Field applications of machine learning, including ANN techniques and sources of power in applied machine learning, were the main sources of discussion in the edited work by Michalski et al. (1999). In supporting biological assessment of stream habitat conditions, machine learning has not completely automated the knowledge engineering process, but it has replaced knowledge engineering with two simpler tasks: characterising the problem and designing a good data representation. This section discusses the most important steps in successfully applying artificial neural networks to solve the given problems.

The first step in using any machine learning technique, including artificial neural networks, to solve any real-world problem is to reformulate the problem in terms that can be dealt with by some induction method (Langley & Simon, 1999). Studying the relations between stream habitat conditions and biological responses such as distributions of macroinvertebrates is a complex task, yet we need to identify components that involve simple classification, a task for which robust induction algorithms exist. In this research, presence/absence of a single taxon might not be an optimally chosen target task. Discrete outputs contain only two values (0 and 1), that might cause difficulties in induction algorithms of ANNs (Rumehart et al., 1986).

Functional feeding groups or abundance of indicator taxa, which are represented by continuous numerical values, may be better alternatives to improve the performance of neural network models.

The second important step is to settle on an effective representation for both training data and the targeted knowledge to be learned. Representation refers to features used to describe examples and to characterise the results of learning. The representation of the problem classified in the research is based on the expert knowledge about the ecological requirement and attributes that are likely to have predictive value. For example, data of oxygen conditions in the systems was crucial in the dirty water approach (Chapter 5). Lack of this information might severely affect the performance of neural network models.

After settling on a task and representation, training data needed to be collected for the induction process. In some cases, this process is straightforward and can even be automated, but in others it can pose a significant challenge. Training data can be very representative for sites or not reliable as a result of technical problems. Most data fall somewhere between these two extremes, and the expert is needed to classify training data or to generate it. Therefore accessing the available data and generating data where it is lacking is an important part of applied work in machine learning.

Rules induced from training data are not necessarily of high quality. A standard approach to evaluation involves dividing the data into two sets, training on the first set, and testing the induced knowledge on the second. This process can be repeated a number of times with different splits, and the results of testing are then averaged to estimate the rules' performance on completely new problems (Wilson & Recknagel, 2001). An important part of the evaluation process is the experts' examination of the learned knowledge. If significant problems emerge at this stage, they may suggest revisions to the problem formulation or representation (Langley & Simon, 1999).

The final stage in application is employing the learned knowledge base. Machine learning can either confirm the expert knowledge or introduce an extension to the knowledge base. Results of sensitivity analyses obtained by neural network models in this study provided many confirmations to findings from the literature on interrelations between habitat conditions and distribution of macroinvertebrates. Apart from their elucidation potential, some problems have risen which need further research, such as the cause of overpopulation of macroinvertebrate taxa in many sites, or the effect of detrital cover on distribution of taxa. Explanation and application capabilities of neural network models depend heavily on the requirements and objectives of users. Therefore, it is extremely important to motivate users and domain experts to participate in the design and application process in order to improve the effectiveness of applied neural networks (Langley & Simons, 1999).

The good performance of ANN models in both clean and dirty water approaches confirmed the potential of neural networks in predicting habitat conditions of freshwater streams. Sensitivity analyses carried out by ANN models allow the determination of major variables that affect ecosystem quality. Method should be considered further investigation in order to be applied in river ecosystem management. Further research is needed to determine the optimal neural network configuration. Moreover, the impact of the applied training algorithms, as well as the risk of overtraining the network, should be further investigated to obtain reliable and meaningful predictions in the long run.

The predictions for moderately frequent taxa such as *Gomphidae* and *Oligochaeta* were less accurate than for common and rare taxa. The results of sensitivity analysis showed that the distributions of these taxa were controlled by a greater number of variables. It is obvious that the more controlling factors, the greater the chance for potential sources of errors – a classic case of the complexity/uncertainty trade-off (Chapter 4). The enormous amount of information available on macroinvertebrate taxa is often too superficial, specific and contradictory, as data collection depends heavily on sampling methods, identification protocols and many other subjective factors such as experiences of samplers. Potential sources of errors could be an inappropriate identification of macroinvertebrate taxa and the spatial–temporal variability of physical and chemical variables as well as natural noise in the data set.

In summary, the potential sources of error in the predictive capabilities of Artificial Neural Networks can be caused by the potential mis-configuration of the ANNs (e.g. overlifting). They can be explained by the fact that most river system database do not contain enough information in order to extract the main relations existing between the structural, physical, chemical and biological variables. However, representation of data can be rectified in many ways in order to improve the performance of neural networks in solving the given problem.

7.2 *Improvement of Input Data Representation*

The most important basis for successful neural network modelling is a reliable and representative database. The networks learn from examples, and the quality of the ANN models depend heavily on the quality of the database. Therefore, representative and compatible data are the main requirement for neural models.

To increase the effectiveness of ANN learning, it is necessary to perform some preprocessing of the data before presentation to the network. Many categorical variables are just empirical presentation, and the numbers are just notation of the information. These input variables need to be split into several categories to avoid misinterpretation of notations as arithmetic numbers (Masters, 1993). Rather than supplying a single category with values such as 1, 2, 3, three input nodes should be created instead, each representing a category and containing 1/0 values only corresponding to whether the category was selected or not. The following variables needed to be sub divided into several inputs:

- Habitat 1-5: 5 nodes
- Substrate categories: 8 nodes
- Reach 0-2: 3 nodes
- Soil Type number: 18 nodes
- Soil Class number: 8 nodes
- Vegetation Type number: 9 nodes

Ordered categorical criteria such as stream order should not be split, as their order clearly affects weights of the connection links in neural networks.

Many water quality variables, such as oxygen concentration, nutrient levels, concentrations of trace metals were found crucial for distribution of macroinvertebrates (See Chapter 5) but were not available for inputs in the database. There was also no information on macrophyte density at sites and some other riparian variables that were considered important for macroinvertebrate assemblages such as riparian cover, deciduous cover (Hawkins et al., 2000). These data need to be supplemented for better representation of the database.

In summary, data representation can be improved by re-expression of categorical inputs into more conducive to ANN learning and collection data of more relevant forcing functions in the input layer.

7.3 *Temporal and Spatial Variations*

Variation in the distribution and activity of aquatic organisms is evident at all spatial and temporal scales, but especially in streams, where biotic differences are often obvious from rock to rock within a reach, from reach to reach within a watershed, and across watersheds. The distributions, abundance, and activities of aquatic biota vary conspicuously with time, as well, over temporal scales raging from second or minutes to years (Stewart and Loar, 1994). The distributions and activities of benthic invertebrates vary greatly, both spatially and temporally, over all scales in polluted and non-polluted flowing-water systems (Hynes 1960, 1970, Cummins, 1979).

7.3.1 Spatial Variance in Invertebrate Data

At small spatial scales, flow strongly affects the spacing patterns and foraging activities of macroinvertebrates. The spacing difference between competitive success or failure, metabolic activities and feeding status of taxa is in the order from millimeters to centimeters or a few metres distance (Stewart & Loar, 1994; Newbury, 1996). At whole-pool and within-pool spatial scales, species-level and ontogenetic shifts in behaviour attributed to predation risk can strongly influence invertebrate communities (Gilliam et al., 1989).

Over larger spatial scales, changes in invertebrate activities and abundances within streams and rivers can be large, even in the absence of human impacts. This expectation emerges naturally in consideration of the river continuum concept (Vannote et al., 1980). Major shifts in species and/or functional groups of aquatic insects occur in response to changes in substrate, temperature, chemical constituents, food supply and predators, with increase in stream order (Stewart & Loar, 1994).

In the clean water approach, the models showed the similar levels of impact at the test sites, probably indicating a high level of redundancy among the habitat types.

The inclusion of more than one habitat in a predictive model may cause confounded assessment of biological impairment. When habitats are included as individual objects, prediction of test sites into groups of equivalent reference sites is made according to the characteristics of particular habitat types rather than general site features. In Australia, edge and woody debris are the most common by occurring habitat in large lowland rivers (Parson & Norris, 1999). The recommendation can be made that separate models need to be developed for each of habitat types; consequently five models in total should be made for riffle, edge, run, pool bed and macrophyte habitats.

However, Parsons & Norris (1999) suggested that riffle and edge are adequate to detect biological impairment. In addition, many of the macrophyte beds are located in the marginal areas, and it is often difficult to distinguish an edge habitat from a macrophyte habitat. These two habitats can be accounted for by only sampling the edge, with no detrimental effect on the outcome of the predictive models and save the cost of sampling and simplify the data processing and model developing. Therefore, four separate models should be developed to improve the performance of neural network models.

7.3.1 Temporal Variance in Macroinvertebrate Data

Macroinvertebrates have a strong seasonal cycle of abundance and/or activities. They also tend to have a shorter life cycle than fish. The shorter life cycles suggest more rapid responses at the community level, which is an important advantage of using macroinvertebrates as bioindicators; however greater temporal variability needed to be considered (Steward & Loar, 1993). The database of the Queensland stream system does not contain any temporal information except seasonal category. It therefore did not allow a deeper study of the long-term effects of any variables, especially water chemistry variables, on macroinvertebrate assemblages.

Although tropical, Queensland has a seasonal climate and does not exhibit the environmental constancy associated with wet equatorial tropics. The time scale of sampling, particularly when it is carried out over more than one season, can significantly affect results of bioassessment (Linke et al., 1999). Season should be explicitly taken into account in bioassessment and monitoring studies, although seasonal variation is currently most often addressed by constraining the time frame of sampling. The recommendation can be made that the neural network model for this database should be developed for each season to study behaviours of macroinvertebrates in wet and dry seasons separately.

7.3.3 Summary

It is very important for management to predict how communities respond to changes in habitat condition after being exposed to disturbance. Communities can respond either progressively with further disturbance or regressively in recovering from pollution (Hellawell, 1986). Time-series predictions are significant for monitoring water quality and deciding on-going management tactics for freshwater ecosystem (Chon et al., *in press*). Analyses of temporal and spatial patterns in community dynamics have been the objectives of many studies in applying ANN models in freshwater bioassessment and have achieved significant success such as: in predictions long-term population of aquatic insects (Schleiter et al., 1999), in patterning community change and short-term prediction of community dynamics (Chon et al., 1996; Chon et al., 2000a; Chon et al., 2000b), and modelling population dynamics (Obach et al., *in press*)

Ecosystem analysis and prediction with empirical statistical and analytical methods are often limited by the spatial complexity and temporal dynamics of ecological processes and typical non-linear interrelations of variables and species, with data not being normally distributed (Schleiter et al., 1999). Artificial neural networks provide specific features such as non-linearity, adaptivity, generalisation and model independence, which allow them to better cope with spatial and temporal variations within freshwater ecosystems.

The neural network models developed in this study were still steady–state models and temporal variation was not considered. An additional effort would be more beneficial to collect data for detecting differences through time and among sites. It is important that methods are developed to characterise how a community varies in space and over time simultaneously, caused either by natural or anthropogenic disturbance. However, improvements in model performance can be achieved by separating neural network models by seasonality and for different habitat types. This can simplify the network performance and avoid potential errors caused by misunderstanding categorical notations designed for habitat types and seasonality.

7.4 Representation of Data on Macroinvertebrate Taxa

7.4.1 Functional Feeding Groups

Family-level identification of macroinvertebrates proved to be appropriate for use as biological indicators for habitat assessment. However, sometimes the taxonomic framework is inadequate to allow identification to this level, such as in the case of *Oligochaeta* at class level, *Acarina* at order level, or time does not permit resolution. Moreover, discrete presence/absence information given by family level of data can make it difficult for neural networks to generate the patterns. Taxa identified at family-level assume that a higher taxonomic category summarises a consistent ecology or behaviour amongst all member species, and indeed this is evident from responses noted in earlier chapters. However, many closely related taxa diverge in their ecology, and higher-level aggregates therefore contain a diversity of responses. By contrast, *functional feeding grouping* requires no taxonomic assumptions but use mouthpart morphology to identify feeding modes (Gullan & Crantons, 2000). The following categories are generally recognised based on feeding mechanisms of macroinvertebrates:

- ***Shredders*** feed on living or decomposing plant tissues, including wood, which they chew, mine or gouge;
- ***Collectors*** feed on fine particulate organic matter by filtering particles from suspension or fine detritus from sediment;
- ***Scraper or grazers*** feed on attached algae and diatoms by grazing solid surfaces; and
- ***Predators*** feed on cells of living animal tissues by engulfing and eating the whole or parts of animals or piercing prey and sucking body fluids (Gullan & Crantons, 2000)

One important ecological observation associated with such functional summary data is the often-observed sequential downstream changes in proportions of functional feeding groups. This aspect of the ***river continuum concept*** relates the sources of energy inputs into the following aquatic system to its inhabitants. In riparian tree-shaded headwaters where light is low, photosynthesis is restricted and energy derives from high inputs of materials such as leaves, woods etc., *shredders* tend to predominate, because they can break up large matter into finer particles. Further downstream, *collectors* filter the fine particles generated upstream and themselves add particles (faeces for example) to the current. Where the waterway becomes broader, with increased available light allowing photosynthesis in the middle reaches, algae, diatoms and macrophytes develop and serve as food for *grazers*. *Predators* tend only to track the localized abundance of food resources (Vannote et al., 1980). Changes in functional feeding groups associated with human activities include:

- Reduction in shredders with loss of riparian habitat, and consequent reduction in autochthonous inputs;

- Increase in grazers with increased periphyton (algae and diatom) development resulting from enhanced light and nutrient entry;

- Increase in filtering collectors below impoundment, such as dams and reservoirs, associated with increased fine particles in upstream standing waters (Gullan & Crantons, 2000).

Use of the functional feeding group approach may be advantageous in that it allows a numerical assessment of the degree to which the invertebrate biota of a given aquatic system is dependent upon particular nutritional resources. This numerical assessment may better suit application to neural network models. As the relative dominance of various food resource categories changes, there is often a corresponding shift in the ratios of the different functional feeding groups. Invertebrate functional group analysis is sensitive to both normal pattern of geomorphic and the biological changes that occur along the *river continuum* (Vannote et al., 1980), as well as to alterations in these patterns resulting from human impacts (Cummins, 1993). In addition, the functional feeding group method is relatively independent of sample size, and its use requires minimal equipment and is accomplished at a level of resolution that can be chosen to be appropriate to the expertise of those performing the analysis (Cummins, 1993). Therefore, for the purpose of rapid assessment, functional feeding group may be a suitable alternative to family-level identification.

7.4.2 Abundance data

Populations of a species might have good potential for environmental monitoring. Many factors associated with population dynamics could be used to assess environmental quality. Environmental conditions which impair growth and mortality or reproduction success would become evident when populations were examined quantitatively (Hellawell, 1986).

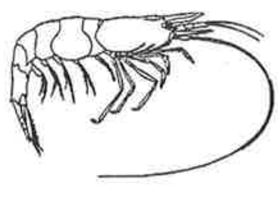
Harswick et al (1995) used *Chironomid* to rapid biological assessment of streams in the Blue Mountain, Australia. The study confirmed that *Chironomids* dominated large proportion in polluted site because they are among the most pollution-tolerant of all stream macroinvertebrates (Hynes, 1960). The abundance rather than presence of *Chironomids* in such rapid assessments may therefore better indicate the ecosystem states.

If the concentration of organic matter is high enough to produce deoxygenation, most of macroinvertebrate families can not survive. Where only very little oxygen remains in the water, or the river bed is completely covered over with organic solids or sewage fungus, the main inhabitants are always the sludge worms of the family *Turbificidae*. In such places, they are particular favoured, they have abundant food in the rich organic mud and they are free from enemies and competitors which cannot stand the low concentrations of oxygen. However, if the organic matter concentration is not high enough, the various members of the macroinvertebrate communities are encouraged such as chironomids, leeches and few others, and occur with *Tubificids*. Species richness therefore does not detect the level of organic enrichment but abundance of *Turbificidae* can be used as more effective indicator for organic enrichment (Hynes, 1960; Norris & Georges, 1993).

Changes in the composition of a macroinvertebrate community are almost certainly due to suspended solids and the resulting slight increase in the siltiness of the environment. However, some mayflies and caddisflies need clean stones on which to attach. Therefore, a slight increase in siltiness would reduce the amount of suitable living space for these creatures. *Tanytarsus*, on the other hand, build their tubes of silts, they therefore are favoured by an increase in the amount of silt in the riverbed and show a marked increase in numbers in this condition (Hynes, 1960; Cranston et al., 1996). Abundance of these macroinvertebrates can better explain the habitat condition than the presence of taxa.

Abundance information seems to be particularly useful for macroinvertebrate data sets with a poor classification structure, and it should be useful to assess disturbances that result in changes in abundance rather than species loss. Especially it seems to be important in cases, when changes in abundance may signal a beginning impact that has not yet reached the level of severity resulting in species loss.

In addition, abundance prediction does not rely on cluster analysis, continuous data of output can help ANN models more effective in learning the patterns and consequently improve the performance of ANN models. Abundance data can be meaningfully used in time series predictions in order to detect the effect of potential impacts. Schleiter et al. (1999) used environmental variables including maximum monthly water temperature, discharge and monthly precipitation to predict long-term population dynamics of aquatic insects in Germany. Chon et al. (2000a, 2000b and *in press*) were very successful in temporal patterning of community changes and short-term prediction of benthic macroinvertebrates in urbanized streams with biological and environmental factors. Obach et al. (*in press*) used feed-forward ANN to predict the abundance of aquatic insect based on the abundance of parental generations and several environmental variables. The results showed that abundance patterns of aquatic insects based on knowledge of their life history and biological traits were related to the patterns of environmental variables. The studies demonstrated that the prediction of aquatic macroinvertebrates with ANN models was promising even though the restricted number of input variables might have limited the quality of the results. Research on abundance of macroinvertebrates is therefore recommended provided the data is available and reliable for developing ANN models.



Freshwater shrimp Atyidae, *popular crustacean in the streams (Giller Malmqvist, 1998))*

8 Thesis Conclusions and Recommendations for Further Research

8.1 *Conclusions*

8.1.1 Usefulness of Macroinvertebrates as Biological Indicators

The results of this study demonstrate that the structure of a macroinvertebrate community can reflect the state of freshwater streams they inhabit. Both *Clean* and *Dirty water approaches* show that there are close relationships between freshwater macroinvertebrate assemblages and their habitat conditions in the Queensland stream system. The distribution of macroinvertebrates at family level is driven by a number of environmental factors. Physical variables strongly determine the habitat where certain macroinvertebrates live. However, macroinvertebrate assemblages are also affected by chemical variables. Both input categories proved to be useful for the prediction distribution of macroinvertebrate assemblages.

The routine use of freshwater macroinvertebrates as indicator organisms to assess ongoing environmental condition requires considerable understanding of the factors involved in determining these conditions. These factors include physical and chemical characteristics of the habitat where the organisms originate: e.g. the chemistry of potential contaminants involved, and the physiological behaviour of taxa exposed to these contaminants need to be taken into consideration. By contrast, research on response of macroinvertebrate assemblages to habitat conditions can improve our understanding of that.

Species-level identification is assumed to have the greatest information content as a result of studies on individual population. However, identification to the species level is sometimes difficult because of the small size of organisms, a lack of adequate species-level keys and descriptions. On the other hand improved keys are now available that allow identifications at the family level. The use of higher-level category at family can be justified depending on the purpose of study, the level of sensitivity required, and the type of index or analysis being used (Resh & McEltravy, 1993). In this study, macroinvertebrate assemblages at family level proved to be appropriate to detect the gross pollution that may have dramatic effects on the fauna and to provide "early warning" of potential problem or changes in communities.

Biotic indices have been extensively used to evaluate pollution stress. The multivariate approach may enhance our understanding of the effect of pollution stress. However, emphasis on improving the efficiency, accuracy, precision, and predictive ability of biotic indices and scoring system is needed.

8.1.2 Usefulness of Artificial Neural Network as Prediction Tools

The results of the study have shown that Artificial Neural Networks (ANNs) can successfully and meaningfully be applied to the analysis of causal relationships including the identification and assessment of complex impact factors and for the prediction of system behaviours. Particularly, they have advantages if the relationships are unknown, very complex or nonlinear as typical for river and stream ecosystems.

Generally, the ANNs performed very well to predict the macroinvertebrate taxa based on physical as well as on chemical predictor variables. This method does not only generate results with low prediction error, but also allows the user to identify associations and general trends in the data. These capabilities make ANNs more appealing than just black box modelling by statistical techniques.

The conclusion can be made that ANNs tend to be grey box prediction techniques, allowing the user to combine a high accurate prediction with getting some information on general trends in the data. Therefore, this methodology can be applied to determine ecological requirements of stream organisms that are not fully understood.

The present study has justified that the data collected in the Queensland stream system could be used for prediction with ANNs. Although many possible improvements in the data representation can be suggested, the existing database proved to be appropriate to develop the ANN models to assess river habitat conditions. This fact demonstrates that ANN models are verified as a powerful tool to investigate and model ecosystems from limited data available. The method therefore can be applied for other stream systems in Australia to assess river health.

Both applied approaches achieved reasonable good results in predicting presence/absence patterns of macroinvertebrate families. The application of these approaches for management purposes requires further researches in detail. A protocol for applying the *Clean Water* approach has already been designed through the multivariate criteria O/E; however, the interpretation of criteria value has not yet been well understood. The *Dirty Water* approach can only be applied for quantitative assessment after further studies on how to determine the best representative taxa can be used for designate chemistry of water they inhabit.

8.1.3 Elucidation by Sensitivity Analyses

Sensitivity analyses allow studying the impact of the input variables on the presence and absence of macroinvertebrate taxa. Many relations were detected by this method that supported previously detected findings by means of related ecological research methods. This indicates that the ANN models perform in meaningful ecological manner. Based on of sensitivity analyses, ANN models allow to determine the major driving variables that affect the stream and river ecosystem quality, and should be taken under direct consideration in the river ecosystem management. The determination of major driving variables improves the generalisation and simplification of the model, and allows a better understanding of interrelationships within systems.

Sensitivity analyses allow a better interpretation of the prediction results by ANN models, easing the cause-allocation of the actual river status and increasing the insight needed to improve assessment system. Relation plots also allow stimulating the effect of potential management options and thus support active decision-making. The development of efficient monitoring networks based on the interpretation of these interrelations is probably another important advantage.

With the aid of sensitivity analyses, data pre-processing methods, ecological prediction and analysis of causal relations can be improved substantially.

The sensitivity analyses not only confirmed expert knowledge on relations between habitat conditions and distribution of macroinvertebrates, but revealed new relations as well. Further study on new and sometimes contradictory findings may enlarge expert knowledge on river ecosystem if they would be confirmed by laboratory and field experiments. Conversely, finding the cause of contradictions may improve the performance of ANN models for better predictions and applications for management purposes.

8.1.4 Limits of Using ANNs

As ANNs learn from examples, the quality of ANN models heavily depends on the quality of the database, in particular whether it is representative for the given problem, the given site or the given study period. Therefore, representative and compatible data are the main prerequisite for ANN models. In reality, interrelation between biological diversity and abiotic factors are highly nonlinear and complex and no monitoring data contains enough information for reflecting these relations.

Time-series analysis, for example, can be a powerful tool to study community dynamics of macroinvertebrates, which in turn provide much information for assessing habitat condition and preparing on-going management tactics for aquatic ecosystem. Time-series analyses can be conducted effectively by ANNs. However, time-series analysis relies on data that are collected regularly at time intervals more frequent than the period of variation among the variables of interest. These requirements are quite stringent to be fulfilled by most monitoring programs.

The training process of ANN models is always very much black box. It does not explicitly provide understanding of the mechanisms it is based on. ANN models extract and generate patterns from data inductively and it is very difficult for modelers to get insight into this process in order to optimise model configuration and performance parameters. Optimisation therefore can be done by experiments and in some cases, the best configuration and performance of ANN models can only be decided intuitively.

8.2 Recommendations for Further Research

Depending on the objectives of the present research project, recommendations can be made as follow in order to improve the ANN performance as well as to have better understanding of ecological processes and phenomena.

8.2.1 Further Research on ANN Performance

- Further research is necessary to determine the optimal neural network configuration. E.g. Waley & Fontama (2000) applied an ANN with two hidden layers for similar simulations. The impact of the algorithm applied for ANN training as well as the risk of overtraining the ANNs should be further analysed to obtain reliable and meaningful predictions in the long run. It is recommended to test algorithm for neuro-genetic training in order to optimise ANNs performance as suggested by Montana & Davis (1989).

- Improve results by using other more continuous value inputs. In the clean water approach, flow regime should be considered. Oxygen concentration, nutrient levels and concentration of some toxicants should be crucial for dirty water approach.

- Procedure to pre-processing data needs to be improved in order to avoid using non-reliable data in the modelling process. Empirical categorical variables should be split into nodes, each node contains two values 1 or 0 depending on whether this category is chosen or not.

- The spatial-temporal variability of physical and chemical variables should be considered. Separate models for different habitats and seasons. As discussed in previous chapters, four models are developed for four different habitats and two seasons for each habitat. As a result, eight models should be developed.

- Improve the performance by applying different input combinations to develop models. Experiments with different sets of input variables do not only results in an alteration of prediction errors but also the complexity of networks as well as the relative importance of some trends can be affected.

- Sensitivity analyses should be carried out several times to obtain further simplification of ANN models by further exclusion of redundant inputs for both approaches.

- Time-series analysis may be used to develop predictive models based on variation in past time series. It may overcome the problem of auto-correction in data between sites or times. The method can determine the occurrence trends, often a primary aim of monitoring studies.

- The next step of pollution or environmental assessment studies should attract not only qualitative but also quantitative analyses of benthic macroinvertebrates. Hypothesis generation through classification, ordination, model construction, and the accuracy of model prediction should be tested.

- Better representation of macroinvertebrate data can be improved by classifying families into functional feeding group or obtaining abundance data. The better representation of macroinvertebrates may overcome difficulties for ANN training caused by the distribution of presence/absence (1/0) data in cases where species had too high or too low probability of occurrence.

- Training and validation of ANNs using databases from other Australian stream systems will contribute to a generalisation of the ANN stream models.

- Better visualisation of results by means of interactive user interface and GIS will help non-specialist to understand interpretation of prediction as well as elucidation from sensitivity analysis. User-friendly software should be next step to apply ANN models for management purposes

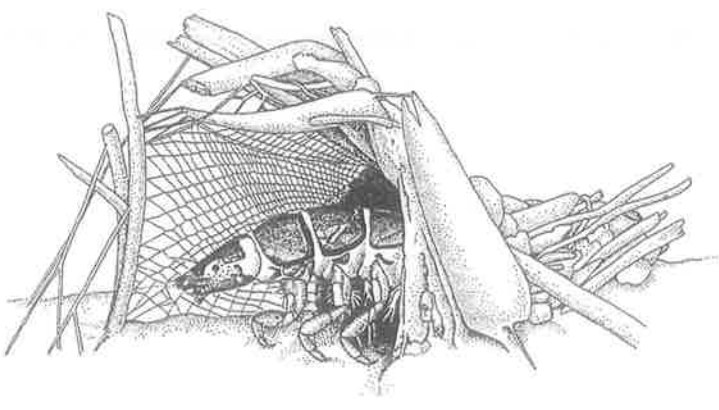
- Other machine learning technique can improve the prediction results. Multivariate analysis of data by Kohonen networks should be very promising alternative.

8.2.2 Further Research on Elucidation of Freshwater Habitat Conditions

- Cause and nature of overpopulation predicted for some test sites in clean water approach needs to be clarified.

- Improve the method of sensitivity analyses by considering pattern-combination of factors preferred by macroinvertebrates.

- Study cause and effect of relations discovered by sensitivity analyses that appeared to be contradictory to expert knowledge.



*A caddisfly larva (*Hydropsychidae*) in its retreat; the silk net is used to catch food (Gullan & Cranston, 2000)*

References

Adeli, H. and Hung, S. L. (1995). *Machine Learning: Neural Networks, Genetic Algorithms and Fuzzy Systems*. Canada, John Wiley& Soms Inc.

Allan, J. D. (1995). *Stream ecology: Structure and Function of Running Water*. London, Chapman and Hall.

Amoros, C., Bornette, G. and Henry, C. P. (2000). "Environmental auditing: A vegetation-based method for ecological diagnosis of riverine wetland." *Environmental Management* **25**(2): 211-227.

Bailey, R. C., Matthew, G. K., Dervish, M. Z. and Taylor, R. M. (1998). "Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yokon streams." *Freshwater Biology* **39**: 765-774.

Bartram, J. and Ballance, R., Eds. (1996). *Water Quality Monitoring: A Practical Guide to the Design of Freshwater Quality Studies and Monitoring Programs.* London, Chapman and Hall.

Beschata, R. L. (1996). Suspended sediment and bedload. *Methods in Stream Ecology*. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Bishop, C. M. (1998). *Neural Network and Machine Learning*. New York, Springer.

Boulton, A. J. and Lake, P. S. (1992). "The ecoloy of two intermittent streams in Victoria, Australia. III. Temporal changes in faunal composition." *Freshwater Biology* **27**: 123-138.

Bunn, S. E., Edward, D. H. and Lonegaran, N. R. (1986). "Spatial and temporal variations in macroinvertebrate fauna of streams of northen jarrah forest, Western Australia: communities structure." *Freshwater Biology* **16**: 67-92.

Bury, K. V. (1975). *Statistical Models in Applied Science*. NewYork, Wiley & Sons Inc.

Cairns, J. J., McCormick, P. V. and Niederlehner, B. R. (1993). "Aproposed framework for developing indicators of ecosystem health." *Hydrobiologia* **263**: 1-44.

Cairns, J. J. and Pratt, J. R. (1993). A history of biological monitoring using benthic macroinvertebrates. *Freshwater Biomonitoring and Benthic Macroinvertebrates*. D. M. Rosenberg and V. H. Resh. NewYork, Chapman & Hall.

Cao, Y., Bark, A. W. and Williams (1997). "Analysing benthic macroinvertebrate community changes along a pollution gradient: a framework for the development of biotic indices." *Water Research* **31**(4): 884-892.

Carbiener, R., Tremolieres, M., Mercier, J. L. and Ortscheit, A. (1990). "Aquatic macrophyte communities as bioindicators of eutrophication in calareous oligoaprobe stream water (Upper Rhine plain, Alsace)." *Vegetation* **86**: 71-88.

Cartwright, D. I. (1997). *Preliminary guide to the identification of late instar larvae of Australian Ecnomidae, Philopotamidae and Tasimiidae (Insecta: Trichoptera)*. Albury, NSW., CRC for Freshwater Ecology. Identification Guide No. 10.,.

Carver, M., Gross, G. F. and Woodward, T. E. (1991). Hemiptera. *Insects of Australia.* CSIRO. Carlton, Victoria., Melbourne University Press. **1**.

Cereghino, R., Giraudel, J. L. and Compin, A. (2000). *Spatial analysis of stream invertebrates* *distribution in Adour-Garonne drainage basin (France), using Kohonen Self Organizing Maps*. Paper submited to 2nd International Conference of Applications of MAchine Learning to Ecological Modelling, Adelaide.

Chessman, B. C. (1999). "Predicting the macroinvertebrate faunas of river by multiple regression of biological and environmental differences." *Freshwater Biology* **41**: 747-757.

Chon., T.-S., Park., Y. S., Moon, K. H. and Cha., E. Y. (1996). "Patternizing communities by using artificial neural network." *Ecological Modelling* **90**: 69-78.

Chon., T.-S., Park., Y. S. and Cha., E. Y. (2000a). Patterning of community changes in benthic macroinvertebrates collected from urbanized streams for the short time prediction by temporal Artificial Neuronal Networks. *In: Artificial Neuronal Networks: Application to Ecology and Evolution.* S. Lek and J. F. Guegan. Berlin, New York., Springer Verlag.

Chon., T.-S., Park., Y.-S. and Park., J. H. (2000b). "Determining temporal pattern of community dynamics by using unsupervised lerarning algorithms." *Ecological Modelling* **132**(151-166).

Chon., T.-S., Kwak., I. S., Park., Y.-S., Kim. T-H. and Kim Y.S. (*in press*). "Patterning and sort-term prediction of community dynamics of benthic macroinvertebrates in urbanized streams with implementation of biological and environmental factors in a recurrent Artificial Neural Network." *Ecological Modelling*.

Choy, S. and Horwitz, P. (1995). A preliminary key to the species of Australian shrimps (*Atyidae*) found in inland waters. *A Preliminary Key to the species of Decapoda (Crustacea: Malacostraca) found in Australian Inland Waters.* P. Horwitz. Albury, NSW., CRC for Freshwater Ecology Identification Guide No. 5.,.

Chutter, F. M. (1969). "The distribution of some stream invertebrates in relation to current speed." *Internationale Revue der gesamten Hydrobiologie* **54**: 413-422.

Clarke, R. T., Furse, M. T., Wright, J. F. and Moss, D. (1996). "Derivation of a biological quality index for river sites: comparison of the observed with expected fauna." *Journal of Applied Statistics* **23**: 311-332.

Connell, J. H. (1980). "Diversity and the coevolution of competitors, or the ghost of competition past." *Oikos* **35**: 131-138.

Conrick, D. L. and Cockayne, B. J., Eds. (2000). *Biological Monitoring and Assessment of Freshwaters using Macroinvertebrates. Background Information, Sampling and Analytical Procedures.* Brisbane, Queensland Department of Natural Resources.

Cooper, S. D., Walde, S. J. and Peckarsky, B. L. (1990). "Prey exchange rates and the impact of predators on prey populations in streams." *Ecology* **71**: 1503-1514.

Corkum, L. D. (1989). "Patterns of benthic invertebrate assemblages in rivers of northwestern North America." *Freshwater Biology* **21**: 235-243.

Coysh, J., Nichols, S., Simpson, S., Norris, R., Barmuta, L., Chessman, B. and Blackman, P. (2000). *Australian River Assessment System - National River Health Program Predictive Model Manual.* Canberra, ACT., CRC Freshwater Ecology, University of Canberra,.

Cranston, P. S., Fairweather, P. and Clarke, G. (1996). Biological indicators of water quality. *Indicators of Catchment Health: A Technical Perspective.* J. Walker and D. J. Reuter. Melbourne., CSIRO Publishing.

Cummins, K. W. (1979). The natural stream ecosystem. _The Ecology of Regulated Streams_. J. V. Ward and J. A. Stanford. New York and London, Plenum Press.

Cummins, K. W., Sedell., J. R., Swanson, F. J., Minshall, G. W., Fisher, S. G., Cushing, C. E., Petersen, R. C. and Vannote R.L. (1983). Organic matter budgets for stream ecosystems. _Stream Ecology: Application and Testing of General Ecological Theory._ J. R. Barnes. and G. W. Minshall. New York, Plenum.

Cummins, K. W., Minshall, G. W., Sedell, J. R., Cushing, C. E. and Petersen, R. C. (1984). "Sream ecosystem theory." _Verh. Internat. Verein. Limnol._ **22**: 1818-1827.

Cummins, K. W. (1993). Bioassessment and analysis of functional organization of running water ecosystems. _Biological Monitoring of Aquatic System_. S. L. Loeb and A. Spacie. Boca Raton, Florida, Lewis Publisher.

d'Angelo, D. J., Howard, L. M., Meyer, J. L., Gregory, S. V. and Ashkenas, L. R. (1995). "Ecological use for genetic algorithms: Predicting fish distributions in complex physical habitatss." _Canadian Journal of Fisheries and Aquatic Science_ **52**: 1893-1908.

Davis, W. S. and Simon, T. P., Eds. (1995). _Biological Assessment Criteria: Tools for Water Resource Planning and Decision Making_. Boca Raton, FL., Lewis Publisher.

Dean, J. C. (1996). Prosopistomatidae. _Mayfly Nymphs of Australia. A guide to Genera._ J. C. Dean and P. J. Suter. Albury, NSW., CRC for Freshwater Ecology, Identification Guide No. 7.

Dinsmore, W. P., Scrimgeour, G. J. and Prepas, E. E. (1999). "Empirical relationship between profundal macroinvertebrate biomass and environmental variables in borel lakes of Alberta, Canada." _Freshwater Biology_ **41**: 91-100.

Dodelec, S. (1989). "Seasonal dynamics of benthic macroinvertebrate communities in the Lower Ardeche River (France)." _Hydrobiologia_ **182**: 73-89.

Dzeroski, S., Grbovic, J. and Walley, W. J. (1999). Machine learning applications in biological classification of river water quality. _Machine Learning and Data Mining Methods and Application_. R. S. Michalski, I. Bratko and M. Kubat. Chichester, John Wiley & Sons Ltd.

Edington, J. M. (1968). "Habitat preferences in net-spinning caddis larvae with special reference to the influence of water velocity." _Journal of Animal Ecology_ **37**: 675-692.

Faith, D. P. and Norris, R. H. (1989). "Correlation of environmental variables with patterns of distribution and abundance if common and rare freshwater macroinvertebrates." _Biological Conservation_ **50**: 77-98.

Fischer, G. W. and Grant, W. E. (1994). "Use of a native predator to control overcrowding in warm water polyculture ponds: simulationof a tucunare and tilapia system." _Ecological Modelling_ **72**: 205-227.

Forstner, U. and Wittmann, G. T. W. (1983). _Metal Pollution in the Aquatic Environment_. Berlin, Springer-Verlag.

Frissell, C. A. and Lonzarich, D. G. (1996). Habitat use and competition among stream fishes. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Furse, M. T., Moss, D., Wright, J. F. and Armitage, P. D. (1984). "The influence of seasonal and taxonomic factors on the ordination and classification of running water sites in Great Britain and on the prediction of their macroinvertebrate communities." _Freshwater Biology_ **14**: 257-280.

Gabriels, W., Goethals, P. L. M., Heylen, S., De Cooman, W. and De Pauw, N. (2000). _Modelling benthic macro-invertebrate communities in Flanders using Artificial Neural Networks_. WATERMATEX, Belgium.

Gammon, J. R. (1970). The effect of iorganic sediment on stream biota. Washington , D.C.,, *Water Poll. Control Res. ser. No.18050 DWC*. US-EPA.

Gee, J. H. R. (1984). "The Microdistribution of Benthic Invertebrates in a Cotwold Stream." _Verh. Internat. Verein. Limnol._ **22**: 1932-1938.

Ghetti, P. F. and Ravera, O. (1994). European perspective on biological monitoring. _Biological Monitoring of Aquatic System_. S. L. Loeb and A. Spacie. Boca Raton, Florida, Lewis Publisher.

Giller, P. S. and Malmqvist, B. (1998). _The Biology of Streams and Rivers_. Oxford, NewYork, Oxford University Press.

Gilliam, J. F., Fraser, D. F. and Sabat, A. M. (1989). "Strong effects of foraging minnows on a stream benthic invertebrate community." _Ecology_ **70**: 445-452.

Gore, J. A. (1996). Discharge measurements and streamflow analysis. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Gozlan, R. E., Mastrorillo, S., Copp, G. H. and Lek, S. (1999). " Predicting the structure and diversity of young-of-the-year fish assemblages in large rivers." _Freshwater Biology_ **41**(4): 809-820.

Gullan, P. J. and Cranston, P. S. (2000). _The Insects: An Outline of Entomology_, Blackwell Science.

Hardwick, R. A., Wright, I. A., Jones, H. A., Chessman, B. C. and Holleley, D. E. (1995). Rapid biological assessment of streams in the Blue Mointains, Australia: characteristics of the Chironomidae (Diptera) fauna. _Chironomids: from Genes to Ecosystems_. P. S. Cranston. Melbourne., CSIRO.

Harrod, J. J. (1965). "Effect of current speed om the cephalic fans of the lavae of *Simulium ornatum var. nitidifrons Edwards*." _Hydrobiologia_ **26**: 8-12.

Hauer, F. R. and Hill, W. R. (1996). Temperature, light, oxygen. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Hauer, F. R. and Resh, V. H. (1996). Benthic macroinvertebrates. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Hauer, F. R. and Lamberti, G. A., Eds. (1996). _Methods in Stream Ecology_. San Diego, Academic Press Inc.

Hawker, H. A. (1997). "Origin and development of the biological monitoring working party score system." _Water Resource_ **32**(**3**): 964-968.

Hawking, J. H. and Smith, F. J. (1997). _Colour Guide to Invertebrates of Australian Inland Water_. Albury, NSW., CRC for Freshwater Ecology Identification Guide No. 8.

Hawkins, C. P., Norris, R. H., Hogue, J. N. and Feminella, J. W. (2000). "Development and evaluation of predictive models for measuring the biological integrity of streams." *Ecological Application* **10**(5): 1456-1477.

Hellawell, J. M. (1986). *Biological indicators of Freshwater Pollution and Environmental Management*. London & NewYork, Elsevier.

Hershey, A. E. and Peterson, B. J. (1996). Stream food web. *Methods in Stream Ecology*. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Hershey, A. E. and Lamberti, G. A. (1998). Stream macroinvertebrate communities. *River Ecology and Management: Lessons from the Pacific Coastal Ecoregion*. R. J. Naiman and R. E. Bilby. New York, Springer-Verlag.

Hildrew, A. G. and Townsend, C. R. (1987). Organization in freshwater benthic communities. *Organization of communities: past and present*. J. H. R. Gee and P. S. Giller. Oxford, Blackwell.

Hoang, H., Recknagel, F., Marshall, J. and Choy, S. (*in press*). "Predictive Modelling of Macroinvertebrate Assemblages for Stream Habitat Assessments in Queensland (Australia)." *Ecological Modelling*.

Hornung, M. and Reynolds, B. (1995). "The effecta of natural and anthropogenic environmental changes on ecosystem processes at catchment scale." *Trends in Ecology and Evolution* **10**: 443-449.

Horwitz, P. (1995). *A Preliminary Key to the species of Decapoda (Crustacea: Malacostraca) found in Australian Inland Waters. Colour Guide to Invertebrates of Australian Inland Water*. Albury, NSW., CRC for Freshwater Ecology. Identification Guide No. 5.

Hughes, R. M., Larsen, D. P. and Omernik, J. M. (1986). "Regional reference sites: amethod for assessing stream potentials." *Enviromental Management* **10**: 629-635.

Hunsaker, C. T. and Carpenter, D. E., Eds. (1990). *Environmental Monitoring and Assessment Program: Ecological Indicators.* NC., Office of Research and Development, US-EPA, Research Triangle Park.

Hynes, H. B. N. (1960). *The Biology of Poluted Water*. Liverpool, Liverpool University Press.

Hynes, H. B. N. (1970). *The Ecology of Running Water*. Liverpool, Liverpool University Press.

Johnson, R. K., Wiederholm, T. and Rosenberg, D. (1993). Freshwater biomonitoring using individual organisms, populations and species assemblages of benthic macroinvertebrates. *Freshwater Biomonitoring and Benthic Macroinvertebrates*. D. M. Rosenberg and V. H. Resh. NewYork, Chapman & Hall.

Johnson, R. K. (1995). The indicator concept in freshwater biomonitoring. *Chironomids: from Genes to Ecosystems*. P. S. Cranston. Melbourne., CSIRO.

Jorgensen, S. E. (1999). "State-of-the-art of ecological modelling with emphasis on development of structural dynamic models." *Ecological Modelling* **120**: 75-96.

Karayiannis, N. B. and Venetsanopoulos, A. N. (1993). *Artificial Neural Networks: Learning Algorithms, Performance Evaluation and Application*. Boston/Dordrecht/London, Kluwer Academic Publishers.

Karr, J. R. (1981). "Assessment of botic integrity using fish communities." *Fisheries* **6**(6): 21-27.

Karr, J. R. (1998). Rivers as sentinels: using the biology of rivers to guide landscape management. _River Ecology and Management: Lessons from the Pacific Coastal Ecoregion_. R. J. Naiman and R. E. Bilby. New York., Springer-Verlag.

Karr, J. R. (1999). "Defining and measuring river health." _Freshwater Biology_ **41**(2): 221-234.

Kartalopoulos, S. V. (1996). _Understanding Neural Networks and Fuzzy Logic: Basic Concept and Application_. New York, IEEE Press.

Karul, C., Soyupak, S., Cileciz, A. F., Akbay, N. and Germen, E. (2000). "Case studies on the use of neural networks in eutrophication modelling." _Ecological Modelling_ **134**: 145-152.

Kay, W. R., Smith, M. J., Pinder, A. M., McRae, J. M., Davis, J. A. and Halse, S. A. (1999). "Patterns of distribution of macroinvertebrate families in rivers of north western Australia." _Freshwater Biology_ **41**: 299-316.

Kent, E. (1980). _The Brains of Men and Machines_. New York, McGraw-Hill Publications Co.

Kompare, B., Bratko, I., Steinman, F. and Dzeroski, S. (1994). "Using machine learning techniques in the construction of models. I. Introduction." _Ecological Modelling_ **75/76**(617-628).

Lamberti, G. A. and Gregory, S. V. (1996). Transport and retention of CPOM. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Lamberti, G. A. and Feminella, J. W. (1996). Plant-herbivore interactions. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Lampert, W. and Sommer, U. (1997). _Limnoecology: The Ecology of Lakes and Streams_. New York, Oxford University Press.

Langley, P. and Simon, H. A. (1999). Fielded applications of macine learning. _Machine Learning and Data Mining Methods and Application_. R. S. Michalski, I. Bratko and M. Kubat. Chichester, John Wiley & Sons Ltd.

Lawrence, J. F. and Britton, E. B. (1991). Coleoptera. _Insects of Australia._ CSIRO. Carlton, Victoria., Melbourne University Press. **2:** 543-683.

Lek, S., Delacose, M., Baran, P., Dimopoulos, I., Lauga, J. and Aulagnier, S. (1996). "Application of neural networks to modelling non-linear relationship in ecology." _Ecological Modelling_ **90**: 39-52.

Lek, S. and Guegan, J. F., Eds. (2000). _Artificial Neuronal Networks: Application to Ecology and Evolution._ Berlin, New York, Springer Verlag.

Lek-Ang, S., L., D. and Lek, S. (1999). "Predictive modelling of *Collembola* diversity and abundance in riparian habitat." _Ecological Modelling_ **120**(197-200).

Lenat, d. R. and Barbour, M. T. (1994). Using benthic macroinvertebrate community structure for rapid, cost-effective, wate quality monitoring: rapid assessment. _Biological Monootoring of Aquatic System_. S. L. Loeb and A. Spacie. Boca Raton, Florida, Lewis Publisher.

Li, H. W. and Li, J. L. (1996). Fish community composition. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Lin, C.-T. and Lee, C. S. (1996). _Neural Fuzzy Systems: A Neuro-Fuzzy Synergism to Intelligent_

Systems. New Jersey, Prentice Hall P T R.

Lindegaard, C. and Brodedersen, K. P. (1995). Distribution of *Chironomidae* in the river continuum. _Chironomids: from Genes to Ecosystems_. P. S. Cranston. Melbourne., CSIRO.

Linke, S., Bailey, R. C. and Schwindt, J. (1999). "Temporal variability of stream bioassessment using benthic macroinvertebrates." _Freshwater Biology_ **42**: 575-584.

Loeb, S. L. and Spacie, A. (1994). _Biological Monitoring of Aquatic System_. Boca Raton, Florida, Lewis Publisher.

Lowe, R. L. and Laliberte, G. D. (1996). Benthic stream algae: distribution and structure. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Luger, G. F. and Stubblefield, W. A. (1993). _Artificial Intelligence: Structure and Strategies for Complex Problem Solving_. California, The Benjamin/Cummings Publishing Company.

MacKay, D. J. C. (1997). Introduction to gaussian processes. _Neural Networks and Machine Learning_. C. M. Bishop. Berlin, Springer-Verlag.

Maier, H. G. and Dandy, G. C. (1996). "The use of Artificial Neural Network for the prediction of water quality parameters." _Water Resources Research_ **32**(4): 1013-1022.

Maier, H. G., Dandy, G. C. and Burch, M. D. (1998). "Use of ANN for modelling cyanobacteria *Anabena* species in the river Murray, SA." _Ecological Modelling_ **105**: 257-272.

Marchant, R., Barmuta, L. A. and Chessman, B. C. (1995). "Influence of sample quantification and taxonomic resolution on the ordination of macroinvertebrates communities from running watersin Victoria Australia." _Marine and Freshwater Research_ **46**: 501-506.

Marchant, R., Hirst, A., Norris, R. and Metzeling, L. (1999). "Classification of macroinvertebrate communities across drainage basins in Victoria, Australia: consequences of sampling on broad spatial scale for predictive modelling." _Freshwater Biology_ **41**: 253-268.

Marshall, J., Hoang, H., Choy, S. and Recknagel, F. (2001). _Relationships between habitat properties and the occurrence of macroinvertebrates in Queensland streams (Australia) discovered by a sensitivity analysis with artificial neural networks_. Paper submited to XXVI congress of Societas Internationalis Limnologiae 2001, Melbourne.

Mason, C. F. (1996). _Biology of Freshwater Pollution_. Harlow, England, Longman.

Masters, T. (1993). _Practical Neural Network Recipes in C++_. Boston, Academic Press.

Matthews., W. J. (1987). Physicochemical tolerance and selectivity of stream fishes as related to their geographic ranges and local distributions. _Community and Evolutionary Ecology of North American Stream Fishes_. W. J. Matthews and D. C. Heins. Norman, Univ. of Oklahoma Press.

McCulloch, W. S. and Piits, W. (1943). "A logical calculus of ideas immanent in nervous activity." _Bull. Math. Biophys._ **5**: 115-133.

Meybeck, M., Friedrich, G., Thomas., R. and Chapman, D. (1996). River. _Water Quality Assessment: A guide to the Use of Biota, sediment and water in Environmental Monitoring._ D. Chapman. London, E&FN Spon.

Michalski, R. S., Bratko, I. and Kubat, M., Eds. (1999). _Machine Learning and Data Mining_

Methods and Application. Chichester, John Wiley & Sons Ltd.

Minshall, G. W., Cummins, K. W., Petersen, R. C., Cushing, C. E., Bruns, D. A., Sedell, J. R. and Vannote R.L. (1985). "Developments in stream ecosystem theory." *Canadian Journal of Fisheries and Aquatic Science* **42**: 1045-1055.

Montana, D. J. and Davis, L. D. (1989). *Training feedforward neural networks using genetic algorithms*. International Joint Conference on Artificial Intelligence, San Francisco.

Moss, D., Furse, M. T., Wright, J. F. and Armitage, P. D. (1987). "The prediction of the macroinvertebrate fauna of unpoluted running water sites in Great Britain using environmental data." *Freshwater Biology* **17**: 41-52.

Moss, D., Wright, J. F., Furse, M. T. and Clarke, R. T. (1999). "A comparison of Alternative technique for Prediction of the fauna of Running-water Sites in Great Britain." *Freshwater Biology* **41**: 167-181.

Neboiss, A. (1991). Trichoptera. *Insects of Australia.* CSIRO. Carlton, Victoria., Melbourne University Press. **2:** 787-816.

Neel, J. K. (1973). "Biotic character as related to stream mineral content." *Transaction of American Microscopical Society* **92**: 404-415.

NeuroDimension (1999). *Neural and Adaptive Systems: Fundamentals Through Simulations. Product information*. NeuroDimension Website.

Newbury, R. W. (1984). Hydrologic determinant of aquatic insect habitats. *The Ecology of Aquatic Insects*. V. H. Resh and D. M. Rosenberg. New York, Praeger.

Newbury, R. W. (1996). Dynamics of flow. *Methods in Stream Ecology*. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Nichols, S., Coysh, J., Sloane, P., Williams, C. and Norris, R. (2000). *ACT AusRivAS. Sampling and Processing Manual*. Canberra, CRC for Freshwater Ecology.

Norris, R. and Georges, A. (1993). Analysis and interpretation of benthic macroinvertebrate surveys. *Freshwater Biomonitoring and Benthic Macroinvertebrates*. D. M. Rosenberg and V. H. Resh. NewYork, Chapman & Hall.

Norris, R. H. (1996). *Predicting water quality using reference conditions and associated communities*. Study design and data analysis in benthic macroinvertebrate assessments of freshwater ecosystems using a reference site approach, Kalispell, Montana, USA, North American Benthological Society.

Norris, R. H. and Thoms, M. C. (1999). "What is river health?" *Freshwater Biology* **41**: 197 - 209.

Norton, R. A., Williams, D. D., Hogg, I. D. and Palmen, S. C. (1988). "Biology of the oribatid mite: *Mucronothrus nasalis (Acari: Oribatida: Trhypochthoniidae)* from a small cold water springbook in eastern Canada." *Canadian Journal of Zoology* **66**: 622-629.

Obach., M., Wagner, R., Werner, H. and Schmidt, H. (*in press*). "Modelling population dynamics of aquatic insects with Artificial Neural Networks." *Ecological Modelling*.

Orth, D. J. and Maughan, O. E. (1983). "Microhabitat preferences of benthic fauna in a woodland stream." *Hydrobiologia* **106**: 157-168.

Otto, C. and Svensson, B. S. (1983). "Properties of acid brown water streams in south Sweden." *Arch. Hydrobiol.* **99**: 15-36.

Paradise, C. J. and Kuhn, K. L. (1999). "Interactive effects of pH and leaf litter on a shredder, the scritid beetles, Helodes pulchella, inhabiting treeholes." *Freshwater Biology* **41**: 43-49.

Parsons, M. and Norris, R. (1996). "The effect of habitat - specific sampling on biological assessment of water quality using a predictive model." *Freshwater Biology* **36**: 419-434.

Paruelo, J. M. and Tomasel, F. (1997). "Prediction of functional characteristic of ecosystem: a comparison of ANN and regression models." *Ecological Modelling* **98**: 173-173.

Patrick, R. (1994). What are the requirements for an effective biomonitor? *Biological Monitoring of Aquatic System*. S. L. Loeb and A. Spacie. Boca Raton, Florida, Lewis Publisher.

Peckarsky, B. L. (1996). Predator-prey interactions. *Methods in Stream Ecology*. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Peterson, B. J., Fry, B., Deegan, L. and Hershey, A. E. (1993). "The trophic significance of epilithic algal production in a fertilized tundra river ecosystem." *Limnology and Oceanography* **38**: 872-878.

Plafkin, J. L., Barbour, M. T., Porter, K. D., Gross, S. K. and Hughes, R. M. (1989). Rapid Bioassessment Protocols for Use in Treams and Rivers, U.S. EPA, Office of Water: EPA/444/4-89-001.

Poff, N. L., Tokar, S. and Johnson, P. (1996). "Stream hydrological and ecological responses to climate change assessed with an artificial neural networks." *Limnology and Oceanography* **4** (15): 857-863.

Pudmenzky, A., Marshall, J. and Satish., C. (1998). *Preliminary application of Artificial Neural Network model for predicting macroivertebrates in rivers*. Queensland, Water Monitoring Group - Queensland Department of Natural Resources.

Quinn, J. M. and Hickey, C. W. (1994). "Hydraulic parameters and benthic invertebrate distributions in two gravel bed new Zealand rivers." *Freshwater Biology* **18**: 521-528.

Reckhow, K. H. (1993). "A random coefficient model for chlorophyll-nutrient relationship in lake." *Ecological Modelling* **70**: 35-50.

Recknagel, F. (1989). *Applied Systems Ecology: Approach and Case Studies in Aquatic Ecology*. Berlin, Academic Press.

Recknagel, F., Petzoldt, T., Jaeke, O. and Krusche, F. (1994). "Hybrid expert system DELAQUA - a toolkit for water quality control of lakes and reservoirs." *Ecological Modelling* **71**(1-3,): 17-36.

Recknagel, F., French, M., Harkonen, P. and Yabunaka, K. (1997). "Artificial neural network approach for modelling and prediction of algal blooms." *Ecological Modelling* **96**: 11-28.

Recknagel, F., Fukushima, T., Hanazato., T., Takamura, N. and Wilson., H. (1998). "Modelling and prediction of phyto- and zooplankton dynamics in Lake Kasumigaura by Artificial Neural Networks." *Lake & Reservoirs* **3**: 123-133.

Recknagel, F. and Wilson., H. (2000). Elucidation and prediction of aquatic ecosystem by Artificial

Neural Networks. _Artificial Neural Networks: Application to Ecology and Evolution._ S. Lek and J. F. Guegan. Berlin, New York, Springer Verlag.

Resh, V. H., Brown, A. V., Covich, A. P., Gurtz, H. W., Li, H. W. and Minshall, G. W. (1988). "The role of disturbance in stream ecology." _Journal of the North American Benthological Society_ **7**: 433-455.

Resh, V. H. and McEltravy, E. P. (1993). Contemporary quantitative approaches to biomonitoring using benthic macroinvertebrates. _Freshwater Biomonitoring and Benthic Macroinvertebrates_. D. M. Rosenberg and V. H. Resh. NewYork, Chapman & Hall.

Resh, V. H. and Jackson, J. K. (1993). Rapid assessment approaches to biomonitoring using benthic macroinvertebrates. _Freshwater Biomonitoring and Benthic Macroinvertebrates_. D. M. Rosenberg and V. H. Resh. NewYork, Chapman & Hall.

Resh, V. H. (1995). Freshwater benthic macroinvertebrates and rapid assessment procedures for water quality monitoring in developing and newly industrialized countries. _Biological Assessment Criteria: Tools for Water Resource Planning and Decision Making_. W. S. Davis and T. P. Simon. Boca Raton, FL., Lewis Publisher.

Resh, V. H., Myers, M. J. and Hannaford, M. J. (1996). Macroinvertebrates as biotic indicators of environmental quality. _Methods in Stream Ecology_. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Reynoldson, T. B., Norris, R. H., Resh, V. H., Day, K. E. and Rosenberg, D. (1997). "The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates." _Journal of North American Benthological Society_ **16**: 833-852.

Richards, C., Host, G. E. and Arthur, J. W. (1993). "Identification of predominant environmental factors structuring stream macroinvertebrate communities within a large argicultural catchment." _Freshwater Biology_ **29**: 295-294.

Robson, B. J. and Chester, E. T. (1999). "Spatial patterns of invertebrate species richness in a river: the relationship between riffles and microhabitats." _Australian Journal of Ecology_ **24**: 599-607.

Rosenberg, D. M. and Resh, V. H. (1993). _Freshwater Biomonitoring and Benthic Macroinvertebrates_. NewYork, Chapman & Hall.

Rumelhart, D. E., Hinton, G. E. and Williams, R. J. (1986). "Learning representations by back-propagating error." _Nature_ **323**: 533-536.

Rumelhart, D. E. and McClelland, J. L., Eds. (1988). _Parallel Distributed Processing: Explorations in the Microstructure of Cognition_. London, The MIT Press.

Scardi, M. (1996). "Artificial neural networks as empirical models for estimating phytoplankton production." _Marine Ecology Progress Series_ **139**: 289-299.

Schleiter, I. M., Borchardt, D., Wagner, R., Dapper, T., Schmidt, K., Schmidt, H. and Werner, H. (1999). "Modelling water quality, bioindication and population dynamics in lotic ecosystems using neural networks." _Ecological Modelling_ **120**: 271-286.

Schofield, N. J. and Davies, P. E. (1996). "Measuring the health of our rivers." _Water_ **23**: 39-43.

Simon, T. P. and Lyons, J. (1995). Appliation of the index of biotic integrity to evaluate water reource integrity in freshwater ecosystems. Biological Assessment Criteria: Tools for Water Resource Planning and Decision Making. W. S. Davis and T. P. Simon. Boca Raton, FL., Lewis Publisher.

Simpson, S., Norris, R., Barmuta, L. and Blackman, P. (1997). Australian River Assessment System - National River Health Program Predictive Model Manual (first draft). Canberra, ACT, CRC Freshwater Ecology, University of Canberra,.

Smith, B. J. (1996). Identification keys to the Families and Genera of Bivalve and Gastropod Molluscs found in Australian Inland Waters. Albury, NSW, Cooperative Research Centre for Freshwater Ecology Identification Guide No. 6.,.

Smith, M. J., Kay, W. R., Edward, D. H. D., Papas, P. J., Richrdson, K. S. J. and Simpson, J. S. (1999). "AusRivAS: using macroinvertebrates to assess ecological condition of rivers in Western Australia." Freshwater Biology **41**: 269-282.

Stanford, J. A., Hauer, F. R. and Ward, J. V. (1988). "Serial discontinuity in a large river system." Verh. Internat. Verein. Limnol. **23**: 1114-1118.

Stanford, J. A. and Ward, J. V. (1993). "An ecosystem perspective of alluvial rivers: connectivity and the hyporheic corridor." Journal of the North American Benthological Society **12**: 48-60.

Statzner, B., Gore, J. A. and Resh, V. H. (1988). "Hydraulic stream ecology: observed patterns and potential applications." Journal of North American Benthological Society **7**: 307-360.

Statzner, B. and Borchardt, D. (1994). Longitudinal patterns and processes along streams: modelling ecological responses to physial gradients. Aquatic Ecology: Scale, Pattern and Process, 34th Symposium of the British Ecological Society. P. S. Giller, A. G. Hildrew and D. G. Raffaelli. Oxford, Blackwell.

Stewart, A. J. and Loar, J. M. (1994). Spatial and temporal variation in biological monitoring data. Biological Monootoring of Aquatic System. S. L. Loeb and A. Spacie. Boca Raton, Florida, Lewis Publisher.

Strahler, A. N. (1957). "Quantitative analysis of watershed geomorphology." Transactions of the American Geophysical Union **38**: 913-920.

Straskaraba, M. and Gnauk, A. H. (1985). Freshwater Ecosystem: Modelling and Simulation. Amsterdam, NewYork, Elsevier.

Straskaraba, M. (1994). "Ecotechnological models for reservoir water quality management." Ecological Modelling **74**: 1-38.

Stream Solute Workshop (1990). "Concepts and methods for assessingsolute dynamics in stream ecosystem." Journal of North American Benthological Society **9**: 95-119.

Suter, P. J. (1996). Baetidae. Mayfly Nymphs of Australia. A guide to Genera. J. C. Dean and P. J. Suter. Albury, NSW., CRC for Freshwater Ecology, Identification Guide No. 7.

Townsend, C. R. (1989). "A patch dynamics concept of stream ecology communities." Journal of North American Benthological Society **8**: 36-50.

Townsend, C. R., Arbuckle, C. J., Crowl, T. A. and Scarsbrook, M. R. (1997). "The Relationship between Land Use and Physicochemistry, Food Resources and Macroinvertebrate Communities in Tributaries of the Taieri River, New Zealand: a Hierarchically Scaled

Approach." *Freshwater Biology* **37**: 177-191.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R. and Cushing, C. E. (1980). "The River Continuum Concepts." *Canadian Journal of Fisheries and Aquatic Science* **37**: 130-137.

Vemuri, V. R., Ed. (1992). *Artificial Neural Network: Concepts and Control Applications*. Los Vaqueros, CA, IEEE Computer Society Press.

Vodopich, D. S. and Moore, C. L. (1984). "Distribution of Benthic Macroinvertebrates in Lake Waco, a Central Texas Reservoir, USA." *Verh. Internat. Verein. Limnol.* **22**: 1470-1474.

Wade, K. R., Ormerod, S. J. and Gee, A. S. (1989). "Classification and ordination of macroinvertebrate assemblages to predict stream acidity in upland Wales." *Hydrobiologia* **171**: 59-78.

Walker, J., Alexander, D., Irons, C., Jones, B., Penridge, H. and Rapport, D. (1996). Catchment health indicators: an overview. *Indicators of Catchment Health: a Technical Perpective.* J. Walker and D. J. Reuter. Melbourne, CSIRO Publishing.

Wallace, J. B., Cuffney, T. F., Webster, J. R., Lugthart, G. J., Chung, K. and Goldowitz, B. S. (1991). "Export of fine organic particles from headwater streams: effect of seasons, extreme discharges, and invertebrate manipulations." *Limnology and Oceanography* **36**: 670-682.

Wallace, J. B. and Grubaugh, J. W. (1996). Transport and storage of FPOM. *Methods in Stream Ecology*. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Walley, W. J. and Fontama, V. N. (1998). "Neural network predictors of average score per taxon and number of families at unpolluted river sites in the Great Britain." *Water Resources Research* **31**(2): 201-210.

Walley, W. J. and Fontama, V. N. (2000). New approaches to river quality classification based upon artificial intelligence. *Assessing the biological quality of freshwaters: RIVPACS and other techniques*. J. F. Wright, D. W. Sutcliffe and M. T. Furse. Ambleside(UK), FBA.

Ward, J. V. and Stanford, J. A. (1982). "Thermal responses in the evolutionary ecology of aquatic insects." *Annual Review of Entomology* **27**: 97-117.

Ward, J. V. and Stanford, J. A. (1983). The intermediate disturbance hypothesis: An explanation for biotic diversity patterns in lotic ecosystems. *Dynamics of lotic ecosystems*. T. D. Fontaine and S. M. Bartell. Michigan, Ann Arbor Sceince.

Ward, J. V. (1992). *Aquatic Insect Ecology 1.Biology and Habitat*. NewYork, John Wiley & Sons, Inc.

Ward, A. K. and Johnson, M. D. (1996). Heterotrophic Microorganisms. *Methods in Stream Ecology*. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Webster, J. R. and Ehrman, T. P. (1996). Solute dynamics. *Methods in Stream Ecology*. F. R. Hauer and G. A. Lamberti. NewYork, Academic Press.

Welsh, A. H., Cunningham, R. B., Donnelly, C. F. and Lindermayer, D. B. (1996). "Modelling the abundance of rare species: statistical models for counts with extra zeros." *Ecological Modelling* **88**: 297-308.

Wen, C. G. and Lee, C. S. (1998). "A neural network approach to multi objective optimisation for water quality management in a river basin." *Water Resources Research* **34**(3): 427-436.

Werbos, P. J. (1992). Backpropagation through time: what it does and how to do it. *Artificial Neural Networks: Concepts and Control Applications*. V. R. Vemuri. Los Alamitos, CA, IEEE Computer Society Press.

Wetzel, R. G. (1983). *Limnology*. Fort Worth, Harcourt Brace College.

Willoughby, L. G. and Mappin, R. G. (1988). "Distribution of *Ephemerella ignita* (Ephemeroptera) in streams: The role of pH and food resources." *Freshwater Biology* **19**: 145-155.

Wilson, H. and Recknagel, F. (*in press*). "Towards a generic Artificical Neural Network model for dynamic predictions of algal abundance in freshwater lakes." *Ecological Modelling*.

Workshop, S. S. (1990). "Concepts and methods for assessing solute dynamicsin stream ecology." *Journal of the North American Benthological Society* **9**: 95-119.

Wright, J. F., Moss, D., Armitage, P. D. and Furse, M. T. (1984). "A preliminary classification of running-watersites in Great Britain based on macro-invertebrate species and the prediction of community type using environmental data." *Freshwater Biology* **14**: 221-256.

Wright, J. F., Furse, M. T., Clarke, R. T. and Moss, D. (1991). Testing and further development of RIVPACS, IFE Interrim report to the National River Authority**:** 1-141.

Wright, J. F. (1995). "Development and use of a system for predicting the macroinvertebrate fauna in flowing waters." *Freshwater Biology* **20**: 181-197.

Wright, J. F., Furse, M. T. and Moss, D. (1998). "River classification using invertebrate: RIVPACS application." *Aquatic Conserv: Marine and Freshwater Ecosystem* **8**: 617-631.

Yabukana, K., Hosomi, M. and Murakami, A. (1997). "Novel application of a backpropagation artificial neural network model formulated to predict algal bloom." *Water Science and Technology* **36**: 89-97.

Yang, M. D., Sykes, R. M. and Merry, C. J. (2000). "Estimation of algal biological parameters using water quality modelling and spot sateline data." *Ecological Modelling* **125**: 1-13.

Yoder, C. O. and Rankin, E. T. (1995). Biological response signatures and the area of degradation value: new tools for interpreting multimetric data. *Biological Assessment Criteria: Tools for Water Resource Planning and Decision Making*. W. S. Davis and T. P. Simon. Boca Raton, FL., Lewis Publisher.