Soft systems analysis of ecosystems

Thesis submitted to Auckland University of Technology in fulfillment of the degree of Doctor of Philosophy

Subana Shanmuganathan

2004

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institution of higher learning, except where due acknowledgement is made herein.

Subana Shanmuganathan

Acknowledgements

I am indebted to Professors Philip Sallis and John Buckeridge for their continuous and tireless efforts in guiding me through this research. Their contribution towards this interdisciplinary research attempt has been significant despite their respective responsibilities and administrative functions at the University as Deputy Vice Chancellor (Research and Development) and Associate Dean, Faculty of Science and Engineering. I wish to extend my sincere gratitude to Professor Nikola Kasabov for his useful advice in the progress of this research and valuable comments in shaping the final thesis.

I thank the Faculty of Science and Engineering for the scholarships awarded to cover my second and third year tuition fees; the office of Deputy Vice Chancellor (R&D) and Software Engineering Research Laboratory (SERL) of the Auckland University of Technology (AUT) for the financial assistance to meet my national and overseas travel expenditure and fees for attending conferences.

The following are acknowledged for their contribution towards this research; Dominic McCarthy, Environmental Scientist and staff of Auckland Regional (ARC) and North Shore City (NSCC) Councils for permission to use their data; Jan Heijs, Wastewater planning engineer for explaining wastewater planning on North Shore, Trish Kirkland-Smith, Environmental Programmes Coordinator for providing the metadata on their water quality sampling; Research staff, namely Dr Russ Babcock, Dr Richard Ford and Jarrod Walker of the Leigh Marine Laboratory of the University of Auckland (UoA) for providing details on ARC's biomonitoring programmes, local knowledge on species assemblages and data collection methods; Staff and undergraduate students of Earth and Oceanic Sciences Research Centre (EOS) for permission to use Long Bay-Okura Marine Reserve data and support; Staff of the National Institute of Water and Atmospheric Research; Neil Binnie, Principal Lecturer, Department of Mathematics for assistance in data analysis with standard statistical methods; Benneth Gilbert for assistance in editing the manuscript; David Parker and Ian Brailsford for their input in editing the final thesis and Ana Krpo for help in incorporating SOM cluster results into a GIS ArcView 8.2.

I am grateful to Josephine Howse for support extended in obtaining information on data sources and also for reviewing the papers submitted at the 5th Annual conference on

Artificial Neural Networks and Expert Systems (ANNES2001). Dave Barry, Patricia Sallis, Paul Doyle and Brendan Dobbs are commended for cheering and helping me at various occasions during the course of this research.

Information Technology staff members of AUT are acknowledged for their efforts in running my computer system for this research. I also wish to thank the administrative staff at the Technology Park of AUT for their efforts in assisting me with various administrative procedures. AUT library staff deserves credit for their valuable services, in particular services from the interloan section staff have been remarkable.

I also wish to express my gratitude to the Doctoral Studies Board, in particular to Professor Ian Shirley, Peter Fleming and Annette Tiaiti.

I take this opportunity to thank Professor Sergei Gulyaev, Keith Ufton the Faculty Registrar, Jenny Walters the Faculty Manager, Warren Ng, Monika Haarman and Leonie Simpkin for their favours.

Finally, I need to mention the support and encouragement extended by my parents, sisters Manchu and Sulo, and brother-in-law Mano.

Abstract

This research is a case study evaluation of the use of self-organising map (SOM) techniques for ecosystem modelling to overcome the perceived inadequacies with conventional ecological data analysis methods. SOMs provide an analytical method within the connectionist paradigms of artificial neural networks (ANNs), developed from concepts that evolved from late twentieth century neuro-physiological experiments on the cortex cells of the human brain.

The rate and extent at which humans influence environmental deterioration with commensurate biodiversity loss is a cause for major concern and to prevent further degradation by human impact, parsimonious models are urgently needed. Indeed, the need for better modelling techniques has never been so great. Ecologists and many national and international bodies see the situation as 'significantly critical' for the conservation of our global ecosystem to foster the continued wellbeing of humanity on this earth.

The thesis investigates and further refines SOM based exploratory data analysis methods for modelling naturally evolving, highly diverse and extremely complex ecosystems. Earlier studies provide evidence on SOM ability to analyse complex forest and freshwater biological community structures at limited scales. On the other hand, growing concerns over conventional methods, their soundness and ability to model large volumes of data are seen as of little use, leading to arguments on the results derived from them.

Case study chapters illustrate how SOM methods could be best applied to analyse often 'cryptic' ecosystems in a manner similar to that applied in modelling highly complex and diverse industrial system dynamics. Furthermore, SOM based data clustering methods, used for financial data analysis are investigated for integrated analysis of ecological and economic system data to study the effects of urbanisation on natural habitats.

SOM approaches prove to be an excellent tool for analysing the changes within physical system variables and their effects on the biological systems analysed. The Long Bay-Okura Marine Reserve case study elaborates on how SOM based approaches could be best applied to model the reserve's intertidal zone with available numeric data. SOM maps depicted the characteristic microclimate within this zone from ecological

monitoring data of physical attributes, without any geographical data being added. This kind of feature extraction from raw data is found to be useful and is applied to two more case studies to study the slow variables of ecosystems, such as population dynamics, and to establish their correlation with environmental variations. SOM maps are found to be capable of distinguishing the human induced variations from that of natural/ global variations, at different scales (site, regional and global) and levels using regional and global data. Hence, SOM approaches prove to be capable of modelling complex natural systems incorporating their spatial and temporal variations using the available monitoring data, this is a major advantage observed with SOM analyses.

In the third case study, potential use of SOM techniques to analyse global trends on the effects of urbanisation in environmental and biological systems are explored using the World Bank's statistical data for different countries. Many state and international institutions, concerned over global environmental issues, have made attempts to develop indicators to assess the conditions of different ecosystems. The enhancements with SOM approaches against the currently recommended indicator system based on information pyramid and pressure-state-response (PSR) models are elaborated upon.

The research results of SOM methods for ecosystem modelling, similar to that applied to industrial process modelling and financial system analysis show potential. SOM approaches (i.e. cluster, dependent component, decision system and trajectories/ time series analyses) provide a means for feature extraction from the available numeric data at different levels and scales, fulfilling the urgent need for modelling tools to conserve our global ecosystem. They can be used to bridge the gap in converting raw data into knowledge to inform sustainable ecosystem management. Increasingly, traditional methods based on Before-After-Control-Impact (BACI) designs and Analysis of Variance (ANOVA) are seen to be unsuitable for ecological data analysis, as they are unable to detect human induced environmental impacts from that of a natural cause. This thesis proves that SOM techniques could be applied to modelling not only a natural systems complexity but also its functioning and dynamics, incorporating spatial as well as temporal variations, to overcome the constraints with conventional methods as applied in other stated disciplines.

Contents

List of Abbreviations

AI	Artificial Intelligence
ANNs	Artificial neural networks
ANOVA	ANalysis Of Variance
ART	Adaptive resonance theory
AUSRIVAS	Australian Rivers Assessment System
AUT	Auckland University of Technology
BACI	Before-After-Control-Impact
BACIPS	Before-After-Control-Impact Paired Series
BMU	Best matching unit
BOD	Biological oxygen in demand
CoA	Correspondence analysis
CPU	Central processing unit
DO	Dissolved Oxygen
DoC	Department of Conservation
DRP	Dissolved reactive phosphorus
ECOSs	Evolving Connectionist Systems
EIA	Environmental impact assessment
EPA	Environmental Protection Agency
EPTC	Ephemeroptera, Plecoptera, Trichoptera and Coleoptera
ESOM	Evolving Self-Organizing Map
EU	European Union
FAO	Food and Agriculture Organization's
GDP	Gross Domestic Product
GIS	Geographical Information Systems
GUI	Graphical user interface
IAM	Integrated assessment and modelling
IEEE	Institute of Electrical and Electronic Engineers
IFC	International Financial Corporation
IS	Intelligent systems
IUCN	© International Union for Conservation of Nature and Natural Resources
LTB	Long-Term Baseline
MDS	Multidimensional scaling

MfE	Ministry for the Environment
MIR-Max	Mutual Information and Regression Maximisation
MLPs	Multi layer perceptrons
MOPED	Modelling Patterns in Environmental Data
NCER	National Center for Environmental Research
NIWA	National Institute of Water and Atmospheric Research
NMC	Nonmetric multidimensional scaling
NMDS	Non metric multidimensional
NNs	Neural networks
NOAA	National Oceanic and Atmospheric Administration
NRC	National Research Council
NRHP	National River Health Program
NSCC	North Shore City Council
NSF	National Science Foundation
OECD	Organisation for Economic Cooperation and Development
ORD	Office of Research and Development
PCA	Principle component analysis
PSR model	Pressure-State-Response model
RAM	Random Access Memories
RICBIS	Repository for Intelligent connectionist-Based Information
Systems	
RIVPACS	River InVertibrate Prediction And Classification System
RPDS	River Pollution Diagnostic System
SOM	Self-organising map
Sp Cond	Specific Conductivity
STAR program	Science To Achieve Results program
SUA	Stochastic urban accretion
TBL	Triple bottom line
u-matrix	Unified distance matrix
UoA	University of Auckland
VTDEC	Vermont Department of Environmental Conservation
WCMC	World Conservation Monitoring Centre
WRI	World Resources Institute
WSJ	Wall Street Journal

List of Figures

Chapter 1

Introduction

This research investigates the use of self-organising map (SOM)[1] techniques for modelling disparate ecological data with a systems approach to overcome the perceived inadequacies with conventional data analysis methods and, in so doing, to address the following research questions:

(i) How could SOM methods be best applied to unravel the structure of highly diverse, extremely complex[2] and naturally evolving ecosystems[3] and to predict their system dynamics?

(ii) How this approach could be applied to transform and disseminate disparate (ecological and socio-economic system) data to a wider community in order to foster conservation of our global ecosystem?

1.1 Hypothesis and assumptions

This research further develops the use of SOM techniques for modelling ecosystem structure, functioning and dynamics, in a manner similar to that applied in industrial process modelling. SOM based approaches, built on measurable process variable data, also referred to as "soft sensors", are successfully applied to monitoring and control of industrial processes without any physical models (Simula et al. 1999; Himberg et al. 2001). They can effectively track the process dynamics of highly complex and diverse industrial systems, provided that a large volume of good quality data is made available. In addition, innovative SOM applications in financial data analysis are investigated for analysis of disparate ecological and economic data to study the effects of urbanisation on the environment including possible economic trade-offs within an ecosystem framework.

---

[1] SOMs, first introduced by Tuevo Kohonen (1982), are feed forward artificial neural networks based on an unsupervised algorithmic training. They are capable of projecting multidimensional input vectors on to a low dimensional, topology preserving output display of self-organising neurons.

[2] The term complex refers to the many interrelating components and mechanisms those make the final outcome of a process, difficult for inference.

[3] An ecosystem is "A biological community *termed as a biological system in this research* and the physical environment, *which in turn is termed as the physical/environmental system* associated with it…" Concise Science Dictionary (1991). Oxford Reference. (*Attention of italics is for specific clarification of this research*).

In theory, the approaches should be viable as SOM methodologies are successfully applied to overcome the issues with traditional methods in the said disciplines.

SOM applications to modelling biological population assemblage using forest (Giraudel and Lek 2001) and freshwater (Ce´re´ghino et al. 2001) system data, at limited scales, provide the basis for this research on the use of SOM based methods in ecology. Giraudel and Lek (2001) concluded that the SOM algorithm could be used for exploratory data analysis in ecology, complementing the existing classical techniques. In (Murray-Bligh 1998) SOMs are considered to have shown "…considerable potential for diagnosing different types of pollution…." In this system, SOM methods are applied to predict the biotic indices based on the distribution and abundance of BMWP[4] taxa for establishing the water quality index of a system. Modelling Patterns in Environmental Data (MOPED) uses SOM techniques for mapping of patterns in freshwater system data, such as fish species distribution and elevation of freshwater systems, and to predict the biological assemblages that should be present in certain streams (Jowett 2001).

1.2 Ecosystem modelling: A need for better techniques

"Humans have profoundly changed the world's ecosystems. Some 40 to 50 percent of land has been transformed (through change in land cover) or degraded by human actions; more than 60 percent of the world's major fisheries are in urgent need of actions to restore overfished stocks or to protect stocks from overfishing; natural forests continue to disappear at a rate of about 14 million hectares each year; and other ecosystems such as wetlands, mangroves, and coral reefs have been substantially reduced or degraded…" (Reid 2000:3).

"Humans have appropriated half of the accessible global freshwater runoff, and this could climb to 70% by the year 2025. Nearly 2/3 of all rivers are regulated in some manner, causing fragmentation, deterioration, and losses of flood plains, wetlands, and riparian ecosystems…" (Clark et al. 2001:1).

---

[4] The BMWP scores are devised for the taxonomic families occurring in British rivers, depending upon their sensitivity to organic pollution, those very sensitive to organic pollution with 10, down to families more tolerant of pollution with 3 or less.

Even with these disturbing statements and predictions, the attitude of resource managers and stakeholders remains unchanged due to lack of sufficient relevant knowledge on ecosystem response to human influenced activities (Bierbaum et al. 2001). This is despite the significant efforts made by national, international and academic research institutions to elucidate how human environmental effects cause damage to natural habitats (Osenberg and Schmitt 1996). In fact, many resource managers and stakeholders still assume that the effects of human impact on the environment as being mitigated or neutralised (Clark et al. 2001). Some state authorities even consider natural resources as a plethora or infinite (Buckeridge 1999) and tend to approve developmental activities that could possibly lead to great losses in ecosystem functioning and biodiversity. On a world scale, state institutions spend US $ 700 billion a year to subsidise environmentally unsound practices in the use of water, agriculture, energy and transport (World Resources Institute 2000).

Understanding complex ecosystems, places more significance on environmental sciences as humans attempts to transform their relationship with the Earth and its natural resources, more sustainable by the latter, argued (Graedel et al. 2001); a seventeen-member committee appointed by the National Science Foundation (NSF) and the National Research Council (NRC). The committee was asked to identify topics of greatest potential for immediate investment and in their report titled 'Grand Challenges in Environmental Sciences', eight chosen topics were elaborated. Under the topic biological diversity and ecosystem functioning, the report reads,
"The challenge is to improve understanding of the factors affecting biological diversity and ecosystem structure and functioning, including the role of human activity. Important research areas include improving tools for rapid assessment of diversity at all scales; producing a quantitative, process-based theory of biological diversity at the largest possible variety of spatial and temporal scales; elucidating the relationship between diversity and ecosystem functioning; and developing and testing techniques for modifying, creating, and managing habitats that can sustain biological diversity, as well as people and their activities" (Graedel et al. 2001)

Similar views on the need for new models to elucidate ecosystem functioning and biodiversity including human activity, to inform sustainable environment management, are expressed in (Ravetz 2000; Parker et al. 2001; Harris 2002). Current methods are inadequate to study the environmental effects of an impact, argued (Thrush et al. 1995). Also, explained the reasons as to why these effects could not be modelled with traditional methods; the environmental effects that are extensively varying even within an ecosystem due to the spatial and temporal variations in the system, may also depend on the nature and extent of the impact being analysed. In addition, species[5] threshold responses[6] and biodiversity[7] (Raimondi and Reed 1996) further complicate ecosystem modelling.

Even the highly complex conventional ecological data analysis methods are not capable of detecting an environmental impact (Stewart-Oaten 1996; Thrush et al. 1995), let alone prediction. For instance, Before-After-Control-Impact (BACI) design methods, solely developed to analyse an impact at a particular site using data before and after an activity occurs, including control sites (and paired sampling BACIPS) may not provide a good assessment for decision making, mainly because of model uncertainty (Stewart-Oaten 1996). Thrush et al. (1995) argued that these complex methods could not be used to describe the formal biological results succinctly and unambiguously within a single general parameter, such as mean abundance.

For instance, Walker et al. (2000), who attempted to analyse the effects of urbanisation on subtidal population dynamics along the northeastern coast of Auckland, based on BACI design with conventional univariate and multivariate analyses, could not establish a link between the causal process and the environmental effects or otherwise from the monitoring data. Conventional methods invariably produce complex matrices of different functions, such as diversity indices, that are difficult to analyse the biological assemblage and its dynamics (Giraudel and Lek 2001).

Bowler (1992) blamed the research approaches of the twentieth century for significantly contributing to the current global environmental issues of pollution and over exploitation of natural resources. In the last century, research in environmental sciences became

---

[5] Species: In systematic biology, species is one of the groups into which a genus is divided. The members of a species are able to interbreed (Collins English Dictionary 1997) and produce viable offspring. In general, species are the final group in the classification of living beings, although in some special cases, sub species are possible within a species.

[6] Threshold responses are the different levels at which deferent species response to the same impact.

[7] Biodiversity is defined as species richness and evenness in a system.

more focused in gaining in-depth knowledge with highly specialised scientific fields, enforcing a fragmented image of nature. Actions based on such narrowly focused research and knowledge are found to be responsible for even altering the Earth's basic chemical cycles (Vitousek et al. 1997; Kirby 2000).

Despite these impeding issues with ecosystem modelling, understanding and prediction of environmental ramifications of human induced impacts are considered to be crucial for sustainable environment management. Blanket restriction on developmental activities would cause undue socio-economic hardship on the current generation. *Vice versa* could cause detrimental efforts on the environment, affecting future generations (Schmitt et al. 1996).

In view of the complexities involved in modelling environmental and biological processes, a holistic approach[8] (Buckeridge 1994; Reid 2000) supported by interdisciplinary research (Soule and Kleppel 1988) has been suggested since the 1980s. The use of new models to elucidate complex ecosystem interactions at different scales and levels, involving data sets, computation and statistics, to assess the state of an ecosystem and to predict the systems response, is critical for sustainable environment management (Soule and Kleppel 1988; Buckeridge 1994; Hammond et al. 1995; Ravetz 2000; Gustavsson 2001; Harris 2002). However, owing to lack of quantitative methods, achieving environmental sustainability seems remote (Shanmuganathan et al. 2003). Conventional models either oversimplify or over complicate the environmental issues (Ravetz 2000; Parker et al. 2001; Harris 2002). They do not facilitate interdisciplinary research on natural habitats with a systems approach, nor do the attitudes of different professionals involved (stakeholders, resource managers and land developers), who mistrust each other and show little enthusiasm even for initiating such efforts (Ravetz 2000; Parker et al. 2001; Harris 2002). The situation leaves the environmental decisions to be made by the expensive and stochastic process of law (Lester 1996).

Thesis case studies illustrate the use of various SOM approaches to quantitative analysis of ecosystem structure, functioning and dynamics, incorporating spatial as well as temporal changes at different scales and levels with a systems approach. This is carried

[8] Holistic approach is sometimes refereed to as 'an integrated' or 'a systems approach'.

out by collectively analysing the available ecological and socio-economic data sets that are not analysed to their potential (Hammond et al. 1995; Vant 1999). Scientists and the general public carryout monitoring programmes, such as the ones by (Vant 1999; Wilcock and Stroud 2000; North Shore City Council - Project care 2002; North Shore City Council - Wai Care 2002) and produce large volumes of data, which can not be integrated for analysis owing to inconsistent labelling. Details on how integrated analysis could be best performed on such dissimilar data sets, to unravel often 'cryptic' ecosystems and to predict their responses for trade-off analysis, using SOM methodologies, are revealed in chapter 4 Experimental methodology.

Unlike the financial analysts, ecologists do not have a single established practice to assess the state of our environment or the natural resources (Hammond et al. 1995). Across the United States (US), policy makers and many others carefully watch every single move of Dow Jones and make predictions on its effects in their daily business activities (Lash 1995). Even global economic trends are predicted based on small changes in such numeric figures, whereas no such indicators are in use to measure the state of the environment or the natural resources to conserve the global ecosystem. Scientists and ecologists are still engaged in research in devising a suite of concise ecological indicators to assess ecosystem changes by integrating the structure, functioning and its biological diversity (National Center for Environmental Research (NCER) Office of Research and Development (ORD) - US Environmental Protection Agency (EPA) 2000 ; Ravetz 2000; Bierbaum et al. 2001; Parker et al. 2001; Harris 2002)).

The shortcomings of the currently recommended indicator system based on information pyramid[9] and pressure-state-response (PSR) concepts are illustrated in chapters 3 and 7. Using this indicator system, local environmental problems, arising from ozone depletion, climate change, acid rain and many more have been studied in the Netherlands for the first time (Hammond et al. 1995). Through this indicator system, a number of related primary indicators are aggregated to form the condensed, composite indices based on well-defined physical processes of the human-environment interaction. The third case study of this research illustrates how SOMs could be used to analyse global data within an integrated framework even with limited prior knowledge on the physical processes.

[9] Information pyramid consists of highly aggregated indices on the top of the pyramid with primary data from monitoring programmes at the bottom. Condensation of data is carried out in a bottom up fashion.

1.3 Artificial neural networks in information processing

Biologically inspired ANNs provide a totally different approach to the conventional computational algorithmic information processing methodologies. Conventional computing methodologies consist of sequential programs with explicit step-by-step instructions to solve a problem whereas no such clear understanding of either the problem or the solution is required in ANN algorithms. The use of ANNs in information processing has been significant since the 1940s.

Deboeck (1998) defined neural networks (NNs) as a collection of mathematical techniques that could be used for signal processing, forecasting and clustering. ANNs could be considered as non-linear, multi layered, parallel regression techniques. NN modelling is like fitting a line, plane or hyper plane through a set of data points to define the relationships that may exist between (in this case) the inputs and the outputs; or it can be fitted for identifying a representation of the data on a smaller scale.

Summary

In this research, SOM techniques are investigated for ecosystem modelling, in a manner similar to that applied in complex industrial system modelling; SOM techniques are seen as being successful in providing an alternative approach to a global system analytical model for complex industrial processes. In addition, innovative SOM approaches of financial data analysis are investigated for integrated analysis of dissimilar ecological and economic system data to model the effects of urbanisation on natural habitats, as conventional methods are perceived to be inadequate in this regard. New methods are urgently required to elucidate ecosystem functioning and biodiversity including human activity, to preserve natural habitats as well as foster human wellbeing on this Earth. Unless the physical interactions are known, the latest pressure-state-response approach, first introduced by the Dutch cannot be applied to modelling natural processes.

Case study chapters illustrate how SOMs could be best applied to modelling ecosystem structure, functioning and dynamics at different scales and levels by making use of the available ecological monitoring and economic system data (from different sources) with a systems approach. Earlier studies produce evidence of SOM use for studying the biological assemblage of forest and freshwater systems at limited scales.

Chapter 2

Self-organising maps: A review

The previous chapter provided an introduction to this research, aimed at exploring the application of SOM techniques to modelling highly complex and diverse ecosystems using biological and environmental monitoring data. SOMs belong to a connectionist paradigm[10] of artificial neural networks (ANNs). This chapter illustrates the literature on ANNs with special emphasis on SOM methodologies and their application to real world problems, reviewed for the research.

2.1 Artificial neural networks

ANNs are biologically inspired approaches to intelligent information processing methodologies. They provide a means to incorporating innovations and flexibility into conventional computing and to solve real world problems (Amari 1995). Complex problems of modern day (Kasabov 1999, 2000) and human expectations from computers (Aleksander 2000) continue to demand innovations in this rapidly changing field.

Efforts to understand the functioning of the human brain and its structure have existed ever since human beings themselves began wondering about their fascinating thinking ability. However, the use of the brain's cognitive abilities and its functioning to construct ANNs and to formulate concepts of artificial intelligence (AI) began only six decades ago. The discovery of actual processing in the human brain consisting of 10^{11} neurons, participating in perhaps 10^{15} interconnections over transmission paths, is unlikely to be made in the near future (Wasserman 1989). However, every breakthrough made by neurologists and neuroanatomists has been modelled and then applied to information processing in knowledge engineering[11]. More recent insights into the brain's capabilities to automatically acquire information-processing algorithms (Matsumato 1999), elucidate the mental growth and the involved factors. They also lay a platform for developing novel information processing methods. The latter is known as artificial neural networks.

---

[10] Connectionist paradigm: A term that refers to the ANN models of the late 1970s.

[11] Knowledge Engineering: A term that refers to the academic research aimed at developing models, methods and basic technologies for representing and processing knowledge for building intelligent knowledge-based systems.

The ANN computational methods have a fundamental difference to the traditional 'conventional computing methods' that basically consist of sequential programming. Conventional computing methods are successfully applied to solve highly laborious, repetitive tasks, such as complex mathematical calculations or 'rote things', without making any mistakes similar to that characteristic of humans, arising from fatigue. However, conventional computers cannot solve what are simple problems for humans, such as remembering patterns, relating and using them for future processing, especially for recognising images and figures that can be handled effortlessly even in low order animal brains (Anderson and McNeill 1992).

One of the many breakthroughs achieved in neurology is to understand how patterns of information are stored in biological nerve cells. Based on such understandings innovative ANN computational methodologies are developed for storing information in the neuronal structures and networks that could be trained through parallel processing. During the training process depending on the neuron type, network architecture and the training algorithm used, the necessary information is transferred into the network in the form of weights and connections, which are later used to solve specific problems; generally found to be impossible by conventional computational methods. Since the 1940s, ANNs have been considerably successful in introducing heuristics into computational algorithmic processing; associated with them are the terms, behave, react, adapt, self-organise, learn, generalise, and forget (Anderson and McNeill 1992).

The kind of parallel processing used in neural computing is different from that of parallel distributed processing. In parallel distributed processing, the main task is subdivided into several fragments, each with its own processor. All these fragments are run simultaneously, whereby the execution of the main task is sped up. Nevertheless in neural computing, the neurons and the network architecture are used to store information. The summation of input into weight and the transformation function are part of the information storing process.

It is widely accepted that the human brain is much more complicated than the available models and many of its cognitive functions are still unknown (Kasabov 1995). Nonetheless the following are the main characteristics that are considered as common in real and artificial networks based on figures 2.1 a & b:

(i) learning and adaptation,

(ii) generalisation,

(iii) massive parallelism,

(iv) robustness,

(v) associative storage of information, and

(vi) spatiotemporal information processing

(Kasabov 1995).

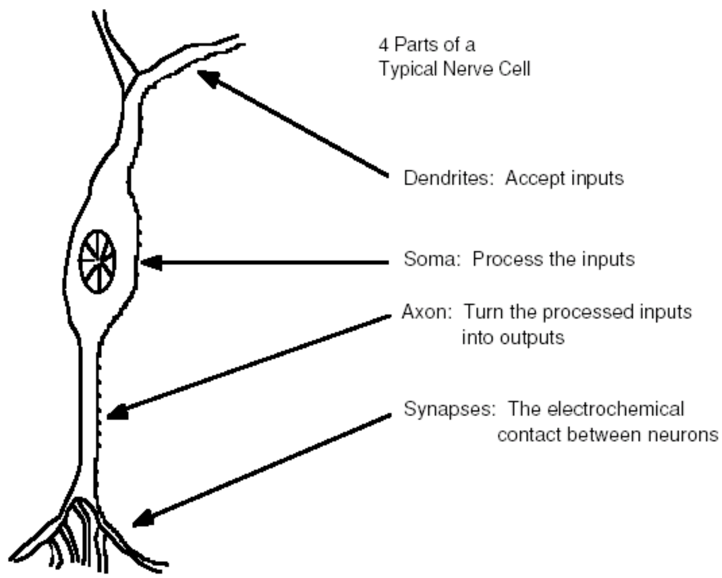
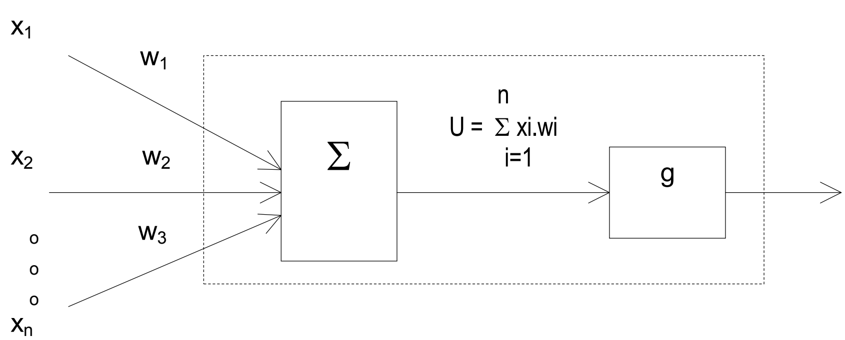


Figure 2.1 a: A schematic diagram of a simple neuron. Source: (Anderson and McNeill 1992).



x_1, x_2 ..... x_n : Input connections, W1, W2....Wn : Weights, U : Summation function, g ; Threshold.

Figure 2.1 b: A model of an artificial neuron. Source: (Kasabov 1995).

ANNs are generally defined by the following four parameters:

(i) Type of neuron (or nodes as a neural network resembles a graph), such as Perceptron of Pitts and McCulloch (1943) and Yamakawa's fuzzy neuron (1990).

(ii) Connectionist or network architecture: The organisation of the connections between the neurons in the network is described as its architecture. The topology of the network, such as fully connected or partially connected, is one way of defining the network architecture. ANNs can also be distinguished based on the number of layers and the number of input and output neurons in the layers:

  a Autoassociative: where input neurons are the output neurons i.e. Hopfield network

  b Heteroassociative: consists of separate input and output neurons i.e. MLP.

  Furthermore, depending on the connections back from the output to the input neurons, two different kinds of architecture are determined:

  a Feed forward architecture: where no connections are found, back from the output neurons to the input neurons. The network does not remember values of its previous output or the activation states of its neurons.

  b Feedback architecture: where there are connections back from the output neurons to the input neurons and as such the network holds the memory of its previous states and the next state depends on the current input signals and the previous states of the network i.e. Hopfield network.

(iii) Learning algorithm: the algorithm used to train a network is referred to as the learning algorithm. Extensive research has been carried out in trying out various concepts and it gives researchers a high degree of flexibility for innovations. Discussing the whole set of learning algorithms is far beyond the scope of this section. However, the learning algorithms so far used could be classified into three groups:

  a. Supervised learning: the training examples that consist of input vectors x and their desired output vectors y, are used in the training. Training is performed until the neural network 'learns' to associate each input vector x to its corresponding output vector y (approximate a function $y = f(x)$). It encodes the example in its internal structure.

b. Unsupervised learning: only input vectors x are supplied and the neural network learns some internal feature of the whole set of the input vectors presented to it. Contemporary unsupervised algorithms are further divided into two i) noncompetitive and ii) competitive.

c. Reinforcement learning: also referred to as reward penalty learning. The input vector is first presented and the neural network is allowed to calculate the corresponding output. If the network calculation is good then the existing connection weights are increased (rewarded), otherwise the connection weights involved are decreased (punished).

(vi) Recall algorithm: the algorithm, by which learned knowledge in a trained network is recalled to solve similar but new problems, is referred to as the recall algorithm.

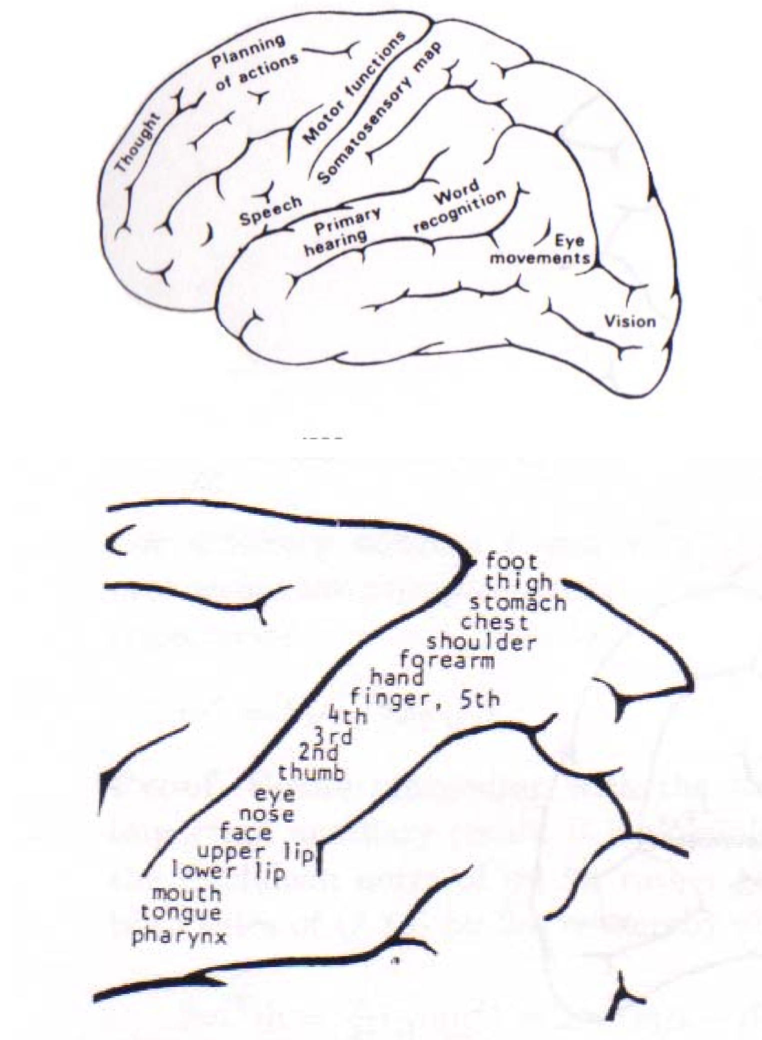
Following this brief introduction to ANNs, in the next section SOMs and their application to real world problems are discussed.

2.2 Self-organising maps

A SOM is a feed forward neural network as shown in figure 2.2 c. It uses an unsupervised training algorithm to perform non-linear regression. Through a process called self-organisation, the network configures the output data into a display of topological representation where similar input data are clustered near each other. At the end of the training, SOM enables analysts to view any novel relationships, patterns or structures in the input vectors. The topology preserving mapping nature of the SOM algorithm is useful in projecting multidimensional data sets into low dimensional displays, generally one- or two-dimensional planes. Thus SOMs can be used for clustering as well as visualisation of multidimensional data sets (Deboeck and Kohonen 1998).

The SOM techniques are successfully applied to many real world problems of large volumes of complex multidimensional data sets, such as pattern recognition, image analysis, process monitoring and control, and fault recognition. As SOM methods are based on an unsupervised training algorithm, they could be used for data clustering without knowing the class membership of the input vectors (Simula et al. 1999).

Traditional methods, such as simple statistical methods, that are useful in summarising low-dimensional data sets (mean value, smallest and highest values), are seen to be less effective in visualising multidimensional (i.e. multivariate) data sets (Deboeck 1998); (Deboeck and Kohonen 1998)



Figures 2.2 a & b: Brain areas and somatosensory map. Source: (Kohonen 1997).

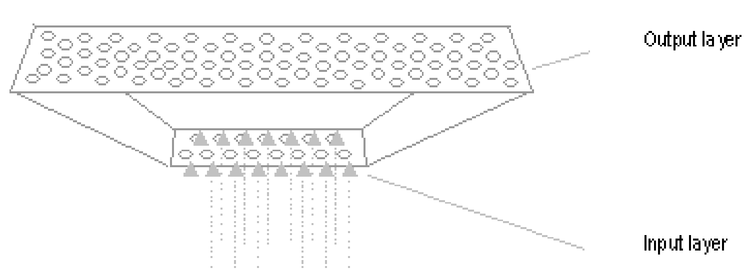


Figure 2.2 c: A simplified diagram of a SOM.

The self-organising map (SOM) algorithm, first introduced by Teuvo Kohonen (1982) was developed from the basic modelling information of the human brain's cortical cells, known from the neuro-physiological experiments of the late twentieth century. The processing of synaptic connections between the cortex cells in the human brain is based upon the nature of the sensorial stimuli. Different patterns of sensorial signals converge at different areas within the brain's cortex cells. Because of this reason different individual neurons or groups of neurons become sensitive to different sensorial stimuli. Neighbouring neurons also learn to respond to similar patterns of signals (figure 2.2 a) including visual, auditory and somatosensory. The somatopic map of such an order is shown in figure 2.2 b. Despite the insights gained in the area of synaptic processing, knowledge on the associative areas of signals and other different tasks involved with the rest of the cortical area is relatively poor. Only ten percent of the total cortical area is described to be involved with the primary sensorial signals. The planning of actions is assumed to take place in the frontal lobe (Kohonen 1997).

Based on Kohonen's SOM, an algorithm of Evolving Self-Organizing Map (ESOM), was developed by (Deng and Kasabov 1999). Unlike the former, ESOM network structure evolves in an on-line adaptive mode with capabilities for exploring large volumes of data flows, updated daily, hourly or every minute. Extracting knowledge from such large and continuously changing data sets, received in an on-line environment, could be of invaluable use for future decision-making, especially in macroeconomic performance of individual countries or country clusters.

2.3 Applications of self-organising maps

SOM applications to real world problems along with traditional data clustering and visualisation methods are examined.

2.3.1 Applications of SOM methods to real world problems

Kohonen's SOM (1996) applications to real world problems have been successful; most of them centred on knowledge discovery. SOM ability to discover implicit knowledge from numerical data is significant. SOMs are capable of displaying the input vectors on low dimensional grid structures while preserving the topology of the original data. However, the main objective of the analysis should be identified before designing ways and means for assessing the effectiveness of the outcome. The following are some of the major identified areas of SOM applications:

(i) Classification, clustering, and/or data reduction
(ii) Visualisation of the data
(iii) Decision-support
(iv) Hypothesis testing
(v) Monitoring system performance
(vi) Lookup table for missing values
(vii) Forecasting

(Deboeck and Kohonen 1998)

In applications aimed at clustering and visualisation of data, consideration of various other traditional statistical clustering methods is advised. Combining traditional methods (explained in the next section) with SOM methods gives an opportunity to obtain some priori knowledge on the data, which would enhance the design on better data reduction, such as how much data reduction would be desired for the design of SOM.

For decision support applications, it is important to precisely define the decisions to be supported such as the scope of decisions and time frame.

In hypothesis testing, it is essential to define the hypothesis and the standards for acceptance and rejection, prior to the analysis.

For monitoring applications, the main goal of the monitoring process, such as quality control, fault detection and standard compliance, should be defined. In occasions where forecasting is the main objective, it is necessary to describe the forecasting window, the preferred accuracy and the methods for performance evaluation.

2.3.2 Traditional methods for data clustering and visualisation

Traditional statistical methods consist of limited abilities for revealing structures, relationships and novel patterns in low dimensional data sets. Two to three dimensional data sets can be visualised using simple two- to three-dimensional graphs. Nonetheless, with multidimensional data sets, plotting a vector or analysing the relationships between different vectors by simple graphs is not possible. Thus other methods are needed to visualise such multidimensional data sets.

In the existing data visualisation methods, the different components contributed by each and every dimension are integrated into the one final result. The major drawback experienced with conventional methods is that they are unable to reduce the amount of data within large sets, as processing becomes incomprehensible. Nonetheless conventional methods can be used to display simple summaries in low dimensional data sets (Deboeck and Kohonen 1998).

Data clustering is an operation, through which similar data items are categorised or grouped together, in the best possible method to reduce the large volumes of data for visualisation purposes. Clustering is described as similar to information processing in humans and preferred to projection methods[12]. Clustering can be automated to classify different categories and automation also reduces bias and errors in the grouping process.

2.3.2.1 Clustering methods

Traditional clustering methods can be classified into two basic types:

i) Hierarchical, and (ii) non-hierarchical.

Hierarchical clustering is proceeded successively by merging the smaller clusters into larger ones, or by splitting the larger clusters into smaller units. In general, clustering methods differ based on the rule that is used to decide on the merger/ splitting of clusters. The end result of the algorithm will be a tree of clusters called as a dendrogram that shows the clusters and their relationships. When two small clusters are merged, a new higher level is created in the dendrogram. The representation of the new level is connected to its respective representations in the lower level clusters. By cutting the dendrogram at an appropriate level the data items could be clustered into different groups (Deboeck and Kohonen 1998).

---

[12] Projection methods: The goal of projection methods is to represent the input data in a chosen low dimensional space, where certain properties of the structure of original data are preserved as faithfully as possible to the original values. Thus these projections can be used to visualise a high dimensional data set if a sufficiently enough low dimensionality is chosen for output display

Non-hierarchical clustering is performed by directly decomposing a data set into disjoint clusters. The different kinds of algorithms used in the clustering vary in assigning clusters to the most densely regions in the data space. The algorithm used defines a cluster, where there are a large number of similar data items are positioned. Another possible approach used for this purpose is minimising the measures of dissimilarities of the samples within each cluster while maximising the dissimilarities between different clusters (Deboeck and Kohonen 1998).

2.3.2.2 Projection methods

Projection methods can be classified into two basic types:

(i) Linear, and (ii) non-linear.

A data set, represented on a *n* dimensional space could be projected on a subspace represented by *m* dimensions (either one-dimensional such as a line or two-dimensional plane), within the *n* dimensional space (where *m* is less than *n*): referred to as the linear projection. The basic concept behind projection methods is that it is possible to represent a data set with a subset of vectors that comprise a linear subspace of a lower dimensionality. Each of the vectors in an *m* dimensional linear subspace would be a linear combination of *m* independently selected basis vectors. However, it is difficult to visualise the structure and distribution of multidimensional data sets, especially the data sets that consist of highly unsymmetrical distribution, on a low dimensional display by using linear projection methods. Principle component analysis (PCA) is a linear projection method.

There are several other approaches to projecting non-linear, highly asymmetric data structures onto low dimensional displays, referred to as non-linear projection methods. In most of these approaches, as a first step, every data item is actually mapped as a point in a lower dimensional space. This mapping is then optimised to make the distances between the image points as similar as possible to the original distances of the corresponding data items. The existing methods vary based on how the different distances are weighted and their representations are optimised. Multidimensional scaling (MDS) is an example of non-linear projection method (Deboeck and Kohonen 1998).

2.3.3 Data mining and knowledge discovery

Data mining is a process by which raw data is sieved through searching for useful information. Information is extracted in the form of patterns, structures or relationships. The misinterpretations that existed over the exact meaning of some terms, such as 'Data mining', 'exploratory data analysis' and 'knowledge discovery' were eventually resolved at a conference [13] in 1995 (Deboeck and Kohonen 1998).

Statisticians blame themselves for being slow in adapting to recent changes and their attitude towards data and technology, which led IT professionals embark on research in data mining techniques (Pregibon and DuMouchel 2001). Furthermore, faster switching, data compaction and better technologies also contributed towards the need for better techniques to analyse the abundant digital data to its potential (Mitchell 1999).

SOM applications in industrial system modelling are evaluated in the next section as the same concept is being experimented in this research to ecosystem modelling, using biological and environmental monitoring data with a systems approach.

2.3.4 Applications of SOMs in industrial process modelling

In the modelling and control of complex industrial systems, it is usually assumed that a global system analytical model could be defined. However, many industrial processes are so complex and diverse it is not possible to build a global model for this purpose. Nonetheless, SOM based ANN models are successfully used to model such complex industrial processes, based directly on their process variable measurements alone. They provide a means to analyse these processes without any physical models, provided that a large volume of good quality, stable, numerical data, describing the process is made available. Similar ANN models, used in the estimation of signal values or process variables, measured indirectly or offline are referred to as 'soft sensors' (Simula et al. 1999).

---

[13] Initially, the terms 'Data mining' and 'exploratory data analysis' were used for knowledge discovery, the whole discovery process of novel patters or structures in the data. It was proposed at the first international conference in Montreal in 1995 that the term 'knowledge discovery' be employed to describe the whole process of knowledge extraction (knowledge means relationships and patters between data elements) from data and the term 'data mining' be used exclusively for the discovery stage of the process (Deboeck and Kohonen 1998b).

The theory behind the SOM applications in data mining are elaborated based on (Simula et al. 1999). The SOM consists of a regular, usually two-dimensional grid of neurons. Each neuron *i* of the SOM is represented by a weight model vector, $m_i = [m_{i1,\ldots\ldots\ldots,} m_{in}]^T$, where *n* is equal to the dimension of the input vectors. The set of weight vectors is called the codebook.

The map neurons are connected to their adjacent neighbours by a neighbourhood relation, which dictates the topology of the map. Usually a rectangular or hexagonal topology is used. Immediate neighbours belong to the neighbourhood N_i of the neuron *i*.

In the basic SOM algorithm, the topological relations and the number of neurons are fixed from the beginning. The number of neurons may vary from a few dozens up to several thousands. It determines the granularity of the mapping, which in turn affects the accuracy and generalisation capacity of the SOM. During an iterative training, the SOM forms an elastic net that folds onto the 'cloud' formed by the input data. The net tends to approximate the probability density of the data; the codebook vectors tend to drift to places where the data is dense, while there would be only a few codebook vectors in places where data is sparse.

At each training step, one sample vector x is randomly chosen from the input data set and the distances (such as the similarities) between the vector x and all codebook vectors are computed. The best matching unit (BMU) denoted here by c, would be the map unit whose weight vector is closest to x: $||x - m_c|| = \min_i \{||x - m_i||\}$

After finding the BMU, the weight vectors are updated. The BMU and its topological neighbours are moved closer to the input vector in the input space. The update rule for the weight vector of unit *i* is:

$$m_i(t+1) = \begin{cases} m_i(t) + \alpha(t)\,[x(t) - m_i(t)], & i \in N_c(t) \\ m_i(t), & i \notin N_c(t) \end{cases}$$

where *t* denotes time. $N_c(t)$ is the non-increasing neighbourhood function around the winner unit c and $0 < \alpha(t) < 1$ is a learning coefficient, a decreasing function of time.

The SOM algorithm performs a topology-preserving mapping, during which the multidimensional input space is transformed onto units in a two-dimensional map, preserving the relative distances between the data points. Data points lying near each other in the input space would be mapped onto nearby map units, thus making the SOM a powerful clustering tool of multidimensional data sets. SOMs are able to generalise data, for example a SOM network can interpolate between the previously encountered inputs. The quality of the mapping is usually determined by the following factors:

(i) Precision: measured by using average quantisation error, which is the average distance between input vectors of the testing set and the corresponding BMUs.

(ii) Topology preservation: several studies have been undertaken on different topology measures (Kaski and Lagus 1996; Kiviluoto 1996). The latter suggested a method called a goodness meter, to measure both precision and topology at the same time.

The SOM can be interpreted by naming the units according to the input vectors whose type or class is known and this is shown in figure 2.3. The labelling gives physical interpretation of the network. If labelled vectors are not available, the map can be interpreted by direct inspection of the weight vectors and clusters on the map by using different visualisation techniques. Automatic interpretation of the map as well is possible with the use of fuzzy rules (Simula et al. 1999).

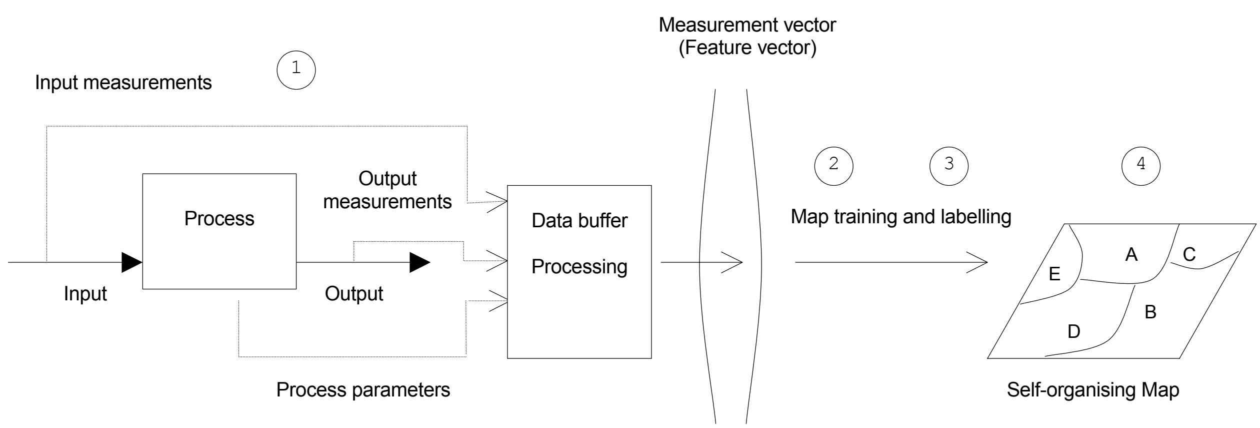
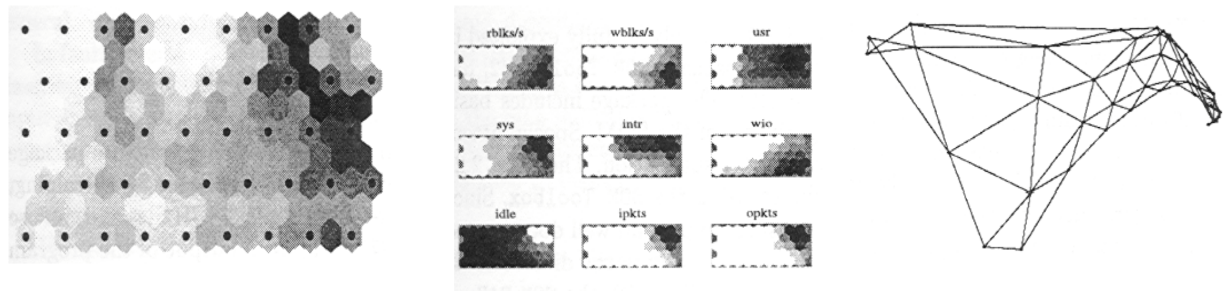


Figure 2.3: A schematic illustration of a SOM application to an industrial process (1) data processing (acquisition, preprocessing, feature extraction, and normalisation), (2) map training, (3) validation and interpretation, (4) visualisation. Source: (Simula et al. 1999).



(i) Unified distance matrix (u-matrix) (ii) Component plane representation (iii) Sammon's mapping

(iv) Data histogram on top of the component plane (v) trajectory on top of the labelled u-matrix

Figure 2.4: Different visualisation of SOM. Source: (Simula et al. 1999).

SOM approaches to visualising a computer system in a network environment using data on utilisation rates of the central processing unit (CPU) and the network traffic volumes are elaborated upon. These visualisation approaches are investigated for integrated analysis of ecosystem variables and economic system data to study the effects of urbanisation on natural habitats and for economic trade-off analysis of developmental activities with a systems approach, in case study chapters 4, 5 and 6:

(i) Unified distance matrix (u-matrix): The u-matrix method (Ultsch 1990) used in figure 2.4 (i), projects the SOM structure on a two-dimensional display. The distance and difference in grey colour represent the variations among map units. The greater the difference in colour the more the distance between the nodes, for instance, the large uniform area on the left of this SOM map, corresponds to an idle state of the computer system.

(ii) Component plane representation: The visualisation approach of figure 2.4 (ii), using component planes enhances the investigation of parameter variations of VLSI circuit (Tryba et al. 1989). By studying the component planes, it is possible to analyse the relative component values of the weight vectors that contribute to the final SOM, which shows the final outcome of CPU usage. This is applied to analyse ecosystem structure, functioning and dynamics by studying the relationships between the environmental parameters and biological assemblage dynamics and to establish the link between the causal processes and their environmental effects as well.

(iii) Sammon's mapping: They can be directly applied to data sets. However, with large data sets the algorithm becomes computationally intensive, making the approach ineffective. With the use of a SOM this limitation could be overcome by quantising the input data to a smaller number of weight vectors whereby, the computation load could be reduced to acceptable levels. Connecting the neighbouring map units in the SOM map enhances the net like structure (figure 2.4 (iii)). Sammon's mappings, based on an iterative algorithm can be used to project high dimensional vectors onto two-dimensional displays. It is a non-linear mapping of data sets that preserves the relative distances between the input vectors.

(iv) Data histogram: A data histogram on a SOM map can be created with a trained SOM and a data set. On the SOM map, when the BMU is determined to each of the data vector, the 'hit counter' for the unit is increased by one. The data histogram of the computer example used in the u-matrix SOM of figure 2.4 (i) is shown in figure 2.4 (iv).

(v) Operating point and trajectory: A trajectory of data points is very useful in tracking the progress of a process in time. The current point of the process is shown as the BMU of the current measurement vector in the SOM map. By watching the progress of the current point in time, its entry towards undesirable areas could be detected in advance. The trajectory of the computer example from normal operation area to disk intensive phase and then to high load area is shown in figure 2.4

(Simula et al. 1999).

A SOM (figure 2.5), created with the measurements of incoming raw material characteristics and process parameter settings, was used to predict the output quality of the manufacturing system process (Hollmen and Simula 1996; Simula et al. 1999).

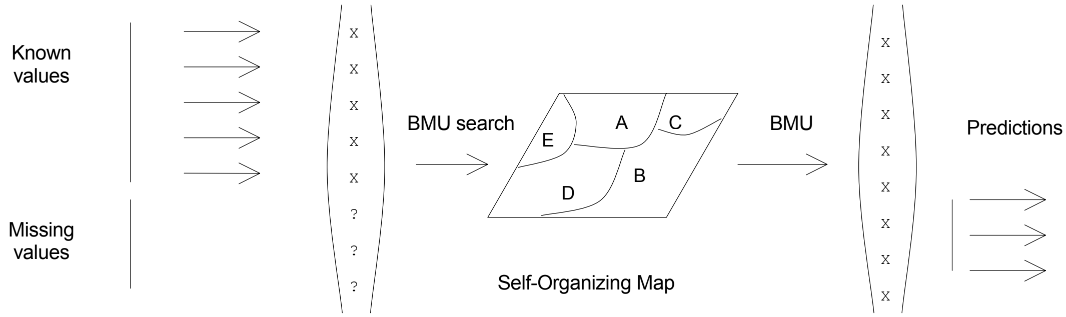


Figure 2.5: Prediction of missing components of the input vector. Source: (Simula et al. 1999:9).

General regression of y on x is usually defined by $y = E(y|\mathrm{x})$ (expectation of the output y stated in the input vector x). To motivate the use of SOM for regression, the codebook vectors represent the local averages of the training data. Regression is achieved by searching for the BMU using the known vector components of x. During the operation the output and approximation of the unknown components of the codebook vector are produced. The approach is applied to predicting species assemblages for particular streams/ freshwater bodies of known altitudes in MOPED (see chapter 3). The system generalises values for species from the vectors of known water bodies.

In industrial process modelling applications, model accuracy could be increased by building local models for the data in the Voronoi sets of the SOM. This method is based on 'divide and conquer' concept, where the input data is divided into subsets (containing points that are nearer to each other in the data space), each of the subset being modelled with an independent local model. This is considered only for simple local linear models and successfully applied in model fitting using Principal Component Analysis (PCA) to promoter recognition issues in DNA analysis (Bajic and Bajic 1999).

SOMs are usefully applied to sensitivity analysis[14] of industrial systems. They are useful in studying the implications of small changes on the whole system. It is generally difficult to observe the effects on the wider system, caused by a small change in one of the system parameters/ components. The observation becomes extremely difficult or often impossible with the presence of noise in the measurements and operating conditions, especially in industrial systems and even more complex in natural systems

[14] Sensitivity analysis means the observation of the whole systems changes, arising from a small change in one or more of system components.

that consists of slow, subdue reactions and compensating mechanisms (Clark et al. 2001). SOM methods are successfully applied to an manufacturing system, where the state of the process was moved towards the desired direction to achieve better quality (Simula et al. 1999). Based on this concept, SOM are applied to natural systems in case study chapters 5, 6 and 7.

Conventional analytical methods cannot be used to evaluate the outcome of a small change on a space with wider ranges; however, SOMs are proven to be successful in this regard (Simula et al. 1999). When a single value in a set of parameters, changes, its BMU in the SOM also changes and by tracking this trend, the final results of the small change can be tracked and even used for optimisation of the operation to improve the quality with minimum cost. This is illustrated in figure 2.6, using a two-dimensional SOM map trained with data originating from a three-dimensional measurement space. When a small change in one of the measurements is imposed, the BMU changes to another map unit. By tracking the change of the BMU caused by a system parameter change, the mutual non-linear dependence of the parameter can be revealed. This is possible only by limiting the space defined by the measurements to a selected characteristic behaviour of the whole system. The approach is applied to track ecosystem dynamics and is explained in chapter 4.

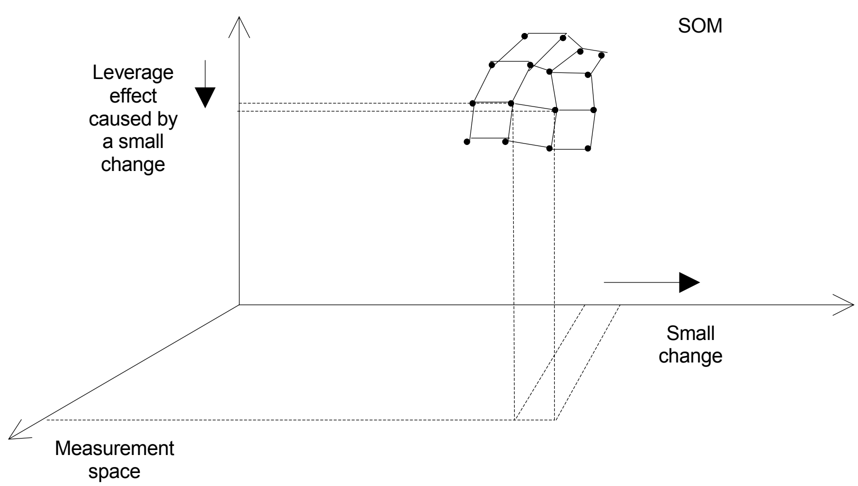


Figure 2.6: A small change along one measurement axis causes a change in the other process parameters. Source: (Simula et al. 1999:11).

In this research, the SOM based sensitivity analysis concept of manufacturing systems is applied to modelling ecosystem functioning and dynamics. In theory the sensitivity analysis concept could be applied to analyse and predict the biological community changes against an environmental parameter (see chapter 4 experimental methodology).

Summary

SOM approaches and their application to real world problems were examined. Biologically inspired ANNs enabled a major breakthrough in introducing heuristics into information processing methodologies with provisions for novel approaches, paradigms and applications.

SOM applications in industrial system process dynamics and financial data analysis were briefly described as they are investigated for cryptic ecosystem modelling in this research. Further details of SOM approaches to real world problems will be discussed in chapter 4. The next chapter reviews the current approaches in ecosystem modelling and the urgent need for better techniques for this purpose. World population growth rates are extending the pressure on the already declining global ecosystems thus increasing the need to preserve them for future generations.

Chapter 3

Ecological modelling: A review

The previous chapter outlined the SOM techniques and their applications to real world problems. In this chapter, the quest for changes in resource management practices along with contemporary ecosystem modelling techniques and biomonitoring concepts reviewed for this research are discussed. The earlier studies on SOM based approaches to modelling the biological assemblages of freshwater and forest systems are elaborated. The main aim of the research is to investigate potential SOM applications to modelling natural habitats in order to bridge the gap in integrated ecological data analysis and this is carried out, based on how industrial process engineers and financial analysts use SOMs in their own disciplines.

3.1 Need for changes in environmental management

Forceful moves by environmentalists from around the world, also supported by the scientific community, to protect our global ecosystem, drive (Hammond et al. 1995a) for a transition in many aspects of human-environment relationship (Reid 2000). The rapid scientific advances achieved in terms of computing power, molecular biology and new techniques to sense biological, physical and chemical phenomena below, on and above the Earth's surface have not been effective, either in warning people of major environmental changes or how human should respond to an environment that is under threat (Graedel et al. 2001). In that context, redirection of research efforts, development of new models and techniques, radical changes in resource management practices and better co-ordination among the professionals involved and the general public, are the key areas identified for human activity on the environment to be sustainable by ecosystem functioning and biodiversity (Harris 2002).

3.1.1 Redirection in future research efforts

Improving our understanding on the human-environment relationship, is vital for the design and management of natural habitats that could support both human uses and natural biota, argued (Graedel et al. 2001). Further identified eight areas as important for future research in this regard. One among them was biological diversity and ecosystem functioning, for which the following recommendation was made for immediate research.

"Develop a comprehensive understanding of the relationship between ecosystem structure and functioning and biological diversity. This initiative would include experiments, observations and theory and should have two interrelated foci: (a) developing the scientific knowledge needed to enable the design and management of habitats that can support both human uses and native biota; and (b) developing a detailed understanding of the effects of habitat alteration and loss on biological diversity, especially those species and ecosystem whose disappearance would likely do disproportionate harm to the ability of ecosystem to meet human needs or set in motion the extinction of many other species." (Graedel et al. 2001:5).

Buckeridge (1994) and Reid (2000) also emphasised the need for a systems/ holistic approach in resource management decisions as opposed to the twentieth century research efforts that became more focused in gaining in-depth knowledge with highly specialised scientific fields (Bowler 1992). Human activities resulting from such narrowly focused research and knowledge are accounted even for altering the Earth's basic chemical cycles (Kirby 2000), leading to global ecosystem failure in its natural functioning that supports the continued existence of humans and other living beings on this Earth (Harris 2002).

Suggestions for the implementation of better planning and decision making based on reliable forecasts on ecosystem state and functioning in order to achieve sustainable environment management has become the theme in recent times. Clark et al. (2001) took a different perspective, in that they argued that the current environmental issues[15], resulting from sectoral resource management that pose unprecedented threats to human civilisation, all of which would have been avoided if the decisions in the past had been based on ecosystem forecasts. Furthermore, suggested the following, aimed at developing co-ordinated working practices based on interdisciplinary linkages;

a. availability of new data sets,
b. together with progress in computation and statistics with increased capacity to forecast ecosystem change,

[15] Environmental issues such as rapid change in climate and chemical cycles, depletion of the natural resources that support regional economies, proliferation of exotic species, spread of disease and deterioration of air, waters and soils

c. approaches for better communication between scientists and decision makers to enforce ecosystem management based on continuos reliable forecasting,
d. suggestions for trade-offs and alternative options, and
e. evaluation of feedbacks.

Changes aimed at developing better co-ordination and, more importantly, better communication among a range of professionals such as scientists, stakeholders and the community in large, play a critical role in solving the growing global environmental issues (Harris 2002). Scientists' predictions that are based on highly complicated principles and hypotheses are generally found to be beyond comprehension by individuals from different professions. This has been a common practice for a long time now. Matters got further complicated with scientists becoming more focused on science and research, being isolated from the rest of the world especially, from the general public, because of their poor communication abilities (Buckeridge 2001). The knowledge divide between these two has resulted in valuable scientific predictions being ignored by stakeholders and the general public (Clark et al. 2001). Vant (1999), Reid (2000), Clark et al. (2001) and Harris (2002) stressed the need for new modelling tools with an integrated approach to create trust between the participants. All these publications describe the need for new models, depicting disparate data (from different disciplines) to understand and predict complex ecosystem behaviour in response to human and other natural/global causes. The need for new approaches is seen as 'significantly critical' for the conservation of our global ecosystem.

3.1.2 Interdisciplinary approaches

The importance of integrated, interdisciplinary environmental research and approaches to improve ecosystem understanding of natural habitats has been stressed since the 1980s, as would be seen in (Mann 1982; Soule and Kleppel 1988; Graedel et al. 2001); Parker et al. 2001; Harris 2002)). In consideration of the then modern analytical technological improvements, Soule and Kleppel (1988) looked into the prospects of bringing scientists together from a broad spectrum of disciplines, such as computing, physical, chemical, biological oceanography, marine ecology and environmental sciences. Further argued that such interdisciplinary research based approaches as vital for the use of indicator species[16] in studying the environmental pollution resulting from anthropogenic causes. This aspect is stressed in the later studies as well. It appears then, that the problems in introducing interdisciplinary approaches to ecosystem modelling have remained the same since the early 1980s until recent times.

With regards to the current quest, Graedel et al. (2001) described the tasks ahead in developing interdisciplinary approaches, as the 'Grand challenges of this era' and highlighted the following four, among the eight areas of 'highest priority' for future research.

"…

- *Biological Diversity and Ecosystem Functioning:* an initiative to develop a comprehensive understanding of the factors that generate, maintain, and diminish biological diversity and their effects on ecosystem functioning.
- *Hydraulic Forecasting:* an initiative to develop a comprehensive hydrological forecasting, specifically including the ecological consequences of changing water regimes.
- *Infectious Disease and the Environment:* an initiative to develop a comprehensive ecological and evolutionary understanding of infectious and environmental diseases.
- *Land-Use Dynamics:* an initiative to develop a systematic, spatially explicit understanding of the changes in land use and land cover that are critical to ecosystem functioning, ecosystem services, and human welfare…"

(Graedel et al. 2001:60)

In consideration of these key issues, data needs within each area and co-ordination with other environmental science research, are emphasised.

"…The overall effort will require interdisciplinary research involving ecologists, ethologists, psychologists, engineers, economists, planners, landscape architects and others. The definitions of data needs and the collection and synthesis of data will require cooperation among physical, biological, and social scientists; engineers and planners; and other associated funding agencies." (Graedel et al. 2001:62).

---

[16] Indicator species/ communities are used in biomonitoring methods to analyse the magnetite and extent of pollution in a particular environment (see section 3.2 biomonitoring).

The following are the areas suggested for co-ordinated research along with the efforts of environmental science, to gain more understating on the controls and means to protect biological diversity:

(i) Co-ordinate research on hydraulic modelling on runoff and subsurface water, which reflect the way living beings, humans inclusive, interact with the landscape.

(ii) Include the effects of human management institutions on ecosystems.

(iii) Include the effects of changing patterns of land use and land cover on potential for habitat redesign.

(iv) Include the effects of climate change in ecosystem functioning assessments and in habitat design to buffer for disturbances and extreme events.

(v) Create partnerships and work with urban long-term ecological research sites.

The scientific publications so far discussed illustrate the future direction to support and inform sustainable environment management by evaluating the state of the environment and its response at different scales and levels. This is increasingly seen as the greatest ever challenge in human history. The existing methods for ecosystem modelling are inadequate, in that they are not even useful in portraying the magnitude of current global environmental issues, as required for the design and management of natural systems. In the next section, sectoral resource management approaches and their consequences on the environment are discussed.

3.1.3 Sectoral approaches of ecosystem management

The conventional approaches of sectoral management, such as reactive resource management practices that consider only the socio-economic benefits of developmental activities, have led to unprecedented consequences on our global ecosystem (Hammond et al. 1995a). Continued use of such approaches will only worsen the situation as the world population growth increases at alarming proportions, the demand on the already declining ecosystem functioning and biological diversity also will increase (Reid 2000).

(Reid 2000:1) described the resource management actions, solely aimed at addressing the social and economic outcome as "… examples abound of vast and uncontrolled ecosystem 'experiments'…". One among these many 'experiments' illustrated in the article, is the state of the native fish, in Lake Victoria, Africa. In this case, the newly introduced species, initially found to be succeeding beyond expectations with a dramatic growth in fish harvests, ultimately became a threat to the fish productivity as well as the native species of the area. The increased harvest of exotic species initially increased land use changes on the limited forest resources surrounding the lake, which in turn increased lake siltation and pollution. This reduced the fish production in the area. In the end Nile perch, a prolific, non-native species also led to the extinction of 350 native species, or reduced them to a fraction of their original size.

A few single need (or secctoral) resource management solutions and their side effects observed world wide are outlined below based upon (Reid 2000);

(i) The expansion of agricultural land into natural habitats around the world increased the food production but also changed the quantity and quality of freshwater runoff tremendously.

(ii) The use of modern fertilisers, as expected increased yield but also caused eutrophication of nearby rivers and estuaries and was found to be responsible for anoxic 'dead zones' seen in coastal areas near major agricultural river basins.

(iii) Timber harvest and the transformation of forestland to agriculture helped many states to meet their needs for food and fibre but also released carbon into the atmosphere that changed the Earth's surface reflectivity, contributing significantly to the risk of global climate change.

Efforts made by the New Zealand and Australian governments and scientists to introduce new policies and methods to protect the environment are elaborated in 3.2.3 New approaches for developing indicators; New Zealand's perspective.

So far in this chapter the prevalent influence of scientists and concerned national and international bodies for changes in environmental management and how this occurred, have been discussed. In the next section, biomonitoring concepts used to study anthropogenic environmental pollution are discussed. Biomonitoring concepts have long been used in ecological modelling despite the controversies in their ability to detect the exact cause of an environmental pollution.

3.2 Biomonitoring

Biomonitoring is the use of biological responses to assess changes in the environment; generally that are anthropogenic. It mainly involves indicator[17] species or indicator communities that accumulate pollutants in their tissues from the surrounding environment and thus reflect the environmental conditions. The extent of environmental effects within an ecosystem on individual organisms, species and communities provides information on the magnitude and ecological effects of the pollutants on the ecosystem despite the expensive processes involved in the initial development of such indicators. Basically there are two types of biomonitoring:

(i) to observe the biological system changes before and after a project is completed or before and after a toxic substance enters the water.

(ii) to ensure compliance with regulations or guidelines or to ensure water quality is maintained (Biological Monitoring 2000).

In the second type, the transformation of biomonitoring data into useful information for decision making has long been a challenge, the key factors for this being:

(i) the identification of robust methods for summarising lots of data, without over-simplification and

(ii) the presentation of the resulting information in an understandable and attractive way to a largely non-technical audience (Vant 1999).

3.2.1 Biomonitoring in environmental pollution studies

A hypothetical model (figure 3.1) (Sastry and Miller 1980), illustrates the time related sequence of possible effects of reduced water quality either pollutant induced or due to natural causes at various levels of biological organisation. The model does not give detailed descriptions of the states in the sequence, however, gives an outline of the effects of different biological organisation.

---

[17] An indicator is a pointer that provides a clue, or a means of measure, to a much more significant event or trend or progress of a more complex set of actions. In ecology, an indicator is a plant or an animal or the whole community, whose existence in an area is strongly indicative of specific environmental conditions. Thus an indicator is significant in extending beyond that which is actually measured to a larger phenomenon of interest (Hammond, A., A. Adriaanse, E. Rodenburg, D. Bryant and R. Woodward 1995; On-line dictionary (2002).

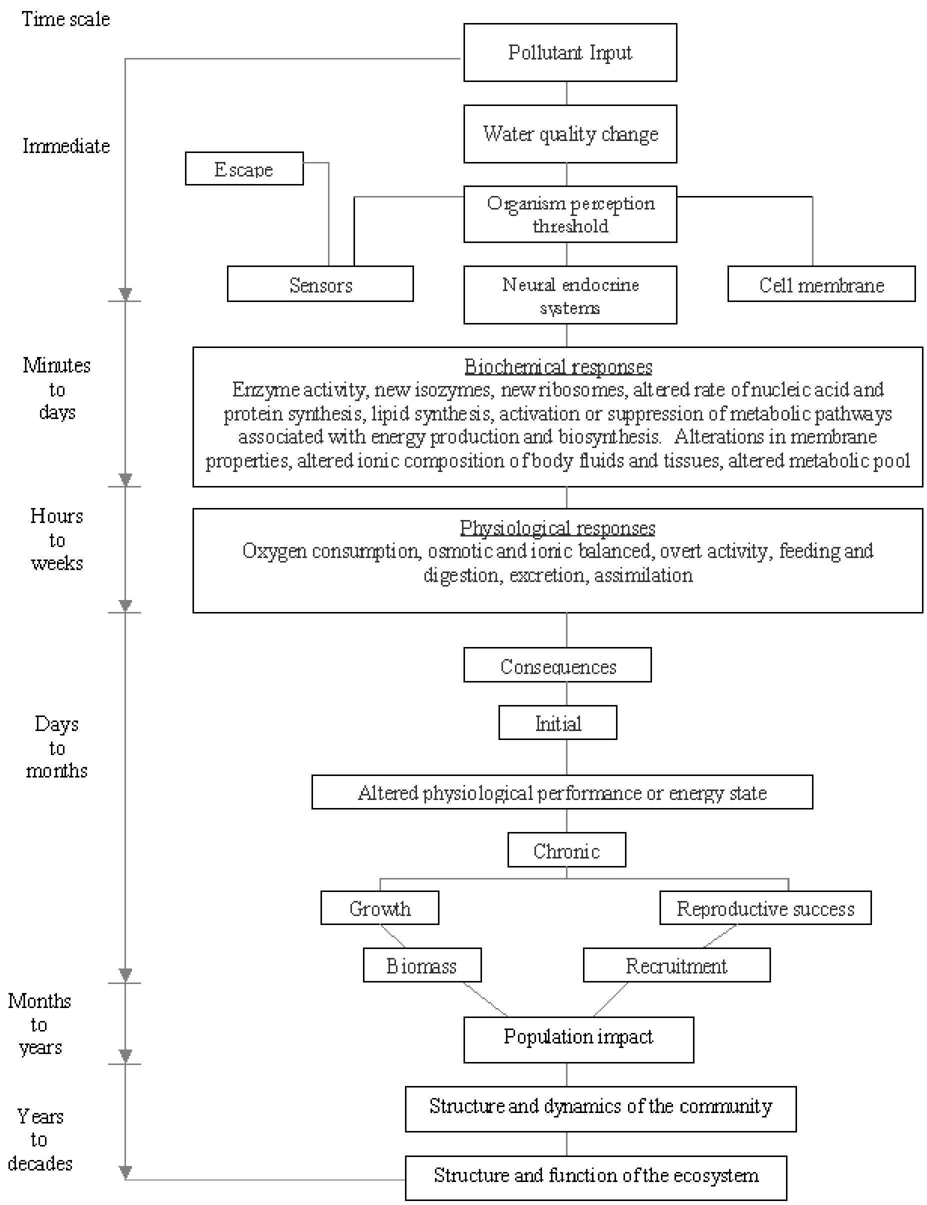


Figure 3.1: A hypothetical model of time related sequence of possible biological effects of reduced water quality. Source: (Sastry and Miller 1980:267).

Even with significant development and broad use of biomonitoring models, in order to study the environmental conditions, a symposium held (in the late 1980s) on this subject for the Southern California Academy of Sciences dealt with a number of issues. Many questions were raised on the concept and the 'rules' administered for appropriate use of indicator organisms, either based on single species or community indicators, to particular problems:

(i) How would the true indicators be distinguished from biological anomalies?

(ii) How would one select the kinds of organisms that would be appropriately associated with conditions and events at various scales in time and space?
(iii) To what extent would any one particular species represent other species in the same environmental setting?
(iv) Could the indicator concept be applied to modern sampling and analytical technology?
(v) How would the anthropogenic disturbances be distinguished from those of natural phenomena?
(vi) How would the differing databases be best matched with deferring scales?

(Soule and Kleppel 1988).

This illustrates the major problem identified in pollution monitoring methods using indicator organisms of the 1980s. An inability to distinguish the exact cause for the observed biological responses, either arising from environmental or from natural causes was a major concern (Soule and Kleppel 1988). More than two decades on, the situation remains the same and is discussed below.

A website, posted by the National Center for Environmental Research (NCER) Office of Research and Development (EPA, United States), consists of a list of science questions and issues, addressed by the institute researchers on the use of ecological assessment and indicators:

"…How Can We Identify and Develop Molecular and Cellular Indicators for Monitoring and Assessing Changes in Genetic Diversity in Response to Environmental Stress?
How Can We Relate Indicators of Population and Community Structure and Function to Exposure to Chemical, Physical and Biological Stressors?
How Can We Assess Ecological Condition Through Chemical Indicators?
How Can We Use Remote Sensing Techniques to Develop Landscape Indicators that Quantify and Characterize the Geographic Extent of Key Attributes as They Relate to a Range of Environmental Values?
How Can We Assess Ecological Condition Using Indicators that Incorporate Multiple Resources and Spatial Scales? …"

(National Center for Environmental Research (NCER) Office of Research and Development (ORD) - US Environmental Protection Agency (EPA) 2000:1)

Funded through this programme, many academic institutions have been conducting research in a wide spectrum of disciplines with an aim of developing integrated approaches for ecosystem modelling. The research effects are focused on developing a new suite of environmental integrative indicators for use in estuarine and other long-term environmental monitoring programs, span from DNA analysis to integrated ecosystem modelling techniques with interfaces to geographical information systems (GISs) (http://es.epa.gov/ncer/starreport.html).

Despite all this investment made by industry, government and academia in reviewing, debating and complying with countless number of environmental regulations, ecologists are still unable to exactly notify the consequences of human induced impact particularly, in marine habitats (Osenberg and Schmitt 1996). The uncertainties on environmental response to various causes remain the same; except for some changes in the terminology. In fact, the more we learn about the ecosystems the greater the complexity revealed despite the advances of science and technology. The major issues for this could be

(i) the inherent complexities of ecosystem structure and functioning (Clark et al. 2001), and
(ii) the twentieth century approaches, based on gaining in-depth knowledge that led to decisions with a fragmented image of nature (Bowler 1992).
(iii) the sectoral resource management approaches that looked only into the social and economic aspects of developmental activities without any concern for the environment and its consequences on biodiversity (Reid 2000).

These considerations will be further elaborated upon in section 3.4 Ecosystem modelling.

3.2.2 Developments in biomonitoring

Biomonitoring that involves the use of biological responses could be classified into two broad categories; either based on the use of indicator species or indicator communities. Monitoring methods of both categories can be carried out at different levels, such as macromolecular, cellular, using organs, organisms, population and biogeocenosis[18]. The

[18] "A combination on a specific area of the Earth's surface of atmosphere, mineral strata, soil, vegetation, animal and microbial life, water-possessing its own specific type of interactions of these components and interchange of their matter and energy among themselves and other natural phenomena…" (Mackey 2003:1).

advanced indicator species methods enable ecologists to analyse biochemical, histological, morphological and physiological changes in individual organisms, however, are successful only in certain specific organisms, such as filter feeders, clams and mussels (Biological Monitoring 2000). The changes observed through these methodologies are then related to the identified environmental stressors, as explained in Sastry's (1980) model; thus using these organisms as true indicators or biomonitoring devices of environmental changes. Such organisms are sometimes referred to as sentinel organisms.

The following is a modern model that provides an overview on the levels of organisation and their associated biomonitoring measures:

(i) Individual - Organism - genetic mutations - reproductive success - physiology - metabolism - oxygen consumption, photosynthesis rate - enzyme/protein activation/inhibition - hormones - growth and development - disease resistance - tissue/organ damage - bioaccumulation

(ii) Population - survival/mortality - sex ratio - abundance/biomass - behaviour (migration) - predation rates - population decline/increase

(iii) Community - Abundance ("evenness") of an organism or organisms - Biomass - Density of an organism or organisms - Richness (variety) - number of species, size classes, or other functional groups, per unit area or volume, or per number of individuals. - Diversity - the richness given the relative abundance of each species or group.

(iv) Ecosystem - Mass balance of nutrients.

(Adams and Brandt, 1990) from (Biological Monitoring 2000).

The next selection looks into the biomonitoring concepts, in New Zealand's perspective.

3.2.3 New approaches for developing indicators; New Zealand's perspective

Efforts by the New Zealand government (ministries, city and regional councils) as well as academic institutions to develop environmental indicators have been significant, in effect both have made considerable progress in introducing a systems approach to preserve natural habitats.

Ministry for the Environment: The Ministry for Environment (MfE) takes responsibility in reporting on the state of New Zealand environment, providing advice to the

Government on practical implications of environmental laws and policies, and also on environmental implications arising from other Government polices. The Ministry is also responsible for initiating any necessary actions for improvements in environmental management. However, most of the responsibilities on the day-to-day environmental management rely on the local government, such as the regional councils. The significant areas of policy on which MfE is responsible for resource management include land, air and water quality; waste, hazardous substances and contaminated sites; protection of the ozone layer; and climate change. The Ministry also contributes towards interdepartmental work on biological diversity, marine environmental issues, energy and transport. It consults with the local government, resource users, resources managers and the others, likely to be affected by changes in policy or legislation and provides information, advice and assistance in this regard (Ministry for the Environment 2002).

Recent efforts by MfE with an aim of developing a set of indicators for its Environmental Performance Indicators programme for implementation in New Zealand are outlined in (Ministry for the Environment 2002; Ministry for the Environment 2002; Ministry for the Environment 2002). The Ministry is working on developing a set of indicators based on a pressure-state-response (PSR) model with cause-effects-social response logic, proposed by the Organisation for Economic Corporation and Development (OECD) (Chapman 1999). This PSR model approach has been adopted by many international institutions (see section 3.5 Contemporary ecosystem modelling techniques).

Department of Conservation: The Department of Conservation (DoC) is responsible for implementing policies for the conservation of the natural and historic heritage of New Zealand. The following statement covers the areas declared as 'conservation land':
"…The Department manages or administers on behalf of New Zealanders:
national parks and forest parks
reserves and conservation areas
protected indigenous forests
protected inland waters and wild and scenic rivers
indigenous/native wildlife
non-commercial freshwater fisheries
historic places on conservation land

marine reserves and protecting marine mammals offshore islands set aside for conservation." (Department of Conservation 2002:1).

These listed categories cover only a third of the main land of this country thus the rest is conserved by the Resource Management Act 1991. Most of the 'conservation land' is steep, inaccessible, mountainous and climatically harsh and does not entirely represent the major ecosystems of New Zealand. Nonetheless the Department's mandate is to preserve the whole country's natural and historic resources on and off conservation land; hence the conservation of private land is facilitated through the Resource Management Act 1991. Through the statutory planning processes under this Act, the regional and territorial local authorities administer the private land. The main purpose of this Act is to promote sustainable management of natural and physical resources. The Act outlines procedures for the recognition and protection of natural and historic values in the preparation of policies and plans by the councils (Department of Conservation 2002).

Meanwhile ecologists and academic researchers also have made significant attempts in developing effective and efficient environmental indicators for sustainable ecosystem management. The new criteria show a shift, in that they place more emphasis towards the protection of ecosystems (Norris 1999). The following guidelines (consisting of three key elements) are set out for developing useful indicators to assess the environmental conditions in New Zealand and Australia; an indicator should,

(i) be more ecosystem specific, rather than focusing on human health as in the past,
(ii) be more focused on the actual issues or problems caused by physical, chemical and biological stressors rather than on individual indicators and
(iii) be risk based.

The new approach aims to develop guideline 'packages' with key performance indicators and trigger levels in these indicators, for each issue and wherever possible for each ecosystem type.

The following are the requirements suggested for an effective environmental indicator based upon (Norris 1999); it should,

(i) be able to quantify and simplify the complex phenomena,

(ii) have scientific validity, operational at appropriate geographic scales,
(iii) be able to respond predictably to environmental changes,
(iv) be easy to understand and
(v) be forward looking or predictive.

None of the conventional methods are able to precisely define the ecosystem response to human activities in a simple, easily understandable format owing to lack of good design with clearly set objectives (Norris 1999). Mann (1982) was correct; the models of the 1980s gave ecologists insights into ecosystem complexities, but were not able to predict ecosystem behaviour as needed for resource management purposes. Even after two decades, the issues remain the same. This shows that the search for ideal environmental indicators that could point out a more complex situation as seen in other fields, such as GDP, still continues.

With that introduction to biomonitoring methods and issues in analysing human induced environmental impact on natural habitats, in the next section traditional data analysis methods, used by local and international researchers to study environmental and biological systems separately or together, are elaborated upon.

3.3 Statistical methods for biomonitoring data analysis

Traditional statistical methods used for community based multivariate analysis to study the relationships between environmental parameters and benthic organisms of marine habitats are explained, based upon (Smith et al. 1988)

The multivariate analytical techniques simultaneously consider more than one single dependent variable in the analysis, whereas univariate methods consider only a single variable at a time. All methods discussed here are based on a model (figure 3.2) that consists of three main steps:

(i) In the first step the biological data is analysed separately to determine the community patterns in them.
(ii) In the second step the environmental data is prepared in a format suitable for use in the next step.
(iii) In the third and final step, the community patterns determined in the first step are related to the environmental factors.

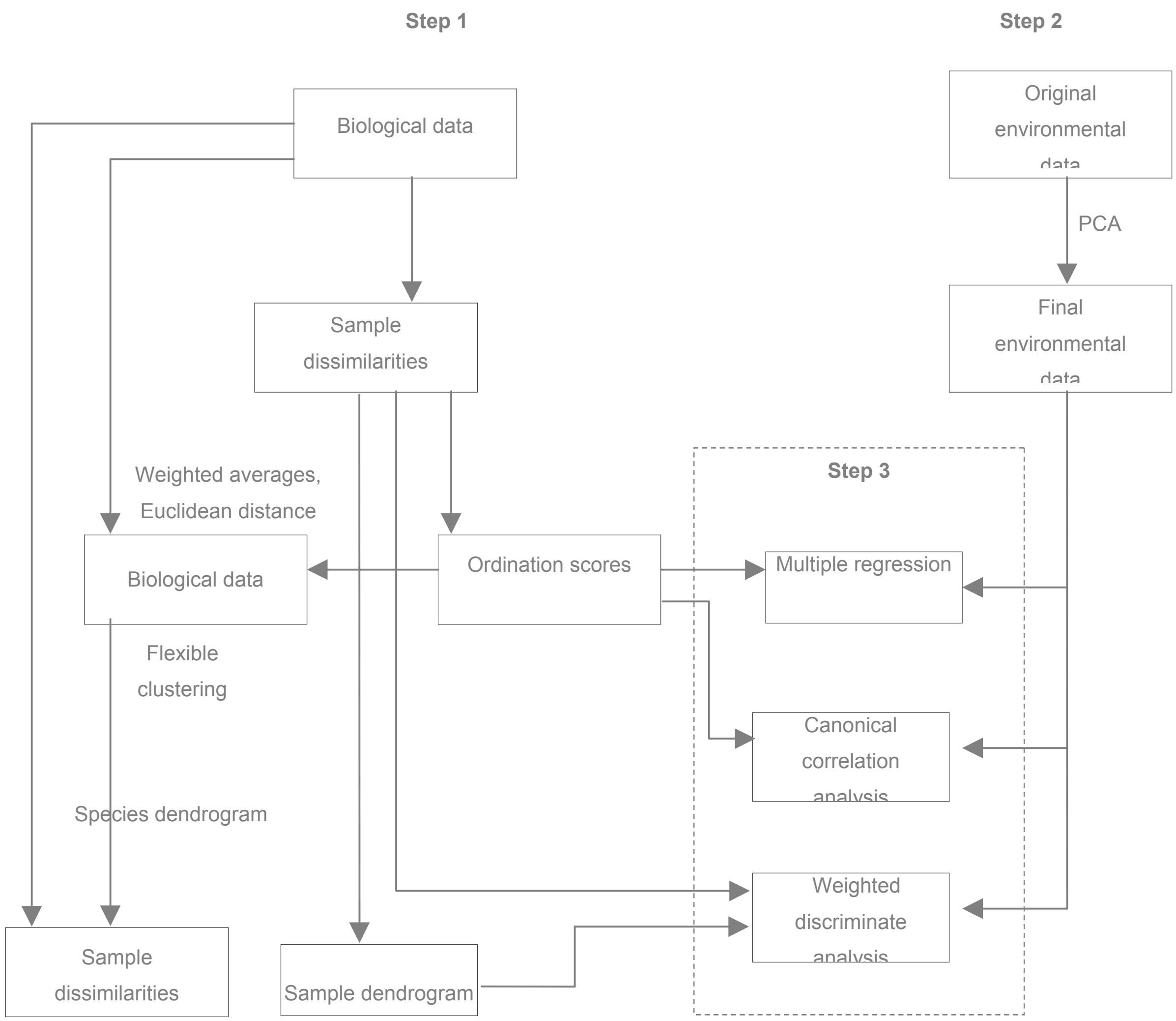


Figure 3.2: Flow diagram showing the interrelationships between the methods used in community biomonitoring of benthic organisms in a marine environment. Source: (Smith et al. 1988:252).

Step 1: Community patterns in the hypothetical biological data of benthic grab samples taken at several points, along a chosen offshore transact at different depths, increased by 10 metres, are determined in this step with the use of dissimilarity index. Indices, such as similarity or dissimilarity, can be used to quantify the community relationships between pairs of samples. A pair of samples, consisting of similar species composition and abundance will be assigned a relatively low dissimilarity value and conversely a relatively high similarity value.

There are some shortcomings in this method; the most important issue is that after a point the dissimilarity values approach an asymptote as the samples being compared show greater amounts of biological change (Smith, 1988). At such a point, the dissimilarity index value reaches the maximum and the biological changes beyond this point could not be measured by any further increase in dissimilarity. The procedures to overcome these issues, such as step-across procedure and ZAD (Smith et al. 1988), only add more complication to the already elaborate operations.

The use of the dissimilarity matrix to determine the community patterns has limited interpretation values. Thus cluster analyses and ordination techniques[19] that utilise the dissimilarity matrix as their starting point, are used to better delineate the community patterns in the data (Smith et al. 1988).

Ordination analyses are generally used to display the biological patterns in a multidimensional space based on the dissimilarity indices. The distance between any two data points would be proportionate to their dissimilarity factor where the axes represent the dimension of the space. The projections of the data points on the axes are scores (Smith et al. 1988).

There are many ordination and classification (or clustering) methods used by ecologists to analyse the correlations in community based biological and environmental data (Palmer 2002; Palmer 2002).

Cluster analysis is considered to be complementing the ordination analysis (Palmer 2002). Among the many cluster analyses, agglomerative hierarchical cluster analysis is often used in benthic ecological studies. It involves successive pairings of the most similar groups of samples based on the dissimilarity matrix. The paring is performed until all samples form into a one large group, generally referred to as a 'dendrogram'. A two-way coincidence table facilitates the task of choosing the groups from a dendrogram.

Step 2: In the next step the environmental data is converted into formats for correlating patterns with the biological data. The multivariate analytical methods used to correlate the environmental and community patterns are data dependent and if used

[19] Ordination techniques are useful operations on community data matrix. Ordination could be defined as the arrangement of species and/or samples along gradients, considering it as a synonym for multivariate gradient analysis (Palmer, M. 2002)

inappropriately, may produce misleading, confusing, unstable or incomputable results (Smith et al. 1988).

Principal component analysis (PCA) is an ordination technique in which a new composite environmental variable is created from the original set of environmental variables. This process is useful in eliminating the problems of data dependent environmental variables. The scores on each PCA axis are new environmental variables. Thus the meaning of these new scores could be studied from the correlations between the original environmental variables and the scores for the axis, and it is important that the new environmental variables from the PCA are interpretable.

Step 3: In this step, as environmental gradients are considered to cause changes in the community, the community patterns expressed by the scores on the ordination axes will be correlated with the environmental gradients.

So far the statistical methods used for ecological data analysis were looked at. These multivariate community biomonitoring methods belong to the hypothesis testing class of analysis. Conversely, data mining techniques belong to exploratory data analysis, the exploration of novel patterns and relationships within the variables from raw data and interpreting them. In consideration of the current need for modern tools to gain more insight into ecosystem response with limited knowledge on the physical interactions of natural processes, the latter could be argued as more appropriate for ecological data analysis. The findings of exploratory analyses may then be used in hypothesis postulation and testing.

In the next section some modelling techniques used to analyse environmental and biological data are explained.

3.4 Ecological modelling

Modelling in ecology became popular and widely applied, since the early 1960s, their growth and popularity being coincided with that of the computers (Mann 1982). Old and the current ecological modelling technique could be classified into two major categories: (i) gaining insights and (ii) making quantitative predictions. The major problem encountered in these old and current techniques is that they all either tend to oversimplify (by ignoring even the very important elements) or very often over sophisticate (resulting in bulky and expensive models) (Schellnhuber 1999). This in turn leads to difficulties in developing useful prediction models for resource management purposes and is elaborated herein.

For instance, the numerical simulation model of energy flow, developed to study the ecosystem process changes during the seasonal variations in Nargansett bay coastal system and the like were found to be successful in some aspects (Mann 1982), the advantages being,

(i) even though borrowed from engineering, they had been successful in gaining better insights into the complex workings of natural systems.
(ii) had given an opportunity to test whether a hypothesis was correct or to rule it out.

Disadvantages

(i) even to develop and run such closed, non evolving models for hypothesis testing purposes it required complicated formulae and procedures with every minute detail, without which such systems could not be programmed to run on computers. Unlike the engineering systems (where variations can be measured and worked out with ultimate precision) the natural systems could not be generalised because of the extensive spatial and temporal variations observed within an environment and the non-linear stochastic, species threshold response (Clark et al. 2001) (see chapter 1). These variations make quantitative predictions practically impossible with the existing methods and our limited knowledge on natural system processes.
(ii) they would not be used to predict or answer management questions, involving large, long-term perturbations of human activities on open, evolving systems, as these simulated systems were deterministic, closed and non-evolving.
(iii) certain hypotheses would not be defined mathematically and there was no way of testing them, as these models needed all involved steps defined.

Despite the above model limitations in predicting ecosystem behaviour, a popular class of simulation models was designed and implemented with considerable success. Among them were the simulation models, used in cleaning up of the tidal portion of the River Thames below London (Longhurst 1978.). The objectives of this study were:

(i) to identify sources of pollution in the River Thames;

(ii) to establish the significance and effects of individual pollutants;

(iii) to develop a mixing equation for the estuary;

(iv) to forecast the effect of changes in balance of the system and to indicate management criteria for the stewardship of the river.

The River Thames models, built using a few general concepts, were used to predict the circumstances that would in turn, return the Thames to a well-oxygenated system. The following was the summarised statement on the physical oceanography of the estuary that pertained to the residence time of sewage effluent. "… A particle of matter introduced into the tidal water at London Bridge may flow 16 km downriver on the ebb-tide and return 15 km on the flood and oscillate in this manner for between 6 weeks and 3 months before reaching water where there is a reasonable interflow with the North Sea…" (Mann 1982:269).

Within a period of three years, the river was turned into a well-oxygenated system due to actions taken following recommendations derived from simple modelling. A biological survey carried out in 1957 showed no fish in the tidal reaches for many kilometres below London. Following this, in the early 1960s, appropriate sewage treatment facilities were designed, constructed and brought into use. By 1965 fishes were seen returning and by 1970, over 50 species had returned (mainly marine) in the lower half of the estuary near London. This led analysts to conclude that the model had served its purpose, even though the data collected during the river's recovery period of fish species was different to that of the predictions by the model. The success of the River Thames model has been elaborated in many later studies (Mann1982).

Longhurst (1976) in (Mann1982), argued that despite the oversimplification of biological oxygen in demand (BOD) and dissolved oxygen (DO) interactions on a validated physical time-dependent model and integrated equations for the conservation of volume and materials, the models had been successful in predicting the real situations. They singled out a few easily handled variables that influenced the important properties of the river's ecosystem. Similarly, if such important properties and their indicators would be identified, they could be successfully applied to analysing the long-term effects on an ecosystem and its biodiversity, for sustainable environment management (Mann 1982). Further argued that ecosystem modellers should focus on developing theories and models to establish the connection between the dynamics of populations and the behaviour of ecosystems, similar to that of the statistical mechanics, which provides a common connection between the motion of particles and the behaviour of a gas.

Clark et al. (2001) as well expressed a similar view in that argued that the slow ecosystem responses that are invariably left out in the current ecological modelling techniques should be included. The 'large inherent uncertainty' arising from strong nonlinearities and stochasticity could not be explained as the impact on ecosystem being neutralised or mitigated. Instead, the 'slow variables', that could be significant in ecological processes should be identified and used for modelling ecosystems.

Conventional methods adopted to model the Earth's ecosystems and to predict their system dynamics, reviewed until now, seem to be aimed at handling either the environmental or biological system changes alone. In the next section modern modelling approaches developed to studying ecosystems with a systems approach are looked at.

3.5 Contemporary ecosystem modelling techniques

Scientific publications (Hammond et al. 1995; Hammond et al. 1995; Hammond et al. 1995; Clark et al. 2001; Clark et al. 2001; Harris 2002), reviewed in section 3.1 Need for changes in environmental management, illustrate the phenomenal quest for new approaches with integrated models, to inform sustainable environment management. However, only Hammond et al. (1995) elaborated upon the pressure-state-response (PSR) and information pyramid models, originally developed by the Dutch government to assess the state, pressure and response of an environment. Details on the application of a suite of indicators, within the PSR framework, to ecosystem modelling are discussed herein and in chapter 7, based on a chapter titled 'Environment indicators: a systematic approach for sustainable development (Hammond et al. 1995).

The PSR framework and information pyramid models provide a means to transform primary (raw) data sets into simple and easily interpretable environmental indicators, especially for non-scientific users, such as stakeholders and the general public, whose involvement has been emphasised in the preservation of our global ecosystem.

The PSR model consists of the following:

i) pressure indicators, as indicators of the stress on the environment or natural resources. The indicators are designed to measure the various pressure sources from human activities to discuss why it is happening,

ii) state indicators or indicators of changes, to measure the changes or trends in the physical and biological system state of the natural world. They are developed to discuss matters pertaining to what is happening to the state of the environment or natural resources and

iii) response indicators to measure the responses resulting from the measures and policies adopted to address the environmental issues or to discuss matters pertaining to what we are doing about it and whether our response activities have improved or worsen the issues.

The information pyramid (figure 3.3) consists of highly aggregated indices on the top of the pyramid with the primary data, from monitoring programmes at the bottom. Based on this pyramid structure raw data are transformed into concise indices, provided that knowledge on the physical processes analysed is available. These indices are then used as indicators of the state, pressure and response of ecosystem aspects at many levels, such as community, sectoral and regional, as they are aggregation of raw data. The use of environmental indicators for national and international decision-making procedures is considered as important for the better use and management of our global ecosystem.

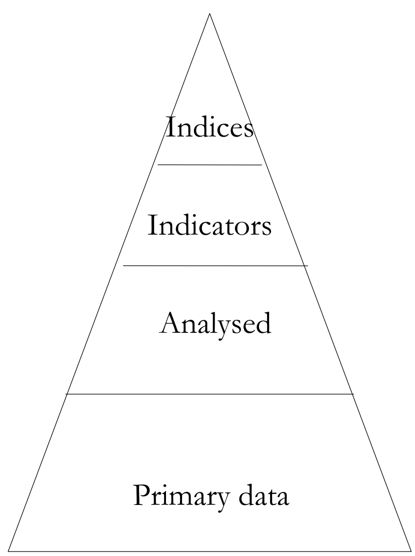


Figure 3.3: The information pyramid. Source: (Hammond et al. 1995:1).

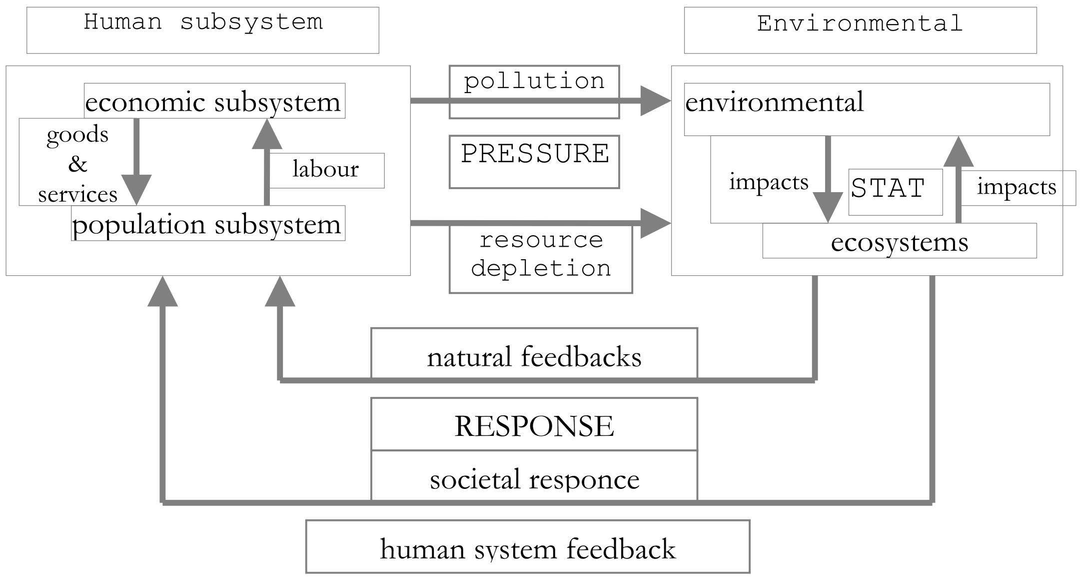


Figure 3.4 The Pressure-State-Response framework indicators (Hammond et al. 1995:11).

Even though the amount of environmental monitoring data has increased over the last six decades, there are wide gaps in them that inhibit the development of an information base argued (Hammond et al. 1995). Further introduced a conceptual model (figure 3.4). to overcome this problem and to direct the development of an information system in easily understandable formats. Such information systems, in their structure consist of a matrix of environmental indicators, developed within the PSR framework. For details on matrices of environmental indicators adopted by WRI and the World Bank, see appendices 3 and 4.

The World Bank's matrix of indicators was developed based on an explicit model (figure 3.5) that illustrates the human interaction with the environment within the PSR framework, recommended by OECD and UNEP.

The issues in the matrix of indicators are classified into four major groups:

i) source indicators
ii) sink of pollution indicators
iii) life support indicators and
iv) human impact indicators.

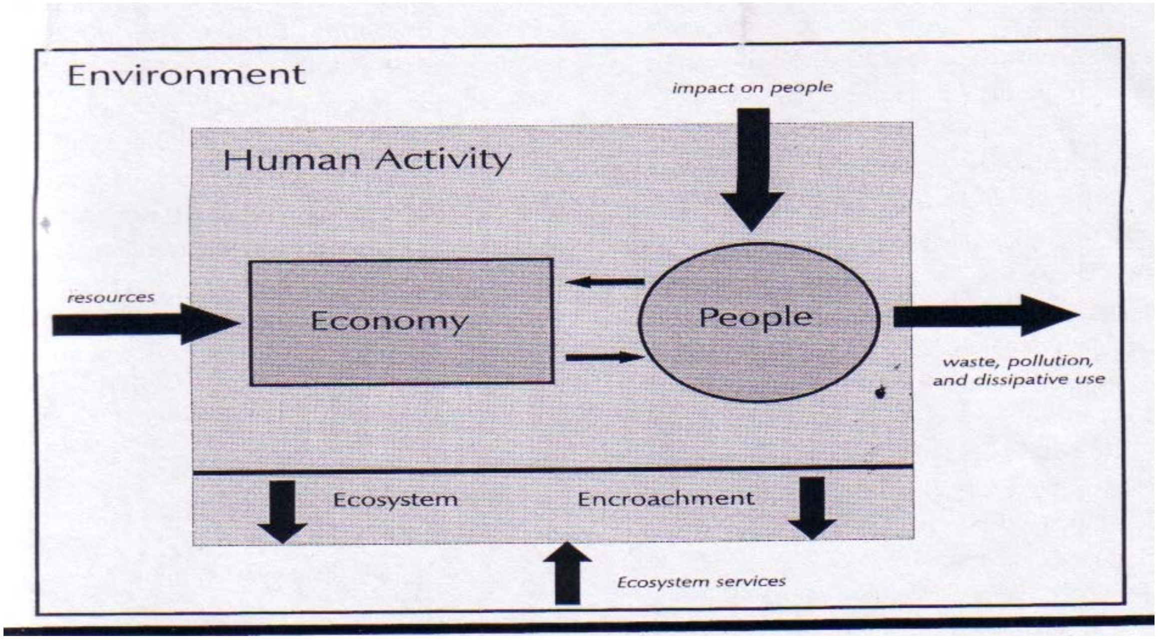


Figure 3.5 A model of human interaction with the environment designed by the World Bank, Source: (Hammond et al. 1995:15).

Life support indicators, the group that is relevant to this research, could be seen classified into further three groups based upon (Hammond et al. 1995); biodiversity, oceans and special lands (such as wetland). However, no aggregation process for the life support indicators, within the PSR framework was set in place, at the time the report was written. This points out the major drawback with this approach, in that the approach cannot be used to develop indices unless sufficient knowledge on the physical processes involved is available. These dubious indicator classes will be elaborated upon in chapter 7. In the next section, literature reviewed on SOM applications in ecological data analysis are elaborated upon.

3.6 Ecological modelling with SOM techniques

MOPED: Modelling Patterns in Environmental Data (MOPED) was developed by NIWA. In MOPED SOM techniques are applied to mapping of patterns within environmental data, such as species distribution and elevation of freshwater bodies. SOMs are found to be successful in predicting the biological assemblages from the available habitat data. They are used to predict the species that should be present in certain streams, when their altitude is presented (Jowett 2001). In this application SOM is used to find the missing values in the input vectors.

In (Ce´re´ghino et al. 2001), SOMs are successfully applied to clustering the community patterning in the regional distribution of 283 lotic macroinvertebrate species within data consisting of four insect orders (Ephemeroptera, Plecoptera, Trichoptera, Coleoptera = EPTC) from the Adour-Garonne drainage basin of Southwest France, covering an area of 116,000 km^2). The aim of the research was to provide a stream classification based on characteristic species assemblages using the occurrence of these species at 252 sampling sites. SOMs were found to be successful in projecting this high dimensional data set onto two-dimensional (U-matrix) display for easy visualisation while preserving the topology of the input vectors. The SOM displays identified the characteristic EPTC distribution underlying the spatial distribution within the raw data, which had no information included in this regard. In this application, SOMs provided a means to analyse the data with four orders (EPTC), covering a relatively larger region that consisted of high mountain to plain and coastal areas whereas, previous studies had been confined to a single taxonomic group (one insect order) and within a single valley or mountain. The study also stated that the SOM classification of EPTC distribution could be extended to detect environmental changes in the region.

In (Giraudel and Lek 2001) an attempt is made to compare the SOM based methods with other conventional data analysis techniques. In this study, a few, more widely used techniques of ordination, such as Polar ordination, Correspondence analysis (CoA), Principle component analysis (PCA) and Non metric multidimensional scaling (NMDS), were compared with SOM analyses, using data from upland forest in Wisconsin, in US. The limitations observed with the above conventional methods are: strong distortions with non linear species abundance relations, PCA's horseshoe effect due to unimodal species response curves, and CoA's arch effect outliers, missing data, and disjoined data matrix. The paper described the SOM algorithm, as fully usable for ecological data as a complementary method to the existing classical techniques, especially for exploratory data analysis to study community ordination.

River InVertebrate Prediction And Classification System (RIVPACS). The Environment Agency researchers in England and Wales use this software, in which biological data (macroinvertebrate samples collected from rivers and streams) from reference sites are used to distinguish the extent of damage caused to indicator species in the polluted sites.

Biological effects of both polluted and reference sites are compared using the environmental parameters unlikely to be impacted (such as chemical and physical characteristics). It is stated in the article that the river data mappings has given the agency biologist insights into the inter-relationships between the river quality monitoring species taxa, referred to as (BMWP). The system demonstrated advantages over traditional river quality assessment systems and received positive feedback during the initial testing by the end users. The software is now modified and used in Australia, New Zealand and Canada as well to analyse rivers and streams (National River Health Program 2002).

A survey on the applications of artificial intelligence in biological surveillance of river quality studies stated that "… An unsupervised-learning, the Self-Organising Map (SOM), was shown to have considerable potential for diagnosing different types of pollution…." (Murray-Bligh 1998) in RIVPACS; software used to predict the biotic indices used for establishing the water quality index based on the distribution and abundance of BMWP taxa (Murray-Bligh 1998).

AUSRIVAS: Australian Rivers Assessment System (AUSRIVAS) was first developed for fresh water systems in Canberra by the National River Health Program (NRHP), to assess river health, based on the British RIVPACS II program. The NRHP was formed in response to the growing concern in Australia for maintaining ecological values (National River Health Program 2002).

These research efforts show that SOMs can be applied to modelling ecological data incorporating spatial variations at regional scales. In the next chapter, how SOMs could be best applied to modelling ecological data at different scales (such as local, regional, global) and levels (such as environmental, ecosystem), incorporating spatial as well as temporal variations using the available data (from monitoring programmes and the World bank's statistical tables) are elaborated. SOM approaches are investigated for this purpose based on their applications in industrial system process modelling and financial data analysis.

Summary

Strong moves, supported by scientists as well as national and international institutions, to enforce changes in environmental management and initiatives for redirection in scientific research, with an aim to improve co-ordination among a range of professionals were elaborated. Various ecological modelling techniques (conventional and contemporary) used to analyse environmental impact on natural systems were discussed. It is concluded that there is an urgent need for new models, to analyse ecosystems and to predict their behaviour to enforce sustainable environment development.

Chapters 2 and 3 revealed the literature reviewed for this research on ANNs, their applications to real world problems, and ecosystem modelling techniques, including SOM methodologies applied to ecological data analysis. The next chapter illustrates the experimental methodology being examined in this research on how SOM methods could be best applied to modelling ecosystem structure, functioning, its dynamics and biodiversity by enhancing the already tested SOM methods, applied in community ordination studies, only of selected systems and at particular levels.

Chapter 4

Experimental methodology

Scientific publications, reviewed in chapter 3 emphasise the need for novel ecosystem modelling techniques and radical changes in resource management practices to prevent our global ecosystem from environmental deterioration. In that context, co-ordination among a range of professionals is needed, in particular between scientists, stakeholders and the general public so that informed decisions on ecosystem use are made. International institutions, such as World Resources Institute (WRI), United Nations and Organisation for Economic Corporation and Development (OECD), reiterate the use of a systems approach to inform sustainable environment management. The experimental methodology being investigated in this research to model complex natural processes using SOM methods, with a systems approach is elaborated.

SOM techniques are applied to modelling cryptic ecosystems in a manner, similar to that applied in highly complex and diverse industrial process monitoring and control. SOM analyses, successfully applied to integrated analysis of financial data, are also examined for collective analysis of disparate data from environmental monitoring programmes and economic systems, to study the effects of urbanisation on natural habitats. Based on the evidence of earlier SOM applications to ecological data analysis at limited scales (see chapter 2), it is assumed that quantitative analysis of environmental and biological monitoring data using SOMs, could be further extended for modelling natural systems. In this research, SOM methods are applied to modelling diverse ecosystems, not only incorporating spatial, as in previous studies, but also temporal variations, at different scales, and levels to gain insight into environmental ramifications of human influenced activities. This could be a viable alternative to physically modelling cryptic ecosystems, and analysing dissimilar data to study economic trade-offs on ecosystem functioning and biodiversity loss.

4.1 SOM applications to real world problems

"The Self-Organizing Map (SOM) with its related extensions is the most popular artificial neural network algorithm for use in unsupervised learning and data visualisation. Over 3,000 applications have been reported in the open literature, and many commercial projects employ the SOMs as the tool for solving hard real-world problems … ” (Allinson 2001:1). The SOM applications could be broadly classified into the following basic categories based on (Deboeck 1998b):

(i) Market and customer profiling
(ii) Customer scoring and behavior analysis
(iii) Financial and economic modelling
(iv) Medical application
(v) Knowledge management and discovery in databases
(vi) Industrial process optimisation and quality control
(vii) Scientific research

Of the above listed SOM applications, financial and economic modelling and industrial process optimisation and quality control, are of great interest to this research. The following are the main steps of SOM applications in financial, economic and marketing disciplines based on (Deboeck and Kohonen 1998a):

(i) Define the purpose of analysis
(ii) Select data source and quality
(iii) Select data scope and variables
(iv) Decide on how each of the variables would be preprocessed
(v) Choose relevant sample data sets that represent the available system
(vi) Select clustering and visualization method(s), giving more consideration for the use of hybrid methods.
(vii) Determine parameters: in the case of SOM, desired display size, map ratio, required degree of detail;
(viii) Tuning of output or map for optimal clustering and visualization;
(ix) Interpretation of results; checking the values of individual nodes and clusters;
(x) Define or paste appropriate map labels;
(xi) Produce summary results that highlight the differences between clusters;
(xii) Documentation and evaluation of the results.

These steps are used as a guide in this research and for further details the original publication (appendix 5) should be consulted. In the next section, the theory behind the experimental methodology of this research is elaborated.

4.2 SOMs for environmental and biological process modelling

The experimental methodology of this research is based on the fact that the dynamic equilibrium of naturally evolving ecosystems (habitats) could be modelled using SOM techniques, as applied in industrial process monitoring and control (see chapter 2). A physical system change that instigates counter balancing changes in the biological system in turn would cause a small shift in the whole systems equilibrium; continued changes make the equilibrium dynamic. Similar complex, industrial system process dynamics is successfully modelled using SOM methodologies. Hence, the approach to apply same methods to modelling ecosystem dynamics should produce promising results as SOMs have already been proven to be useful in depicting the biological assemblages of freshwater and forest habitats (Ce´re´ghino et al. 2001; Giraudel and Lek 2001).

The possible SOM application to modelling the dynamic equilibrium of the Long Bay - Okura Marine Reserve's intertidal zone, is illustrated herein. The dynamics of this system equilibrium depends on the changes its physical and biological systems may go through, either induced by humans or owing to natural causes (figure 4.1). Hence, it is possible to track the intertidal system dynamics using SOMs created with appropriate measurable physical parameters and their corresponding biological indicators (that show response to the chosen physical changes), and are investigated in this study. Using the changes on a SOM, created with the system variable data, Simula et al. (1999) studied the process dynamics of diverse and complex industrial systems, referred to as sensitivity analysis (see section 2.3.4 Applications of SOMs in industrial process modelling). A change in an environmental parameter that instigates a chain of reactions in the biological system would move its current position (node) to a different one, in the SOM. By following the changes on the SOM map, Long Bay intertidal zone dynamics could be studied. The SOM component planes of the physical parameters and biological species should enable the detection of potential indicators within the reserve's coastal habitat. Decision support systems with the environmental and biological parameters could be developed to depict the system scenarios such as highly polluted or improving situations.

In consideration of above theory, SOM methodologies using ecological monitoring data sets should be able to track diverse complex natural system dynamics without any physical models, as used in industrial process monitoring and control. In cases where knowledge on ecosystem dynamics is limited, map component planes could be studied to gain insights into the contributing factors, such as various human activities, or any correlations within ecological parameters, and even with socio-economic indicators.

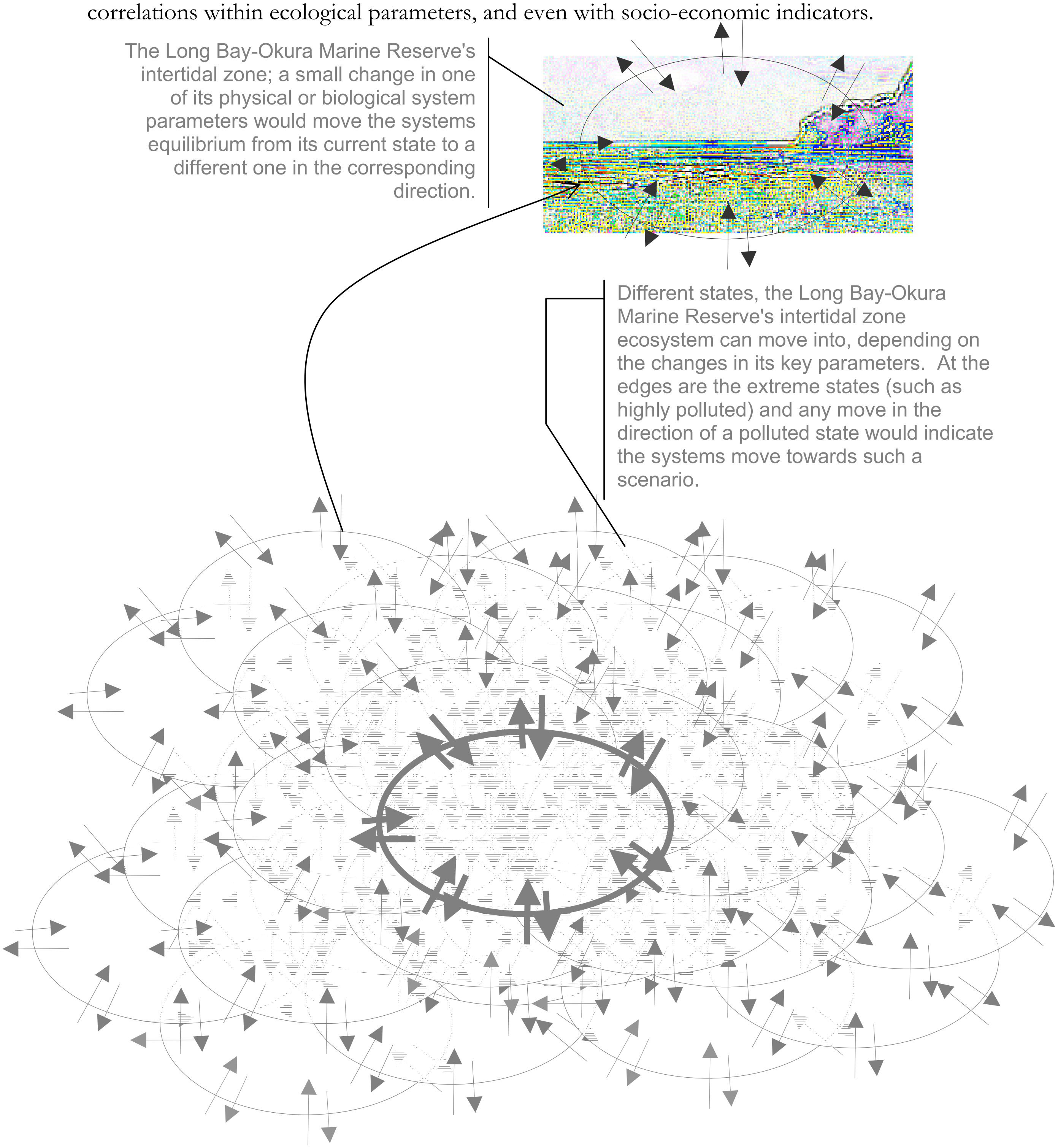


Figure 4.1: Dynamic equilibrium of the Long Bay-Okura Marine Reserve's intertidal ecosystem that depends on the changes its physical and biological systems may go through either induced by humans or owing to natural causes or global variations. A small change in an environmental parameter will cause a change in the biological system that would be reflected in its biological assemblage. Any such change would move its position on the SOM, as seen in the industrial systems sensitivity analysis (see figure 2.6).

The following are the assumptions made in the application of SOM techniques to ecosystem modelling:

(i) There is insufficient overlap to natural and industrial systems, as analysts of both disciplines could not yet accurately elucidate or model the physical process of these systems. Scientists do now agree that despite the significant knowledge, gained in different advanced areas of environmental and other scientific disciples, future research needs to be redirected to accurately predict our global ecosystem response to human activities and natural causes, and to aid sustainable environment management. This is discussed in the literature review chapter, Ecological modelling: A review.

(ii) Financial examples (in section 4.6 SOM applications) with little prior knowledge on the data were successfully analysed using SOM based tools, hence the approach should produce good results with ecosystems. Again, ecologists do not have a clear knowledge on the long-term effects of environmental and biological processes; in particular on the effects of human induced causes. Similar to the financial data analyses, SOMs of ecological data could provide an alternative means to gain more insight into ecosystem structure, functioning and biodiversity.

(iii) Historically, ecologists have been successful in applying engineering, statistical and mathematical models to ecosystem modelling. These borrowed models permitted ecologists to gain insight into many complex ecosystems (Mann 1982). Similarly, the use of SOM modelling tools based on the approaches of industrial process engineering could provide a means to track the complex interactions within natural processes.

(iv) Ecological monitoring data sets are not analysed to their potential (Vant 1999) and the use of SOM based data mining approaches can enhance the extraction of useful knowledge from the monitoring data. SOM based data mining tools have enabled analysts of other disciplines and areas, such as medicine (Shalvi and Declaris 2000), text mining (Sallis et al. 1998), financial and commercial sectors (Deboeck and Kohonen 1998b), to extract useful knowledge from raw data. Thus should produce success in ecological monitoring data mining as well.

(v) SOMs use numerical data, hence they could provide a form of quantitative analysis, a method for converting abundant ecological data into useful information, fulfilling the urgent need, described to be a challenge (Vant 1999) in this field.

The following are the four types of SOM techniques experimented in this research for modelling ecosystems:

(i) cluster analysis,
(ii) component dependency analysis,
(iii) decision support systems and
(iv) time series analysis or trajectory.

As this research is aimed at exploring the SOM based approaches for natural habitat modelling, hybrid methods are not included for clustering and visualisation, however, SOM results are validated based on their cluster statistics. Map parameters are decided depending on the volume of data and clarity needed for the analysis. For clustering and component analyses, few nodes and for trajectories more nodes are used. The best results are achieved by fine tuning the maps on a trial and error basis.

So far the SOM approaches to be experimented for ecosystem modelling were elaborated. In the next section, the generic practical limitations, encountered with conventional data analysis methods of ecosystem modelling and the remedial measures taken to overcome them are discussed. More details with regard to these issues will be discussed in the relevant case study chapters

4.3 Environmental pollution modelling issues

The major issue with the currently used conventional data analysis methods for analysing environmental pollution, is the detection of correlations between the observed environmental and biological system changes within the monitoring data, especially in establishing the exact cause of pollution (Sastry and Miller 1980:267). In most of the conventional methods environmental data and biological data are analysed separately and then the results (the gradients) of both analyses are compared for any correlation between them (see chapter 3). However, such conventional methods with complex statistical and mathematical formulae, such as Before-After-Contorl-Impact design (BACI), are described to be incapable of establishing the link between the causal processes and their environmental effects succinctly and unambiguously (Stewart-Oaten

1996; Thrush et al. 1995). For instance, a study by Walker et al. (2000), where BACI design methods were used to analyse the environmental and biological monitoring data obtained from selected beaches, along the northeastern coast of Auckland in New Zealand, failed to produce conclusive results. The aim of the monitoring programme was to detect any deviation on the annual changes in the coastal population dynamics and if so, to find whether its was a result of the urbanisation in near shore or not.

The additional problems encountered in ecosystem modelling other than the major issue discussed above, are:

(i) Dimension of the data sets: There is a need to reduce the dimensionality of the data set without loosing useful information. This is seen to be impossible with traditional ecosystem modelling methods (Thrush et al. 1995).

(ii) Inability to understand complex mathematical formulae: Models borrowed from engineering consist of very complex formulae. They are difficult to comprehend and also show limitations (Stewart-Oaten 1996).

(iii) Confidence problems in prediction: The currently used conventional models have very high confidence measures that are beyond acceptable ranges. These measures are introduced to overcome the gaps arising from our lack of understanding on ecosystem structure and functioning (Thrush et al. 1995).

(iv) Difficulties in finding the effects of alternative approaches: The need to predict potential ecosystem trade-offs with alternative approaches is becoming more imperative for sustainable resource management, which is found to be impossible with the existing methods.

(v) Defining physical models owing to lack of fundamental knowledge (Ambrose et al. 1996)

(vi) Defining properties of ecosystem. (Ambrose et al. 1996)

The following studies of SOM based approaches to ecosystem modelling using species assemblage data, have been successful in overcoming the above stated limitations;

(i) Giraudel and Lek (2001) analysed forest species composition data using SOMs to look for any patterns in them. In this study no environmental data was incorporated.

(ii) In (Ce´re´ghino et al. 2001) a stream classification based on characteristic species assemblages was developed, again with SOMs, using data that consisted of 283 species at 252 sampling sites. The approach enabled the analysts to study the spatial distribution of 282 lotic macroinvertibrates species from four insect orders: Ephemeroptera, Plecoptera, Trichoptera and Coleoptera = EPTC) in the Adour-Garonne drainage basin in south-western France. The four major FPTC regions were characterised within the drainage basin along with the theoretical species assemblages. The study established a number of species characterising each region ranging from 45-159, also correlated to the spatial difference in EPTC assemblages. The analysts intend to use this technique as a means to compare the stability of theoretical assemblages for biological surveillance, assuming that any change in the species composition within a given EPTC region as a indication of changes in the environment.

(iii) In (Jowett 2001), fresh water body attributes along with species were modelled using a SOM based approach (MOPED). The SOM application, developed to find the missing values of fish species for freshwater systems of different altitudes, produced promising results. In effect, MOPED was able to find the ideal species composition for a particular freshwater body when fed with its altitude.

(iv) In (Walley et al. 2001; Walley and O'Connor 2001) SOMs are used to analyse the river water monitoring data. The report concluded that the SOM methods of classification/ diagnosis to be consisting of considerable potential not only in river quality monitoring but also in other environmental fields.

(v) Techniques similar to SOM clustering were utilised to depict the species community changes of freshwater systems in (O'Connor and Walley 2001; Walley et al. 2001). The system uses SOMs to produce information on the extent of the pollution at a river site by comparing its biological assemblage with that of similar unimpacted reference cites, using environmental parameters unlikely to be affected.

See chapter 3 for details on these SOM applications.

4.4 Procedure for SOM applications in ecosystem modelling

The following are the steps (figure 4.2), followed to collectively analyse the environmental and biological data of natural habitats analysed herein:

(i) Initially, the biological, physical/ environmental and economic system data are preprocessed for time synchronisation and then combined into one file. By doing this the problem of environmental readings not coinciding with biological system readings is overcome.

(ii) Missing values, such as missing data in biological sampling, are deduced by interpolation based on domain expertise. In addition, software that is capable of handling missing values is used for analysing incomplete fused data sets.

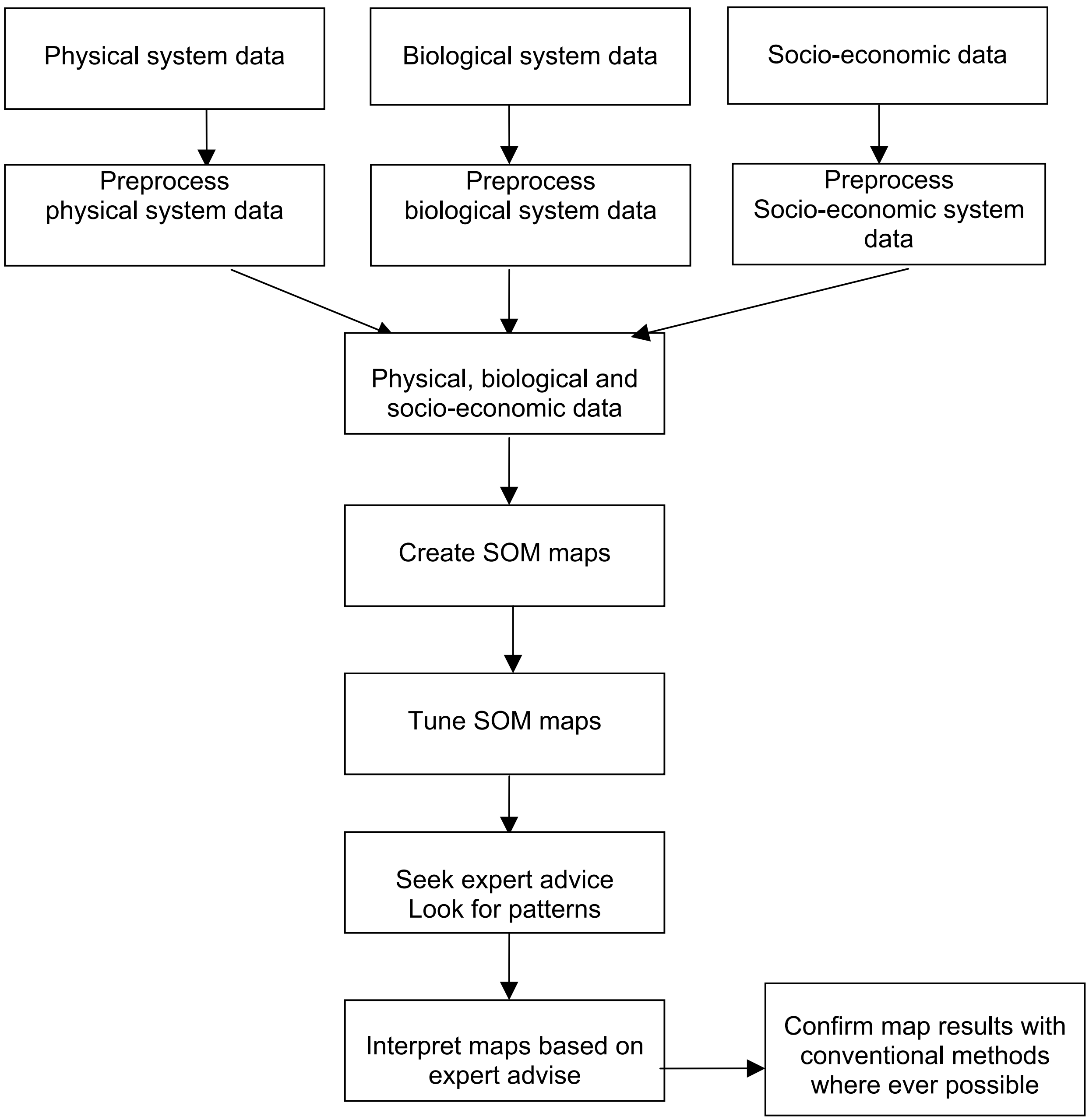


Figure 4.2: Flow chart showing the measures taken for collective analysis of environmental and biological system data in this research.

The SOM approaches based on the steps (figure 4.2) to overcome the problems with conventional methods are further illustrated in detail in case study chapters.

So far the approaches and assumptions to apply SOMs to ecosystem modelling were discussed. In the next section, Kohonen's SOM principles and their applications in industrial engineering and financial data analysis are elaborated.

4.5 Kohonen's SOM algorithm

Kohonen's SOMs (1982) are two layered, feed forward artificial neural nets based on an unsupervised training algorithm that enables the output layer nodes to self organise themselves to preserve the patterns of input vectors. The abstract relation between the input (sensorial) signal and the synaptic adaptation of neurons was first mathematically resolved by Tuevo Kohonen and the learning rule he put forward was a quite simple one, named after him (see chapter 2). The neurons on the output layer are uniformly spread out, functionally connected to their neighbouring nodes. Each node in the network has a weight factor w^{γ}. The components of this vector represent the strength of the synapse connections to the input neurons. In the SOM algorithm these weights can adapt themselves in response to the input signals. During the training process, the nodes in the output layer are presented with the same input signal x^{γ} and the node with the strongest response is assigned as the winner.

The response of a node is defined as the distance $| w^{\gamma} - x^{\gamma} |$ between the vectors w^{γ} and x^{γ}. The closer the weight vector w^{γ} of a node to the input vector x^{γ}, the greater the response. Hence, Kohonen's algorithm could be defined as follows:

(i) The winner node c's weight w^{γ} is made more similar to the input vector x^{γ}.

(ii) All neighbours of c, found within a predefined distance to the winner node, also change their weight vectors similar to that of x^{γ}.

This modification is proportional (factor α is referred to as the *learning rate*) to the difference between the input vector x^{γ}, and the corresponding weight vector. Once the training is completed the Kohonen's net will become ordered with similar input vectors of the original data being concentrated to the neighbouring nodes. In mathematical terms this process could be defined as a non-linear, non-parametric regression. The corresponding error function $E(\tilde{w})$ with an expectation value converging to a minimum during the training process, could be defined as

$$E = \int \sum h_{ci} \mid {}^{\rho}w_i - {}^{\rho}x \mid^2 g({}^{\rho}x) d^n x$$

Where h_{ci} is the neighboring function of node i to the corresponding winner $c(\vec{x})$, an exponential function and $g(\vec{x})$, the density function of the vectors $\vec{x}$ in the n-dimensional data space. The Kohonen algorithm is obtained in a discrete data space by computing the optimal weight vectors (for minimising $E(\tilde{w})$) by gradient descent (Eudaptics software gmbh, 1998:114-115).

So far the motivation behind the hypothesis to experiment SOMs for ecosystem modelling as applied in industrial systems and financial data analysis was explained. In the next section, SOM applications from the latter two disciplines are outlined.

4.6 SOM applications

Possible SOM applications to ecosystem modelling based on industrial engineering and financial data analysis methodologies are discussed below.

4.6.1 Cluster analysis

SOMs clustering abilities have been successfully applied to grouping of data classes in financial data analyses by studying the clustering patterns, even when combined with dissimilar data sets and are elaborated upon. This method is applied to analyse dissimilar data sets of natural systems in case study chapters.

SOM cluster analyses (CAs) provide a useful tool for initial analysis (IA) (Serrano-Cinca 1998). In this study, the SOM clustering patterns within the economic indicators of financial institutions, revealed the economic status of the institutions analysed. Data obtained from financial information systems, such as data banks, was used in the SOM analyses. IAs that are found to be useful in summarising and scrutinising large volumes of multidimensional data sets as well as in serious analyses of model formulation are often undervalued by academics and professional analysts.

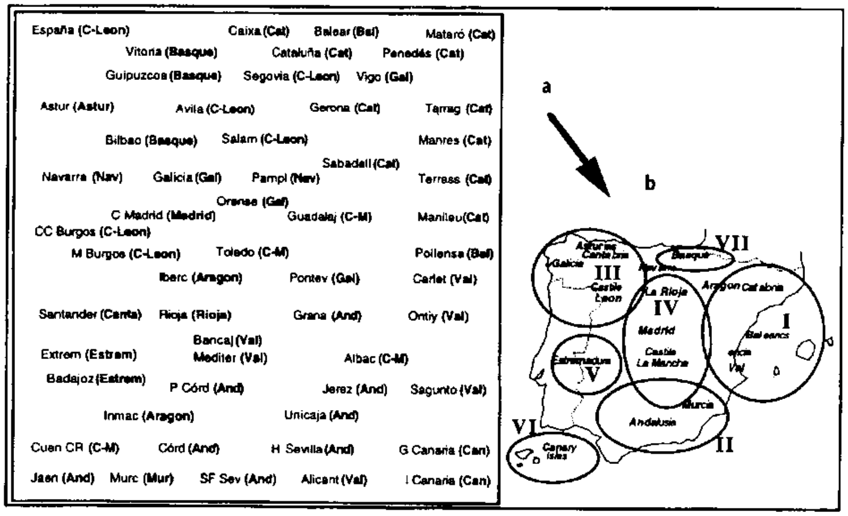


Figure 4.3 a: The Spanish Savings Banks have self-organised according to their geographical distribution. b: Strategic groups. Source: (Deboeck and Kohonen 1998:6).

In (Serrano-Cinca 1998), the use of SOM techniques to study complex information on Spanish saving banks, was discussed. In one of the five examples, SOM based IAs were found to be successful in analysing the strategic positioning of Spanish saving banks and in formulating a corporate strategy for these banks.

In the study, SOMs created with financial information published by The Spanish Confederation of Saving Banks, enabled analysts to study the territorial pattern within the complex data, which had no explicit codes in this regard. The strategic groups of these saving banks differed in their operations and implementation strategies based on the region they were located in.

SOMs (figures 4.3 a & b) not only confirmed these 'territorial' divisions, but also revealed more facts, depicting greater precision such as divisions within the regions. This kind of SOM feature extraction from raw data can be applied to modelling complex ecosystem dynamics and are explored in case study chapters using ecological monitoring and economic system data to study the effects of urbanisation on natural habitats.

By selecting appropriate environmental and biological variables, an ecosystems trends, could be studied as applied in the European Union (EU) economic trends analysis. In a different example of (Serrano-Cinca 1998), the economic convergence of the EU member states was studied before the merger, using two distinct sets of dissimilar data.

The EU merger was an historic event as far as the whole world was concerned and many business institutions were interested in knowing about the economic situation of these EU countries, in advance to the merger. Unlike the earlier example, in this analysis there was an urgency to analyse the macro economic variables of different countries whose accounting practices were different. To overcome this situation, initially, the dissimilarities between pairs of countries were calculated as Euclidean distances within the standardised macro economic variables proposed by the Maastrich Treaty, provided by the EU, corresponding to 1995.

A second SOM map was created using company financial information that consisted details on productivity, profitability, etc. of the EU member states, obtained from the balance sheets and profit and loss accounts of these companies. Here again, it was found to be impossible to create homogeneous financial information for comparison in an international context, as the companies from different countries had had different accounting practices. In fact, comparison of company financial information of such different countries with different accounting practices was considered to be dangerous. Hence, the analysts used information obtained from the BACH database that had homogeneous financial data information of different countries, however, these variables were different to the earlier study. The latter SOM of 16 financial ratios consisted of company economic results in relation to the resources employed such as gross profit, net profit and financial return, and their relative costs such as intermediate consumption, personal costs and financial charges and their financial structure equity, indebtedness, debt structure and provisions. Interestingly, the first and the second SOM maps showed different cluster groupings.

The EU analyses gave details on how data with inconsistent labelling could be collectively analysed for trends and features, inherent to the issues by studying the SOM clustering patterns and groupings in the data. These approaches that enabled financial analysts to overcome the limitations imposed by conventional data analysis methods could be applied to modelling ecological data from different sources, at different scales and levels. It could provide an alternative means to overcome the difficulties in physically modelling the highly diverse and complex natural systems. The SOM approaches can even used to overcome the issues encountered with the latest information pyramid model approach within the pressure-state-response (PSR) framework (see chapters 2 & 7). Ecosystem processes (natural habitats) at different geographic scales (such as within an area, region, national) and levels (such as individual species, communities or environmental system) could be modelled, using SOMs, created with fused data sets that consists of inconsistent labelling. As SOMs are capable of clustering input vectors without knowing their class membership, ecological data can be modelled even without comprehensive knowledge on the complex, interrelated ecosystem reactions. SOM components could be used to gain useful information on environmental and biological parameters by analysing the contributing factors of the observed SOM clustering patterns in the fused data and the correlations within them. In instances where data is not available on a daily basis, monthly averages or totals could be used for the analysis and is experimented in case study chapter 7.

SOMs could be used to study the emerging trends at higher levels (such as global trends in urbanisation and its effects on natural habitats) as used in stock market analysis. The use of SOM techniques to study the patterns across emerging stock markets in 30 countries (figure 4.4) is discussed in (Deboeck 1998). In this analysis, SOMs were used to study the patterns in the evolution of emerging stock markets by reducing the dimensionality of the publicly available data (at the end of 1996) on fundamental and technical indicators of the International Financial Corporation (IFC). In addition, SOM maps were also created to compare some individual companies (such as banking, telecommunication and construction companies) around the globe. Overall, Deboeck (1998) illustrated how SOM techniques could be applied to grouping countries of similar features using the similarities and dissimilarities between the various markets.

The SOM grouping of the emerging markets was described to be an important factor in the allocation of assets between markets. It not only created improved benchmarks but also reduced the maintenance cost of an emerging market investment portfolio. The results of the SOM analyses were compared with those of classical asset allocation strategies and described to have been demonstrated improvements on the return of the portfolio. The classic methods were normally based on a single criterion, whereas in the SOM analyses, many indicators were collectively analysed.

Similar to the emerging stock market analysis, SOM techniques using dissimilar data sets could permit collective analysis of physical system changes and community dynamics to improve our understanding on trends and human relationships with natural habitats. The clustering patterns of SOM maps created with fused data could unravel the common and distinctive features of the natural systems analysed, even if the data has been originally gathered by different groups of researchers. This is investigated in case study chapter 5, where SOMs are examined to analyse the Long Bay-Okura Marine Reserve's intertidal data of physical and biological systems from different sources and in chapter 6, with different monitoring sites along the northeastern coast of Auckland.

SOMs also permit for collective analysis of different sets of variables to study any possible emerging trends that may arise from the current situation. Analysis on clustering patterns and relationships in the observed data could help in developing new hypotheses for further analysis.

A SOM application of financial data analysis at global scale is presented in (Deboeck 2002). The market groupings of countries with similar risk patterns on stock market investment were studied using data containing financial, economic and stock market information from 52 markets around the world. This approach is investigated in case study chapter 7, using the World Bank data to study the global environmental trends arising from urbanisation and its effects on biodiversity. If proven, the issues with the currently used information pyramid and PSR concepts (see chapters 3 & 7) could be overcome using SOM techniques as applied in financial data analysis; it as well confirms the hypothesis of this research (chapter 1 & 4).

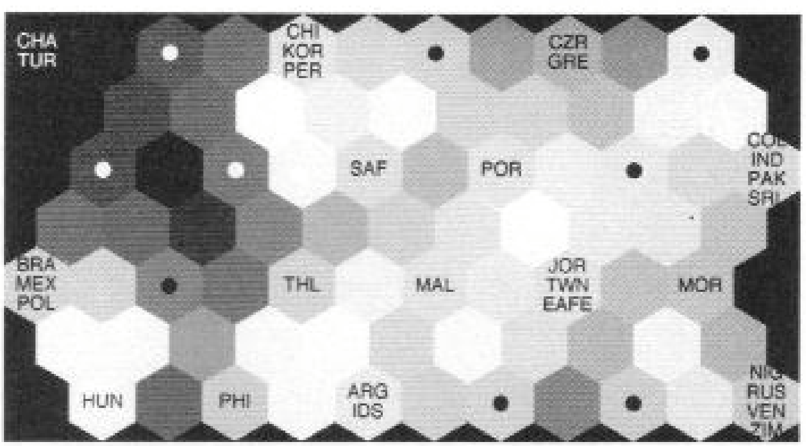


Figure 4.4: SOM of the weekly returns of 30 emerging markets, based on market price indices (market capitalisation, market dividend yield and P/E and P/B rations for 1996). Source: (Deboeck 1998:91).

It was concluded in (Deboeck 2002) that the SOM grouping of country risk indicators and stock market characteristics were to be more logical than the Wall Street Journal's (WSJ) five groupings of the 52 countries. In addition it is argued in the article that despite the fact that domain expertise plays a significant role for SOM methods to be realistic, especially in variable selection and interpretation, still SOM results do provide a better rationale than conventional methods. More importantly, SOM techniques are ideal for exploratory data analysis in areas where prior knowledge is insufficient for hypothesis postulation and testing.

Thesis case studies are aimed at analysing complex natural systems within an ecosystem framework, using the available disparate monitoring data sets. The first two case studies are based on intertidal (physical and biological system) monitoring data collected by AUT undergraduates and sediment deposition rates with subtidal population dynamics data gathered by the Auckland Regional Council. The original projects of both case studies were undertaken to study the effects of urbanisation on marine life at Long Bay and the latter was eventually extended to cover the northeastern coast of Auckland in New Zealand.   The SOM approaches are used to produce visual displays of these fused data sets to look for hidden patterns (such as annual or any other variations resulting from the urbanisation on North Shore), in them. The third case study is based on disparate global data sets on urbanisation and biodiversity. The financial sector examples discussed, showed that by changing the variables for SOM analyses, one could be able to visualise the different patterns and their contributing factors along with the correlations, for the observed patterns. This feature could be applied to distinguishing different groupings within world countries and to study the reasons for such groupings as well as any common factors in human influence and their effects on natural systems.

In consideration of the above, SOM analyses could be possibly used as a tool for integrated data analysis of complex natural systems to study the effects of urbanisation on natural habitats using dissimilar data at different levels and scales, as applied in financial analyses. Conventional ecological data analysis methods including the complex BACI and BACIPS series design methods are unsuccessful in discerning the anthropogenic environmental effects from those of any natural causes (Soule and Kleppel 1988; Osenberg and Schmitt 1996). Chapter 3 provided details of these studies, pointing out the inadequacies with these conventional methods.

4.2.3 Component dependency analysis

The component planes of SOM maps could be analysed to reveal the relationships between the components used in the SOM analysis and is successfully applied in (Simula et al. 1999), to a continuous pulp digester of a pulp mill. In this analysis, 11 temperature sensor readings from the digester side covering a period of half year, were used as SOM input vectors with one output on quality measurement, the keppa number. These process variables were selected to study the correlation between the keppa number and the digester temperatures. The influence of process temperature was known to be the most important factor in the digester operation.

The component planes of this map are shown in figure 4.5. The black colour patches of the temperature (# 1 - # 11) components depict high temperature and white correspondingly low. Accordingly, high temperatures in the first eight measurements #1 - #8 (black spot in the middle of each plane) could be seen reflected in the last plane, the keppa number by small values (white spot in the middle). The kappa number roughly correlates (inversely) with the first eight temperature measurements. This enabled the analysts to confirm the phenomenon, with clear explanations; when the cooking temperature was high, the delignification reactions were fast and in turn reduced the keppa number.

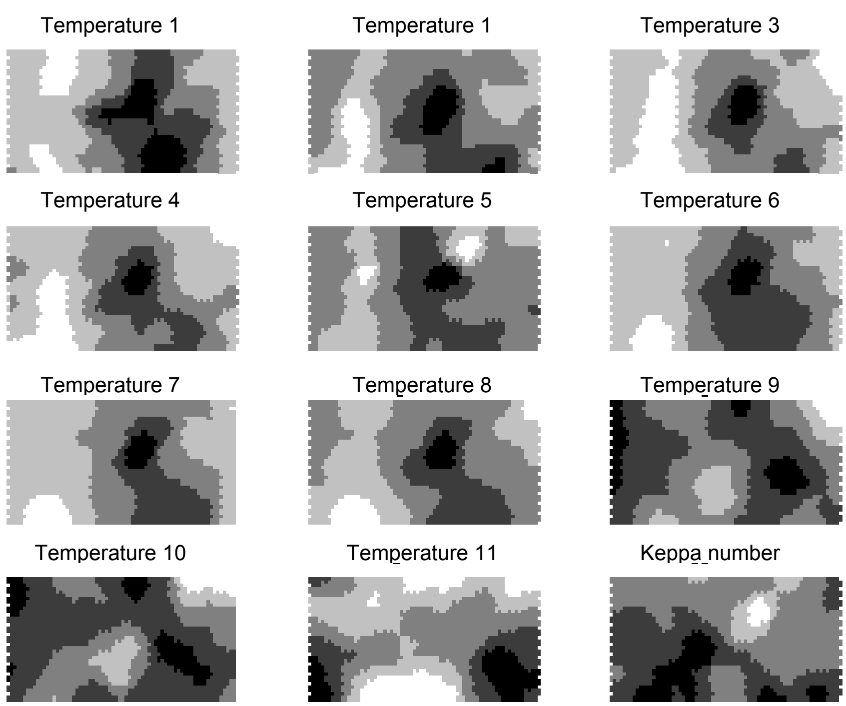


Figure 4.5: Component planes of the continuous pulp digester. Source: (Simula et al. 1999:12)

It was further concluded that the keppa number was also dependent upon factors that were not included in the analysis, for instance the concentration of cooking chemicals participating in the delignification reaction and keppa number variations could not be explained.

SOMs are used to analyse the dominant variable in industrial systems. In (Eudaptics software gmbh 1998), the dominant process variables of a simulated industrial process were identified from the 16 analysed process variables (such as temperature, speeds and pressure). Of the 16 components included in the analysis only four are shown in figure 4.6, of which three variables: EB1x Temperature, EHB Temperature and ESx Speed, carry gently over the map windows. It was concluded that these were the dominant variables of the above process. These three were stated that together defining a complete order of the data space. In contrast, Ehm Temperature seems to be distributed non-uniformly and was described to be a non major role player in the overall process operation

Similarly, environmental and biological system variables could be collectively analysed to establish any correlations between them. Component plane analysis on SOMs, created using fused data sets of environmental and biological system data should reveal the relationships between the two systems. Furthermore, could be experimented with SOMs, created with biological system variable data (different species) to study the dominant factors (indicator species) in them.

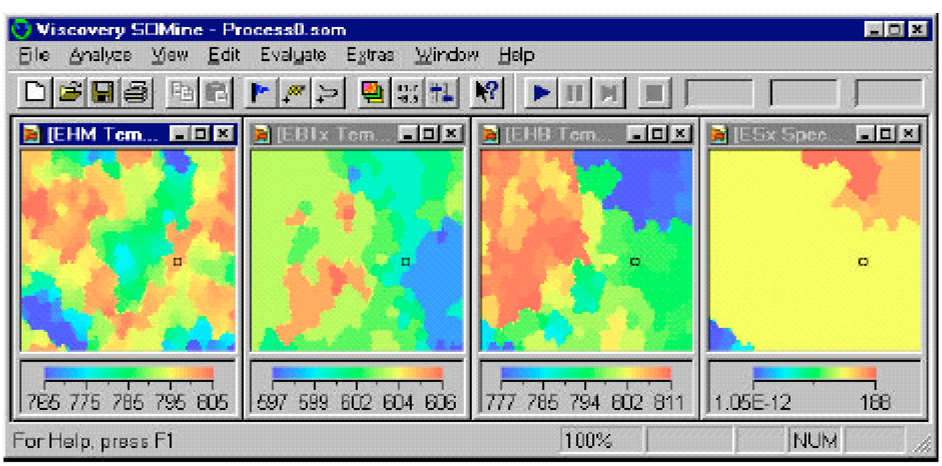


Figure 4.6: Distribution of process variables in SOM map created with 16 contributing variables. Source: (Eudaptics software gmbh 1998:51).

A SOM map (Figure 4.7 a) was created using the macro economic data in the US covering a period of twenty years. The variables used in the analysis were: S&P 500 / CPI, Prime Rate, Treasury Bill (90D), Treasury Bond (30Y), S&P dividend, S&P P/E ratio, Gold Price / CPI, CPI (rel. change) and unemployment. During the SOM training, the priority of gold price was set to 0.0001 in order to associate it with the map.

In Viscovery (a commercial software developed based on Kohonen's SOM algorithm), by making the priority of a factor as 0.0001 (whose values are deemed totally dependent on the other factors), this factor could be associated to the map's other variables. Using this SOM map, values of the dependent variable could then be predicted, such as the gold price example from (Eudaptics software gmbh 1998).

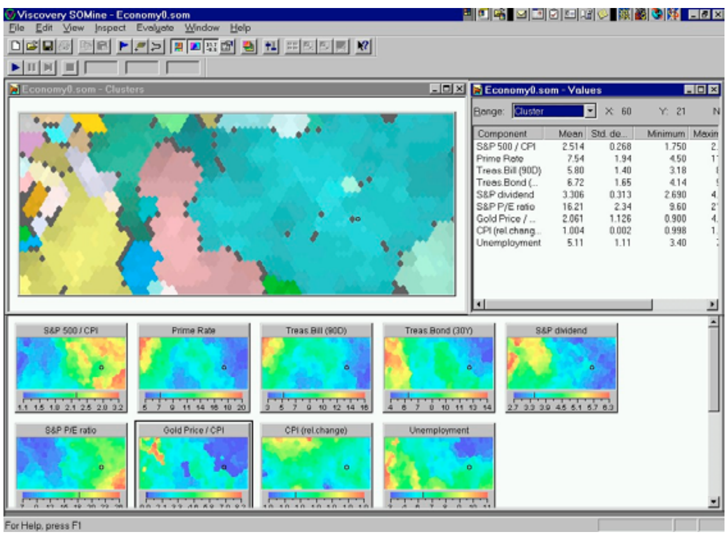


Figure 4.7 a: Distribution of US macro economic data on the SOM map created using Viscovery. Source: (Eudaptics software gmbh 1998:32,33 &47).

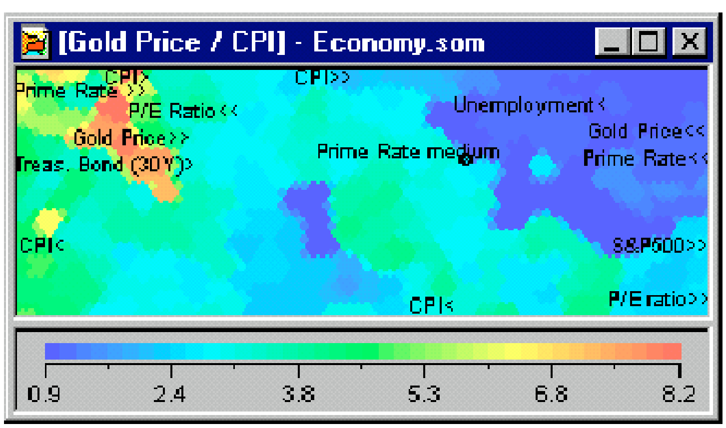


Figure 4. 7 b: Gold price component window of the SOM created with US macro economic data using Viscovery. Source: (Eudaptics software gmbh 1998:33).

The SOM clusters (figure 4.7) were used to analyse the different economical phases during the analysed time period. In the gold price component window (figures 4.7 a & b), gold prices at different economical phases are nicely separated into an area with high values and an area with low values. This was described that the gold price to be totally dependent upon other factors and be reasonably deduced from the other variables. Hence, using the map data records with high gold price (such as selected area) were filtered out. Further, using the map gold price was predicted based on other variables, in which this was left out in the input vectors. If the high and low values were randomly distributed over the whole map, the conclusion would not have suggested the gold price to be related to the other variables. In the similar way, data records representing a time series could be used to predict the values of a dependent variable. On the other hand, it as well could be applied to modelling links between different variables or a combined effect of certain variables, such as causal processes and the environmental effects and is applied in case study chapter 7, to modelling the link between sedimentation rates and subtidal community dynamics.

4.2.2 Decision support systems

In addition to the first two cluster analyses, discussed earlier in the chapter, Serrano-Cinca (1998) further produced SOM maps to demonstrate the financial situation of companies in a graphic and intuitive form by studying the synaptic weight of the maps. In the map (figures 4.8 a & b), the variable that provokes the greatest response for each neuron is shown by it synaptic weight.

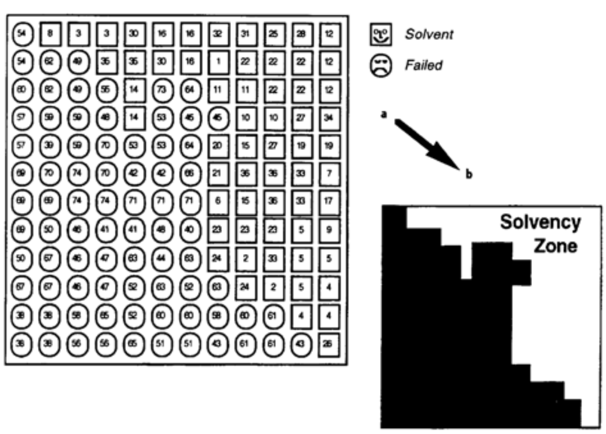


Figure 4.8 a: SOM showing each of its neurons with the firm that gives the strongest response. 1 to 36 are solvent firms. 37 to 74 contain information about the year prior to the occurrence of bankruptcy. The two main areas show the neurons with bankrupt companies and the ones with solvent companies. b - The solvency map. Source: (Deboeck and Kohonen 1998:19).

Using this graphic representation the main features of the SOM, were summarised and then the whole map was eventually turned into a decision support system showing the solvent and failed companies.

Financial data from Moody's Industrial Manual from 1975 through to 1985 for a total of 129 US firms, of which 65 were bankrupt and the rest in solvent stages, was used for this purpose in (Serrano-Cinca 1998). The SOM map (figure 4.5) shows how the two regions were delineated with sufficient clarity, one corresponding to the solvent firms and the other to the bankrupt firms. By finding the positioning of a company data in the map, its financial situation could be revealed. This approach is applied to analysing polluted scenarios/ areas of ecosystems in case study chapters 5 and 6.

4.2.4 Trajectories

With the use of SOM trajectories complex process dynamics has been successfully studied in process industrial process monitoring and control (Simula et al. 1998; Simula et al. 1999). These publications illustrate how SOM trajectories could be applied to tracking the system process dynamics. A process control unit of a reactor, depicted in relation to the output measurements at discrete times (figure 4.9) in (Eudaptics software gmbh 2002) is discussed herein. The data that consisted of the different stages of a particular reactor in an ordered way, was used to create the SOM map. The time series analysis of the measurements on the process was seen to be useful in tracing the process states in time. In occasions where data on faulty situations are difficult to obtain, simulated values could be used and the process entry towards such adverse conditions could be detected well ahead in time.

The theory behind these time series analyses is that in a SOM map, the current operating point of the process would be assigned to the BMU for this current measurement vector. Similarly, all input vectors would be assigned to their respective BMUs. The trajectory of the process dynamics on a topographically ordered map with labels on it could enable analysts to track the process dynamics on a time series. By watching the animation of a trajectory on the map, it is possible to detect the process entry towards any unfavourable conditions. Time series analysis could be carried out at any time intervals, depending on the availability of input vectors, such as at hourly, 24 hrs, monthly or on a yearly basis.

In (Deboeck 1998:94) the evolution patterns of emerging stock markets of 30 countries over time, was studied on SOMs (figure 4.10) using combined data sets of 1988, 1990, 1993 and 1996 SOMs

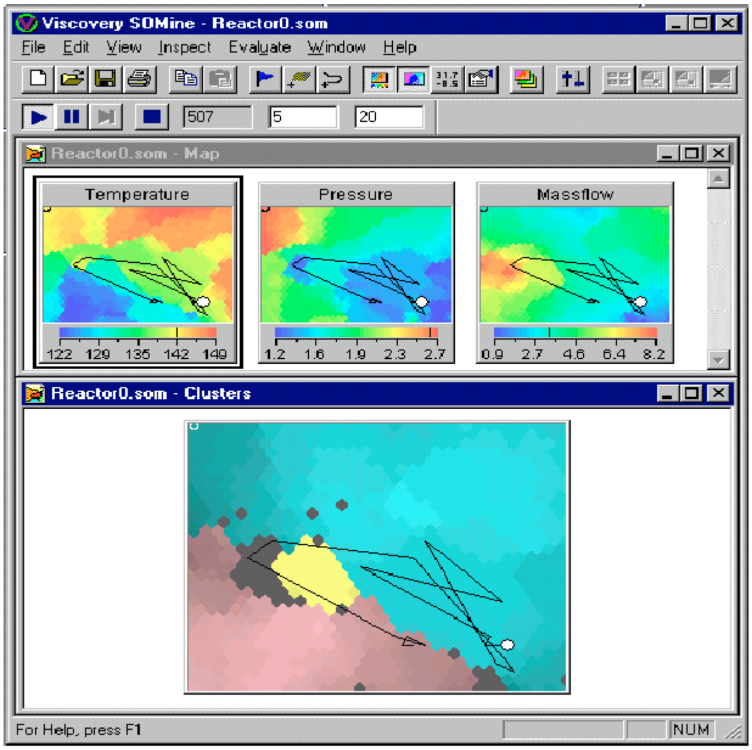


Figure 4.9: SOM map showing how trajectories could be used to trace the process states in time. Source: (Eudaptics software gmbh 2002).

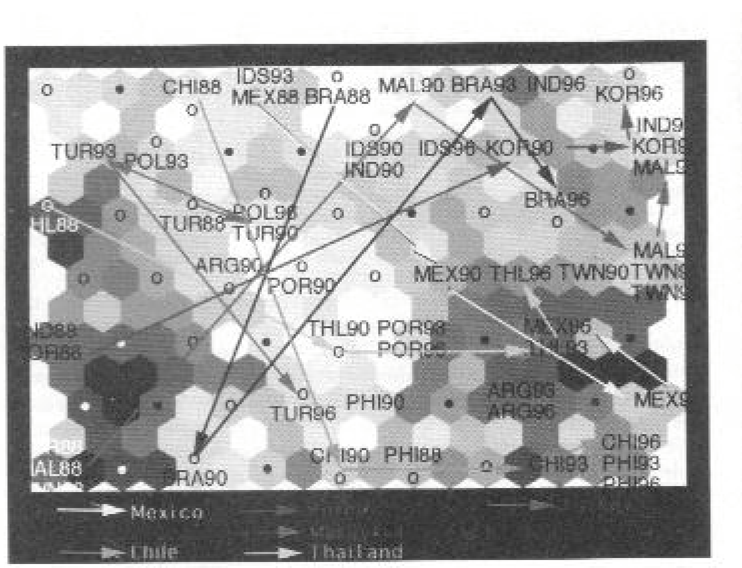


Figure 4.10: SOM showing the evolution patterns of emerging stock markets of 30 countries over time using combined data sets of 1988, 1990, 1993 and 1996. Source: (Deboeck 1998:94).

SOM maps created with measurable data of natural habitat monitoring programmes could be used to create decision support systems to analyse scenarios and trajectories to predict the system process entry towards ideal or polluted scenarios. SOM maps of normal and simulated data on adverse conditions, such as highly polluted or high temperature, could be used to analyse an ecosystems dynamics at desired intervals. This approach could be applied to restrictions encountered with conventional methods that require clear understanding of physical processes for modelling and prediction of natural systems behaviour. It could as well used for transforming abundant data into useful information for resource management purposes.

A SOM based neural network trained in (Harris 1993) with the vibration data recorded from a machine operating under normal operation, was used for diagnosing faulty conditions, such as detect the machine operation leading towards faulty conditions. The entry towards faulty conditions was studied by calculating the increase in the error of the input vectors. Similarly, using the error, methods could be developed to detect a natural systems move towards unusual scenarios.

4.2.5 Software tools

There are software tools for SOM based exploratory data analytical methods, such as clustering, visualisation, component plane analysis and trajectories, discussed in the earlier sections. These software products are based on the basic SOM algorithm first introduced by Kohonen (1982). Currently, there are three methods for the implementation of SOM based approaches: (i) public domain software (ii) self-coded software and (iii) commercial software packages with SOM capability. Self-corded software packages are common in academic research applications. The following are the software tools generally used by academics:

(i) SOM_PAK: Was developed a few years ago by (Deboeck 1998). It is an extensive program package that facilitates the main steps in SOM analysis including selection of the map size and format, proper initialisation, monitoring of computational process, analysis and interpretation of the resulting mapping. It is a public domain software, available for scientific research at *http://nucleus.hut.fi/nnrc/nnrc-programs.html*

(ii) Repository for Intelligent connectionist-Based Information Systems (RICBIS) (Deng et al. 1999): RICBIS is aimed at creating a repository with software tools for the implementation of connectionist based intelligent information systems, for academics.

(iii) Viscovery: Commercial software developed by Eudaptics software gmph in Austria (Eudaptics software gmbh 1998).

4.2.5.1 SOM_PAK

Corded by (Deboeck 1998), especially for scientific work and applications in UNIX machines. Even though later an MS-DOS version was made available it could not be used in Windows 95 or NT environments. Eventually due to the popularity of MatLab in scientific computing, SOM_PAK was rewritten in MatLab (version 5) with many additional facilities such as graphical user interfaces (GUIs) and improved display formats for SOM visualisation. The MatLab version of SOM_PAK is referred to as SOM Toolbox, which is described to be complimenting to the SOM_PAK and could be downloaded from *http://www.cis.hut.fi/projects/somtoolbox*

4.2.5.2 RICBIS

Of the many RICBIS modules, made available for academic use, a java applet developed by (Deng et al. 1999) with abilities to cluster data using SOM and ESOM techniques are used in this research. The applet also consists of tools for other projection methods such as sammon projection and PCA. However, there are not any modules in RICBIS for component plane or time series analysis.

4.2.5.3 Viscovery

Viscovery SOMine is a user friendly commercial software developed by (Eudaptics software gmbh 1998) based on a modified Kohonen's self-organising map algorithm, namely Batch SOM, which is a robust variant of unsupervised neural networks. The software was further enhanced with new scaling techniques and learning algorithms to speed up the learning process. Last year, a fourth version with added facilities was released. With the latest Viscovery version, up to fifty components could be analysed and it also consists of better clustering abilities. Using Viscovery, multidimensional numeric data sets could be projected on to a two dimensional display with a hexagonal grid structure. During the training process the input vectors, the nodes on the gird gradually adapt to the intrinsic shape of the input data distribution. Once the training process is over, the data represented in the trained SOM map could then be converted into various different visualisation formats depending on the requirement. For instance, cluster analysis for refining and evaluation of clusters, component plane analysis for studying the correlations between components and trajectories for process monitoring. At the same time, numerical statistics of a cluster or a node can also be retrieved on demand through a new window for developing decision support systems. Labels could be added to the map for easy reference.

Summary

The experimental methodology adopted to investigate SOM methods for modelling highly complex and diverse natural systems without any physical models, using numerical data alone, as modelled by industrial engineers and financial data analysts was explained. SOM applications in complex industrial system modelling (control and monitoring) and in initial analysis of financial information systems with multidimensional, disparate data sets were discussed. The possibilities of using the SOM approaches to better understand ecosystem structure, functioning and process dynamics, using environmental and biological monitoring data are as follows:

- (i) Cluster analysis, (ii) component analysis and (iii) decision support systems to look for potential relationships between environmental and biological data, which is generally found to be difficult with conventional methods.
- (iii) Time series analysis (trajectory) to track ecosystem dynamics on SOMs created with monitoring data.
- (iv) Models for integrated data analysis, to investigate the correlations between developmental, socio-economic and environmental parameters with biodiversity indicators. Global statistical data, compiled by international institutions are used in this analysis.

The next chapter illustrates the first case study of this research using AUT and NSCC data from the Long Bay-Okura Marine Reserve in northern New Zealand. The examples of the second case study utilise data from beaches, north of Auckland in New Zealand. The final case study examines some possible ways of using SOM methods to analyse global data on biodiversity, economics and socio-economic aspects, within an integrated framework to study ecosystem functioning, use and economic trade-offs. Further details on the experimental methodology will be discussed in the relevant chapters.

Chapter 5

SOMs to analyse Long Bay-Okura Marine Reserve data

Chapter 4 gave details on the SOM techniques being experimented in this research for modelling highly complex and often 'cryptic' ecosystems. Gaining more knowledge on natural system processes (environmental and biological) to predict their response to a range of human activities is seen as vital for sustainable environment management. The main objective of this case study is, to analyse dissimilar ecological monitoring data sets with a systems approach to distinguish complex ecosystem response to human influence from that of natural causes.

The SOM methods are applied to ecosystem modelling, based on the approaches of complex industrial system process dynamics modelling and financial data analysis, using numerical data (see chapters 3 and 4) and could be divided into four major categories:

(i) cluster analysis
(ii) dependent component analysis
(iii) decision support systems and
(iv) trajectory (time series analysis).

5.1 Background

The Long Bay-Okura Marine Reserve was established in 1995 as a direct result of the pressure imposed by the North Shore's environmental groups and the general public who were concerned over the massive environmental degradation at Long Bay. Buckeridge (1999) described the environmental situation as "stochastic urban accretion (SUA)", in reference to the apparent lack of planning within city development. When granting resource consent for proposed development, the authorities failed to give required attention to the impact caused upon the surrounding environment. Poor monitoring of environmental change and improper impact assessment on proposed development have contributed to this situation. It is argued that an ever-increasing load on the existing public utility systems is the main cause for the observable biodegradation at the reserve. Buckeridge (1999) discusses how silt runoff and sewage infiltration contribute to the degradation in biodiversity at Long Bay. Until very recently, no measures were taken to improve the services such as ageing sewage, storm water systems and roading. The sewage and storm water systems are becoming increasingly overloaded resulting in

(i) more wastewater pumping station overflows,
(ii) storm water leaks into wastewater systems,
(iii) storm water infiltration into wastewater systems and
(iv) wastewater leak into ground water.

(Couriel et al. 2000).

All of this results in increased amount of untreated water entering the sea, causing degradation in coastal and marine biodiversity (Couriel et al. 2000).

Being New Zealand's first urban marine reserve, setting up of the Reserve itself was subjected to severe criticism from local land developers. Even though it was controversial at the time, local authorities, academic and research institutions took advantage of an opportunity to evaluate the effectiveness of the existing, aging infrastructure (as part of the Resource Management Process). Research involving biomonitoring and environmental monitoring carried out by staff and students of Auckland University of Technology (AUT) and other institutions provide evidence of environmental changes as the main cause for biodegradation. Their research and monitoring programmes gave a formal approach to assess the effects of urban development on marine life at the Long Bay-Okura Marine Reserve. The research findings emphasise that the current development practice on the North Shore is having a deleterious effect upon the marine environment. This effect is ongoing and thus unsustainable, if biodiversity is to be maintained.

5.1 Objective

The following are objectives of this case study being investigated for SOM applications in ecosystem modelling using Long Bay-Okura Marine Reserve data;

(i) to develop a means to visualise the reserve's system dynamics using the existing biological and environmental monitoring data from different sources.
(ii) to model the possible relationships within and between biological and environmental system parameters again using the reserve's monitoring data. In this research, SOMs are explored to reveal the correlations between physical, biological, economic and social aspects at different levels and scales

(regional, national, and global). Environmentally unsustainable anthropogenic activities lead to certain conditions such as high biological oxygen in demand (BOD), high nutrient levels (nitrate, ammonia, referred to as eutrophic conditions) that affect the growth of marine organisms causing an alternation in the ecosystem structure. Methods that are able to establish the link between such causal processes and their environmental effects using dissimilar data sets (that consist of inconsistent labelling) could provide an approach to analysing these complex natural systems within an integrated framework.

(iii) to develop a decision support system by superimposing all SOM cluster details, such as environmental parameter and bio cover ranges. This can be used to visualise the scenarios within dissimilar monitoring data sets.

(iv) to establish a method to observe the process dynamics with the use of a trajectory on a SOM, created with the monitoring data. The SOM clustering map can distinguish the areas of different conditions, such as high *Enterococci* counts, monitoring station and the time these conditions were experienced. The animation of a trajectory (time series analysis) on this map could show the systems dynamics and its environmental effects on the wider system.

SOM trajectories could be used with simulated values to predict bio cover percentages under different environmental conditions, such as sediment deposition rates or development scenarios. For this, data on sediment deposition has to be incorporated as a component in the SOMs. By including simulated values that reflect highly polluted scenarios, the systems entry towards these areas on a SOM trajectory could be used to predict the environmental trends and their effects as applied to monitoring and control of industrial system processes (see chapter 2). Similarly, SOM trajectories could also be used to measure and make improvements on activities that tend to cause pollution. They could also be applied to modelling coastal habitats, within a region, using dissimilar data, collectively or individually at desired time intervals.

5.3 Data from Auckland University of Technology research projects

From 25 March 1999 to 4 April 2001 continuous monitoring of Reserve's ecology was carried out as part of undergraduate research programmes at AUT (Higgs 1999; Snowdon 1999; Meyer 2000; Scharader 2000; Hecht 2001). The following tidal zones (figure 5.1) had slate tiles positioned along a transect from uppermost to upper sub tidal zones:

(i) Lower supralittoral: S1

(ii) Upper littoral: S2

(iii) Mid littoral: S3

(iv) Lower littoral: S4

(v) Upper sublittoral: S5

The five intertidal divisions have characteristic zoning of sciaphilic organisms (those organisms that encrust the rock surfaces). Observations were made on a monthly basis by taking photographs to monitor the colonisation patterns (Buckeridge and Tapp 1999);

Zone	Key species observed on slate tiles
Lower supralittoral-S1	*Chamaesipho columna*
Upper littoral-S2	*Chamaesipho columna*
	Epopella plicata
Mid littoral-S3	*Austrominius modestus*
	Crassostrea gigas
	Pomatoceras caeruleus
	Cyanobacteria
Lower littoral-S4	*Austrominius modestus*
	Pomatoceras caeruleus
	Xenostrobus pulex
Upper sub littoral-S5	*Austrominius modestus*
	Balanus trigonus
	Cyanobacteria
	Corallina officinalis

AUT students analysed the bio cover on these slate tiles to study the sciaphilic colonisation dynamics over time by determining the percentages of different species (live and dead) on the tiles (Higgs 1999; Snowdon 1999; Meyer 2000; Scharader 2000; Hecht 2001). The total bio cover percentages of sciaphilic colonisation were calculated based on a chart (see appendix 4 for chart). No data was available on the upper sub littoral (S5), as it was not included in the monitoring programme.

Figure 5.1: Photograph showing the data collecting monitoring stations at the Long Bay-Okura Marine Reserve, north of Auckland, New Zealand

A different group of AUT students (Scharader 2000) collected the following data to measure the environmental changes in the Long Bay-Okura Marine Reserve caused by urban developmental activities in near shore:

(i) Date
(ii) Place (Data collection monitoring stations)
(iii) Time
(iv) Temperature
(v) Specific Conductivity (Sp Cond) [mS/cm]
(vi) Dissolved Oxygen [mg/l] / DO
(vii) Air temperature.
(viii) Special remarks
(ix) pH
(x) Ammonia [mg/l]
(xi) Nitrate [mg/l]
(xii) Turbidity [mg/l]
(xiii) High Tide
(xiv) BOD-5 [mg/l] (biological Oxygen in demand)
(xv) Phosphate [mg/l]

The above-described data sets are analysed collectively and are explained in the next section.

5.4 Methodology

Using the SOM techniques, data collected by AUT research groups to analyse the observed biodegradation and the environmental parameters, is analysed together with NSCC *Enterococci* test results (in *Enterococci* count/100mls) to look for patterns and relationships in these variables. The AUT data consists of information on the two different studies, namely, physical and biological (sciaphilic colonisation) systems along the chosen transect (figure 5.1).

The physical system parameters of the tidal zones, shown in figure 5.1 (S1, S2, S3 and S4), were monitored 5-6 times a month. Photographing of these plates to study the bio cover was carried out once a month. This created gaps in the data sets as the physical system data and photographing of the sciaphilic colonisation did not always coincide. In order to overcome this problem the following measures were taken:

(i) Missing values in the bio cover of sciaphilic colonisation data were filled with interpolated values by looking into their growth sequence from the photographs and based on domain expertise. For instance, bio cover on a day in week 2 was determined as 10% if the bio cover during week 1 and week 3 were 10%, assuming that no organisms grew and died in between or the died ones were replaced by same number of new ones.

(ii) An assumption was made that the changes in bio cover of live organisms on the slates were in proportion to their growth rates, meaning that all organisms on the plate were alive.

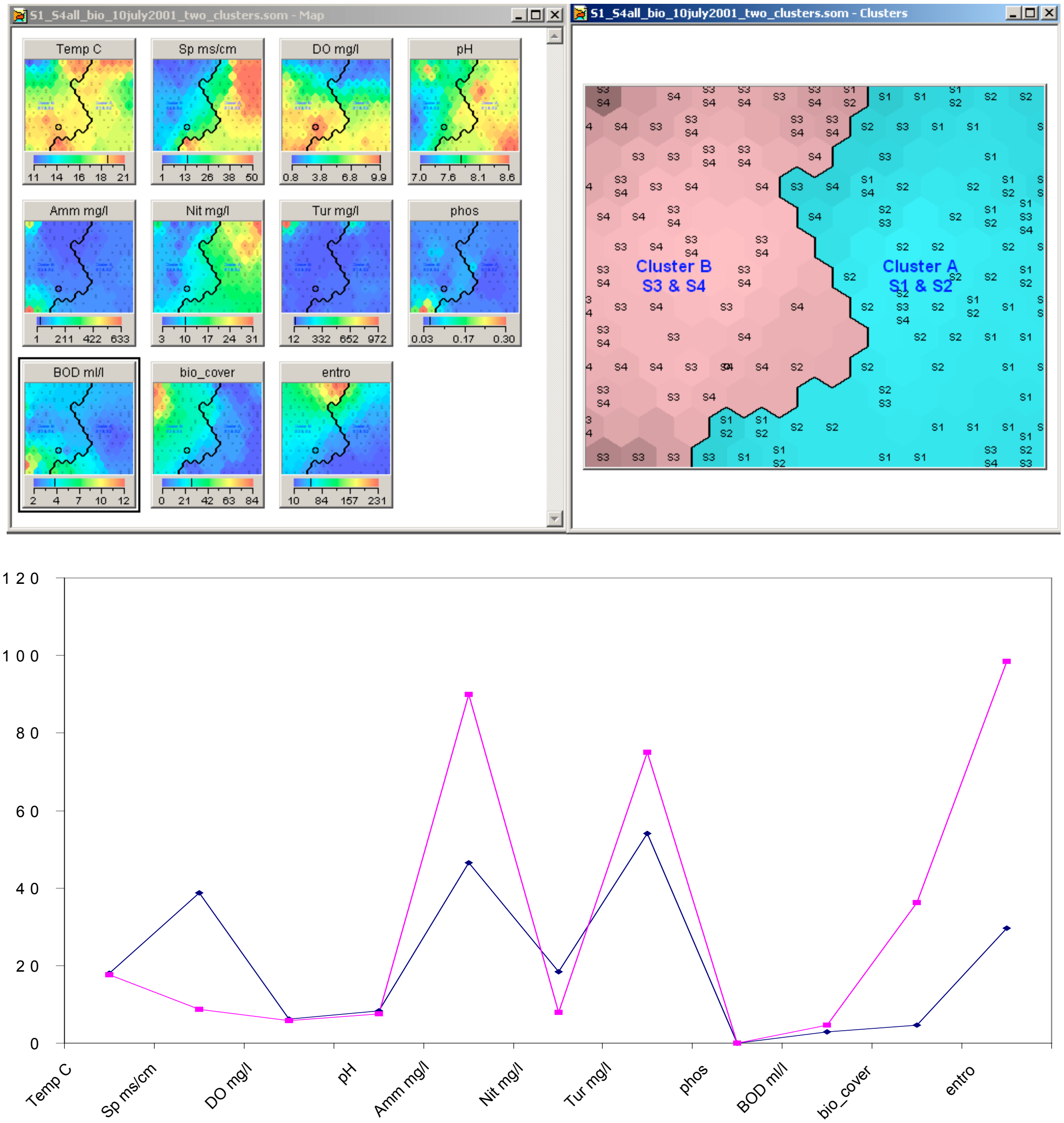
(iii) Visovery® SOMine lite version 2.1 by eudaptics software gmph package was used as it could create maps even with a few missing values.

5.5 SOM results

SOM based analyses (cluster, component plane, decision support and trajectory) are discussed herein.

5.5.1 Cluster analysis

Initially, a cluster analysis was carried out to see the clustering patterns in the S1-S4 slate tile data. A SOM (figure 5.2 a) was created using Viscovery with 200 nodes, priority for all components set to 1 and all other map creation parameters set to default values.



Cluster	S1 & S2 (Blue)	S3 & S4 (Red)
Component	Mean	Mean
Temp C	18.04	17.66
Specific Conductivity ms/cm	38.72	8.65
DO mg/l	6.16	5.89
pH	8.31	7.57
Ammonia mg/l	46.6	90
Nit mg/l	18.47	8.04
Turbidity mg/l	54	75
Phosphorous	0.0413	0.0671
BOD ml/l	2.89	4.62
Biocover	4.73	36.22
Entrococci	29.6	98.5

Figure 5.2 a: SOM of the Long Bay-Okura Marine Reserve's physical and biological system data with all components. b: Graph showing the SOM clusters details c SOM cluster details.

Sub divisions (S1-S4) of the intertidal zone (figure 5.1) have different topography and microclimate, thus possess characteristic zoning of sciaphilic organisms. Interestingly, the SOM (figure 5.2) has picked up these environmental variations, categorising the data into two distinct areas. It should be noted that S1 is the closest station to the land and as the number increases exposure time to seawater increases. The SOM component windows (figures 5.2 a) illustrate the relationships among the system variables. Interpretations derived from this map are:

(i) The data collecting monitoring stations S1 and S2 (cluster A) are separated from S3 and S4 (cluster B) by a diagonal line running from the upper right to the lower left of the map except a few S3 and S4 data points in the S1 and S2 area and a few more crossovers along the separation line. No data on their location was included in the map creation process, yet the self-organisation of the SOM shows them separated into two categories, which confirms the microclimatic conditions within this zone as distinguishable. Also shows SOM feature extraction abilities from raw data.

(ii) The classification of S3 & S4 data in the S1 & S2 cluster of the map (25-Nov-99 S3, 27-Oct-99 S3, 05-Oct-99 S3, 27-Oct-99 S4, 25-Mar-99 S4, 06-Oct-99 S4, 06-Oct-99 S3, 28-Mar-01 S3 28-Mar-01 S4), indicates that during these days S3 & S4 show similar attributes to S1 & S2 monitoring stations.

(iii) Component planes of specific conductivity (Sp Cond), pH, Nitrate, bio cover and *Entrococci* count show a corresponding correlation to the clusters map, implying that specific conductivity and pH are higher in S1 and S2 monitoring stations than in S3 and S4. Under normal conditions pH and specific conductivity of S1 and S2 should be lesser than that of S3 and S4 as the waters near land get more diluted and tend to exhibit attributes more similar to that of freshwaters. Expert advice in this regard clarified the unusual nature as being caused by the evaporation of water due to the high temperatures experienced in S1 and S2, in turn resulting in higher concentrations of salts, with increased pH and specific conductivity values.

(iv) By comparison, DO values are high in S1 & S2 area than in S3 & S4. DO values depend on the temperature and amounts of nutrients (largely nitrogen) in aquatic systems. DO levels in aquatic systems decrease with increasing water temperature, however, temperature readings in this case are 18.04 at (S1 & S2) and 17.66 at (S3 & S4). Hence, the biological decomposition of the nutrients in coastal waters (such as nitrogen and phosphorus received from runoff) could have possibly caused the lowering of DO in S3 & S4. During the biological decomposition, ammonia is transformed into nitrates utilising dissolved oxygen. High ammonia at S3 & S4 with (90 mg/l) and at S1 & S2 with (46.6 mg/l) supports the argument. Therefore it could be concluded that the need for excess oxygen to breakdown the high ammonia into nitrate in the former had resulted in low DO levels or hypoxia. Hypoxia, especially below 5mg/l could harm larval life stages of fish and shellfish species (Wilcock and Stroud 2000). Sessile species will not survive.

(iv) Bio cover is higher in S3 and S4 than in the other two. This shows that the conditions in S1 & S2 as unfavourable for the key species *Chamaesipho columna*, *Chamaesipho columna* and *Epopella plicata* or the colonisation was disturbed. Nonetheless, conditions in S3 & S4 seem to be favouring one or more key species, *Austrominius modestus, Crassostrea gigas, Pomatoceras caeruleus,* Cyanobacteria, *Austrominius modestus, Pomatoceras caeruleus* and *Xenostrobus pulex* of this division. From the SOM, no conclusion could be arrived, as to exactly what species thrived under these conditions, as community structural details were not included in the analysis. However, Buckeridge (1999) stated that it is the cyanobacteria that seem to be flourishing at the reserve. The study pointed out that in the past, on two different occasions between 1996-1999, during warmer periods, high nutrient levels produced eutrophic conditions, which in turn had affected the presence of *E. plicata* and the normally prolific intertidal algae *Corallina officinalis* both resulting in cyanobactria colonies dominating the reserve's ecosystem.

5.4.2.2 Decision support system

By superimposing the interpretation details of the cluster map, a decision support system (figures 5.3 c & d) was developed to visualise the relationships between the collected environmental and bio cover data. The following were derived from the SOM component analysis and clustering details (figures 5.2 a & b):

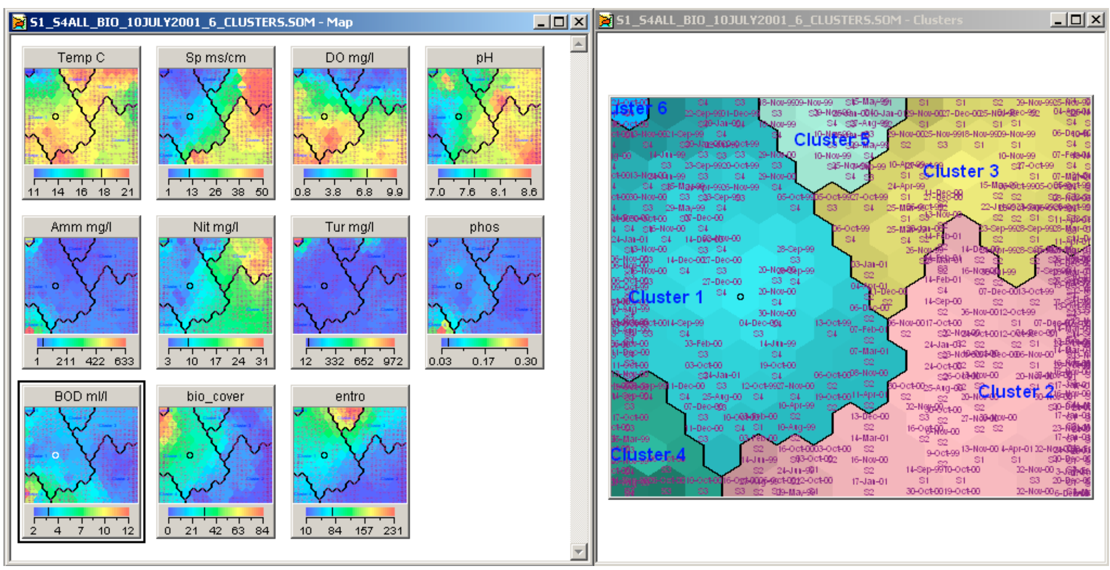
(i) S1 & S2 of SOM (figure 5.2 a) can be further divided into two major clusters (clusters 2 & 3). The two show the subtle changes within the variables in this littoral zones.

(ii) S3 & S4, the left diagonal half is subdivided into four clusters (1, 4, 5 and 6 in figure 5.3 a) with remarkable variations in ammonia and nitrate levels.

(iii) Cluster 4 shows the highest values for BOD (8.08ml/l), DO (8.17 mg/l) and phosphate (0.12), reasonably high temperature, ammonia and ammonia/ nitrate ratio, indicating a polluted scenario within S3 & S4, at the reserve.

(iv) Cluster 5 consists of the lowest ammonia, nitrate and DO levels with the highest temperature and *Entrococci* values. The cluster also shows the highest temperature and *Entrococci* in the map; depicting the eutropic conditions experienced during warmer periods as argued by Buckeridge (1999).

(v) Clusters 4 and 5 reflect two different polluted scenarios observed in the reserve during summer.



Cluster	C1	C2	C3	C4	C5	C6
Component	S3&S4	S1&S2	S1&S2	S3&S4	S3&S4	S3&S4
Temp C	17.1	17.99	18.2	18.69	19.61	11.3
Sp ms/cm	8.73	38.12	40.55	11.22	5.92	1.55
DO mg/l	6.31	7.28	2.7	8.17	1.5	2.45
pH	7.62	8.34	8.23	7.25	7.77	7.79
Ammonia mg/l	50	50.6	34.1	232.3	4.6	428.9
Nit mg/l	7.67	16.96	23.17	8.14	8.15	15.93
Tur mg/l	57	55	53	44	49	972
phosphate	0.054	0.0395	0.0469	0.12	0.0524	0.0489
BOD ml/l	3.77	2.92	2.8	8.08	3.5	4.16
biocover	40.5	4.01	6.97	28.65	22.13	75.24
Enterococci	87.1	17.2	67.9	56.7	203.8	105.5

Figure 5.3 a: SOM of S1 - S4 data with six clusters and components. b: SOM cluster details

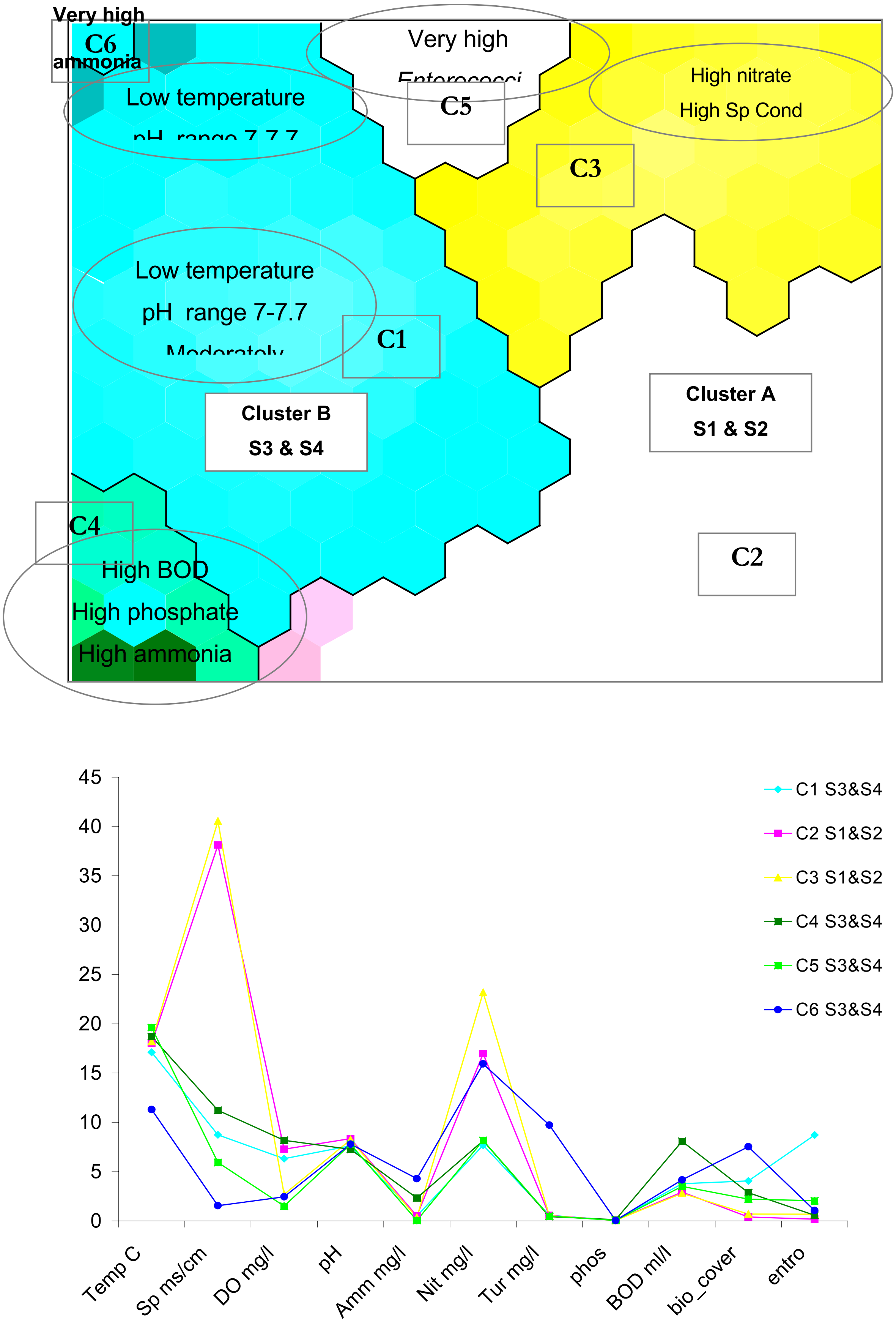


Figure 5.3 c: Superimposition of the (S1 - S4) SOM cluster details. d: Graph of the SOM cluster details.

(vi) Cluster 5 with 19.61° C temperature and very low DO (1.5 mg/l) has the highest *Enterococci* count (203.8), depicting a third polluted scenario of summer at the reserve. Please note that *Enterococci* tests are carried out on water samples from mid sea and not on water samples from the monitoring stations. Still considered as an indicator for pollution in seawaters in this study as by NSCC (North Shore City Council - 0800SAFESWIM). Figure 5.3 c, could be used as a look up table to analyse the reserve's intertidal divisions in that by finding the position of new data, the possible future trends could be assessed for early warnings on the reserve's ecosystem changes.

(vii) Cluster 6 shows the map's lowest temperature and the highest ammonia, fairly high nitrate, ammonia/nitrate ratio and *Entrococci* depicting a polluted scenario at the reserve during winter times.

5.4.2.3 Dependent component analysis

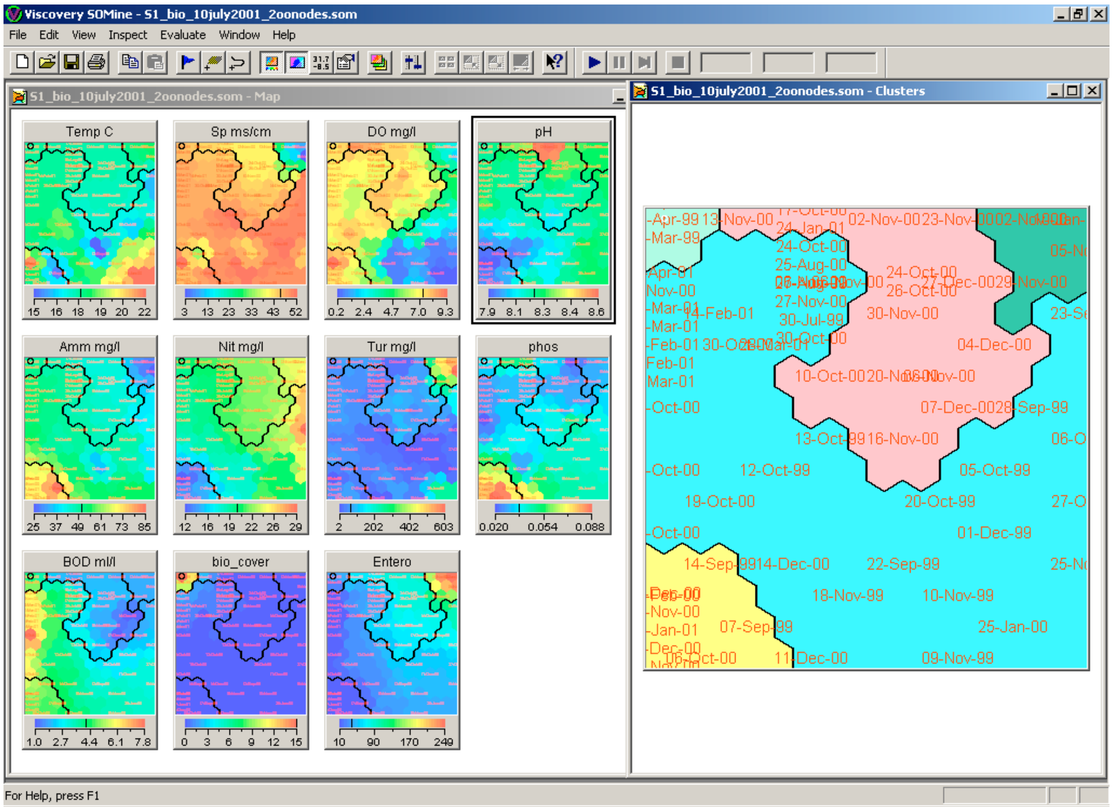
The SOMs and their components of the individual monitoring stations can be used to analyse the effects of urban development on the sciaphilic organisms (figure 5.4 a, b, c & d) within each individual intertidal division. The environmental changes that cause changes in the growth of these organisms in turn cause equilibrium shifts in the reserve's intertidal zone structure. These shifts could be revealed through the relationships in the SOM created with physical and biological data. The following are the results derived from the SOM dependent component analyses of the individual intertidal zone divisions S1 to S4:

SOM (figure 5.4 a) of the S1 data collecting monitoring station (25 March 1999 to 11 April 2001):

(i) Cluster 4 has the lowest ammonia/ nitrate ratio and the lowest DO in the map. It also consists of the highest *Enterococci* counts. The data points consist of 02-Nov-00, 10-Jan-01, 05-Nov-99 and 29-Nov-00.

(ii) Cluster 3 (14-Sep-99, 03-Feb-00, 07-Sep-99, 13-Dec-00, 16-Nov-00, 17-Jan-01, 20-Dec-00, 23-Nov-00, 24-Jan-01, 3-Jan-01, 6-Dec-00 and 06-Oct-00) has the highest values for ammonia, nitrate, ammonia/ nitrate ratio and DO.

(iii) Clusters 3 & 5 have high ammonia/ nitrate ratios. The difference between the two is the level of phosphorus, which may have caused the 0% bio cover in cluster 3, whereas in cluster 5 bio cover is 15%.



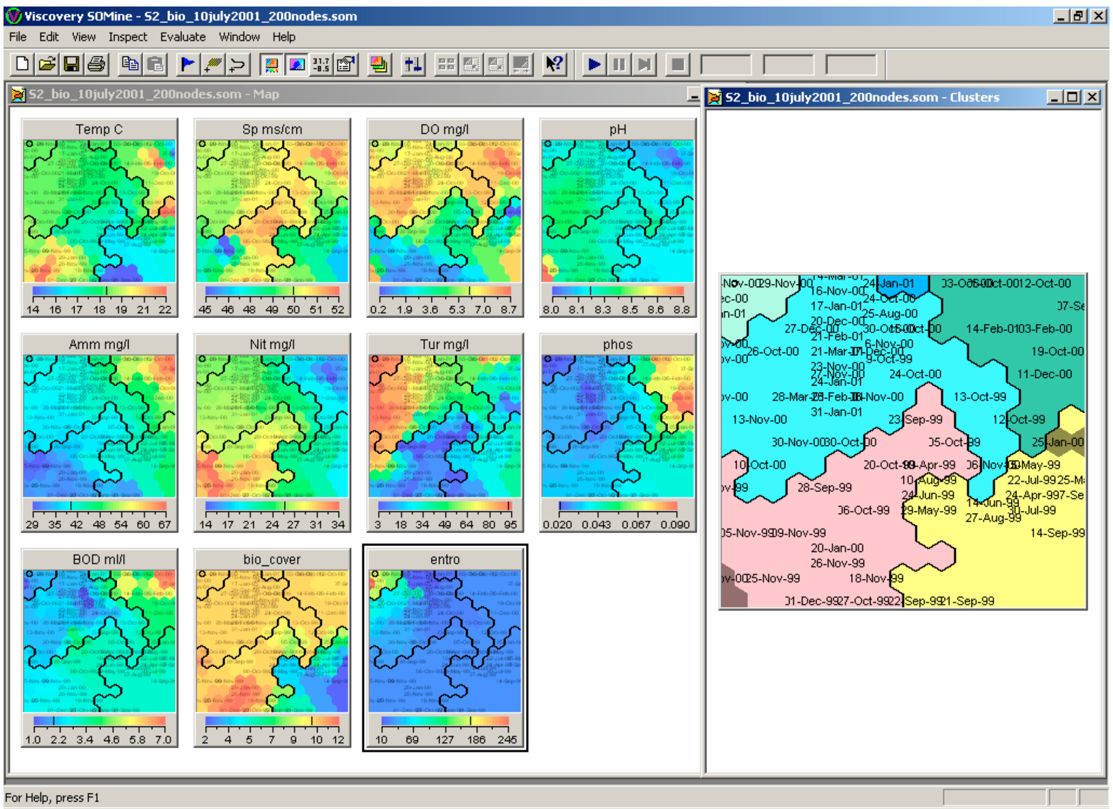
Cluster	C1	C2	C3	C4	C5
Component	Mean	Mean	Mean	Mean	Mean
Temp C	17.85	17.41	17.75	17.98	17.63
Sp ms/cm	47.94	42.56	45.38	13.29	45.08
DO mg/l	4.963	7.556	7.424	3.122	7.306
pH	8.193	8.306	7.988	8.283	8.215
Amm mg/l	49.83	45.32	71.8	29.64	50.4
Nit mg/l	20.51	21.91	13.9	24.3	20.1
Tur mg/l	52.7	74.5	51.7	471.9	60.9
phos	0.0371	0.0425	0.0759	0.0301	0.0367
BOD ml/l	4.225	1.985	4.868	1.995	4.188
bio_cover	0.39	0.09	0	0	15
Enterococci	36.4	73.2	10.7	214.3	37

Figure 5.4 a: SOM of lower supralittoral S1 zone data. Map parameters used are: 200 nodes, priority for all components set to 1 and other parameters were set to default values.

SOM (figure 5.4 b) of the S2 data collecting monitoring station (25 March 1999 to 11 April 2001):

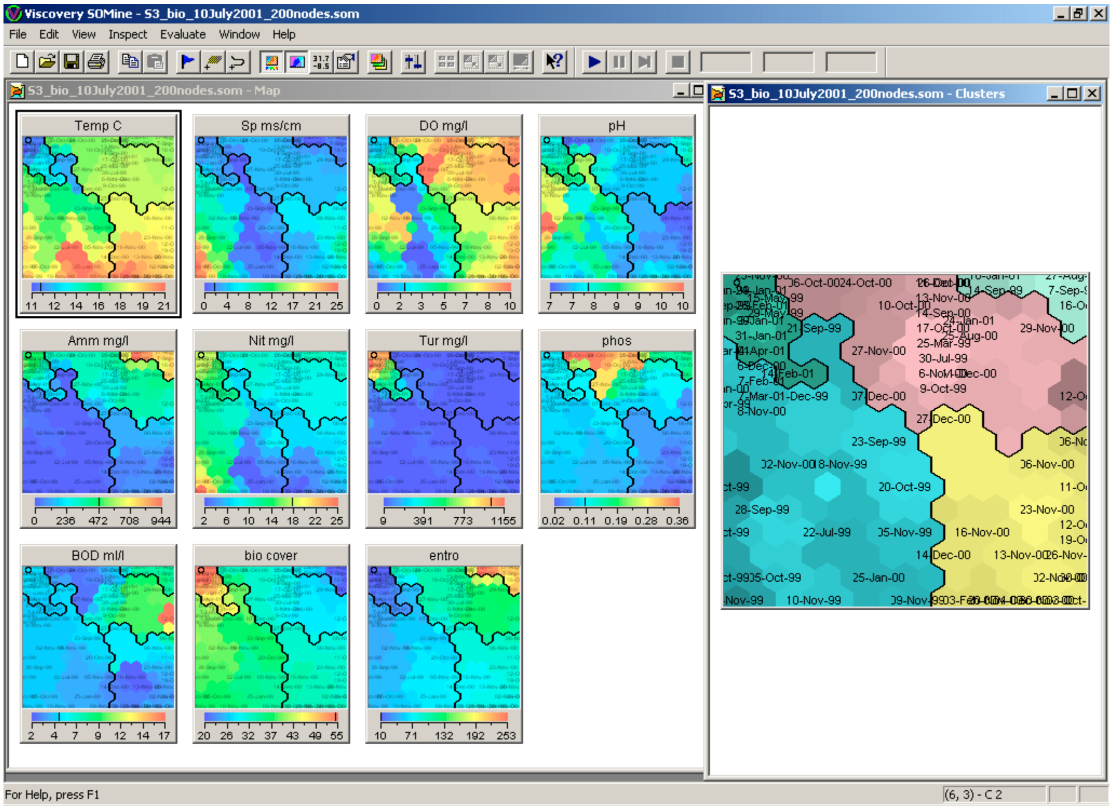
(i) Clusters 2 and 5 show the lowest ammonia/ nitrate ratio, however, only cluster 2 shows low DO at 2.639mg/l (harmful to aquatic life) and high *Entercocci* count (214.3), whereas cluster 5 shows high DO (6.513) and high turbidity (93.28 mg/l).

(ii) Cluster 4 has the highest ammonia/ nitrate ratio (55.71/17.71) combined with the highest DO (7.156 mg/l). It shows that the ammonia in this area has not been decomposed into nitrates yet, leaving the DO values high.



Cluster	C1	C2	C3	C4	C5	C6
Component	Mean	Mean	Mean	Mean	Mean	Mean
Temp C	18.4	18.87	16.83	18.79	18.53	18.4
Sp ms/cm	50.08	49.73	48.44	50.81	49.92	50.44
DO mg/l	6.771	2.639	3.799	7.156	6.513	5.95
pH	8.265	8.321	8.187	8.125	8.205	8.1
Ammonia mg/l	40.41	34.75	42.82	55.71	39.76	45.52
Nitrate mg/l	25.61	29.97	21.99	17.71	25.99	22.65
Turbidity mg/l	69.41	22.16	26.71	40.51	93.28	56
phosphate	0.03044	0.03253	0.04434	0.07277	0.02074	0.03
BOD ml/l	2.116	2.809	3.132	4.792	1.524	2.98
biocover	10	10.43	4.2	10	10	10
Entrococci	26.7	31.9	31.7	31	202.1	21.5

Figure 5.4 b: SOM of upper littoral S2 zone data. Map parameters used are: 200 nodes, priority for all components set to 1 and other parameters were set to default values.



Cluster	C 1	C 2	C 3	C 4	C 5
Component	Mean	Mean	Mean	Mean	Mean
Temp C	17.62	17.32	19.06	11.6	16.98
Sp ms/cm	8.27	4.09	6.43	2.03	0.75
DO mg/l	3.7	8.25	7.57	2.3	8.66
pH	8.3	7.46	7.29	7.84	7.72
Ammonia mg/l	12.2	317.1	86.8	459.2	835.9
Nitrate mg/l	10.36	8.92	6.06	17.18	10.45
Turbidity mg/l	39	29	30	990	18
phosphate	0.0896	0.1768	0.0527	0.1067	0.1687
BOD ml/l	4.54	6.94	4.05	5	8.04
bio cover	37.73	30	32.17	53.38	23.44
Enterococci	55.2	110.8	83.2	11.5	229.3

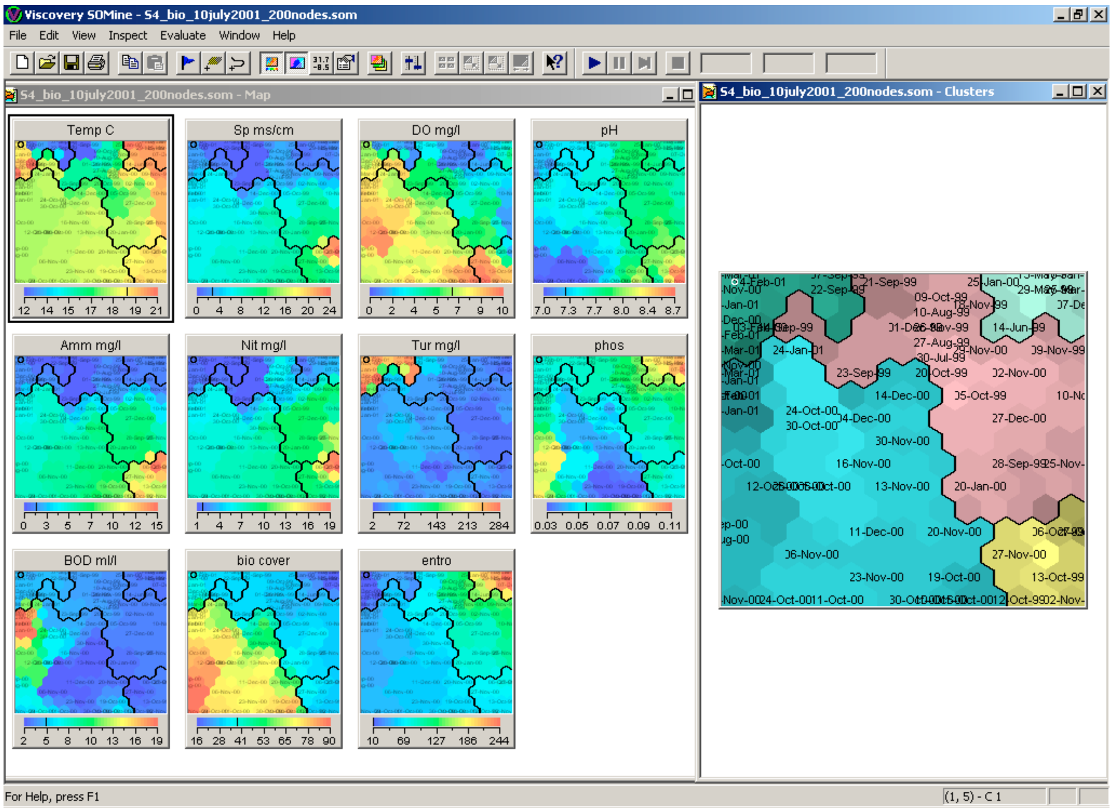
Figure 5.4 c: SOM of mid littoral S3 zone data. Map parameters used are: 200 nodes, priority for all components set to 1 and other parameters set to default values.

SOM (figure 5.4 c) of the S3 data collecting monitoring station (25 March 1999 to 11 April 2001):

(i) Clusters 2, 4 and 5 exhibit an unusually high ammonia, BOD and phosphorus compared with the other littoral division of the zone. Cluster 4 shows very high ammonia and turbidity with low temperature, showing rainy conditions.

Clusters 2 and 5 show similar conditions, high BOD and *Entrococci* counts except that cluster 5 has the extremes. Cluster 5 has the lowest specific conductivity within the reserve.

(ii) Cluster 1 has the lowest ammonia/ nitrate ratio along with low DO.



Cluster	C 1	C 2	C 3	C 4	C 5
Component	Mean	Mean	Mean	Mean	Mean
Temp C	18	17.02	19.01	18.16	20.28
Sp ms/cm	8.03	5.06	15.15	2.85	2.05
DO mg/l	7.7	2.95	7.11	5.82	2.62
pH	7.414	7.741	8.172	7.294	7.645
Ammonia mg/l	5.91	2.91	11.49	1.37	1.18
Nitrate mg/l	7.28	6.67	12.1	1.8	3.51
Turbidity mg/l	34.6	25.7	22.1	250	18.9
phosphate	0.0502	0.0653	0.0543	0.0554	0.0956
BOD ml/l	5.7	3.42	2.98	4.72	2.75
biocover	71.51	30.55	31.3	37.13	19.33
Entrococci	53	102.8	70.9	13	216.1

Figure 5.4 d: SOM of lower littoral S4 zone data. Map parameters used are: 200 nodes, priority for all components set to 1 and other parameters set to default values.

SOM (figure 5,4 d) of the S4 data collecting monitoring station (25 March 1999 to11 April 2001):

(i) Ammonia, nitrate levels and the ratio are very low, compared to the other intertidal divisions. Clusters 2 and 5 exhibit the lowest ratio along with low DO and high *Enterococci* counts and phosphorus values. They also show low bio cover within the division indicating that such conditions as harmful to sciaphilic growth.

(ii) Cluster 1 has the highest bio cover even though BOD is high as 5.7 ml/l.

(iii) Clusters 3 and 4 similar attributes except for specific conductivity at 15.15, 2.85 ms/cm and BOD at 2.98, 4.72 ml/l.

5.4.2.4 Trajectories

The reserve's ecosystem dynamics could be analysed with the use of a trajectory on the SOM of S1-S4 monitoring station data. Shown in figure 5.5 is the trajectory depicting the process dynamics of S1 monitoring station in the overall SOM of S1-S4.

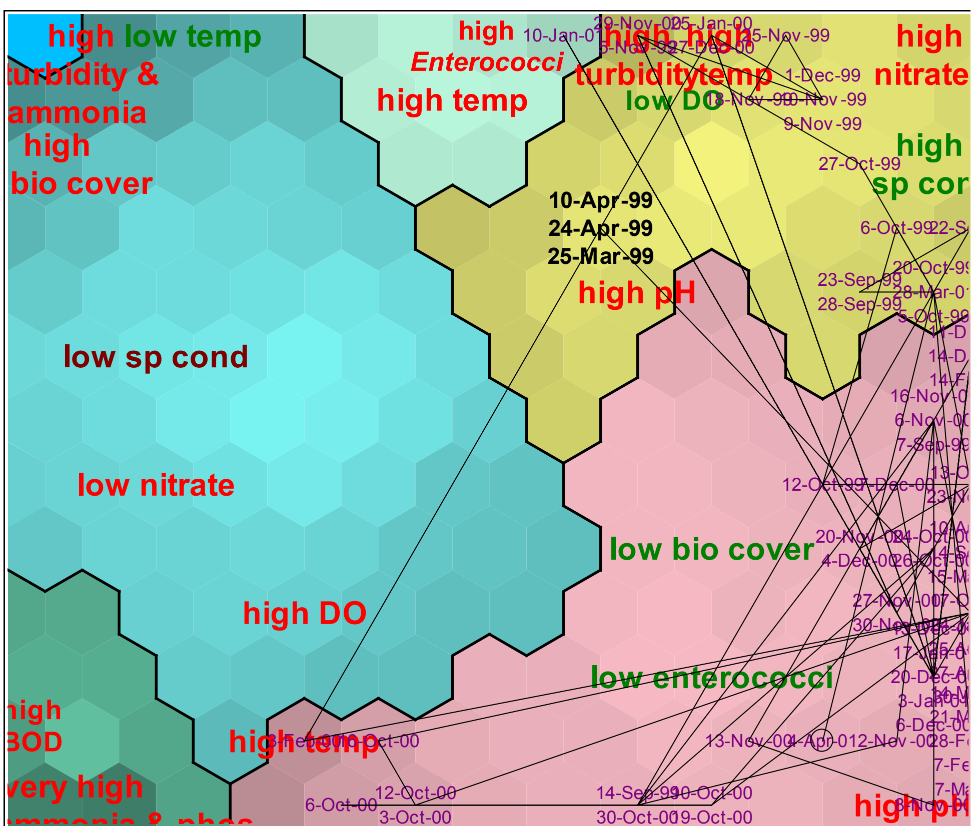
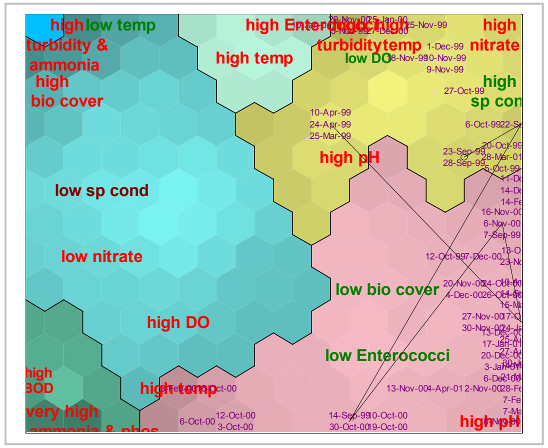
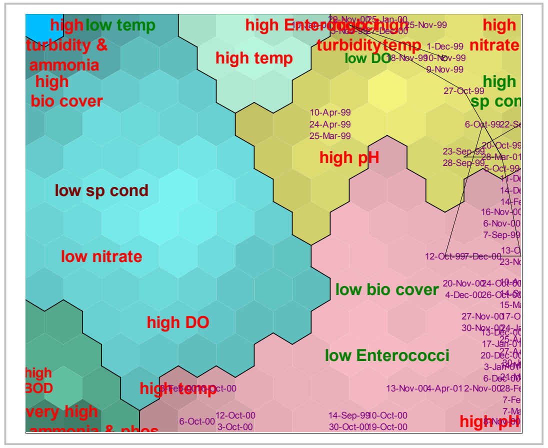


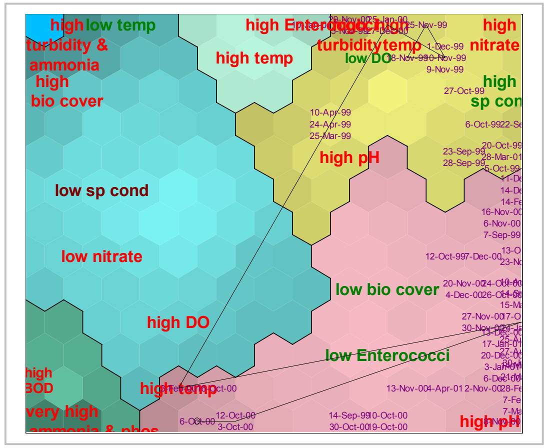
Figure 5.5 a: Trajectory of S1 data on S1- S4 SOM. Map parameters used for creation of maps are: 200 nodes, priority for all components set to 1 and other parameters set to default values.



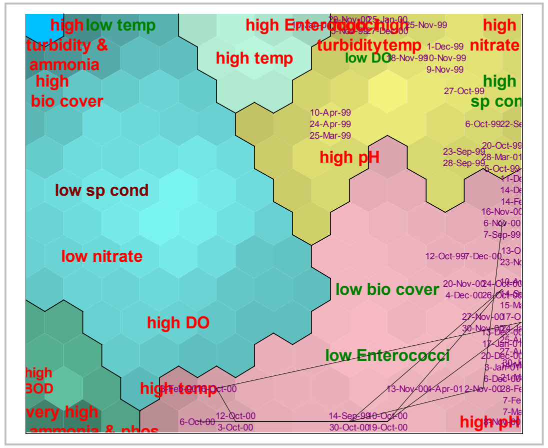
1-10



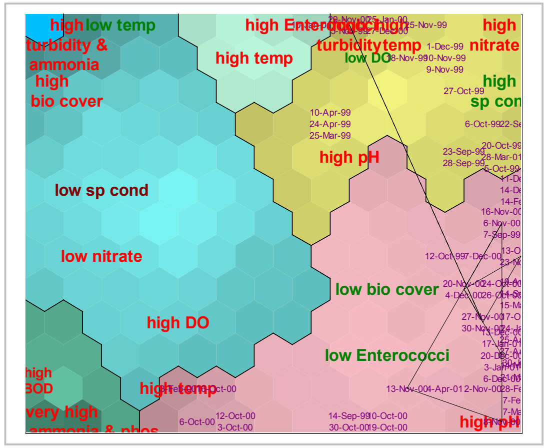
11-20



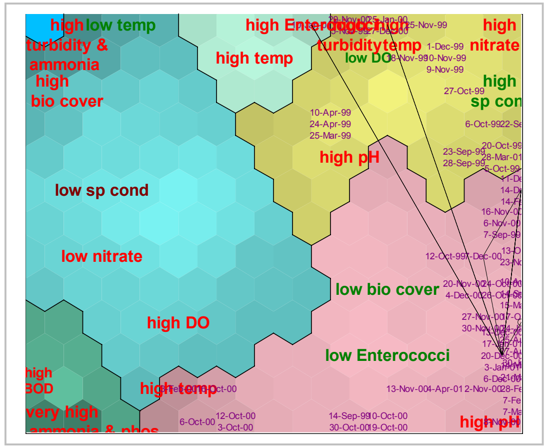
21-30



31-40



41-50



51-60

Figure 5.5 b: Trajectory of S1 data on S1- S4 SOM with 10 days per map

(1-10) 5-Mar-99, 10-Apr-99, 24-Apr-99, 15-May-99, 30-Jul-99, 10-Aug-99, 27-Aug-99, 7-Sep-99, 14-Sep-99, 22-Sep-99,

(11-20) 23-Sep-99, 28-Sep-99, 5-Oct-99, 6-Oct-99, 12-Oct-99, 3-Oct-99, 20-Oct-99, 27-Oct-99, 5-Nov-99, 9-Nov-99,

(21-30) 10-Nov-99, 18-Nov-99, 25-Nov-99, 1-Dec-99, 25-Jan-00, 3-Feb-00, 25-Aug-00, 14-Sep-00, 3-Oct-00, 6-Oct-00,

(31-40) 10-Oct-00, 12-Oct-00, 16-Oct-00, 17-Oct-00, 19-Oct-00, 24-Oct-00, 26-Oct-00, 30-Oct-00, 2-Nov-00, 6-Nov-00,

(41-50) 8-Nov-00, 13-Nov-00, 16-Nov-00, 20-Nov-00, 23-Nov-00, 27-Nov-00, 29-Nov-00, 30-Nov-00, 4-Dec-00, 6-Dec-00,

(51-60) 7-Dec-00, 11-Dec-00, 13-Dec-00, 14-Dec-00, 20-Dec-00, 27-Dec-00, 3-Jan-01, 10-Jan-01, 17-Jan-01, 24-Jan-01,

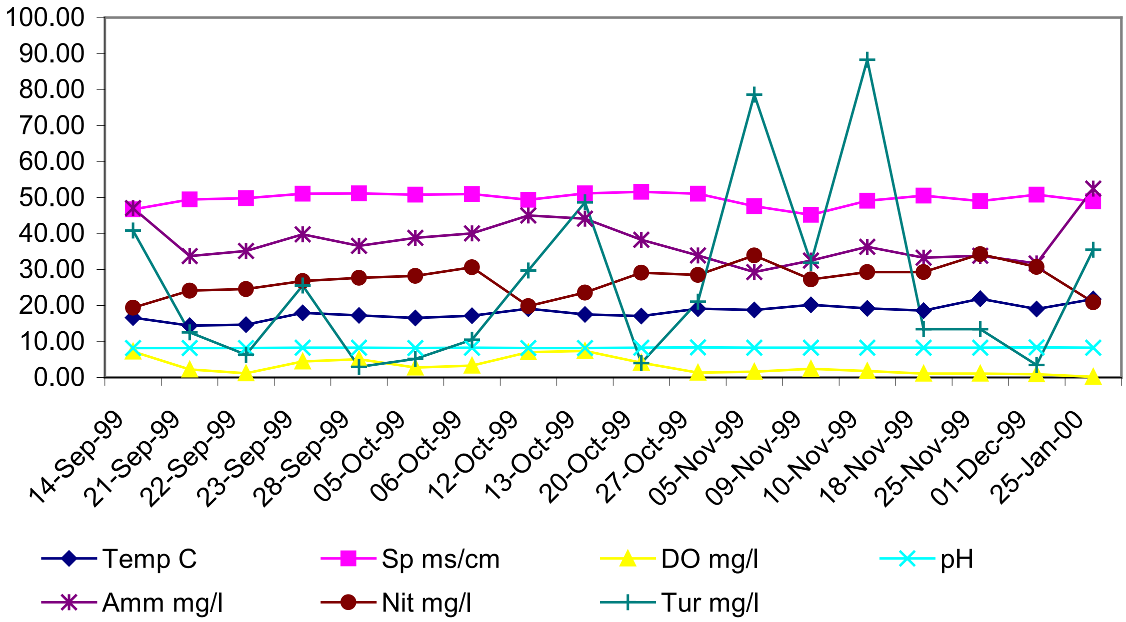
(61-69) 7-Feb-01, 14-Feb-01, 28-Feb-01, 7-Mar-01, 14-Mar-01, 21-Mar-01, 28-Mar-01, 4-Apr-01

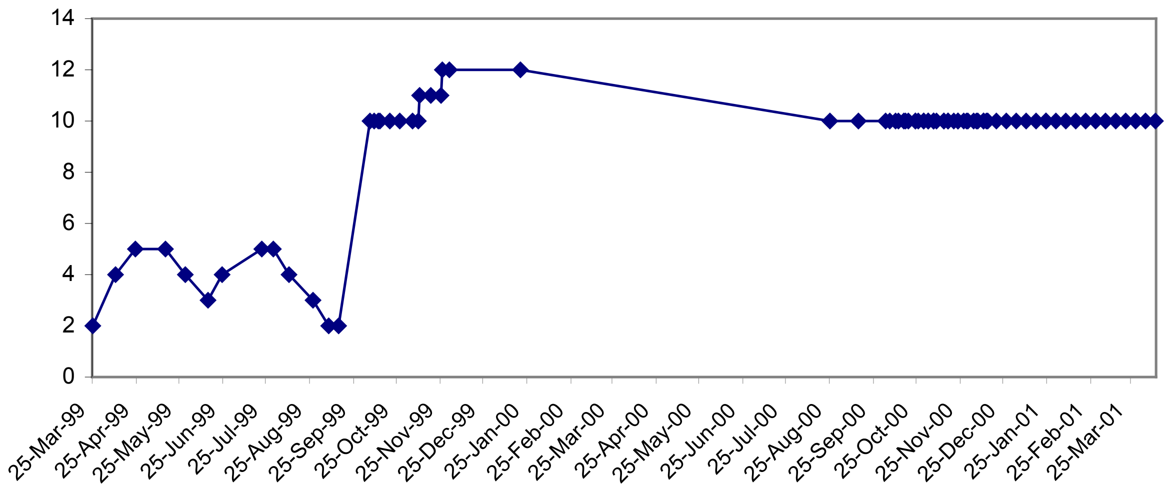
Figure 5.5 c: Trajectory of S1 data for the last 9 days on S1- S4 SOM, d: Days included in the SOM trajectories of figures 5.5 b and c.

The SOM trajectories (figures 5.5 a-d) show how they could be applied to modelling the reserve's intertidal ecosystem dynamics. Time intervals (11-20), (31-40) and (41-50) depict a smooth flow on the S1 data on the SOM of S1-S4 data. In (11-20) 27-Oct-99 shows the trajectory moving towards low DO with high *Enterococci* area. The other intervals do not show a smooth pattern as in the earlier; this is because of the data gaps that can be seen in figure 5.5 d. It could be stated that data collected on a regular basis could be analysed to study the systems dynamics using SOMs.

5.5 Conventional analyses

Data collected from the Long Bay-Okura Marine Reserve for a period between 25 March 1999 and 11 April 2001, used in this case study, was originally collected for projects carried out by student groups (Higgs 1999; Snowdon 1999; Meyer 2000; Scharader 2000; Hecht 2001). Using these data sets the students carried out their own analyses to study the reserve's physical and biological system changes separately with conventional methods. Hence, analysed these system data sets (of interdital divisions S1 to S4) along the chosen transect (figure 5.1) by plotting against time separately and discussed the trends in them during this period. Figures 5.6 a & b are examples of such graphs of the physical and biological system data of plate S2.





Figures 5.6 a & b: Graphs of the physical and biological system data of S2.

Due to huge gaps in these data sets they could not be used to study the reserve's dynamics with a systems approach using conventional analyses. Because of this reason, the reserve's physical and biological system data sets, collected by AUT student groups, NSCC and ARC have never been collectively analysed. However, (Higgs 1999; Snowdon 1999; Meyer 2000; Scharader 2000; Hecht 2001) studied the physical and biological systems by visually comparing the observed patterns in the graphs.

5.6 Discussion

SOMs can be used as an excellent tool for ecosystem modelling, as they are capable of depicting spatial and temporal variations within the monitoring data. This was illustrated by the segregation of S1 and S2 from S3 and S4, within the reserve's intertidal littoral divisions based on the system variables represented in numeric monitoring data, without any direct information. The segregation of the reserve's intertidal zones into two distinctive areas shows SOM feature extraction, which can be used for modelling the slow ecosystem variables that cannot be analysed using conventional methods. The conventional methods have in fact led ecologists and decision makers to assume that environmental effects as being mitigated or neutralised by the system (Clark et al. 2001) (see chapter 3).

SOMs provide a means to collectively analyse the physical and biological systems changes of a natural habitat with a systems approach. Relationships among the physical and biological system variables of the reserve's four intertidal littoral zones (S1 to S4) were analysed with visual formats.

(i) The lower supralittoral and Upper littoral (S1 and S2) zones showed low ammonia/ nitrate ratios in comparison to the other two (S3 and S4) zones, linking the biological decomposition of ammonia into nitrate using DO.

(ii) The correlation between low bio cover and high BOD, DO, temperature, phosphate and ammonia shows that the biodiversity is directly affected by the physical changes in the mid and lower littoral zones (S3 and S4 within the intertidal zone). The bio cover in lower supralittoral and Upper littoral zones (S1 and S2) that shows no significant change over the period (even with changing temperatures). This may have been due to either of the following:

a. time needed for settling of organisms as the S1 tile was reaffixed due to vandalism, or

b. the physical changes did not harm the biological system.

The former is considered the most likely.

(iii) The S3 and S4 zones, during warmer periods, with high DO, BOD and phosphate, seen with low bio cover, as opposed to the same zones with lower temperature values, DO and BOD seen with high bio cover, confirm earlier studies of the reserve's seasonal variation. During warmer conditions, waters entering the sea, with increased nitrogen wastes have led to algal and bacterial booms and ultimately resulted in eutrophication, leading to structural changes in reserve's biological diversity (Buckeridge 1999).

(iv) In the S3 and S4 zones, high ammonia/ nitrate at low temperatures are seen with increased bio cover, and also with increased *Entrcocci* (105.5/100ml).

Using SOMs the different scenarios within the intertidal zone (S1 to S4 littoral divisions) were analysed collectively and individually. The individual maps enabled the detection of S3 as the division with the highest ammonia/ nitrate ratio within the intertidal zone. This was indicated in the collectively analysed map, which had three distinctive clusters within the S3 & S4 cluster.

SOMs could also be used to analyse ecosystem trends where full data sets are unavailable. It should be noted that although BOD tests were not carried out in 1999, the SOM generalisation abilities have clustered them, with appropriate BOD based on their other available attributes.

5.7 Conclusion

SOM analyses provide a means to relate and analyse the environmental changes with biological responses in visual formats. The patterns in the data could be analysed directly linking the causal processes and the environmental effects within an ecosystem framework.

The results support the hypothesis to apply SOM based complex industrial process modelling to ecosystem processes (environmental and biological) with a systems approach. The fused data of physical and biological systems used in this study had missing values at critical points as different research groups had collected them on different days. Not withstanding, the SOMs created with interpolated values were able to reveal the relationships among the variables and their changes. It should be noted that any other constructive analysis could not be carried out on these fused data sets using conventional statistical methods. Previous research groups carried out their own analyses of data collected by themselves and studied the physical and biological systems separately.

5.8 Future work

Efforts are being made to study the biological assemblages of the S1-S4 data collecting monitoring stations at the Long Bay-Okura Marine Reserve with the use of SOM techniques to reveal the trends in biodiversity over the years, since 1996 to present time. Also adding rainfall data could reveal more relationships and insights into the interactions between the discharge of nitrogen waste and *Enterococci* counts at the reserve.

Summary

The approach to use different SOM methods as used in highly complex, industrial system dynamics modelling and dissimilar financial data analysis (based on the hypothesis explained in experimental methodology chapter), produced promising results. In particular, the cluster analyses, dependent component analyses and decision support systems offered methods to collectively analyse numerical data sets of physical and biological systems of the Long Bay-Okura Marine Reserve without any physical models. In this example, the trajectories did not produce results similar to those of industrial system process dynamics modelling. This was mainly due to the high time intervals of the reserve's fused data set, used in the case study.

The next chapter illustrates a second case study for this research, where data sets from Auckland Regional Council, are analysed to further investigate the use of SOM methods in coastal environmental and biological system modelling. These methods provide a means for quantitative analysis in ecosystem dynamics modelling by relating the urbanisation indicators (developmental activities) and their effects on the coastal environment. The observed biodegradation from selected northeastern beaches of Auckland, New Zealand is also studied using monitoring data. The examples of the next chapter are at regional scales, wider than the example of this chapter.

Chapter 6

SOM techniques in ecosystem modelling

The previous chapter illustrated SOM abilities to analyse a complex natural system, the Long Bay-Okura Marine Reserve, north of Auckland, using dissimilar data sets. SOM applications to analysing correlations between environmental parameters and biological indicators within regional monitoring data from northern New Zealand are elaborated in this chapter. The SOM results are then compared with those of the conventional analyses carried out either in this research or from the respective reports by the institutions originally gathered the data.

With the escalating human influenced environmental deterioration, the need for enhanced techniques to model highly complex and diverse natural systems and to predict their process dynamics appears to be critical for the continued wellbeing of humanity. A large volume of past and recent scientific papers reviewed in chapter 3, reveal the issues and the need for enhanced modelling techniques to analyse cryptic ecosystems. SOM methods are investigated for ecosystem modelling in this research as applied in industrial system modelling and financial data analysis based on earlier SOM approaches in ecological studies at limited scales and levels.

6.1 Background

The Auckland Regional Council (ARC), North Shore City Council (NSCC), National Institute of Water and Atmospheric Research (NIWA), and the University of Auckland (UoA) have been carrying out monitoring programmes for various purposes, such as to monitor the water quality, to develop prediction models and to detect the effects of urbanisation on the receiving coastal environment. These programmes are mainly aimed at either gaining more knowledge or making predictions of the analysed ecosystems under different developmental scenarios. Data sets collected through the monitoring programmes by these local authorities as well as research and academic institutions on the environmental and biological systems of the coastal habitats of northern New Zealand are elaborated upon.

6.1.1 Data from city councils

ARC and its predecessors have been monitoring the water quality of freshwater streams and saline harbour sites as part of their Long-Term Baseline (LTB) programme since the mid 1980s to study the trends and effects of human activities on these water sources. Of the documents released on this programme, in (Wilcock and Stroud 2000) a report on the monitoring of 16 streams, 18 saline water sites in Manukau, Waitemata and Kaipara Harbours, and seven lakes, is included. Of this LTB programme, only the saline water quality data, sampled on a monthly basis from the 11 sites listed below, covering a period of ten years, from May 1991 to December 2000 was made available for this research. These beaches lie on the east coast and some within the Waitemata Harbour (figure 6.1), north of Auckland:

- (i) Browns Bay
- (ii) Chelsea
- (iii) Goat Island
- (iv) Henderson
- (v) Hobsonville
- (vi) Kaipara (Shelly Beach)
- (vii) Kawau Bay
- (viii) Mahurangi
- (ix) Orewa
- (x) Ti Point
- (xi) Wha

In addition, deep sea water samplings collected from different sites also have been included in the monitoring data. The following are the numeric data elements included in the ARC's LTB programme:

- (i) Site
- (ii) Site#
- (iii) pH
- (iv) Temperature (deg C)
- (v) Suspended solids (SS) (mg/l)
- (vi) Turbidity (NTU)
- (vii) Chloride (mg/l)
- (viii) Salinity (ppt)
- (ix) Total phosphorus (mgP/l)
- (x) Dissolved reactive phosphorus DRP (mgP/l)
- (xi) Nitrite (mgN/l)
- (xii) Ammonia (mgN/l)
- (xiii) BOD (mgO/l)
- (xiv) Total Coliforms (mpn/100ml)
- (xv) Faecal Coliforms (mpn/100ml)
- (xvi) Dissolved oxygen (DO %)
- (xvii) Dissolved oxygen (DO ppm)
- (xviii) Secchi disk depth*
- (xix) Chloride*
- (xx) *Enterococci* ME* (cfu/100mL)
- (xxi) NO_2*
- (xxii) NO_3 NO*

* Refers to the tests that were not carried out throughout the monitoring period

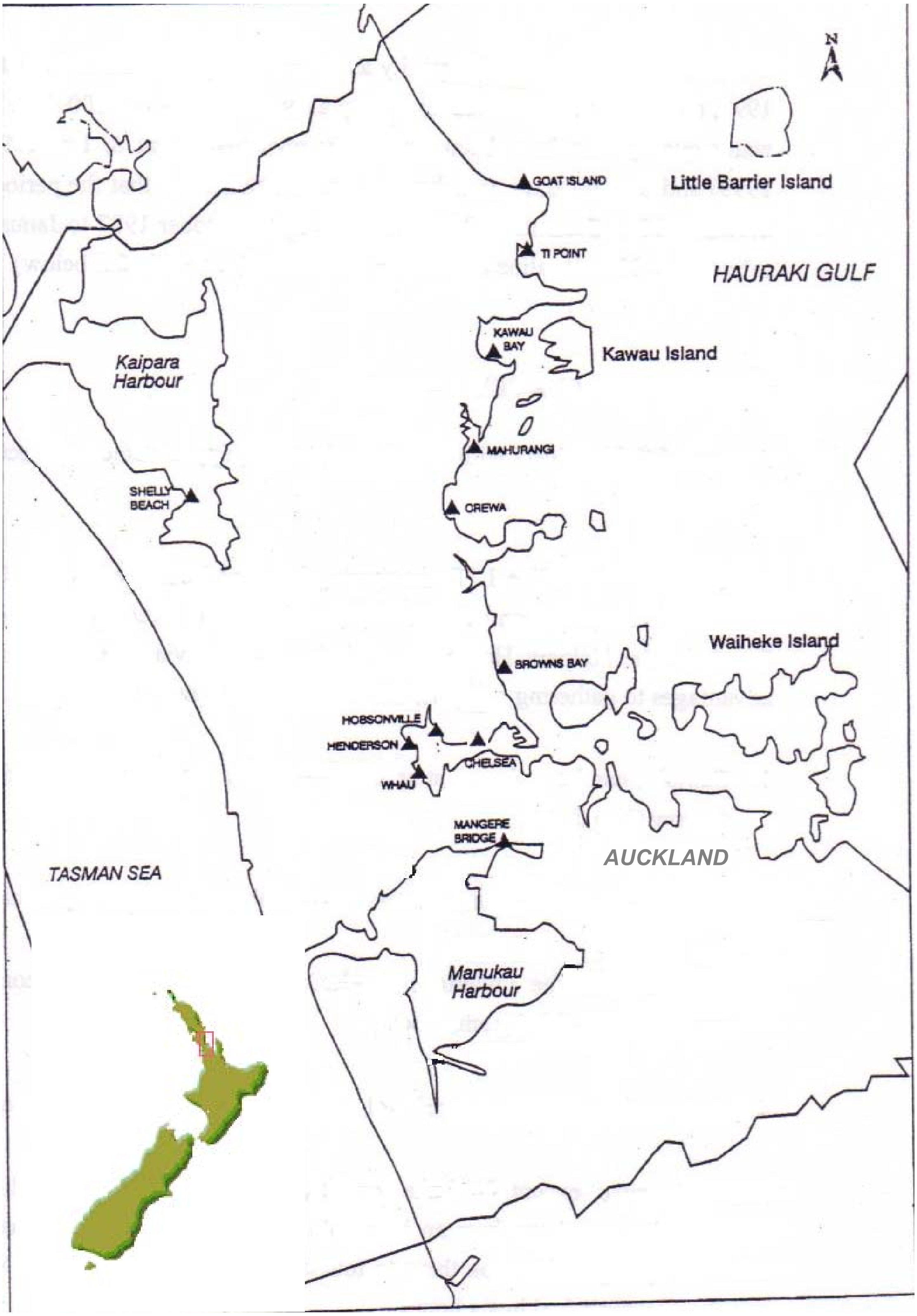


Figure 6.1: Auckland saline water quality monitoring sites. Source (Wilcock and Stroud 2000:3)

Many regional and city councils carry out *Entrococci* tests on water samples taken from beaches within their administrative confines. The Ministry of Environment (1999) and these state institutions consider this bacteria as the indicator most closely correlated with health effects in New Zealand marine waters. The Ministry aims to have a running median of less than 35 *Enterococci* per 100 mls of water, and less than 277 *Enterococci* per single water sample. Beach water that exceeds two consecutive levels of 277 *Enterococci* per 100 mls of seawater in a single sample is considered as unsafe for bathing (North Shore City Council - 0800SAFESWIM). *Enterococci* bacterial levels are regarded as the primary indicator of levels of pollution in this analysis.

Bacteria are derived from a range of authorised activities such as sewage and control, storm water discharges, land use, sediment control, works in watercourse. Details on these permits and other permits such as earthworks, vegetation removal, quarrying and other activities in the coastal marine area are considered as indicators for developmental activities. Data relating to these activities was obtained from the relevant local and regional government authorities, such as building consent.

In addition to the saline water quality tests, ARC also carries out biomonitoring programmes. Of particular interest to this research are the subtidal and intertidal monitoring programmes carried out by UoA's postgraduate and postdoctoral researchers attached to the Leigh Marine Laboratory. The aim of the programme was to analyse the rates of sediment deposition and their impact on species abundance and composition in the coastal habitats from Waiwera to Campbells Bay (figure 6.11). An ARC report by Walker et al. (2000) concluded that Long Bay and Campbells Bay were having the highest sediment deposition rates among the monitored sites. Further, noted a continued lack of bivalve recruitment, such as pipi and tuatua, at Long Bay and low numbers of these species recorded across the six monitoring sites.

6.1.2 Data from NIWA's research

NIWA has produced models to predict the sediment runoff during the current and two different future development scenarios at Long Bay. The two scenarios discussed in the report were:

(i) mixed development: including areas of intense development (one dwelling per 200 to 325 m^2), a few areas at conventional densities (one dwelling per 600 m^2) and major areas at large lots and lower rural residential densities (one dwelling per 2000 m^2 and up).

(ii) fully urbanised: with one dwelling for per 600 m^2.

The study made predictions on the daily sediment generation for different return periods (such as 0 to 30 years) and for different development stages (such as 10 and 20 years) and is based upon (Green et al. 2000).

6.2 Objectives

The following are the objectives of this case study on the use of SOM techniques for ecosystem modelling with data from the above illustrated coastal system, north of Auckland:

(i) to develop a visualisation means to analyse the environmental changes and the observed biological responses such as population dynamics, and

(ii) to study the relationships in the above data, in particular to establish any relationships between the variables. This could be applied to analysing ecosystem responses and trends between the causal process and the environmental effects, in the form of patterns, structures and interactions within the data. Using this approach environmental and biological system data along with economic data could be modelled within an integrated framework.

The first example illustrates the implementation of the SOM approaches investigated here to analysing the LTB programme data on beaches, north of Auckland. The second example examines the possibilities of studying the correlations between ARC's sediment deposition rates and subtidal community dynamics from selected beaches of northeastern coast, off Auckland using SOMs. Both examples analysed in this chapter consist of data from larger areas than that of the Long Bay Okura Marine Reserve, studied in the earlier chapter.

Within an ecosystem, a change in the environmental or physical process is mediated through parameters such as sediment deposition rates, chemical and physical property changes of water. Such an environmental change invariably causes a chain of changes on the biological system, which could be observed in the form of changes in biodiversity, such as species composition, and species motility rates (at the community level). Biological system changes could also be studied in the form of deformed organs, lesions, pollutant accumulation in tissues (in individual organisms), and in cells from changes in

DNA structures (at levels further below within an individual organism). Even though many models and concepts of indictor species and community dynamics are capable of detecting these changes, a need for methods to analyse the ecological data at ecosystem level has been emphasised for sustainable environment management (see chapter 3).

6.3 ARC saline water quality data analysis

The ARC data on water samples, collected on a monthly basis from 11 beaches lying northeast of Auckland (figure 6.1), through the LTB saline water quality programme is used in this section. The SOM analyses on this data to study the patterns in them are explained. Following this, SOM results are compared with those of conventional data analysis methods carried out on the same data.

6.3.1 Methodology

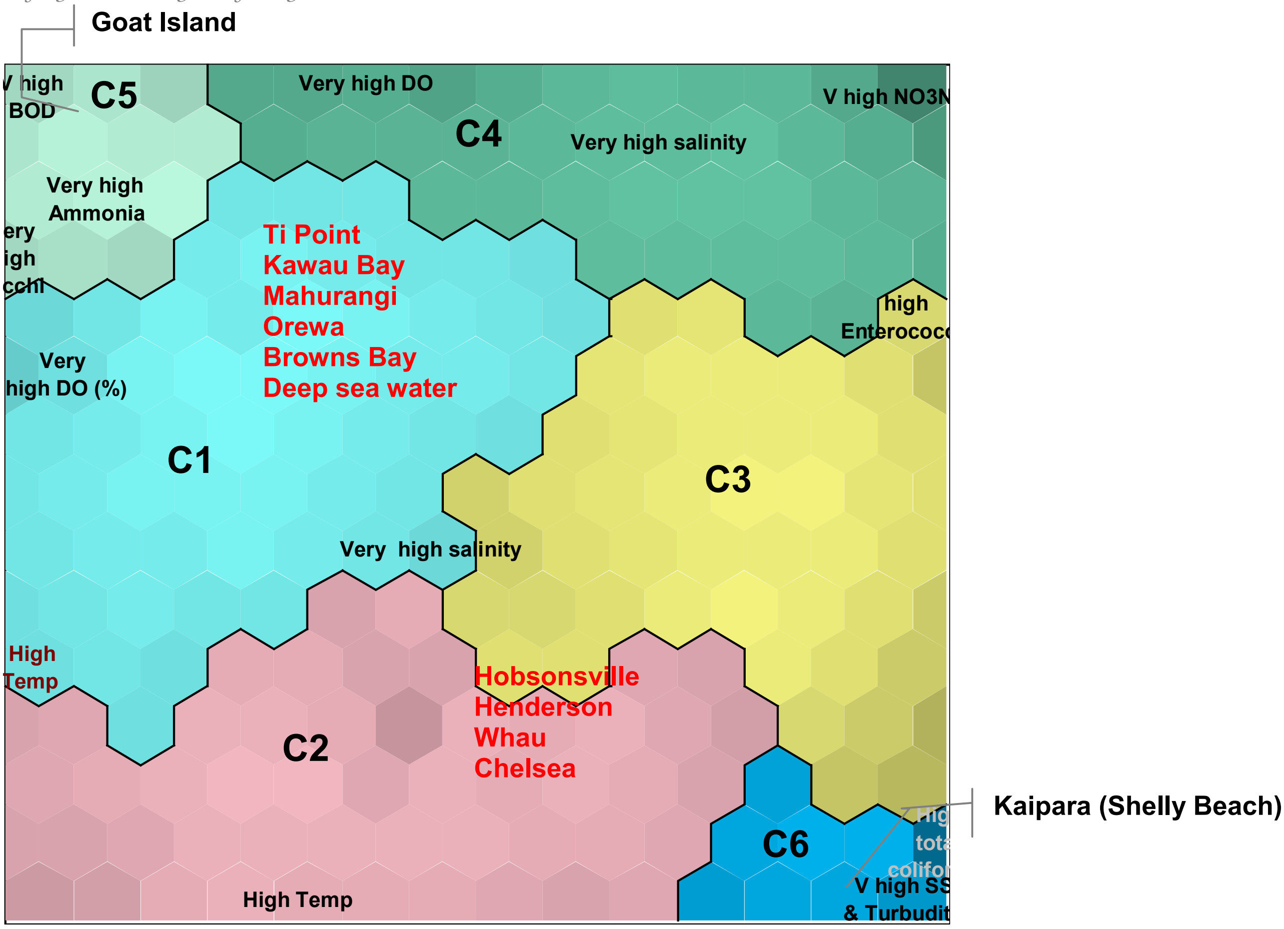
Initially, the SOM clustering patterns in the region's beach water quality monitoring data (LTB programme) are analysed based on their statistics. The trends in these beach water quality data over a period of 19 May 1991 to 12 Dec 2000 are then analysed. Hence, the types of SOM techniques experimented in this analysis are; cluster analysis, component plane analysis, decision support system and time series analyses (trajectories) and could be classified as initial exploratory data analysis. Finally, results of conventional analysis are compared with that of SOMs.

6.3.2 Results and discussion

SOM maps were created with Visovery® SOMine lite version 4.0 by eudaptics software gmph package. Due to the inconsistencies and gaps in the ARC's saline water quality test data, any other software could not be used for SOM analyses. Map creation parameters are set to default values unless stated in the text.

6.3.2.1 Cluster and dependent component analyses

Initially, a SOM (figure 6.2 a & b) was created with 200 nodes, priority of all components set to 1 and all other map parameters set to default values. The clustering patterns show the monthly trends in the environmental parameters as water sampling has been carried out once a month. The six different clusters in the SOM of the monitoring data coincide with their geographical location without any such specific details being added in the data.



Cluster	C1	C2	C3	C4	C5	C6
pH	8.191	8.09	8.102	8.13	8.166	8.078
Temp (deg C)	19.02	20.07	13.64	14.17	16.41	17.68
SS (mg/l)	7.2	20.5	19.7	7.3	4	145.6
Turb (NTU)	2	7.1	6.8	1.8	0.6	45.7
Chloride (mg/l)	19653	18871	17495	19194	19855	17603
Salinity (ppt)	32.72	31.41	29.79	33.87	33.07	29.34
TP (mgP/l)	0.0308	0.0498	0.0442	0.0308	0.0187	0.1302
DRP (mgP/l)	0.0151	0.0216	0.0224	0.0192	0.0168	0.0208
Nitrate (mgN/l)	0.0091	0.0118	0.0371	0.0197	0.0111	0.035
Ammonia (mgN/l)	0.022	0.012	0.018	0.019	1.837	0.032
BOD (mgO/l)	1.94	1.88	1.93	1.74	2.23	2.11
T Coliforms (mpn/100ml)	23	63	374	34	2	481
F Coliforms (mpn/100ml)	7	16	178	15	73	436
DO (%)	98.4	87.2	87.8	105.2	104.7	95.9
DO (ppm)	7.29	6.27	7.78	8.7	7.5	7.77
Secchi*	1.541	0.928	0.93	1.298	2.042	0.569
CHLORO*	0.00304	0.00304	0.00431	0.00785	0.00221	0.00398
ENTER-ME*	2.67	4.83	10.75	5.98	2.52	5.52
NO2*	0.00216	0.00224	0.00249	0.00344	0.00209	0.00213
NO3NO2*	0.0084	0.0075	0.0198	0.0234	0.0082	0.0136

Figure 6.2 a: SOM of saline water quality data form the 11 beach water sampling locations included in ARC's programme. b: Chart showing SOM cluster details.

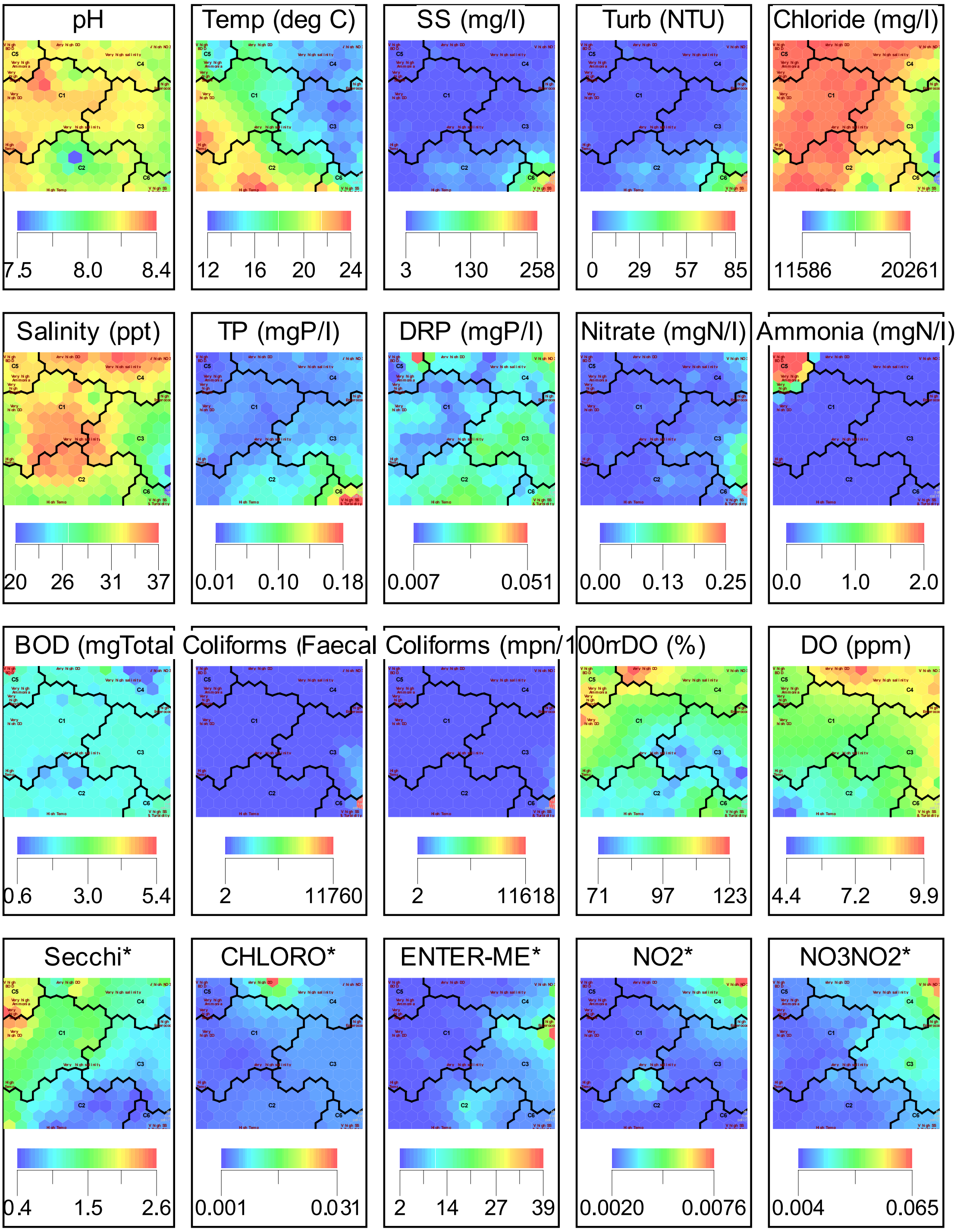
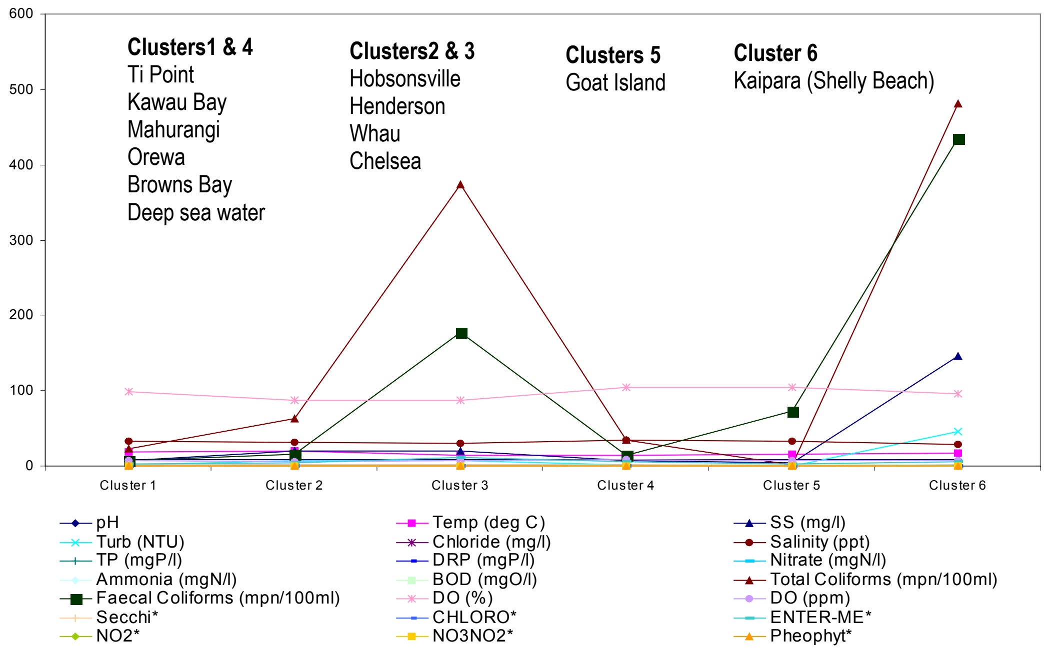


Figure 6.2 c: Component planes of the SOM (figure 6.2 a) of beach water quality monitoring data.



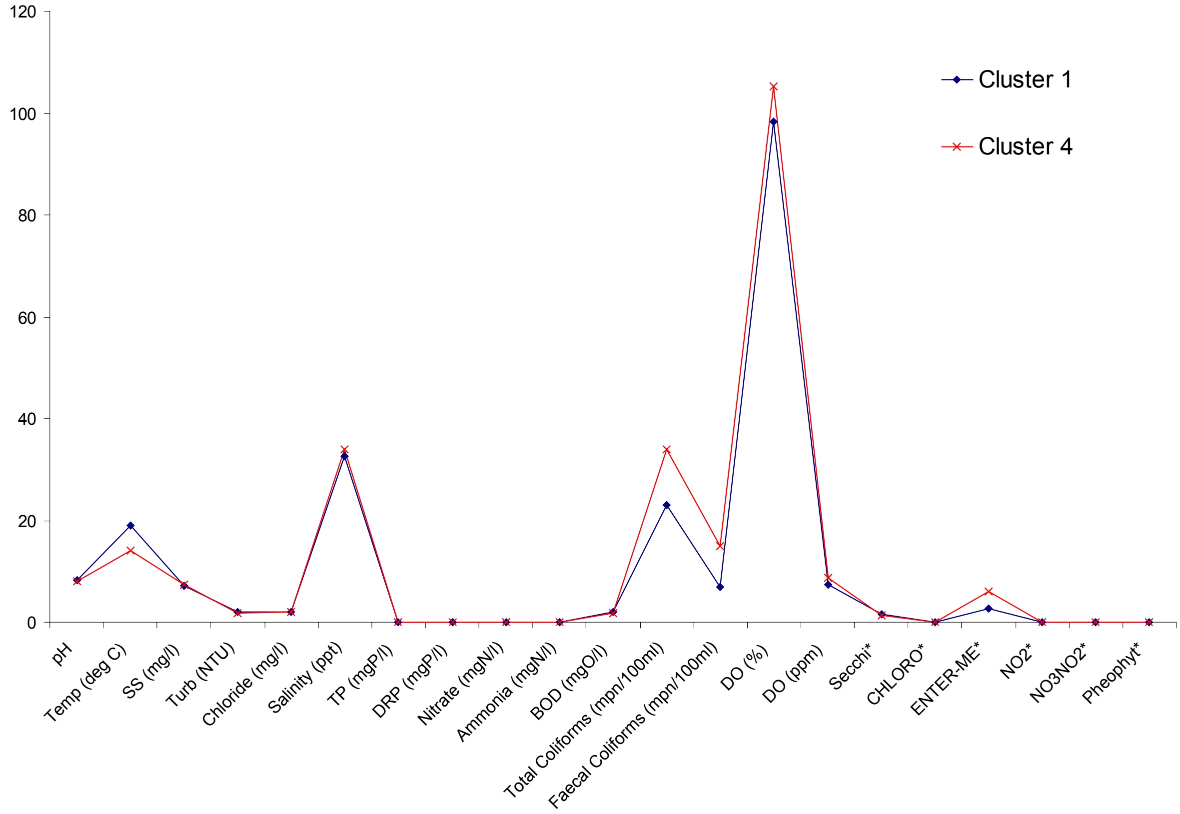


Figure 6.2 d: Graph showing SOM (figure 6.2 a) cluster details of beach water quality monitoring data. e: Graph showing the difference between clusters 1 and 4.

The following are the cluster analysis interpretations derived from the SOM and component planes (figures 6.2 a, b & c) that illustrate the patterns within LTB monitoring programme data;

(i) Clusters 1, 4 and 5 show high secchi disc depth values and low values for suspended solids and turbidity compared to other clusters. This means that Goat Island, Ti Point, Kawau Bay, Mahurangi, Orewa, Browns Bay and Deep sea waters seem to be better in quality in that were able to let sunlight through deeper into water columns, compared with the waters of Hobsonsville, Henderson, Whau, Chelsea and Kaipara (Shelly beach).

(ii) Clusters 1, 4 and 5 also show high chloride and salinity compared to the other beaches. Of this group of clusters, cluster 5 consisting of Goat Island data varies from the rest in that it has high ammonia, BOD and secchi disk depth values. Clusters 1 and 4 consisting of Ti Point, Kawau Bay, Mahurangi, Orewa, Browns Bay and Deep sea water, within them vary in temperature, total, faecal coliform and *Entercocci* (see figure 6.2 b).

(iii) Clusters 2 and 3 consisting of Hobsonsville, Henderson, Whau and Chelsea show similarities in many attributes except for temperature, total and faecal coliform values in which they show remarkable variations. Cluster 2 shows high temperature (20^{o} C mean) and low total and faecal coliform values (63 and 16 mpn/100ml), depicting patterns of summer, whereas cluster 3 shows low temperature (16^{o} C mean) along with high total and faecal coliform values (374 and 178 mpn/100ml), depicting wintry conditions.

(iv) Cluster 6 as well shows similar physical attributes to the above, however in addition, it exhibits elevated values of faecal Coliform (73mpn/100ml). All data points pertaining to Kaipara (Shelly beach) fall in this cluster.

SOM component planes could be used for comparative analysis of multidimensional data with easily understandable visual displays. The SOM created with ARC's LTB saline water quality data of the 11 beaches can be seen as more effective in comparative analysis of the beach water samples. For instance, all 20 attributes of the 11 beaches can be viewed and studied in the SOM component planes (figure 6.2 c), in contrast to the time consuming effort of analysing 11 into 20 graphs (similar to the ones in figures 6.8 a & b).

SOMs can be used to convert data into useful information without losing much of the details in the raw data. SOM component planes (figure 6.2 c) show the 20 attributes analysed on a single page. Each and every attribute values belonging to the beaches are shown as points on the plane. The scale below each plane shows the values. Hence, the problems encountered in plotting multiple, highly varying ranges of numeric values on a limited space are overcome.

For example, in the pH plane (figure 6.2 c & f), the scale underneath the plane shows the pH range of the 11 beaches ranging from 7.5 to 8.4. The blue patch in the plane shows both the days and the beaches that had water samples with low pH values. In a trajectory (time series analysis) data points closer to that point indicate the possibility of beach sample data nearing this value. In figure 6.3, on 18 January 1995, the pH of Browns Bay waters was 7.8, below the usual range. In a similar manner, other attributes may be analysed from the same SOM (figure 6.2 c).

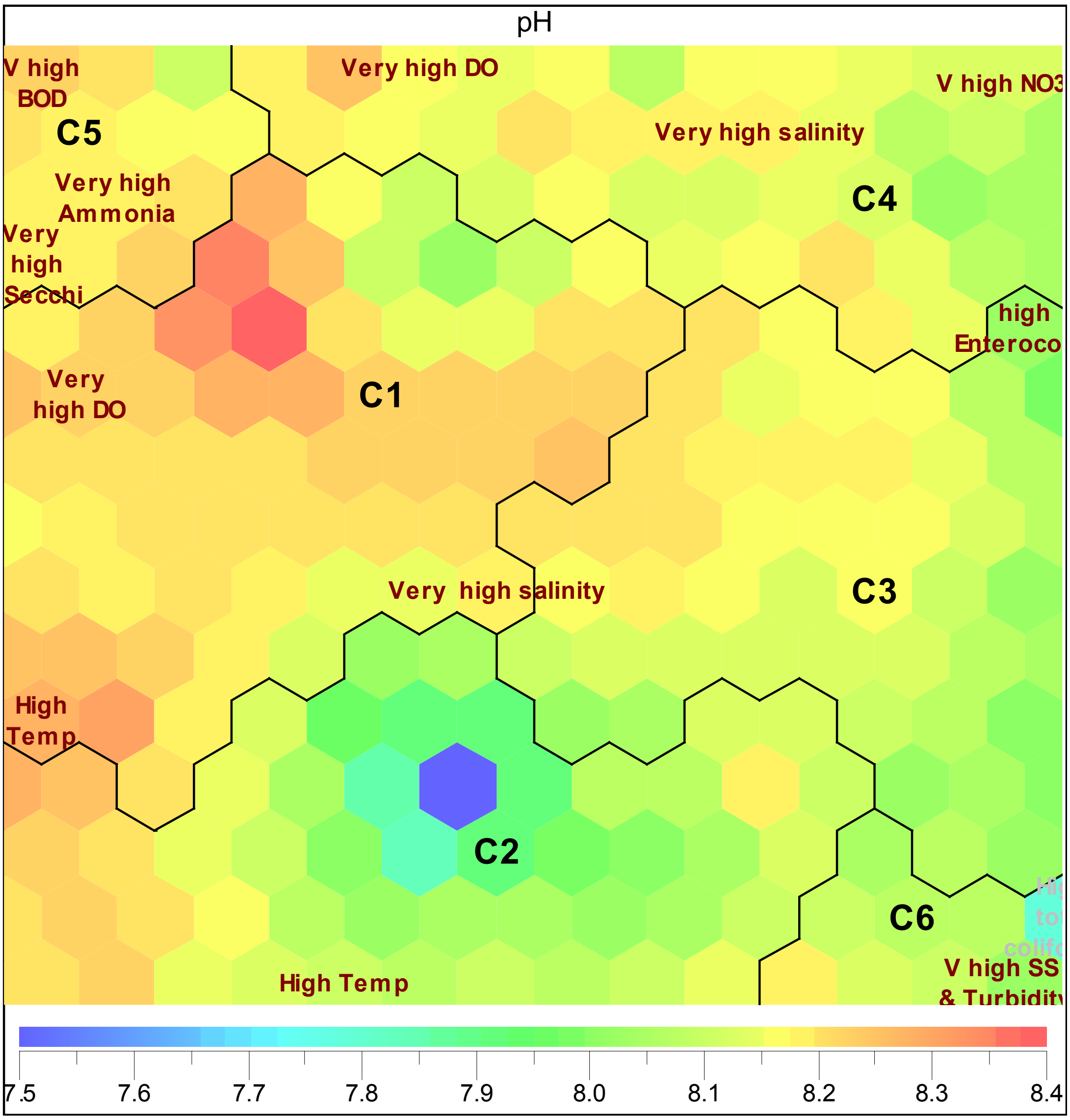


Figure 6. 2 f: pH component plane of SOM (figure 6.2 a). The scale underneath the plane shows the pH range of the 11 beaches ranging from 7.5 to 8.4

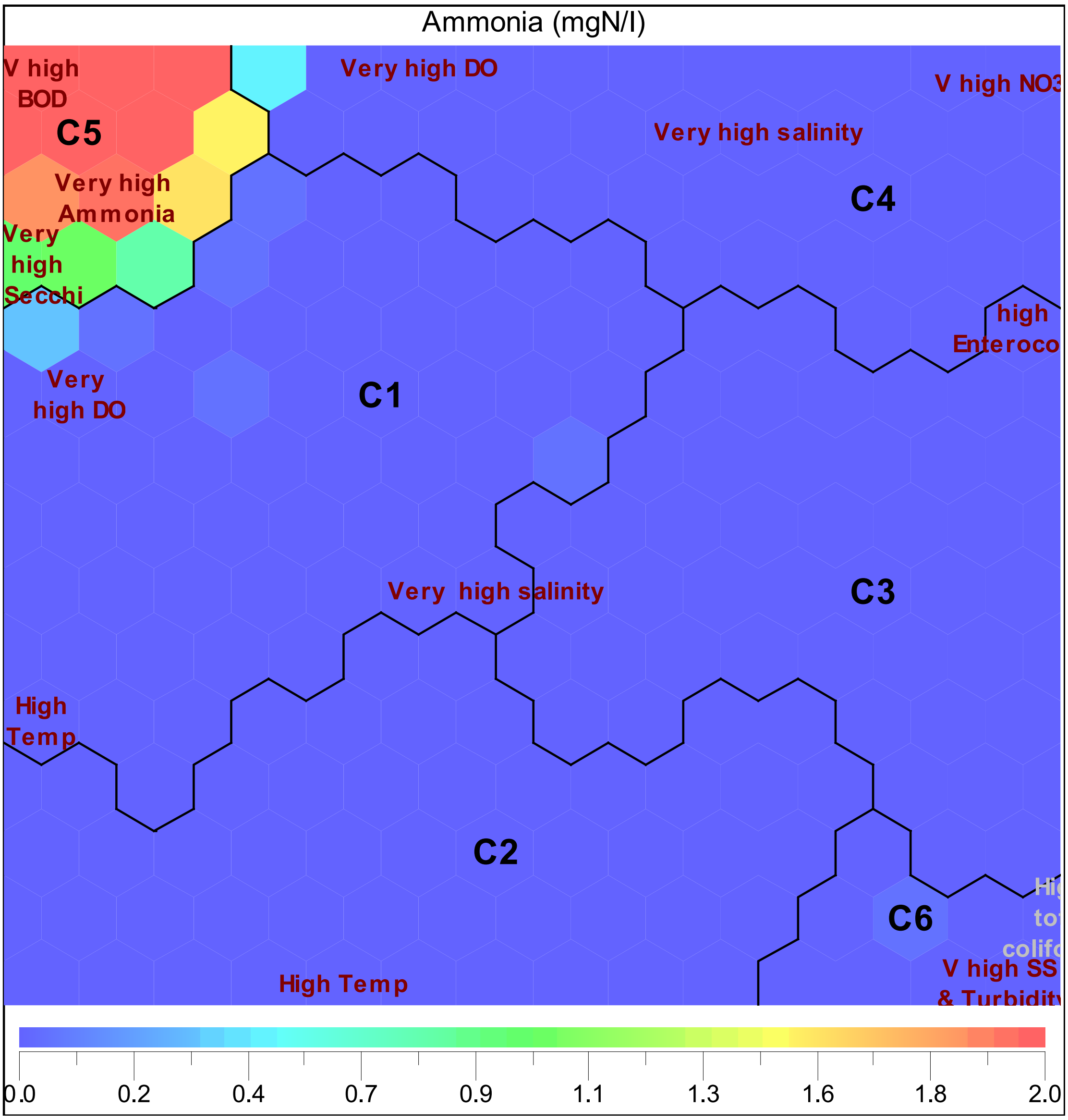


Figure 6. 2 g: Ammonia component plane of SOM (figure 6.2 a).

In the same SOM (figure 6.2 c), details of ammonia levels of all the beaches in the region are analysed. The red patch in the top left corner of this plane (figure 6.2 g) consists of high concentrations of ammonia (2mgN) with Goat Island beach water samples. Also, the component plane well displays the fact that Goat Island beach water samples have the highest ammonia values.

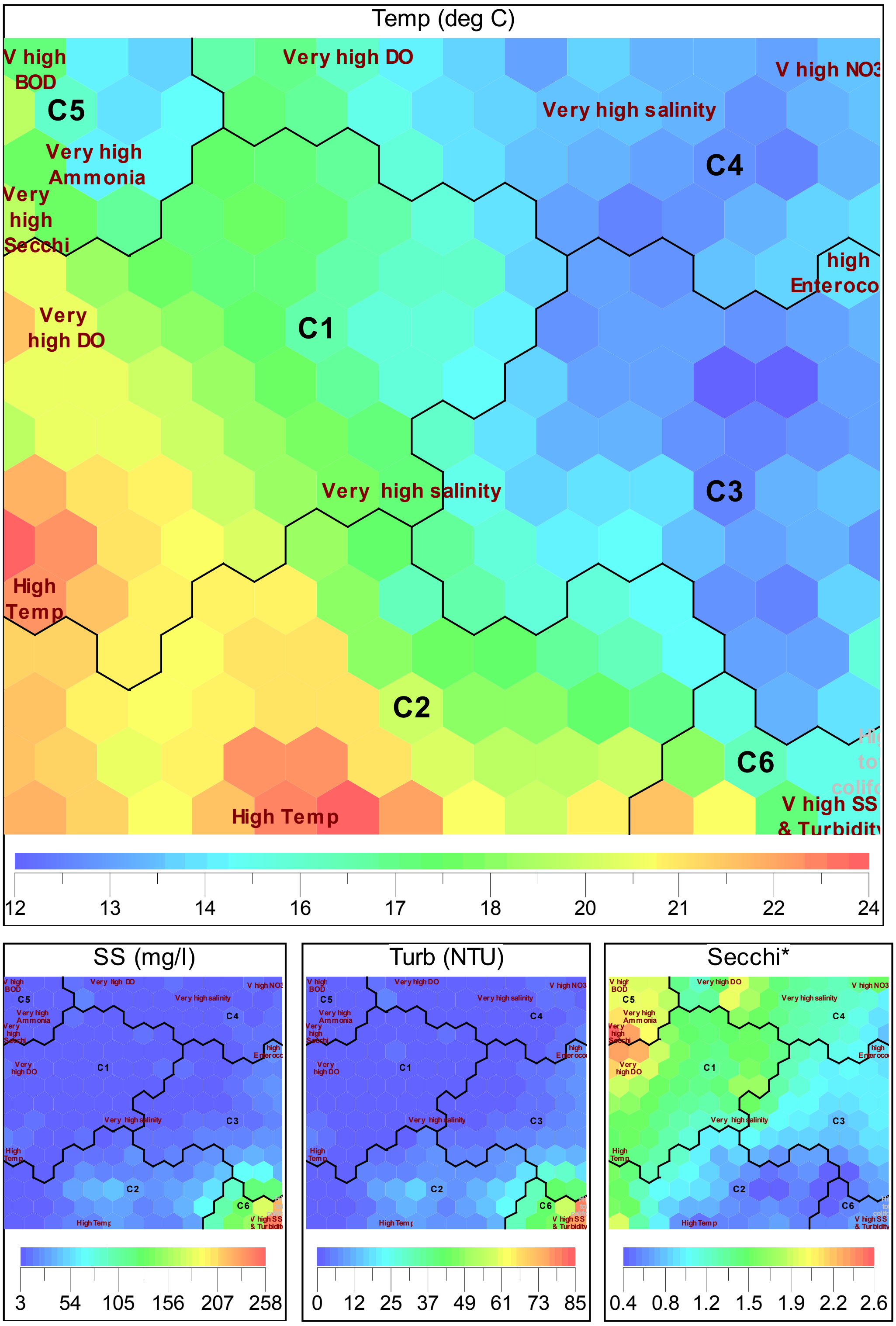
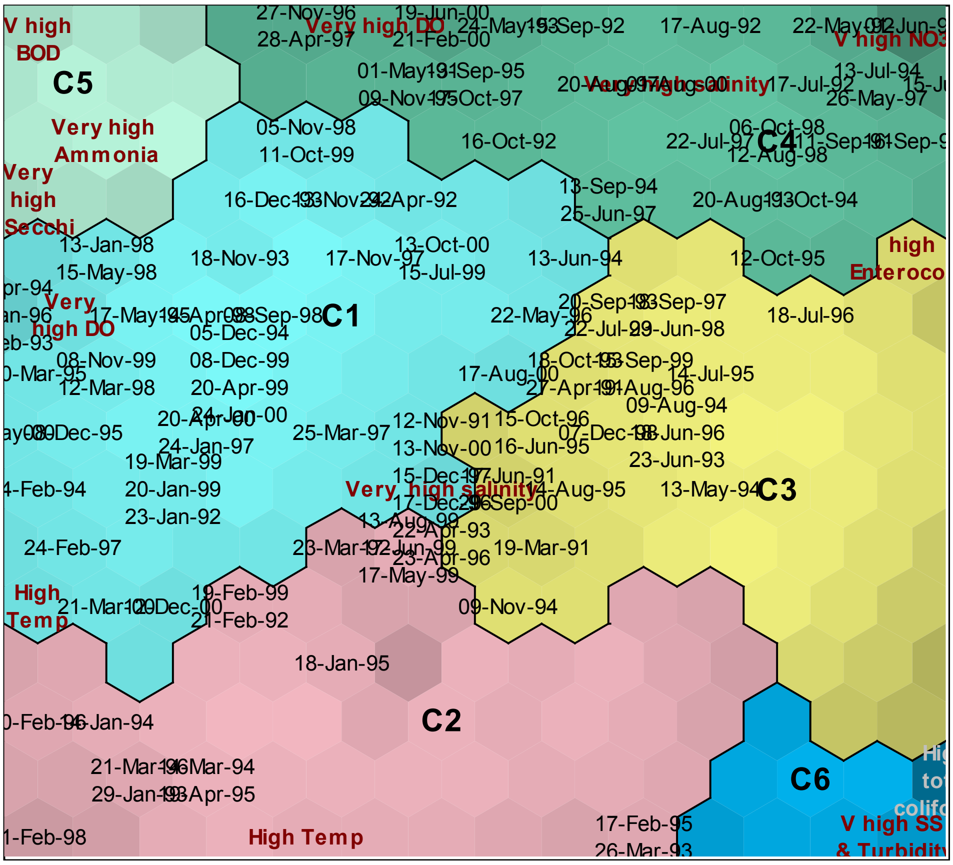
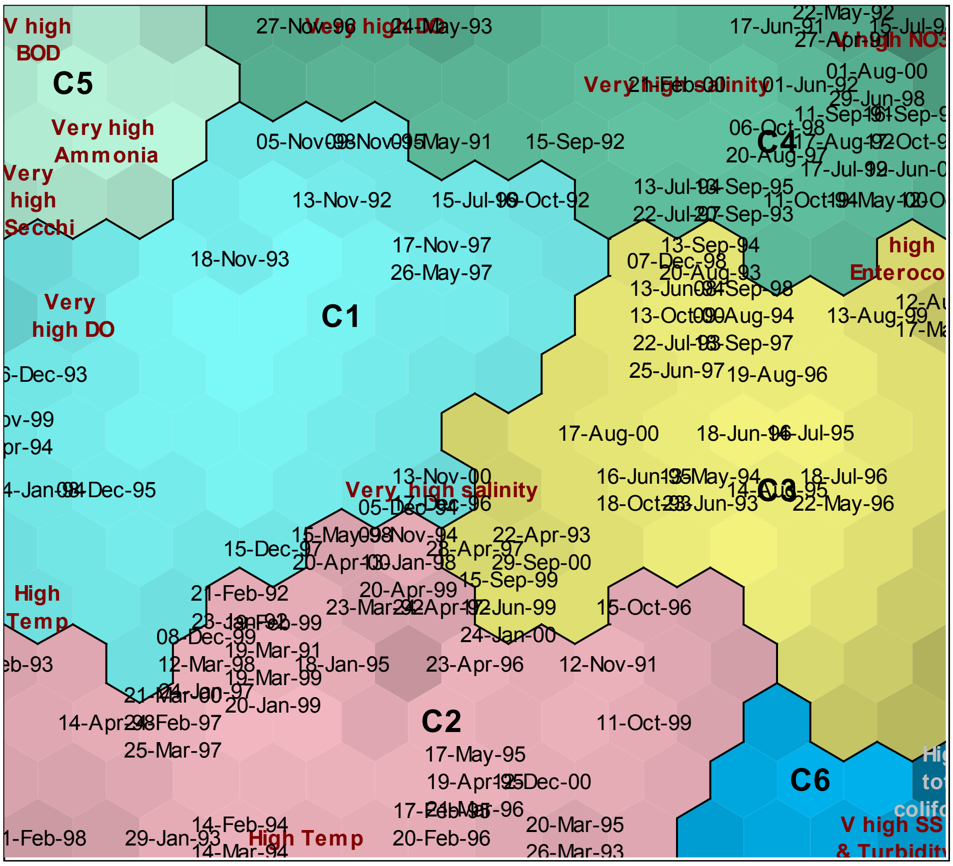


Figure 6. 2 h : Temperature component plane of SOM (figure 6.2 a). i: SOM component planes of suspended solids (SS), Turbidity and secchi disc depth values.

The temperature plane (figure 6.2 h) depicts both summer (15-25 deg C) and winter (10-15 deg C) water temperature ranges. Furthermore, secchi disc depth, suspended solids and turbidity planes (figure 6.2 i) show the correlations in these attributes.





Figures 6.3 a & b: SOM of saline water quality data of the 11 beaches of ARC's LTB programme with a: with Browns Bay data and b: Chelsea data.

cluster	year									
C1	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
C1	12-Nov-91	23-Jan-92	24-Feb-93	14-Feb-94	20-Mar-95	23-Jan-96	24-Jan-97	13-Jan-98	20-Jan-99	24-Jan-00
C1		24-Apr-92	22-Apr-93	18-Apr-94	17-May-95	23-Apr-96	24-Feb-97	12-Mar-98	19-Mar-99	21-Mar-00
C1		13-Nov-92	18-Nov-93	13-Jun-94	8-Dec-95	22-May-96	25-Mar-97	14-Apr-98	20-Apr-99	20-Apr-00
C1			16-Dec-93	5-Dec-94		17-Dec-96	17-Nov-97	15-May-98	15-Jul-99	19-May-00
C1							15-Dec-97	8-Sep-98	11-Oct-99	17-Aug-00
C1								5-Nov-98	8-Nov-99	13-Oct-00
C1									8-Dec-99	13-Nov-00
C1										12-Dec-00
	1	3	4	4	3	4	5	6	7	8
C2		21-Feb-92	29-Jan-93	14-Jan-94	18-Jan-95	20-Feb-96		11-Feb-98	19-Feb-99	
C2		23-Mar-92	26-Mar-93	14-Mar-94	17-Feb-95	21-Mar-96			17-May-99	
C2					19-Apr-95				17-Jun-99	
C2									13-Aug-99	
	0	2	2	2	3	2	0	1	4	0
C3	19-Mar-91		23-Jun-93	13-May-94	16-Jun-95	18-Jun-96	18-Sep-97	29-Jun-98	15-Sep-99	29-Sep-00
C3	27-Apr-91		22-Jul-93	9-Aug-94	14-Jul-95	18-Jul-96		7-Dec-98		
C3	17-Jun-91		20-Sep-93	9-Nov-94	14-Aug-95	19-Aug-96				
C3			18-Oct-93			15-Oct-96				
	3	0	4	3	3	4	1	2	1	1
C4	1-May-91	22-May-92	24-May-93	13-Jul-94	13-Sep-95	16-Sep-96	28-Apr-97	15-Jul-98		21-Feb-00
C4	11-Sep-91	1-Jun-92	20-Aug-93	13-Sep-94	12-Oct-95	27-Nov-96	26-May-97	12-Aug-98		19-Jun-00
C4		17-Jul-92		11-Oct-94	9-Nov-95		25-Jun-97	6-Oct-98		1-Aug-00
C4		17-Aug-92					22-Jul-97			
C4		15-Sep-92					20-Aug-97			
C4		16-Oct-92					17-Oct-97			
	2	6	2	3	3	2	6	3	0	3

Figure 6.3 c: Browns Bay data grouping details from the SOM of the 11 beach water sampling, included in ARC's saline water quality monitoring programme.

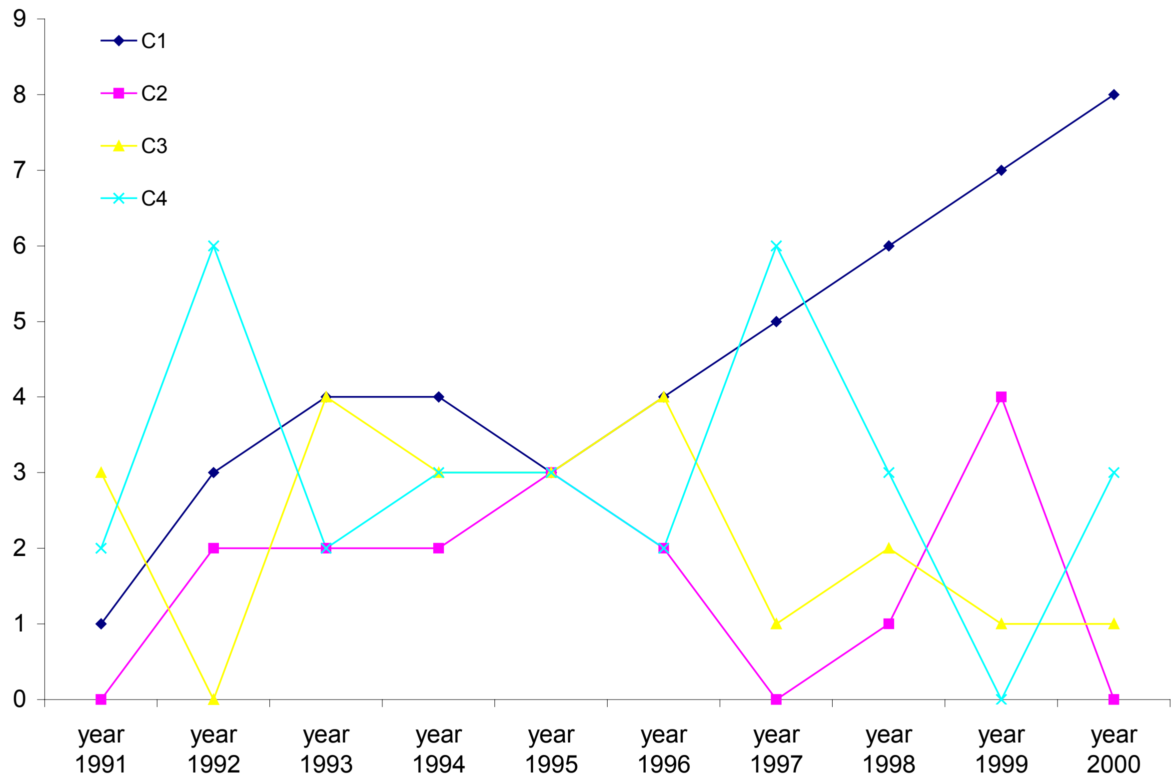


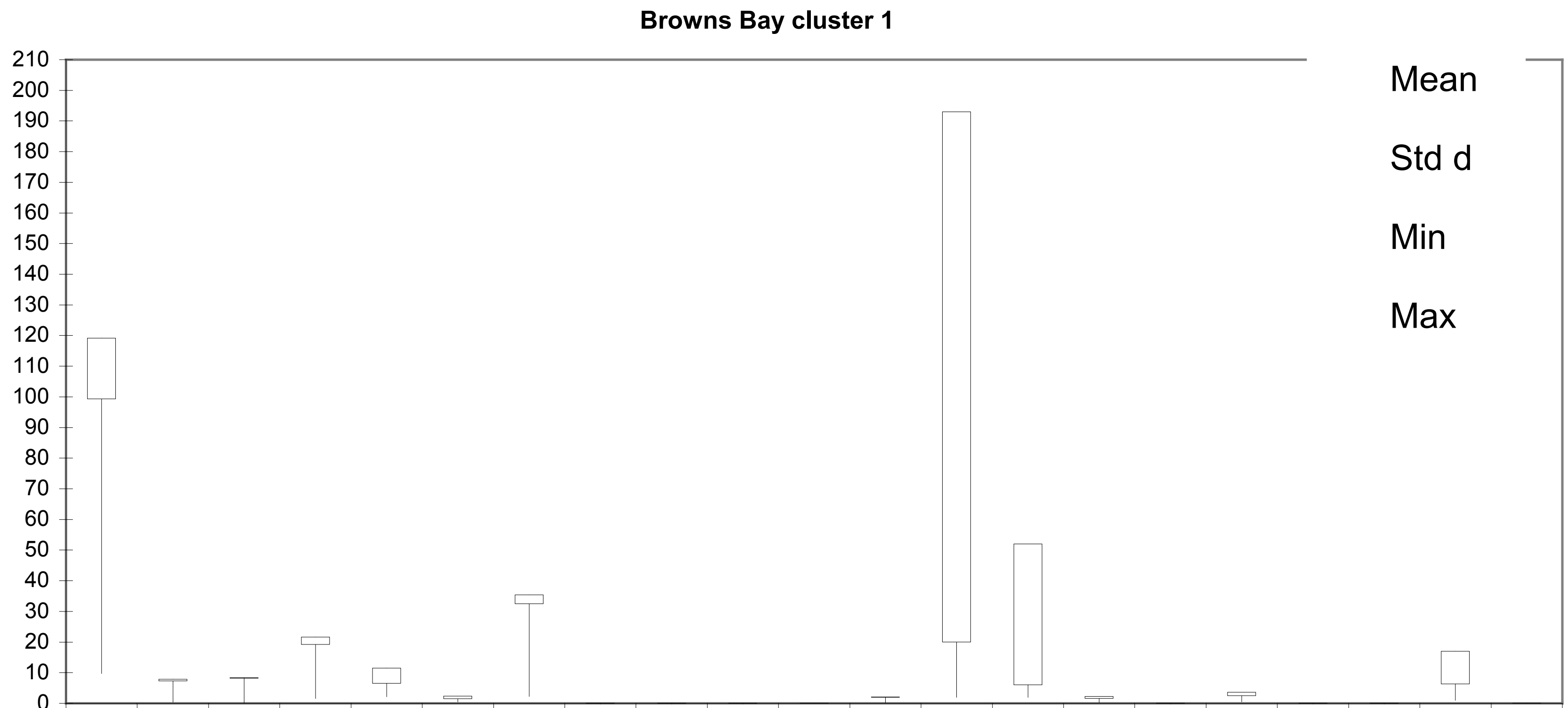
Figure 6.3 d: Graph of SOM cluster details (C1 to C4) showing Browns Bay's annual variations within the monthly sampling of ARC's saline water quality monitoring programme data.

The SOM with Browns Bay data labels (figures 6.3 a, c, d & e) illustrates the annual variations in the water quality along this coast over the ten year period analysed.

(i) Almost all the data points lie in the left upper half on the SOM, except for the corner. Since 1991 onwards, increasing number of data points are seen to be showing cluster 1 attributes. The annual trend in figures 6.3 c, the data summary chart and 6.3 d, graph of Brown Bay clusters as well confirm the trend. Browns Bay data summary graphs for cluster details (figures 6.3 e & f) show the range for each attribute analysed herein.

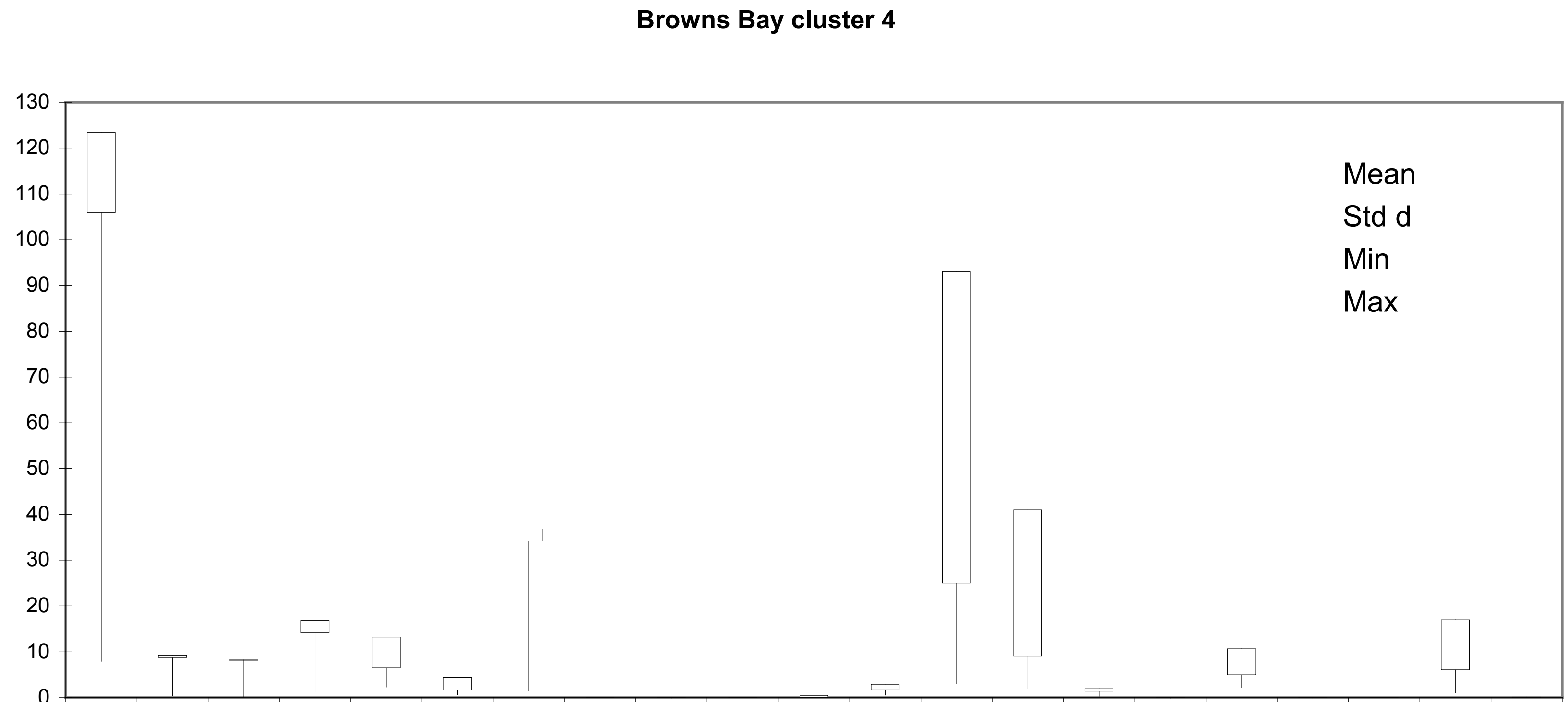
(ii) More data points (May - Oct 1992 and April - October 1997 are seen in cluster 4, where total and faecal coliform counts are higher than in Cluster 1.

Labelled SOMs (figures A 6.1 - 6.11 of appendix 6) illustrate the positioning of individual beach data points and their spread on the SOM of all 11 beach data.



	DO (%)	DO (ppm)	pH	Temp (deg C)	SS (mg/l)	Turb (NTU)	Salinity (ppt)	TP (mgP/l)	DRP (mgP/l)	Nitrate (mgN/l)	Ammonia (mgN/l)	BOD (mgO/l)	Total Coliforms	Faecal Coliforms	Secchi*	CHLORO*	ENTER-ME*	NO2*	NO3NO2*	Frequency	Quantization error
Mean	99.3	7.29	8.188	19.2	6.5	1.5	32.49	0.0305	0.0161	0.0081	0.019	1.93	20	6	1.606	0.0026	2.48	0.00212	0.0076	6.278	0.021
Std d	9.7	0.4	0.057	1.59	2.2	0.5	2.3	0.0055	0.004	0.0033	0.014	0.15	35	9	0.339	0.00106	0.5	0.00013	0.002	3.525	0.015
Min	79.9	6.36	8.088	15.94	3.9	0.6	28.41	0.0214	0.0086	0.0033	0.008	1.41	2	2	1.143	0.0018	2.04	0.00201	0.0045	1	0.005
Max	119.2	7.88	8.362	21.66	11.5	2.4	35.4	0.0377	0.0204	0.016	0.061	2.06	193	52	2.271	0.006	3.68	0.00254	0.0119	17	0.056

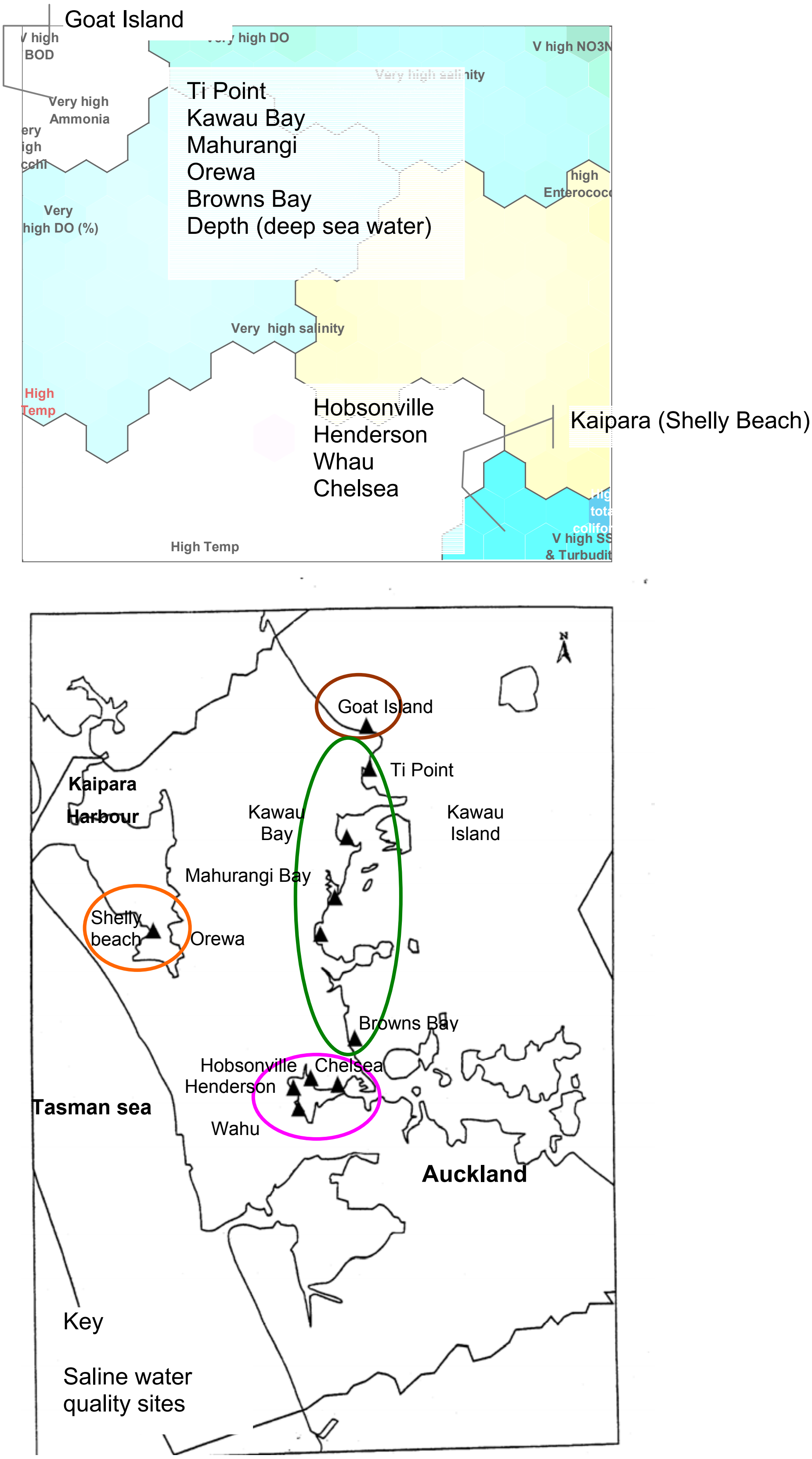
Figure 6.3 d: Graph showing the physical attribute ranges in cluster 1of the SOM for Browns Bay data



	DO (%)	DO (ppm)	pH	Temp (deg C)	SS (mg/l)	Turb (NTU)	Salinity (ppt)	TP (mgP/l)	DRP (mgP/l)	Nitrate (mgN/l)	Ammonia (mgN/l)	BOD (mgO/l)	Total Coliforms	Faecal Coliforms	Secchi*	CHLORO*	ENTER-ME*	NO2*	NO3NO2*	Frequency	Quantization error
Mean	105.9	8.7	8.137	14.17	6.4	1.6	34.14	0.0293	0.0184	0.018	0.019	1.65	25	9	1.344	0.00861	4.95	0.0033	0.0207	6.043	0.031
Std d	7.9	0.31	0.052	1.27	2.3	0.8	1.44	0.0061	0.0046	0.0118	0.041	0.58	28	9	0.272	0.00632	2.48	0.00107	0.0087	3.329	0.031
Min	96.5	8.2	8.03	12.77	3.1	0.6	31.16	0.022	0.008	0.0044	0.008	0.55	3	2	0.958	0.00429	2.15	0.00202	0.0126	1	0
Max	123.4	9.25	8.22	16.91	13.2	4.4	36.87	0.0442	0.0312	0.0539	0.495	2.88	93	41	1.967	0.03073	10.63	0.00558	0.0437	17	0.137

Figure 6.3 e: Graph showing the physical attribute ranges in cluster 4 of the SOM for Browns Bay data

6.3.1.2.2 Decision support systems



Figures 6.4 a & b: Decision support system superimposed on a SOM of ARC's saline beach water quality monitoring data consisting of the 11 beaches, north of Auckland.

The map component labelling on the cluster map shows the general spread of the data collected from the beaches analysed through the LTB saline water quality programme. The map clustering has separated the data into four major groups that coincided with their geographical locations (figures 6.4 a & b):

(i) Goat Island in the top left corner (cluster 5).

(ii) Ti Point, Kawau Bay, Mahurangi, Orewa, Browns Bay and Depth (deep sea water collected from different locations), all beaches monitored from northeastern coast of Auckland fall in the top left half of the map (clusters 1 & 4). Cluster 1 describes the typical summer along the northeastern coast of Auckland, whereas cluster 4 depicts the wintry conditions along these beaches.

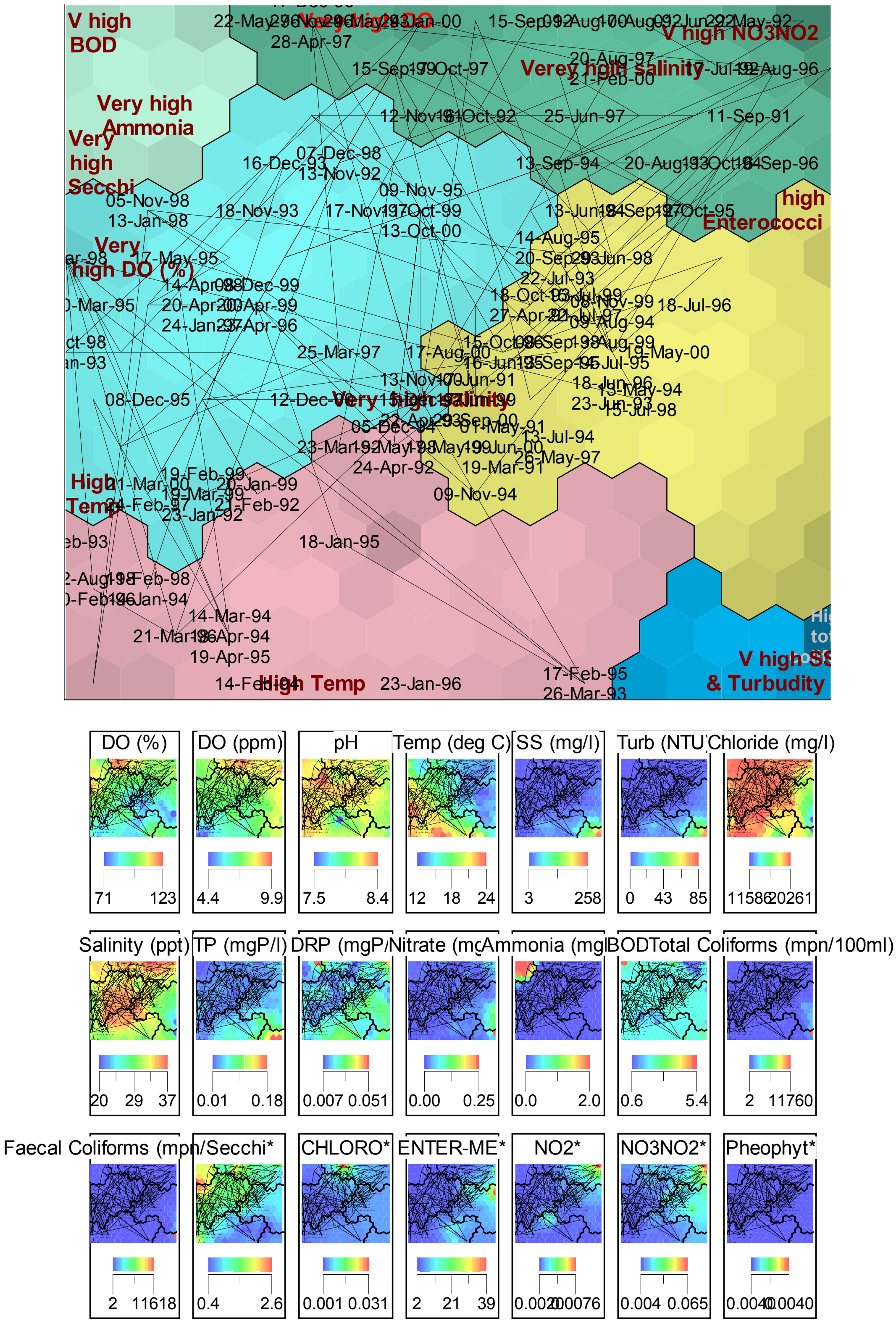
(iii) Hobsonsville, Henderson, Whau and Chelsea in the Waitemata Harbour fall in the right bottom half of the map (clusters 2 & 3). Here again, clusters 2 and 3 depict the warmer and wintry conditions at the Waitemata Habour. Cluster 3 also exhibits higher total and faecal values, indicating high bacterial discharge at the sites as concluded in (Wilcock and Stroud 2000).

(iv) Kaipara (Shelly beach, cluster 5) in the bottom right corner of the map has the highest total and faecal coliform values, indicating that the Kaipara Harbour gets the highest bacterial discharge of all the beaches included in the monitoring programme, again confirms (Wilcock and Stroud 2000).

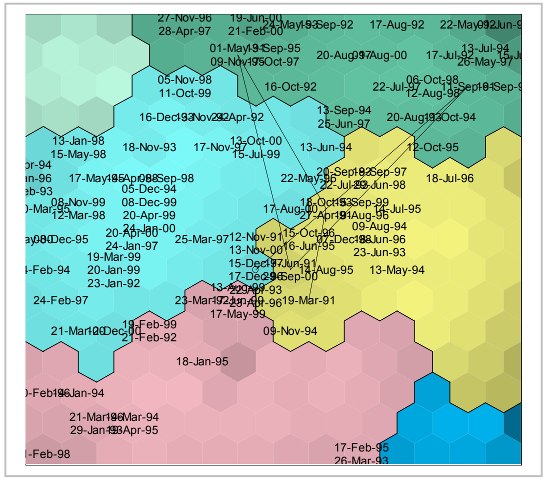
6.3.1.2.3 Time series (trajectory) analysis

The animation for the trajectory (figures 6.5 a & b) created on this SOM, using Browns Bay data shows the water quality dynamics over this period along this beach.

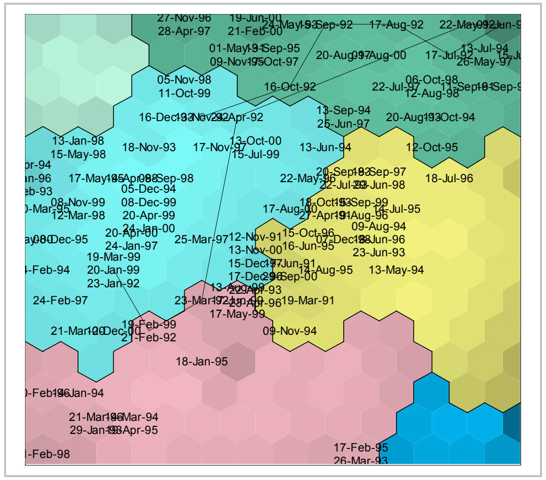
11 September 1991, 13 September 1994, 11 October 1994, 18 July 1996, 12 October, 15 July 1995 (in winter) and 9 November 1994 (in summer) are getting closer to high *Eterococci* patches. Similarly, using the respective beach data, water quality dynamics of the other sites also could be analysed in a time series. The data used in this example consists of monthly test results thus the trajectory does not look smooth as the trajectories of industrial processes. By using environmental data sampled at shorter intervals, such as hourly, daily, etc, ecologists could be able to follow the process dynamics at closer intervals.



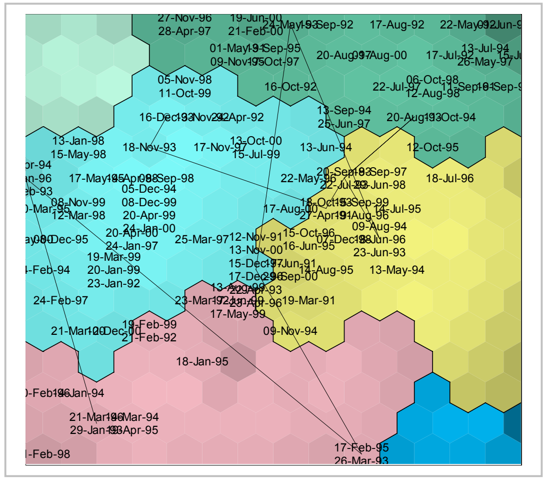
Figures 6.5 a & b: SOM and component planes with the trajectory of Browns Bay data on SOM created with ARC's LTB saline water quality data from the 11 beach sites with 200 nodes and all other map parameters set to default values.



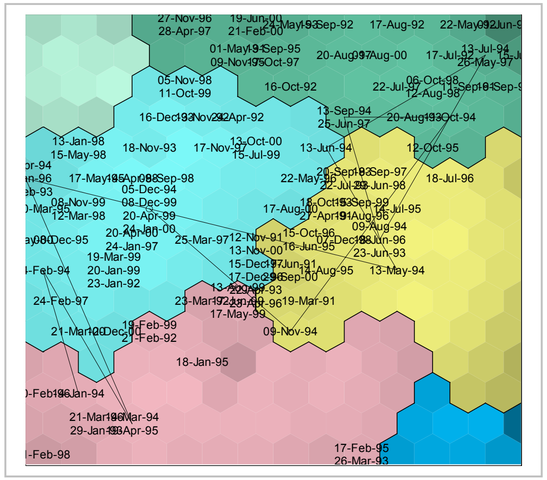
1-6



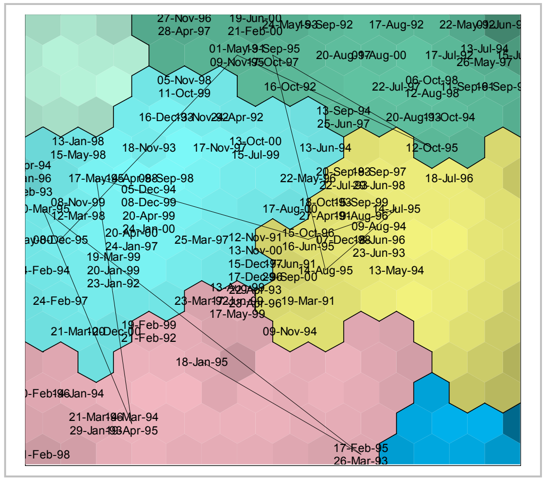
7-18



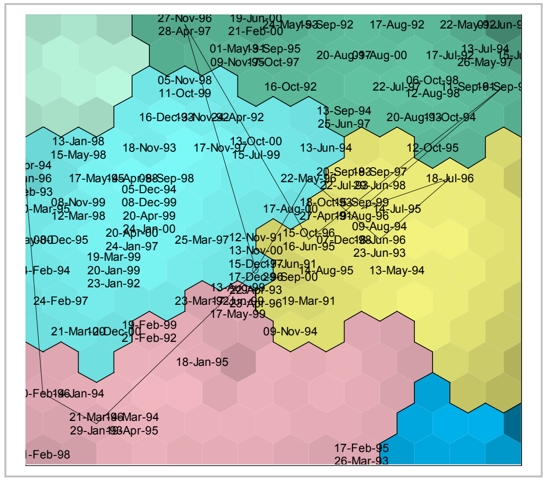
19-20



31-42

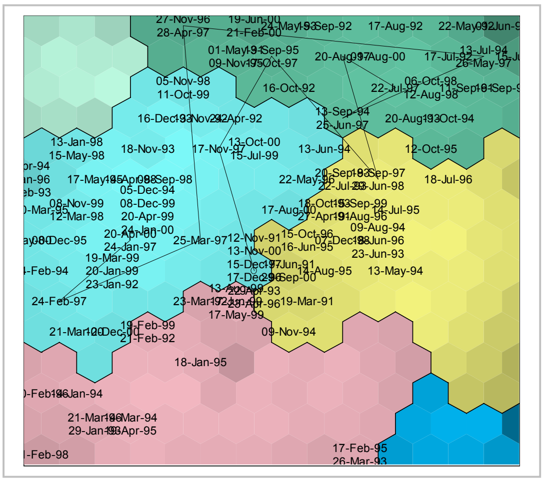


43-54

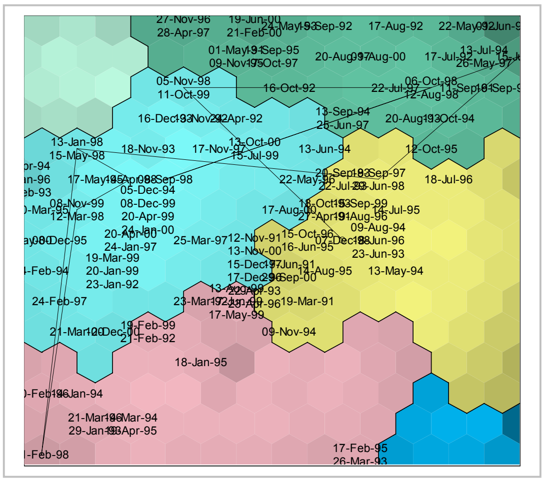


55-66

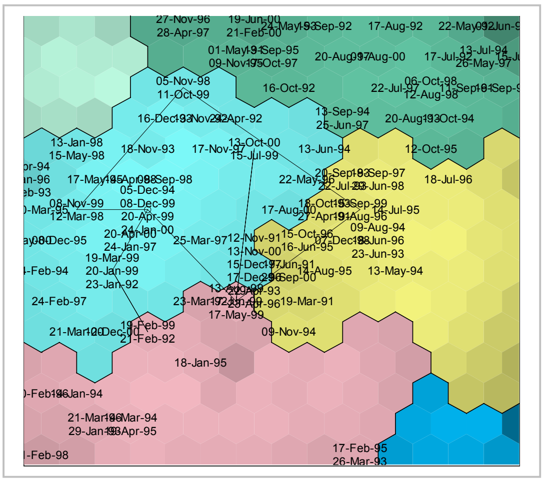
Figure 6.5c: Trajectories of Browns Bay monthly data on SOM of the ARC's LTB saline water quality data from the 11 monitored beach sites.



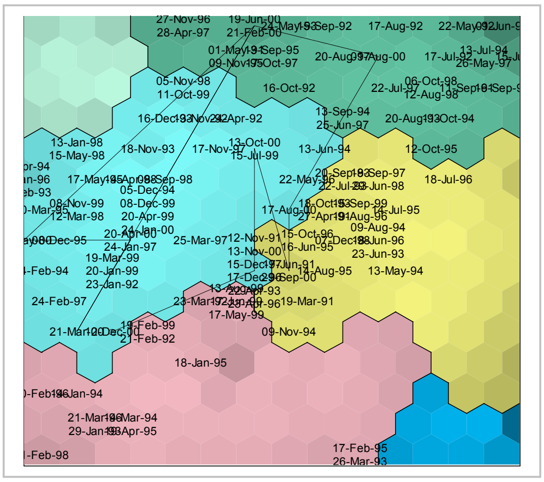
67-78



79-90



91-102



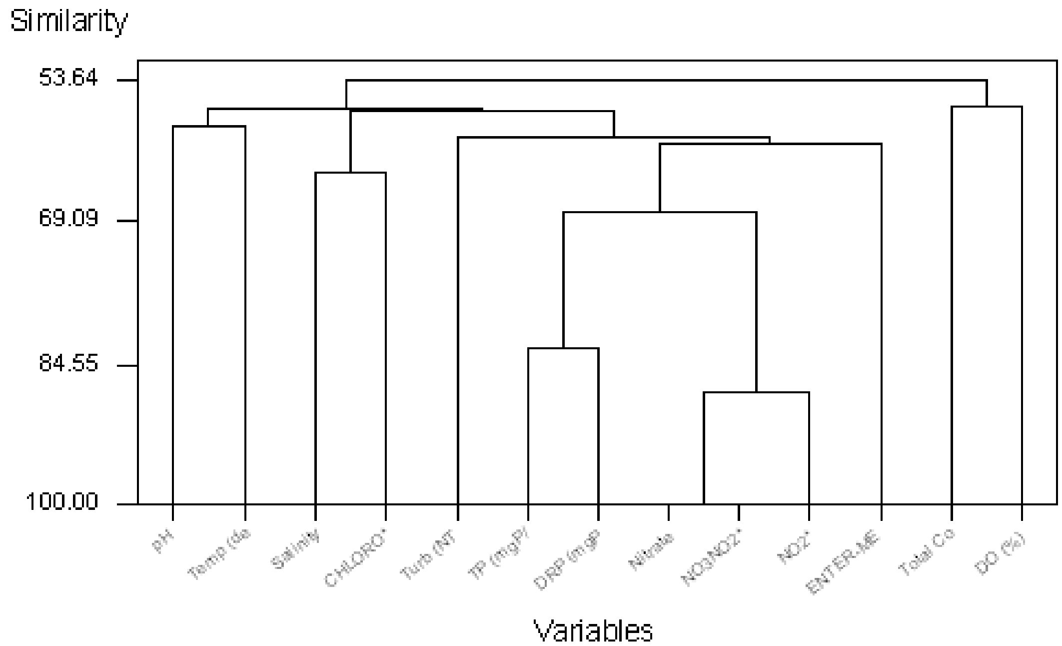
103-114

Figure 6.5 d: Trajectory of Browns Bay data on the SOM of ARC's LTB saline water quality data from the 11 beach sites with 200 nodes and all other map parameters set to default values.

The time series analysis carried out on Browns Bay data (figures 6.5 a - d) do not show a smooth flow as the monthly sampling of LTB programme consists of remarkable variations. SOM maps (figures 6.5 a & b) show the whole 10 year Browns Bay data on the SOM of all the 11 beaches monitored. SOM trajectories (figures 6.5 c & d) show the animation of Browns Bay's monthly and annual variations on the SOM of all beaches.

6.3.2 Conventional data analysis

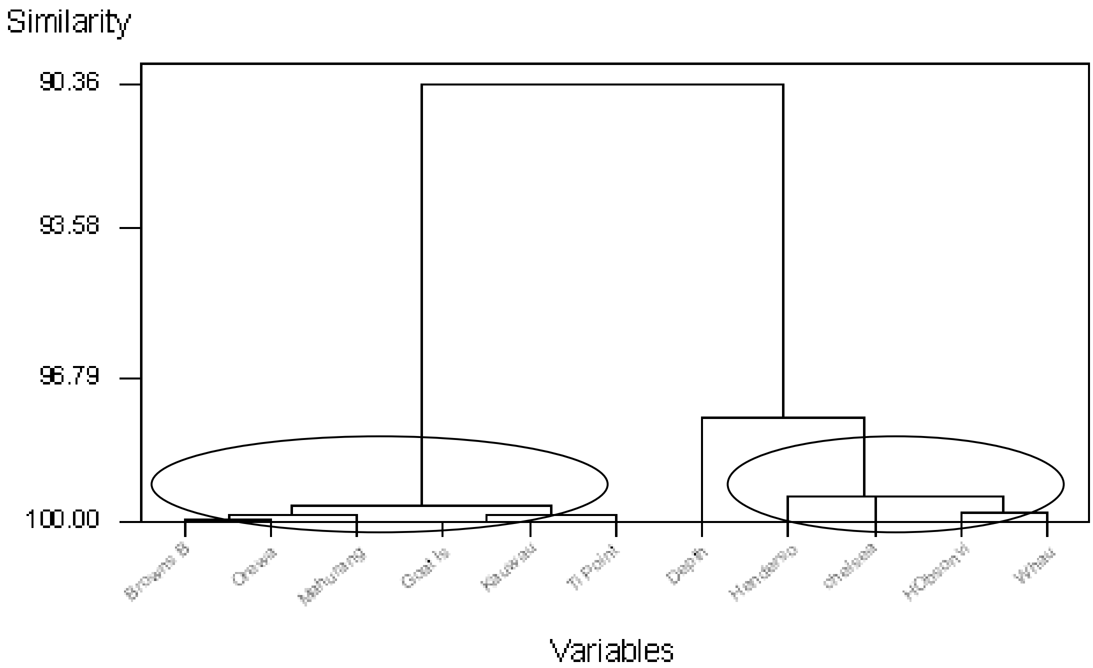
A Cluster analysis on the saline water quality data was carried out, initially by variables and then by sites to see the similarities among them using Minitab, a software package generally used by statisticians. The results of this analysis are displayed along with their correlation coefficient distance tables (figure 6.6 a & b and 6.7 a & b). Even though Minitab and SOM approaches produced the same results, the latter are seen to outperform the conventional analyses, as SOMs are visual analyses in that they provide a means to study the trends and variations within all 20 variables (attributes) of the 11 beaches monitored from year 1991-2000, in one SOM (figure 6.2).



Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	12	99.71	0.006	7	13	7	2
2	11	87.62	0.248	7	12	7	3
3	10	82.84	0.343	5	6	5	2
4	9	68.00	0.640	5	7	5	5
5	8	63.54	0.729	4	10	4	2
6	7	60.43	0.791	5	11	5	6
7	6	59.90	0.802	3	5	3	7
8	5	58.60	0.828	1	2	1	2
9	4	56.80	0.864	3	4	3	9
10	3	56.58	0.868	1	3	1	11
11	2	56.46	0.871	8	9	8	2
12	1	53.64	0.927	1	8	1	13

Figure 6.6 a: Cluster analysis of variables (saline water quality data) b: correlation coefficient distance table.

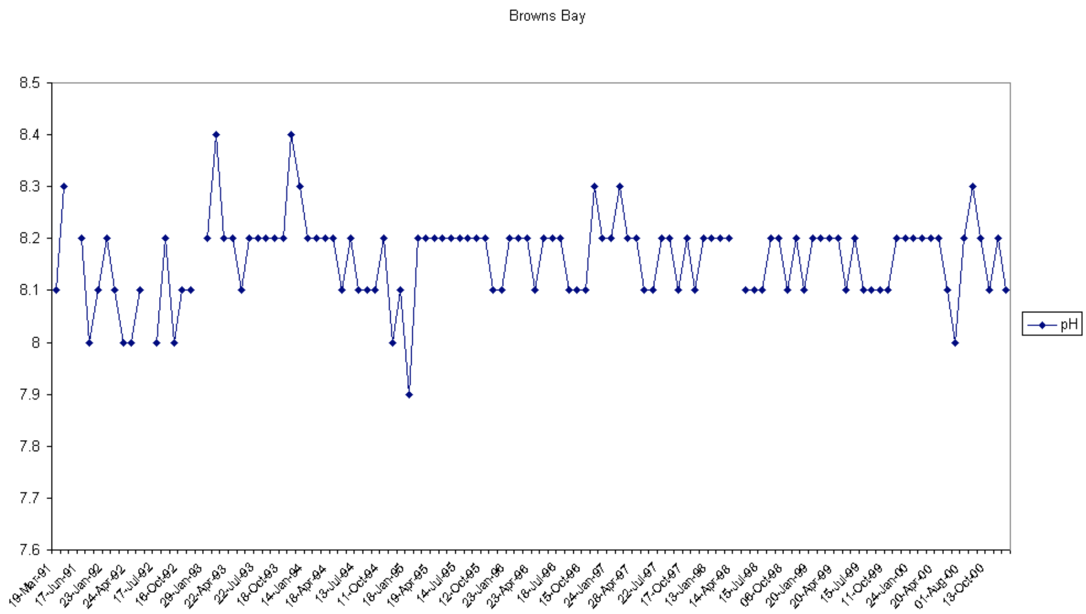


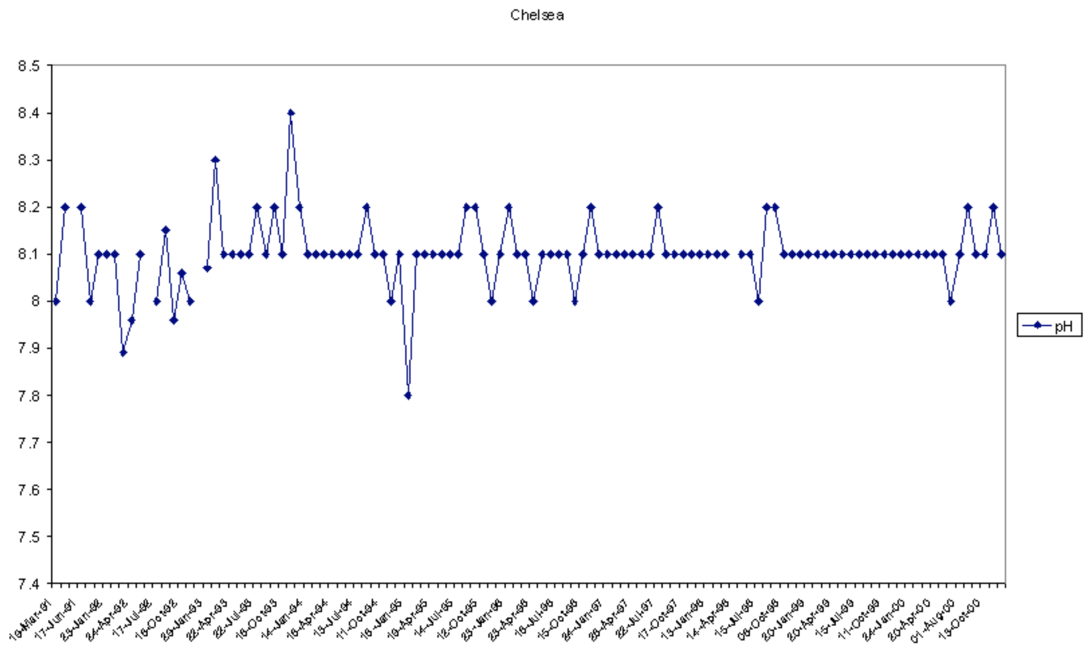
Amalgamation Steps

Step	Number of clusters	Similarity level	Distance level	Clusters joined		New cluster	Number of obs. in new cluster
1	10	99.98	0.000	4	7	4	2
2	9	99.93	0.001	1	9	1	2
3	8	99.87	0.003	4	10	4	3
4	7	99.86	0.003	1	8	1	3
5	6	99.79	0.004	6	11	6	2
6	5	99.64	0.007	1	4	1	6
7	4	99.42	0.012	2	6	2	3
8	3	99.42	0.012	2	5	2	4
9	2	97.73	0.045	2	3	2	5
10	1	90.36	0.193	1	2	1	11

Figure 6.7 a: Cluster analysis of saline water quality sites b: correlation coefficient distance table.

Wilcock and Stroud (2000) analysed the annual and seasonal variations in the LTB saline water quality monitoring programme data using statistical tests with non-parametric techniques contained in WQSTAT PLUS. Using this programme, statistical calculations such as median, normality, seasonally, trend and slope were generated and these values were then compared with graphs of each and every attribute of individual beaches all separately, covering a period 10 years. Similar graphs for pH values plotted against time are shown in figures 6.8 a & b, for Browns Bay and Chelsea.



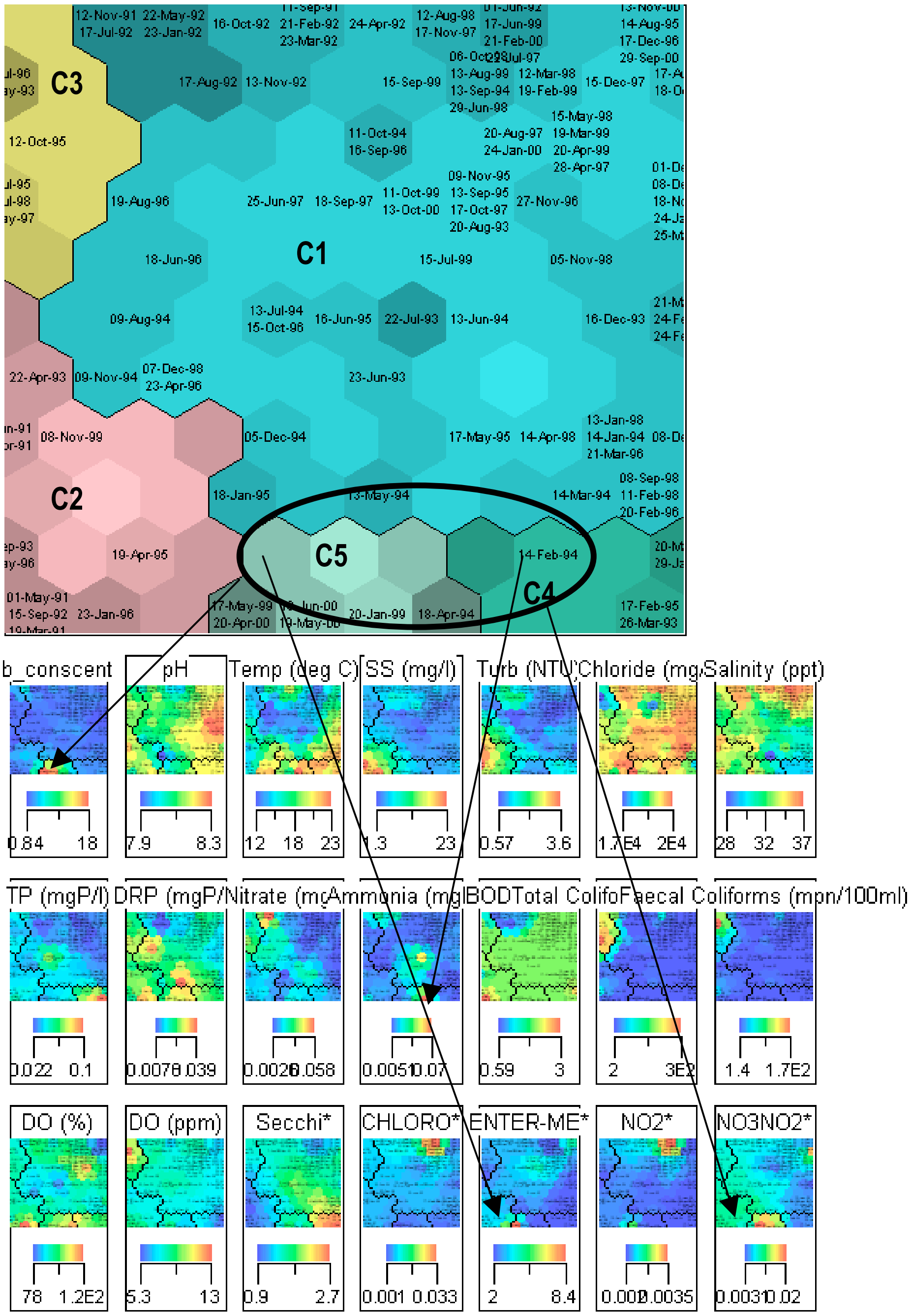


Figures 6.8 a &b: Graphs showing the pH trends in saline water quality water sample at Browns Bay and b: Chelsea from 19 March 1991 to 15 October 2000.

Wilcock and Stroud (2000) analysed the same monitoring data with conventional methods. The report consists of 11 graphs for each and every attribute, making any analysis of the big picture very difficult or impossible. On the contrary, SOMs were seen to be useful in viewing the whole regional data covering a period of ten years in one map. It is also possible to analyse the urbanisation along with the monitoring data and is elaborated in the next section for Browns Bay.

6.4 SOM analysis on integrated data

A SOM map (figure 6.9 a and b) was created with NSCC's building consent data fused with ARC's saline water quality data for Browns Bay in order to explore SOM abilities for modelling ecosystems with ecological and economic data within an integrated framework. In this example, the building consent data is considered as an indicator of developmental activities taking place on North Shore.



Figures 6.9 a & b: SOM and component planes of Browns Bay saline water quality monitoring data with NSCC's building consent data.

Component	C1	C2	C3	C4	C5
b_conscent	1.97	2.38	1.18	2.56	12.28
pH	8.162	8.157	8.157	8.199	8.114
Temp (deg C)	16.44	16.36	15.12	20.42	20.26
SS (mg/l)	4.3	14.84	4.53	4.8	6.04
Turb (NTU)	1.047	2.813	1.308	1.333	0.982
Chloride (mg/l)	19648	19560	18375	19021	19569
Salinity (ppt)	33.29	33.92	32.32	29.76	32.44
TP (mgP/l)	0.0335	0.0429	0.0404	0.0817	0.0407
DRP (mgP/l)	0.01936	0.02513	0.02674	0.02399	0.02909
Nitrate (mgN/l)	0.0111	0.01393	0.02728	0.00997	0.01541
Ammonia (mgN/l)	0.01276	0.01291	0.01023	0.00907	0.0305
BOD (mgO/l)	1.84	2.118	2	2.089	2
Total Coliforms (mpn/100ml)	14.1	33.8	259.1	2.2	2.3
Faecal Coliforms (mpn/100ml)	4.3	13.7	85.3	2.2	2
DO (%)	93.6	99.8	96.5	104.2	108.5
DO (ppm)	7.42	7.82	10.21	6.3	7.44
Secchi*	1.567	1.032	1.34	2.585	1.59
CHLORO*	0.00838	0.00603	0.00737	0.00261	0.00258
ENTER-ME*	2.91	3.29	2.52	2.02	3.17
NO2*	0.00223	0.002116	0.002108	0.002009	0.002128
NO3NO2*	0.00719	0.01013	0.00871	0.00465	0.01623

Figure 6.9 c: SOM cluster details of Browns Bay saline water quality monitoring data and NSCC's building consent data

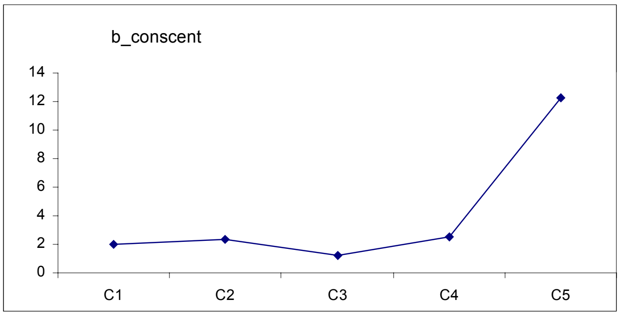
The building consent data was obtained as quarterly summaries whereas the saline water quality sampling was carried out on a monthly basis. Owing to the different time intervals, the two data sets could not be analysed integrated using conventional data analysis methods. To overcome this problem the building consent data was equally apportioned to the three respective months of that quarter and a SOM (figures 6.9 a & b) was created using the fused data set to look for relationships between the building consent data and saline water quality test results at monthly intervals. The following are the interpretations arrived at from the SOM:

(i) Cluster 5 in the bottom centre left with 17-May-99, 20-Apr-00, 19-Jun-00, 19-May-00, 20-Jan-99 and 18-Apr-94 with high building consent values also show the highest NO_3NO_2 (0.01623), ammonia (0.033305 mg/l), nitrate (0.01541 mg/l) and *Enterococci* count (3.17) values in the map.

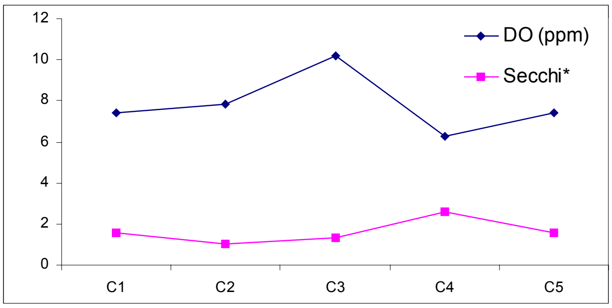
(ii) Cluster 4 data, 14-Feb-94, 20-Mar-95, 29-Jan-93, 17-Feb-95, 26-Mar-93 at 20.42°C temperature, shows the typical summer scenario.

(iii) Clusters 1, 2 and 3 describe three different wintry scenarios along the beach.

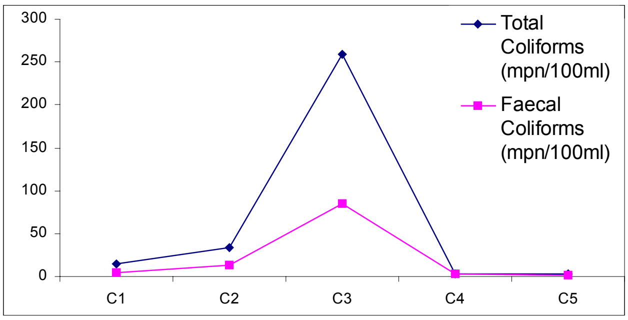
(iv) Within cluster 5, on18 April 1994 ammonia is seen with the highest value of the whole SOM map.



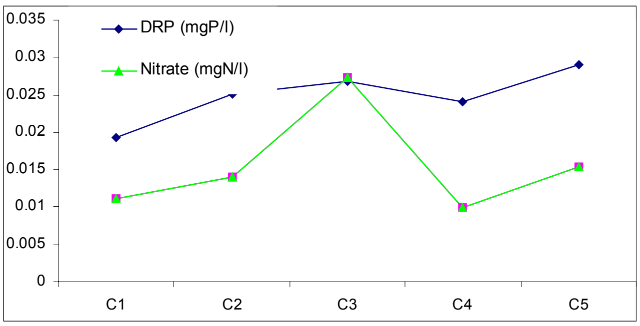
Cluster 5 shows high building consent values.



Cluster 3 consists of high DO with the lowest temperature readings (see chart).



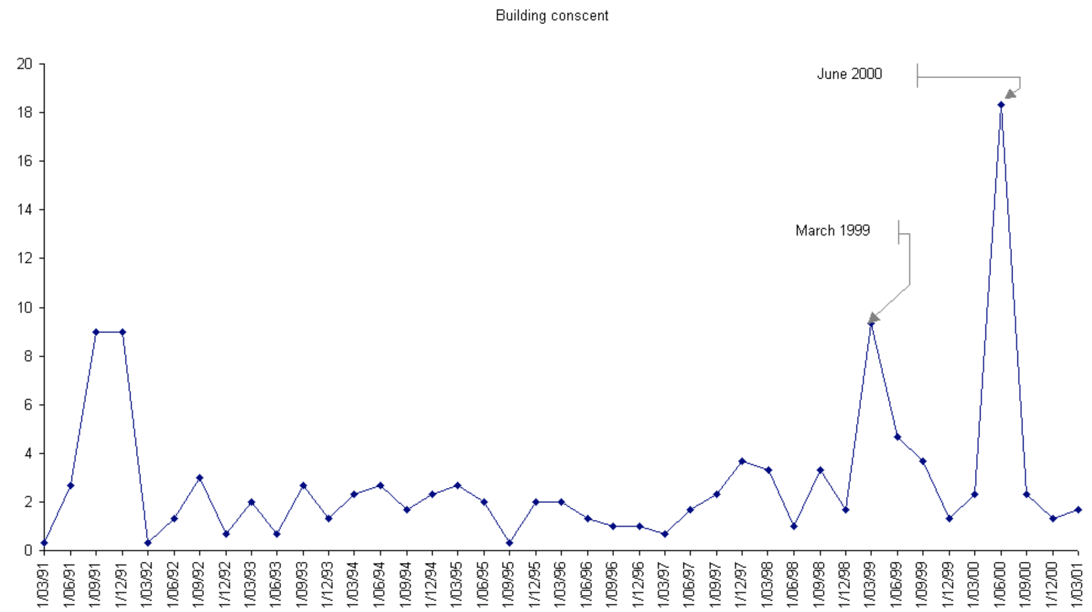
Cluster 3 with high total and faecal coliform values show high bacterial discharges along these sites.

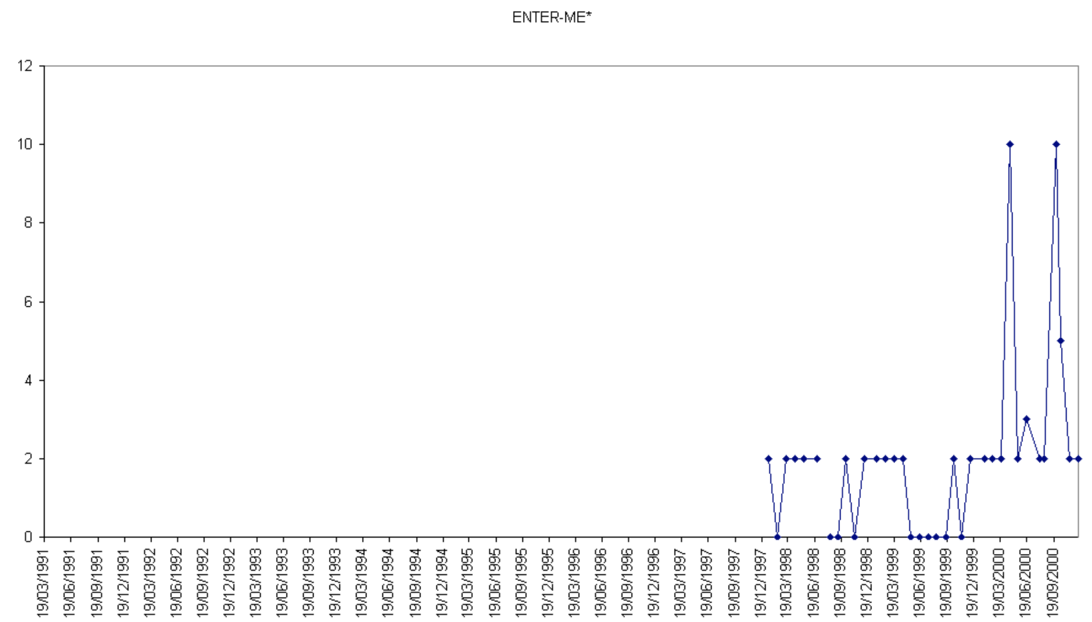


Cluster 3 with high nitrate values

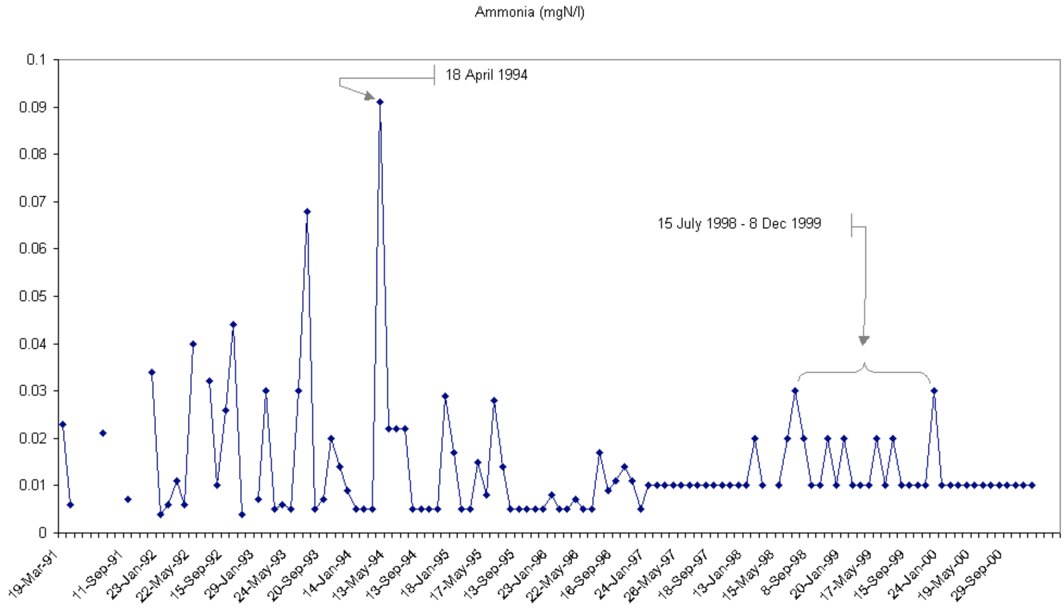
Figure 6.9 d: Graphs showing the cluster details of the SOM created with LTB water quality monitoring data and NSCC's building consent data.

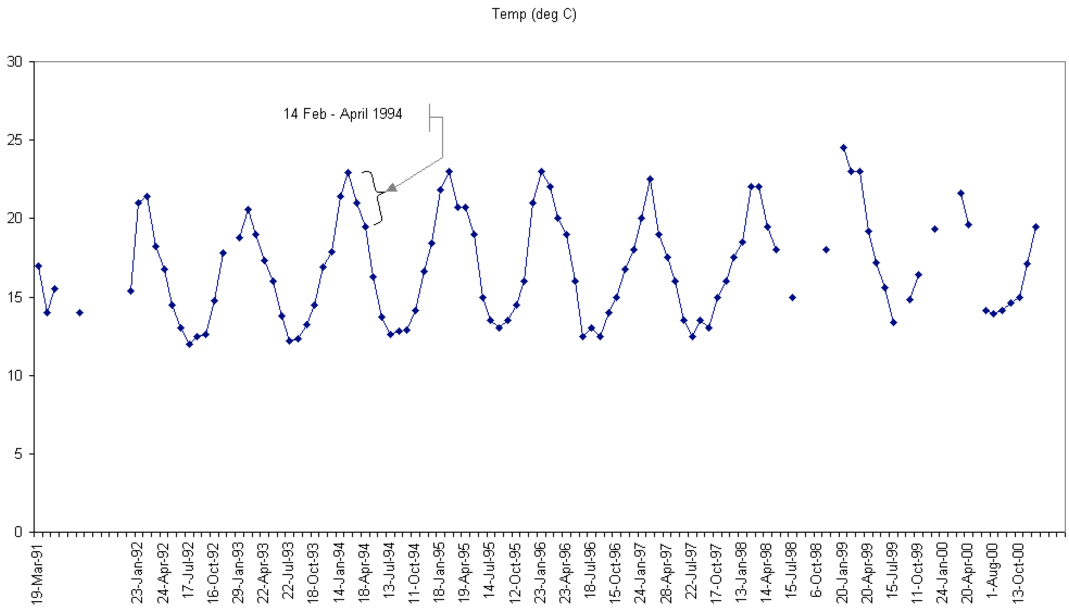
Analysis on the individual graphs of each site's attribute makes comparison among the sites and their seasonal variations within the data laborious, whereas SOM displays produce visual representation of the data, also depicting the spatial and temporal variations within the monitoring data. More details of SOM advantages against conventional methods are elaborated upon in section 6.6 conclusions. Building consent data does not include the lapse or lead time involved in actual building construction and consent approval.





Figures 6.10 a & b: Graph showing a: building consent and b: Enterococci count data against time.





Figures 6.10 c & d: Graphs showing c: ammonia and d: temperature data against time.

Graphs on building consent and *Entercocci* values of Browns Bay (figures 6.10 a & b) show a correlation in the time period analysed.

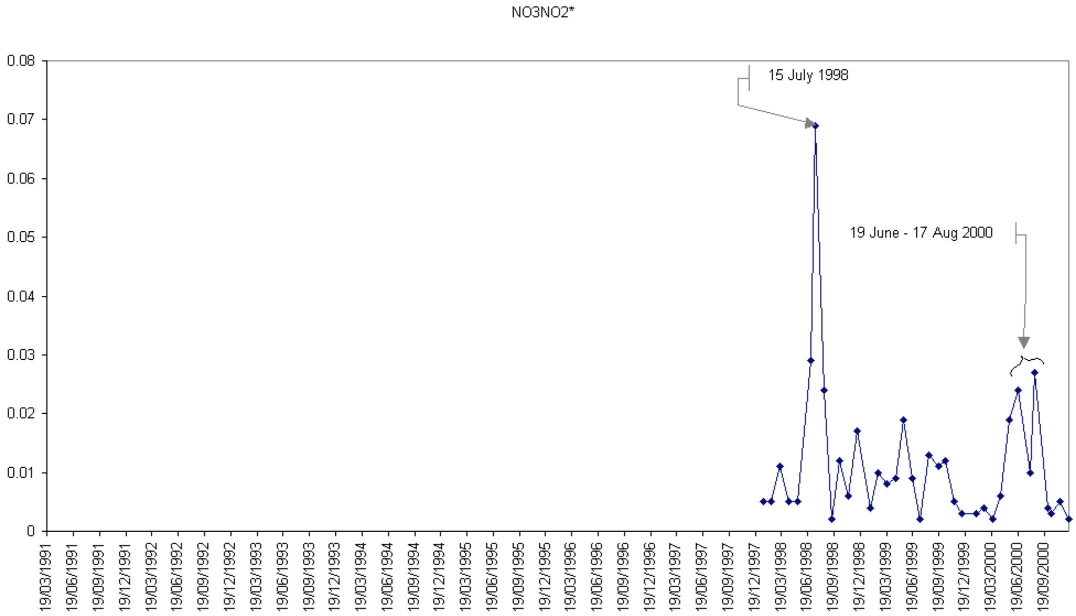


Figure 6.10 e: Graph showing NO_3NO_2 values against time.

In the next section, the second example of the chapter is elaborated upon.

6.5 ARC subtidal data Analysis

In 1998, ARC began conducting sub and intertidal biological monitoring programmes to study the effects of urbanisation on the marine habitats of the Waitemata Harbour and northeastern coasts of Auckland. The main aim of the programme was to analyse the population dynamics (such as species composition changes in the monitoring data), to study the environmental change and its effects on the coastal habitats under study. Initially, the programme was limited to Long Bay only, but eventually was extended to cover areas from Campbell Bay to Waiwera (figure 6.11), fulfilling the BACI design method requirement; to see whether the Long Bay's species composition changes were confined only to its coastal systems, affected by the urbanisation in near shore. Researchers of UoA based at the Leigh Marine Laboratory have been carrying out the monitoring programmes for ARC, since 1998. For more information on the methods, the original report (Walker et al. 2000) should be consulted.

6.5.1 SOM analysis and Methodology

Of the above monitoring data collected by UoA research staff, only the subtidal species composition changes (consisting of 42 species count average data of 30 sites from six beaches Campbells bay, Torbay, Long Bay, Manly, Stanmore and Waiwera (figure 6.11) along with sedimentation data are analysed with SOMs in this section.

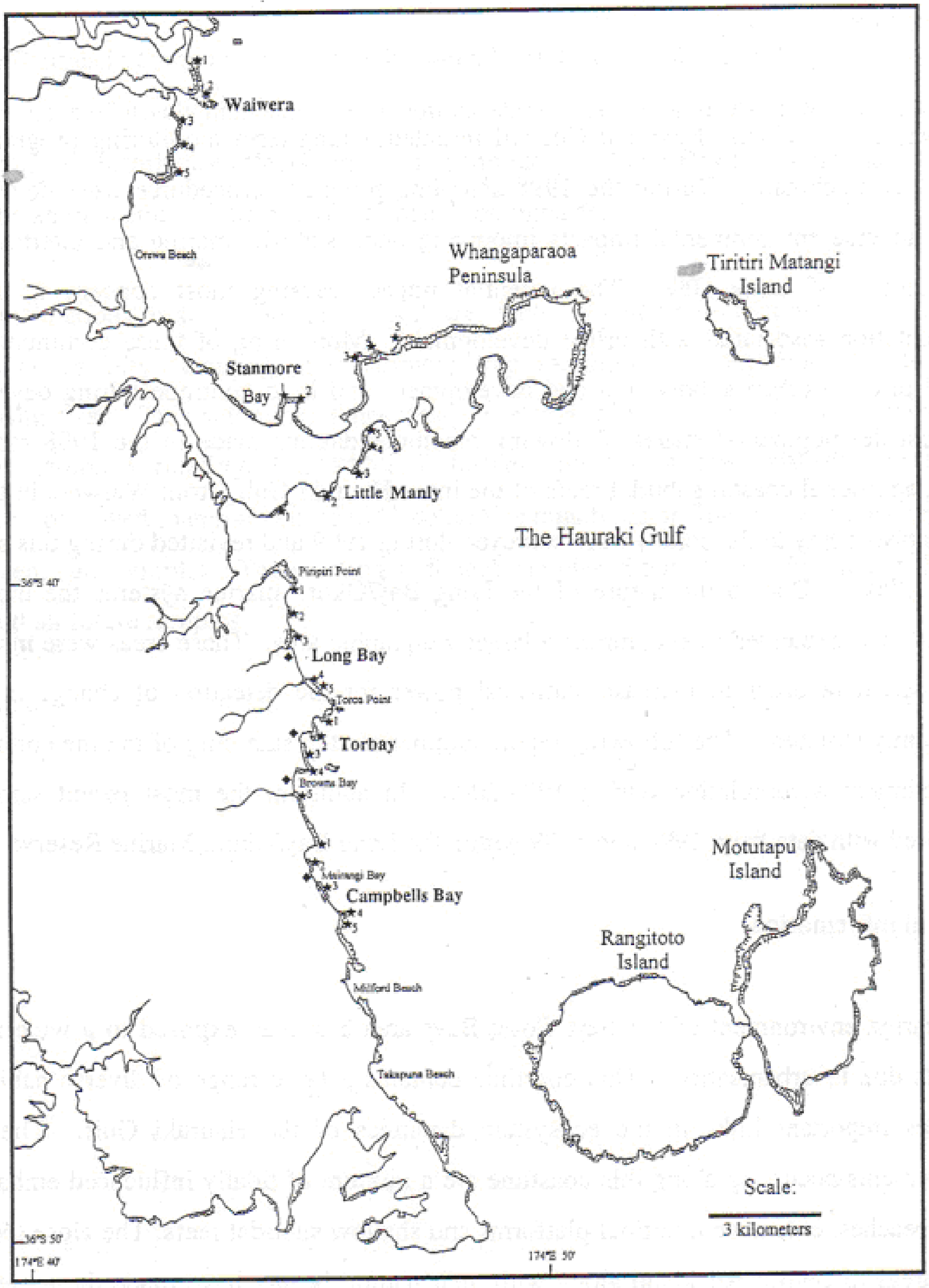


Figure 6.11: Map of the Hauraki Gulf with sites and areas where monitoring of subtidal (with asterisk) and intertidal (with diamonds) communities was carried out during March 2000. Source: (Walker et al. 2000:5).

The following were the main problems encountered with conventional ecological data analysis methods based upon (Walker et al. 2000):

(i) Size of the data: Due to the presence of a large number of species in the quadrants, analysis with standard ecological methods using conventional statistical calculations produced large matrices, which caused difficulties in studying the community patterns in the data. This is a common problem faced by ecological data analysts (Giraudel and Lek 2001).

(ii) The researchers did not have any prior knowledge of any indicator species for the sites studied, hence, analysed all the species collected.

(iii) Many of the species have zero at most of the sampled quadrants. However, elimination of rare species in similar multi species biomonitoring solutions invariably lead to false conclusions argued (Cao et al. 2002). Hence, Walker et al. (2000) did not eliminate the rare species in their analysis and advised the same in this research.

The following are the measures taken to overcome the above-motioned problems:

(i) Viscovery® SOMine lite version 4.0 by eudaptics software gmph package is used to create SOM maps as it has the capacity to handle up to 50 variables.

(ii) Initially the average counts of all 42 species, calculated from the collected data covering a period of three years from 1999 to 2001 are used in the SOM analysis. Species average count values are used in conventional multivariate analyses, including (Walker et al. 2000), hence adopted in SOM analyses too.

(iii) Later, count average data on 25 species, selected by Walker et al. (2000) as sensitive to environmental changes along the beaches analysed, is analysed alone and then with sedimentation data from the same monitoring programme. This is done to see whether these species could be used to model coastal ecosystem changes caused by sedimentation arising from urbanisation on North Shore. If proven this could facilitate the design of biological indicators based on subtidal population dynamics to analyse the effects of urbanisation on coastal and marine life at Long Bay and the other beaches along the coast.

(iv) Sedimentation rates of less than 63 micron particles are included in the studies as per expert advice obtained from the Leigh laboratory researchers.

6.5.1.2 Results and discussion

Firstly, a SOM (figures 6.12 a and b) was created with the 42 subtidal species average count data, from the 30 sites (5 sites from each of the six beaches selected along the northeastern coast of Auckland), from 1999 to 2001. In order to look for the correlations between sediment deposition rates (percentage of <63 micron sedimentation) and subtidal population dynamics, another SOM (figures 6.13 a, b and c) was created with these two data sets (the sedimentation data linked), which were actually analysed separately in (Walker et al. 2000). However, the clustering patterns seen in these two SOMs (figures 6.12 a b c, 6.13 a b & c) without and with the percentages of sedimentation (<63 microns) linked with the subtidal population dynamics show similar patterns. This confirmed the correlation between the two data sets (the sedimentation deposition rates and their effects on the subtidal population dynamics observed through experiments along these beaches). SOMs displayed the variables on visual formats through which the analysis was enhanced significantly. Viscovery's ability to link a component to the SOM (see chapter 4), in this case the sedimentation data with other components (the species dynamics data), show potential for developing prediction models on the subtidal population dynamics with NIWA's simulated sedimentation data.

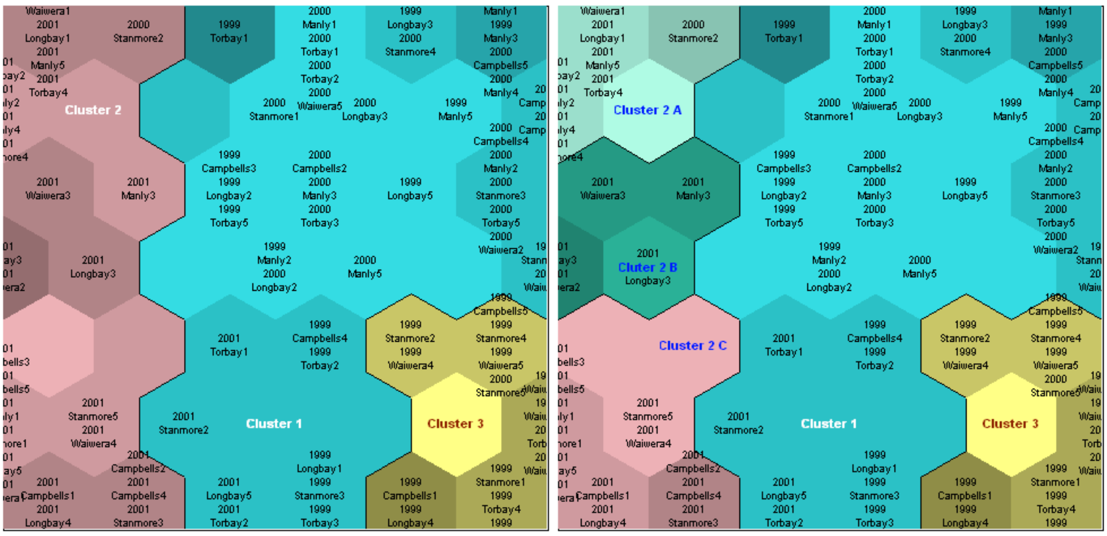


Figure 6.12 a: SOM of average count of all 42 species, for three years from 1999 to 2001. Map parameters used are 50 nodes, priority of all components set to 1 and all other parameters set to default values. b: SOM with five clusters

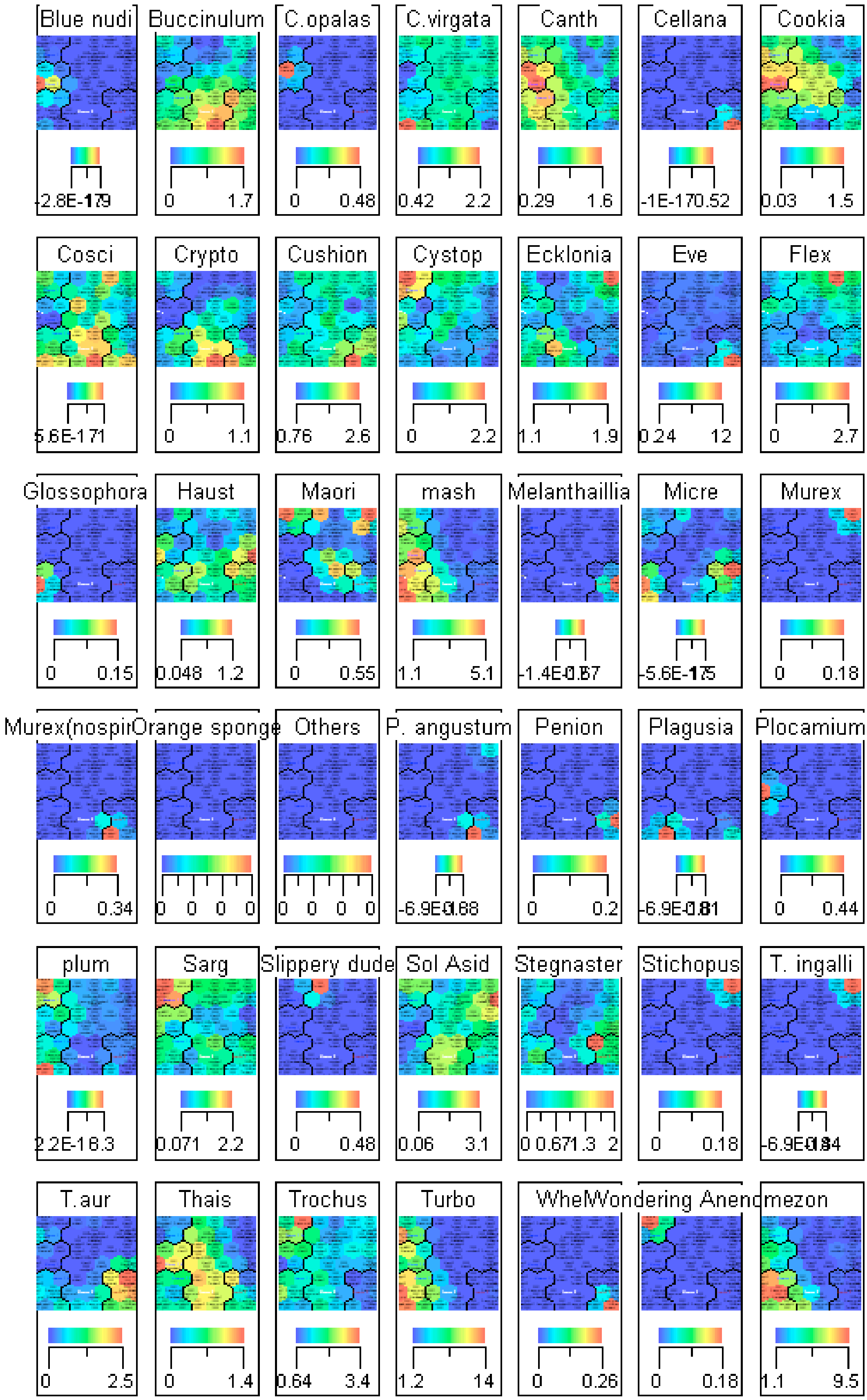
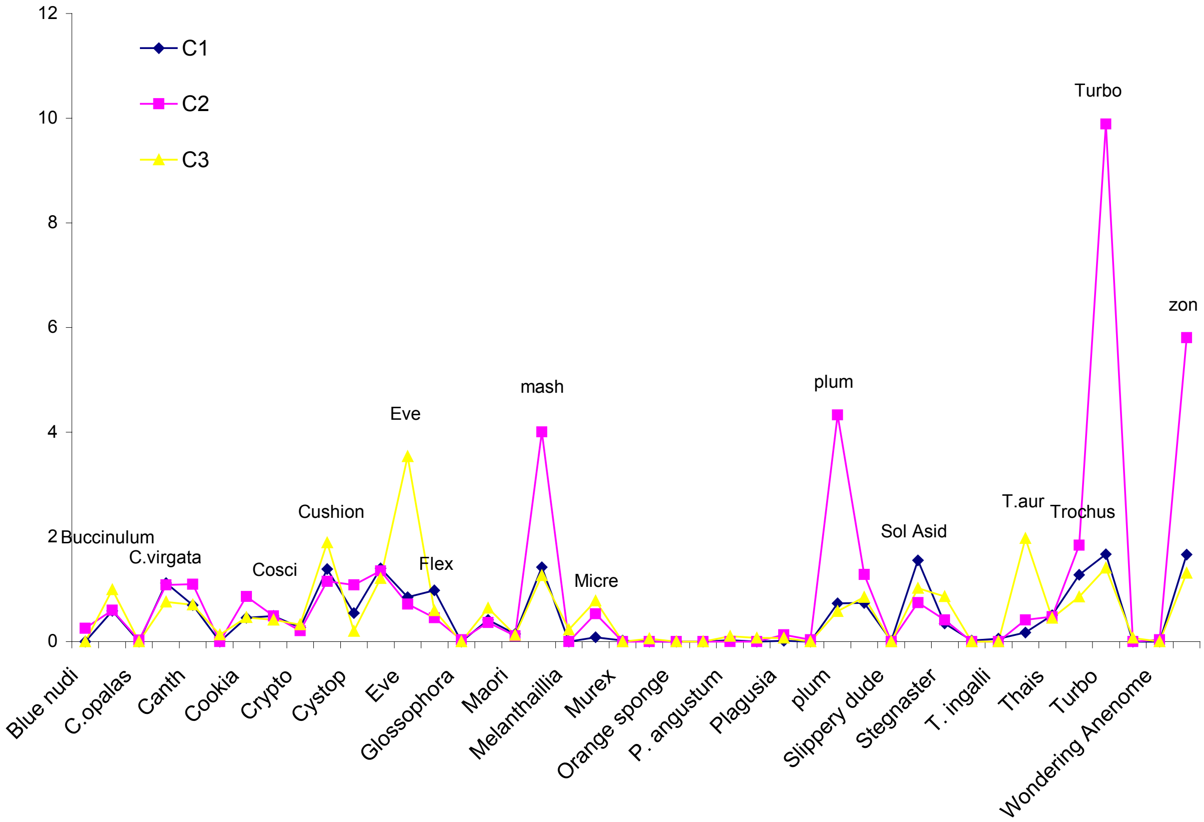
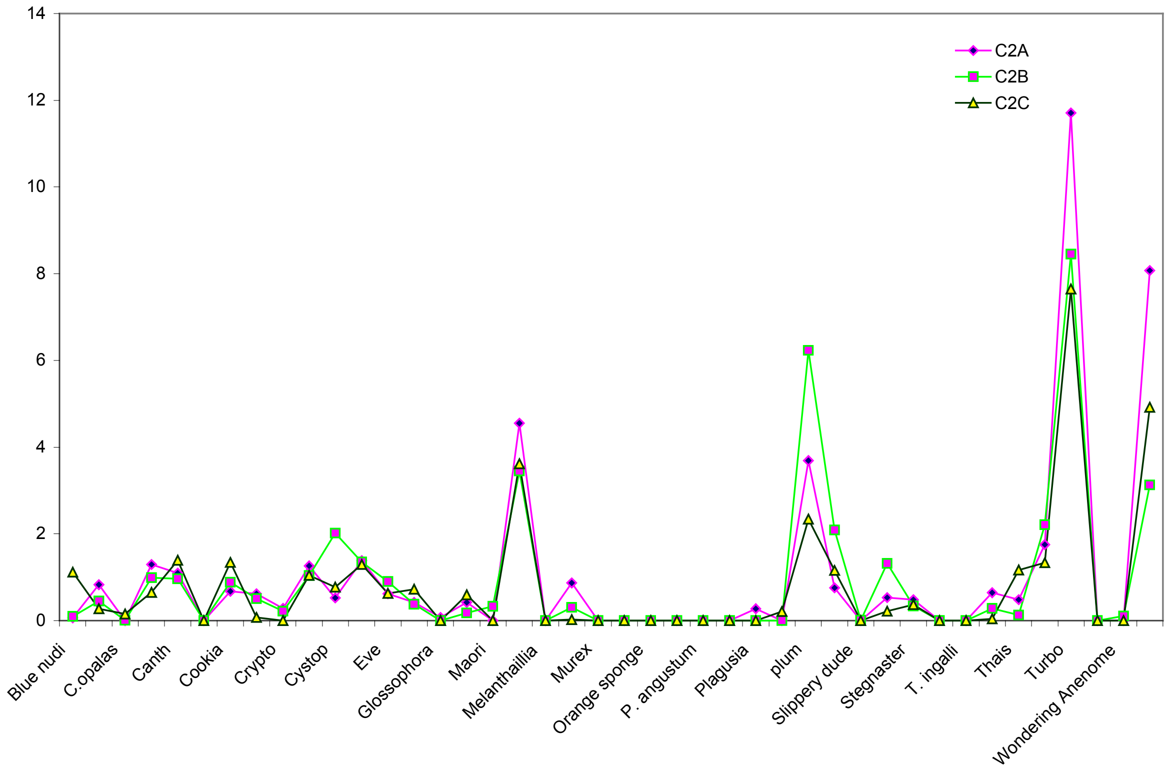
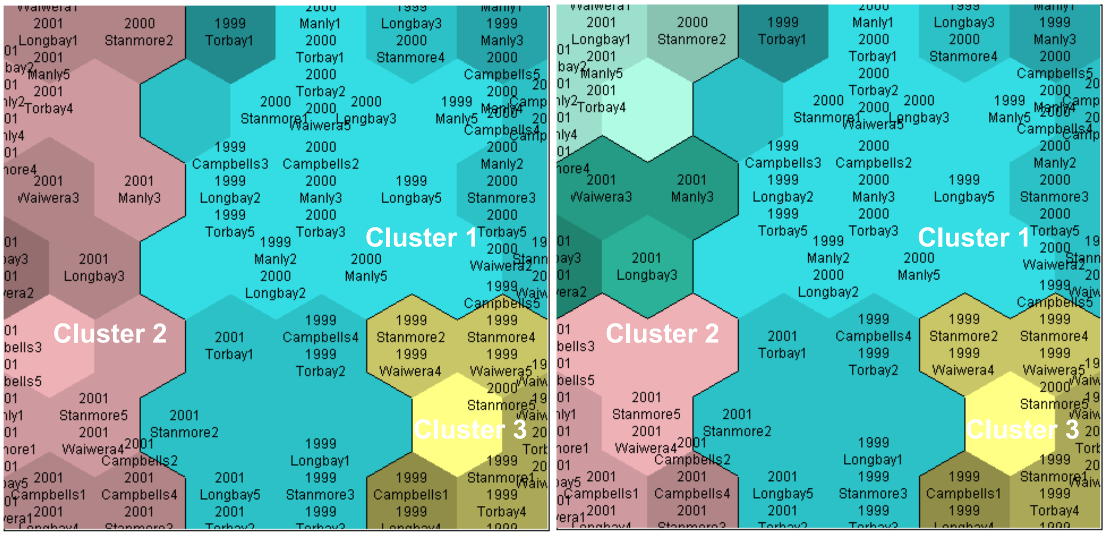


Figure 6.12 c: SOM component planes (with five clusters) of average count data of all 42 species, for a period of three years from 1999 to 2001.





Figures 6.12 d & e: Graphs showing the cluster details of the SOM of 42 intertidal species average count data d: three clusters and e: subdivisions of cluster 2 (2 A, B and C).



Cluster no.	1999	2000	2001	Grand Total
1 (cyan)	18	25	3	45
2 (magenta)	1	1	27	30
3 (yellow)	12	3	-	15
	30	30	30	90

Figures 6.13 a & b: SOM of the 42 species average count data with sedimentation data linked, from 1999 to 2001 a: with 3 clusters and b: with five clusters

The following is a summary of the observations seen in these two SOM maps (figures 6.12 a b c, 6.13 a b & c):

(i) In both maps the clustering has picked up the annual variations among the beaches. Clusters 1, 2 and 3 with many of year 2000, 2001 and 1999 data respectively. Walker et al. (2000) also identified the consistent variations and concluded that the monitoring frequency of subtidal samples as adequate to incorporate natural variations. However, the percentage of 2001 beach sites in cluster 2 is seen prominent than that of 2000 and 1999 in clusters 1 and 3 respectively. This leads to the conclusion that the deviation in year 2001 to be more than that of annual. SOM cluster data superimposed on a GIS Arc View 8.2 (figure 6.13 d) illustrates this fact.

(ii) When divided into five clusters, cluster 2 was further subdivided into three clusters; 2A, 2B and 2C, all five sites of Campbells bay for year 2001 falling in the left bottom corner, in cluster 2C. This shows that the inter beach variations to be more than that of inter annual.

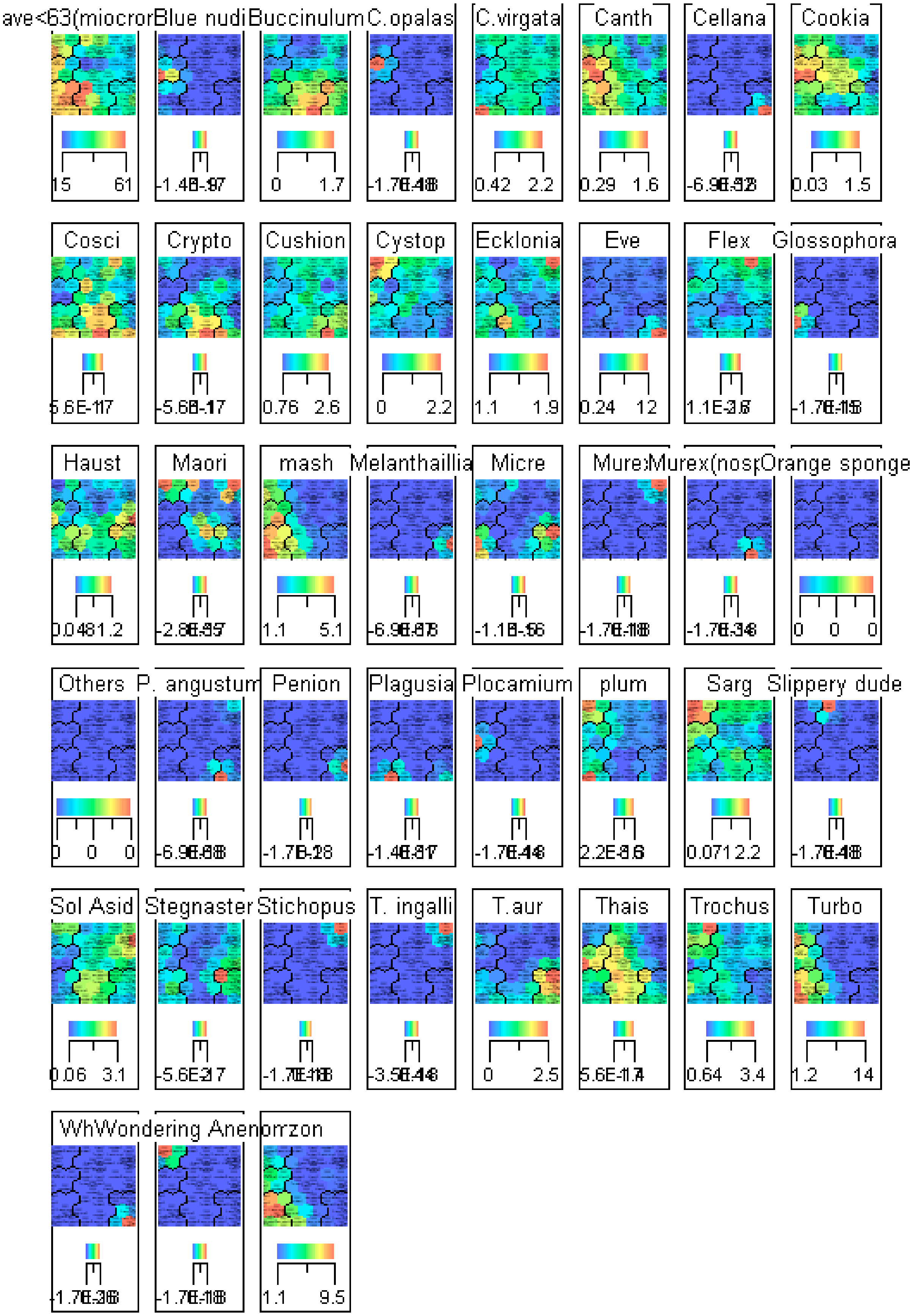


Figure 6.13 c: SOM component planes of 42 intertidal species average count data with sedimentation data linked, from 1999 to 2001.

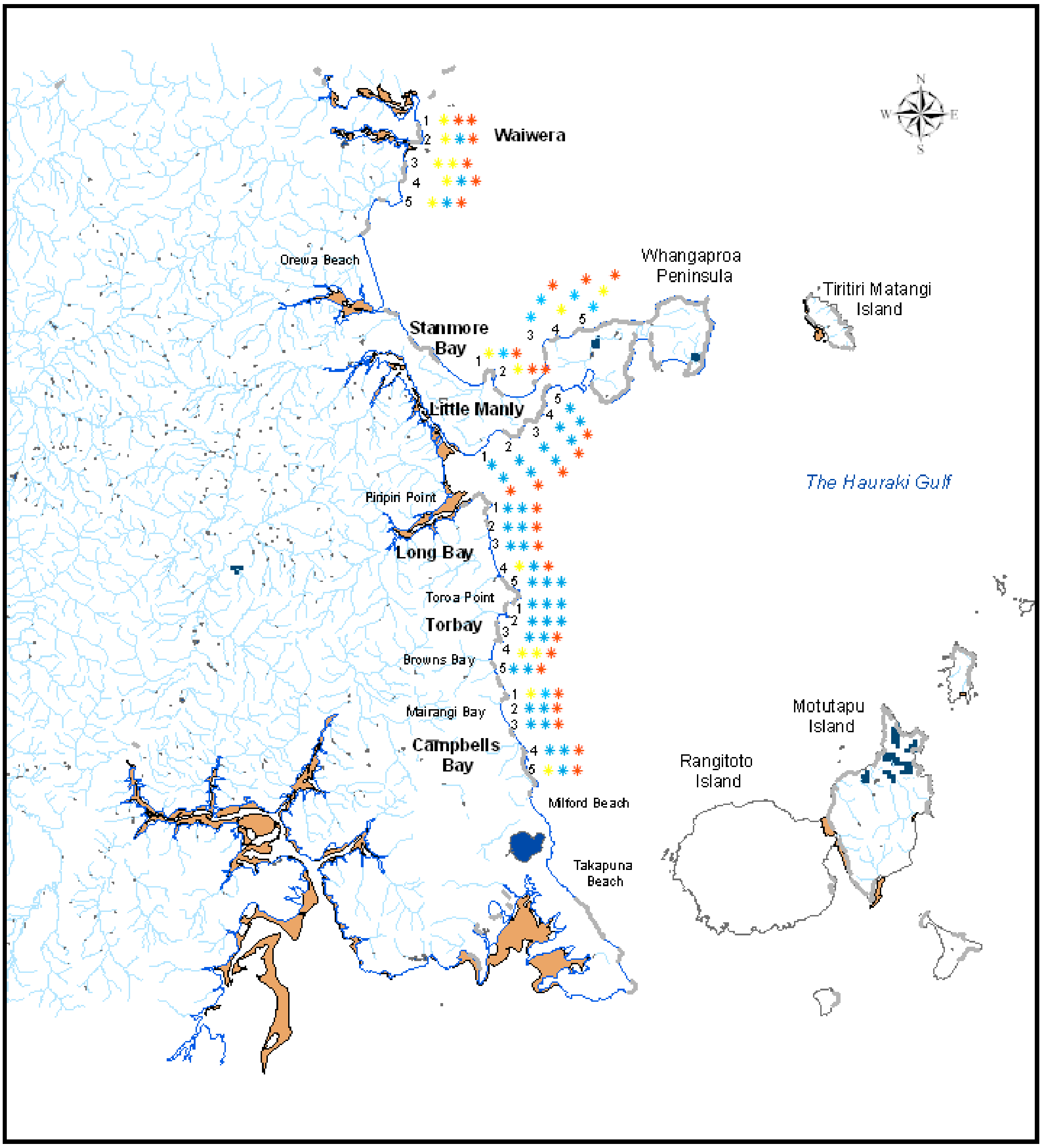
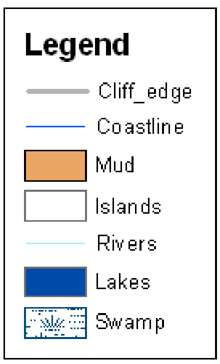
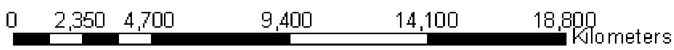
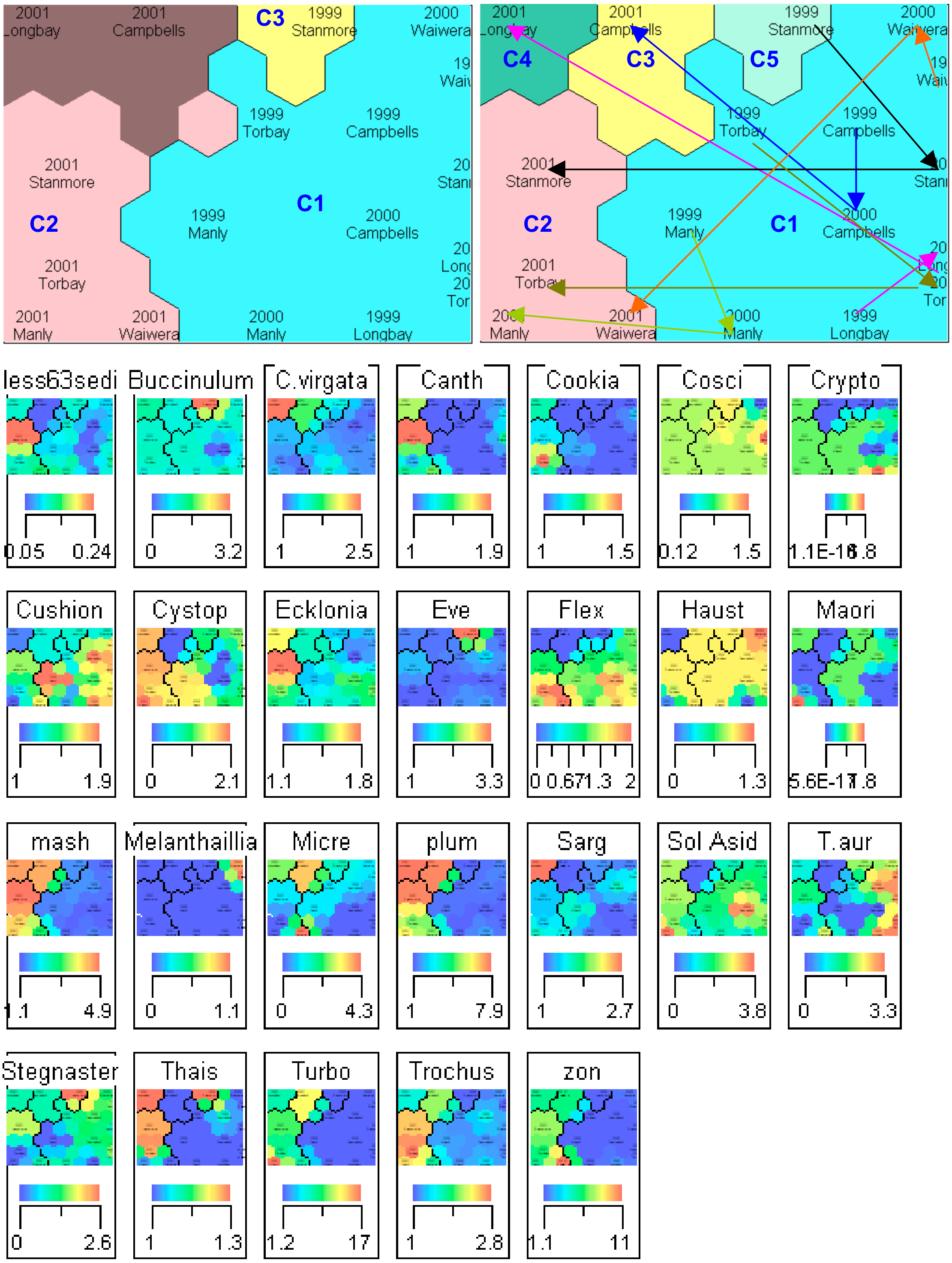


Figure 6.13 d: SOM clustering results (all 42 species average count data with sedimentation data linked, for three years from 1999 to 2001- left to right) superimposed on a GIS Arc View 8.2. The colours relate to the SOM clustering details in figure 6.13 a.

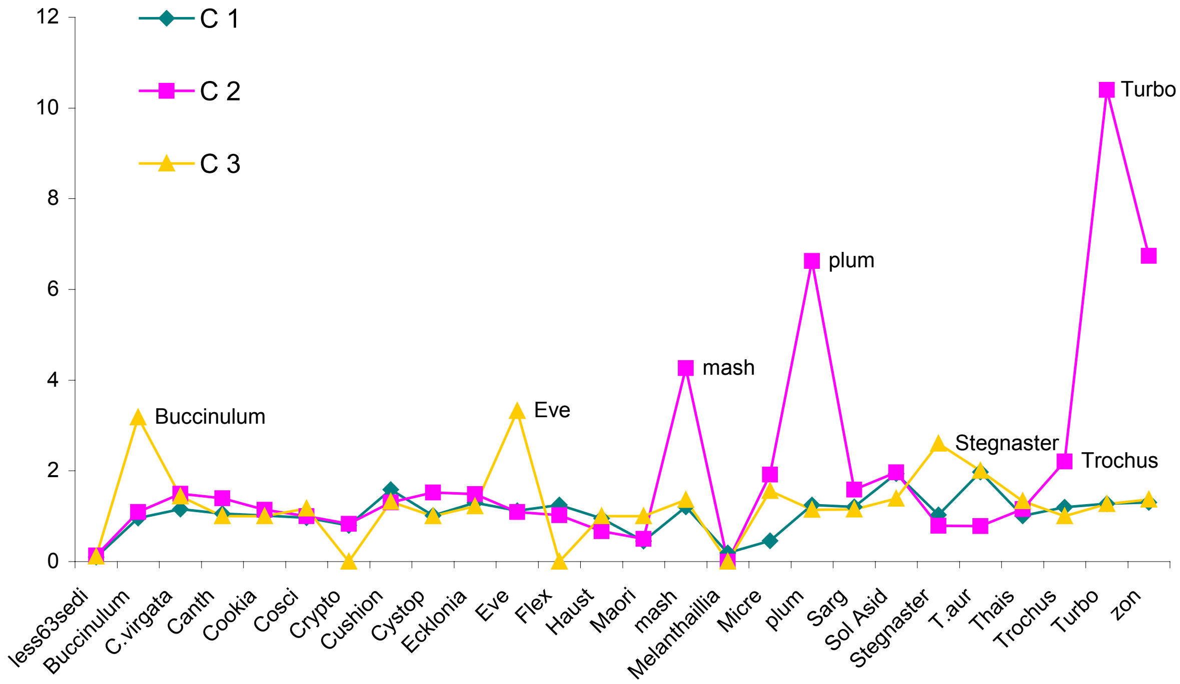


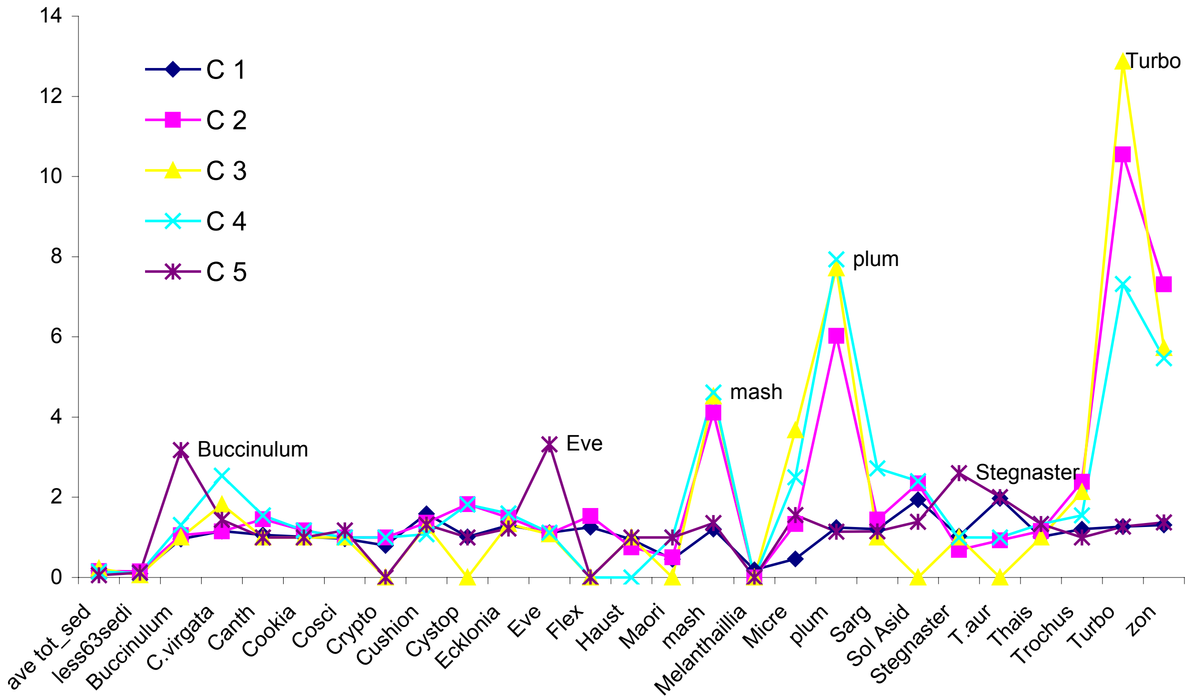


In the next stage, SOMs were created with a summarised list of 25 species average count data for the six beaches covered in the monitoring programme with (figures 6.14 a b & c) and without sedimentation data (figures 6.15 a, b, and c).

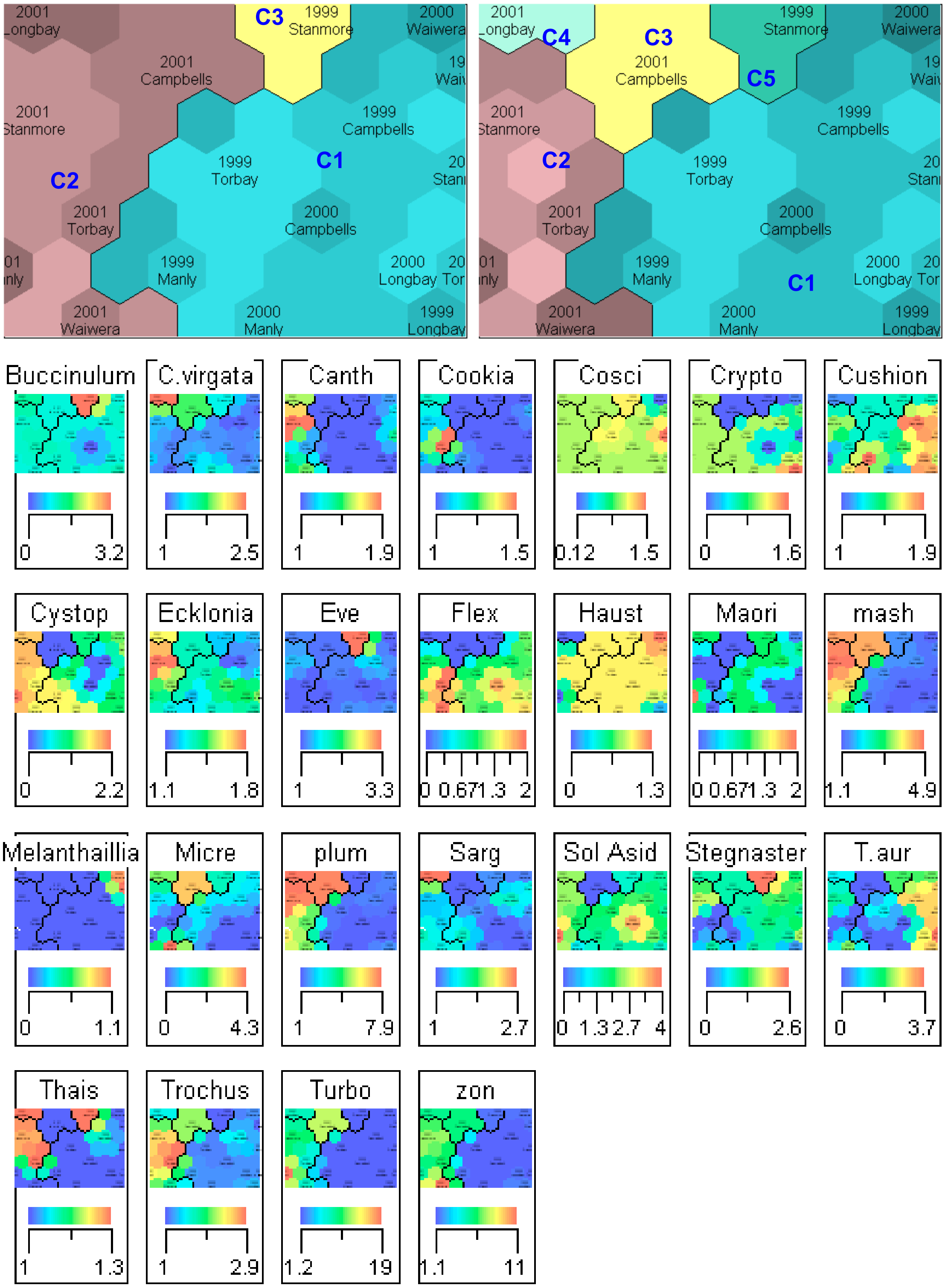


Figures 6.14 a, b & c: SOM created with a summarised list of 25 species average count data with sedimentation data. Map creation parameters: 50 nodes priority of all components set to 1 and all other map parameters set to default values. a: three cluster SOM, b: five cluster SOM and c: SOM component planes.

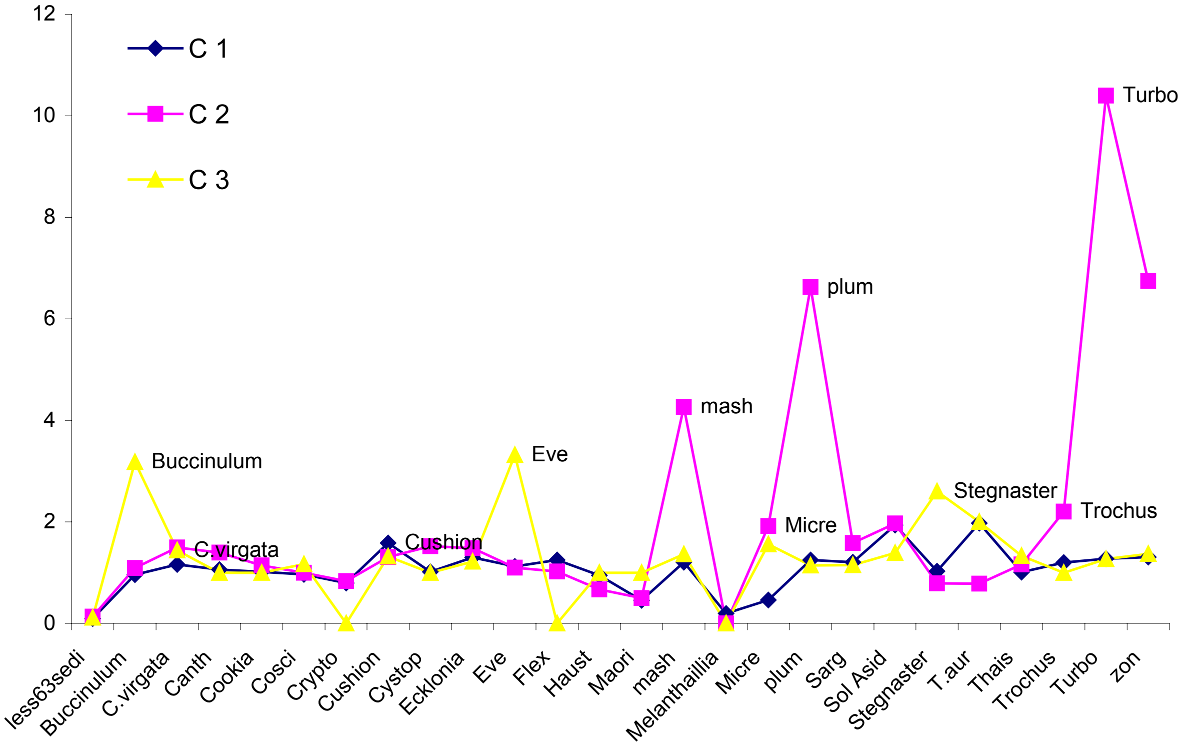


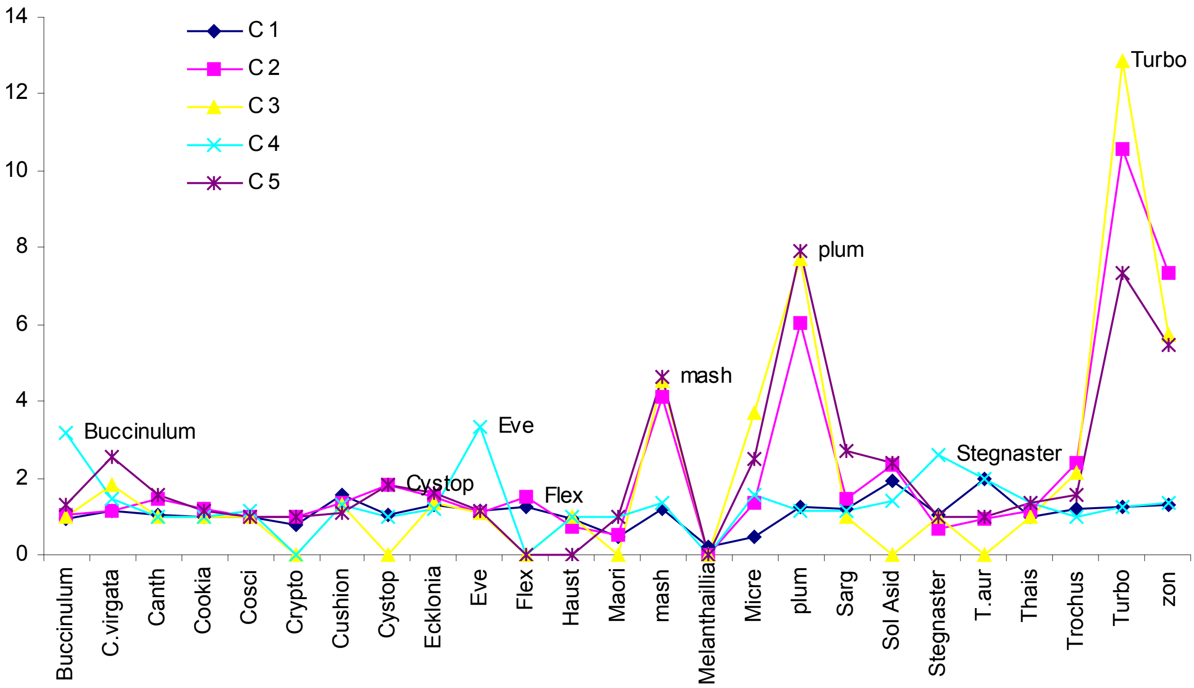


Figures 6.14 d & e: Graphs showing SOM cluster details d: three clusters and e: five clusters. The SOM was created with a summarised list of 25 species average count data with sedimentation data.



Figures 6.15 a, b and c: SOM created with a summarised list of 25 species average count data without sedimentation (<63 microns) data. Map creation parameters: 50 nodes priority of all components set to 1 and all other map parameters set to default values. a: three cluster SOM, b: five cluster SOM and c: SOM component planes.





Figures 6.15 d & e: Graphs showing SOM cluster details d: three cluster and e: five clusters. The SOM was created with a summarised list of 25 species average count data without sedimentation data.

The interpretations derived from the SOMs of 25 species with sedimentation (figures 6.14 a b & c) and without sedimentation data (figures 6.15 a, b, and c) are:

(i) Data points were separated into 3 major clusters with the year 2001 data in the left side on both maps, (with and without sedimentation data). Similarly, most of the year 1999 points were gathered in the centre along with the year 2000 data in the right side of the maps. Stanmore 1999 is seen alone in one cluster. This shows that the annual variations among the subtidal population dynamics along these beaches can be differentiated by SOM cluster analyses.

(ii) When divided into five clusters, year 2001 cluster got further divided into three divisions. It could be interpreted that the variations among beaches in year 2001 as higher than that of experienced in the earlier years.

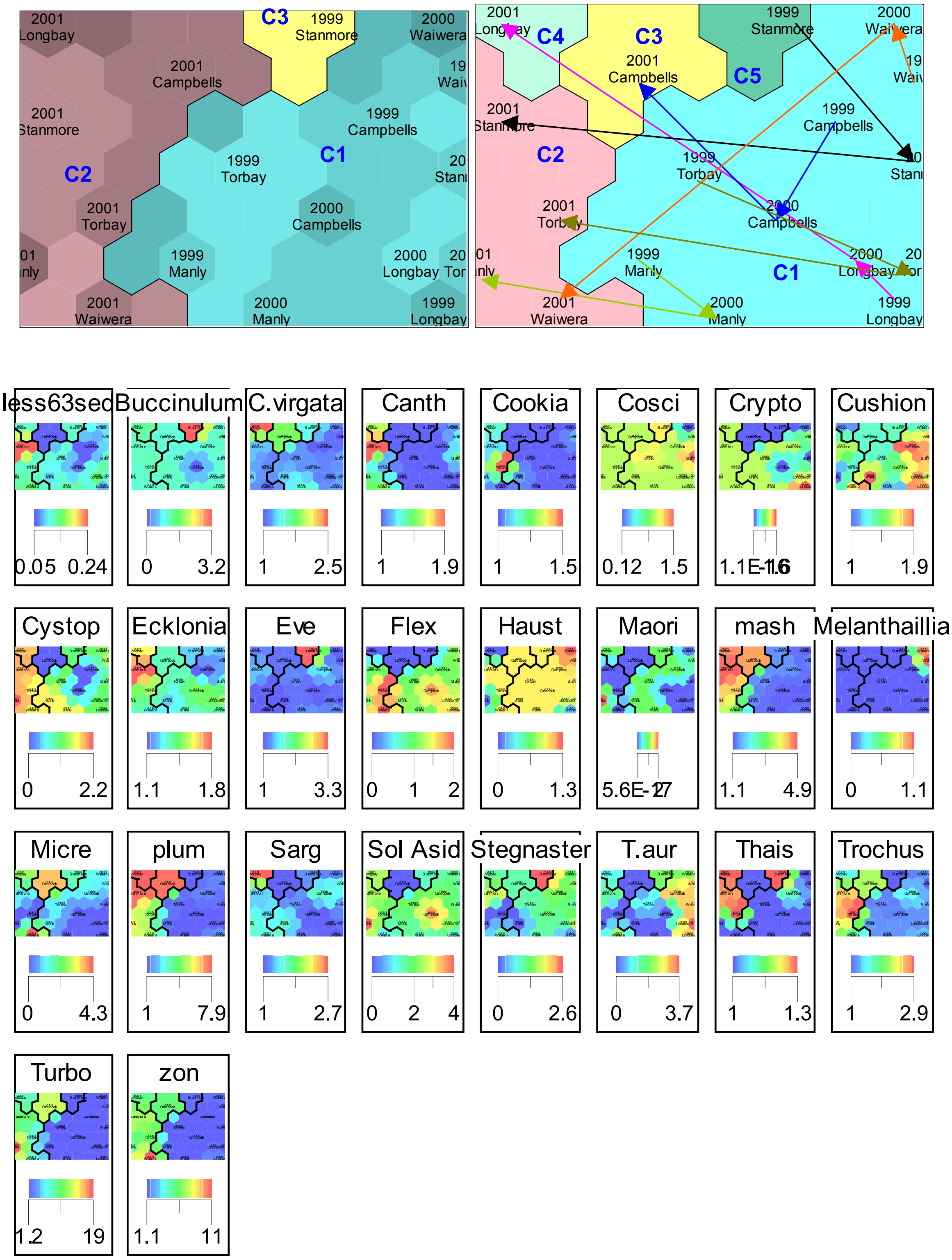
(iii) Of the five beaches Campbells bay 2001 and Long bay 2001 fall into different clusters. All the other sites Manly, Torbay, Stanmore and Waiwera for 2001 fall into one cluster. In figure 6.12 b, all 5 sites of Champbells bay fall in 2C

Arrows in figure 6.14 b show the trajectory for each and every beach over the three year period monitored.

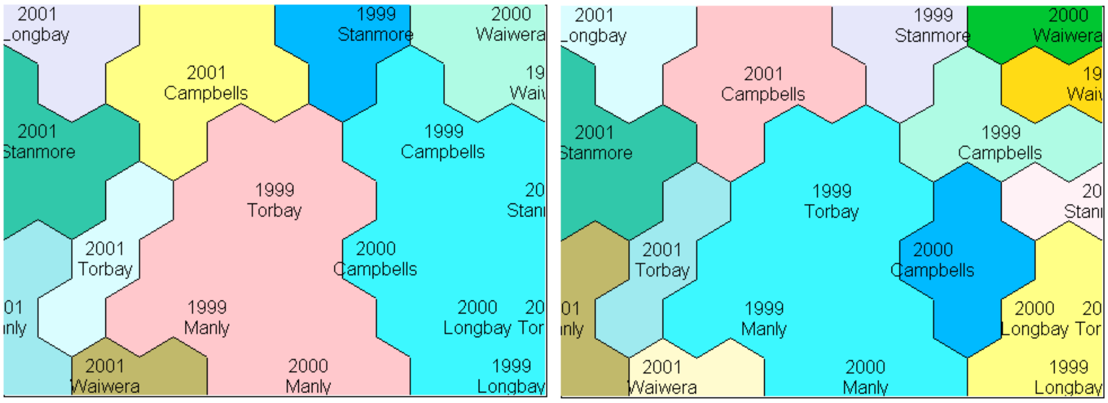
Finally, a SOM map (figures 6.16 a & b) was created with sedimentation data linked, by setting the priority of this component to 0.0001. This would enable making predictions on the effects of sedimentation on the population dynamics of these subtidal habitats (see chapter 4 Gold price prediction analysis). The following are the interpretations arrived at from this map:

(i) When the SOM map was divided into ten and fourteen clusters (figures 6.16 e and f), the clustering showed a good correlation to five species average count data: Echinoderms: *Patiriella regularis, Evechinus chloroticus*, *Stegnaster inflatus,* and a sponges species *Tethya aurantium* . Macroalgae *Carpophyllum flexuosum* showed a negative correlation to an extent.

(ii) A few more species showed positive response during the year 2001. Macro algae *Carpophyllum maschalocarpum, Carpophyllum plumosum*, *Sargassum sinclairii*, *Zonaria turneriana* along with herbivorous gastropods *Turbo smaragdus, Trochus viridus,* Predatory whelk *Thais orbita, Cookia sulcata* and *Cantharidus purpureus*.



Figures 6.16 a, b and c: SOM created with a summarised list of 25 species average count data with sedimentation (<63 microns) data linked. Map creation parameters: 50 nodes priority of sedimentation set to 0.0001 and all other components set to 1 and all other map parameters set to default values. a: three cluster SOM, b: five cluster SOM and c: SOM component planes.



Figures 6.16 e & f: SOM created with a summarised list of 25 species average count data with sedimentation (<63 microns) data linked. e: ten cluster SOM, f: fourteen cluster SOM and c: SOM component planes.

In order to confirm the correlations between the observed sediment deposition rates and the subtidal species average count data from the above SOM analyses, significance tests were carried out and the following species showed significant correlation at less than 0.05p value;

`Correlations: % ave<63(mio, tot ave(g/d), Blue nudi, Buccinulum, C.opalas, C.vir….`

`Flex mash Sarg Thalis, Thais, Trochus, Turbo, Zon`

Flex	-0.250	0.159	-0.045	-0.157	0.172	-0.114	0.026	-0.130
	0.025	0.158	0.674	0.140	0.104	0.284	0.805	0.220
mash	0.535	0.120	0.208	0.057	0.039	0.149	0.322	-0.084
	0.000	0.289	0.049	0.593	0.717	0.162	0.002	0.433
Sarg	0.301	-0.120	0.118	-0.008	0.010	-0.062	0.084	-0.081
	0.007	0.287	0.269	0.944	0.929	0.559	0.431	0.449
Thais	0.253	0.156	0.132	0.175	0.095	0.136	0.150	0.003
	0.023	0.167	0.214	0.098	0.372	0.201	0.159	0.977
Trochus	0.312	-0.032	-0.055	-0.094	0.137	-0.031	0.227	-0.158
	0.005	0.778	0.608	0.381	0.197	0.773	0.031	0.137
Turbo	0.395	0.099	0.100	-0.014	0.115	0.086	0.224	-0.079
	0.000	0.382	0.351	0.897	0.281	0.422	0.034	0.457
zon	0.502	0.092	0.148	-0.036	0.097	-0.052	0.361	-0.080
	0.000	0.416	0.163	0.733	0.364	0.623	0.000	0.454

Of the several subtidal species that exhibited negative association with sedimentation in the SOM analyses, only one macro algal species *Carpophyllum flexuosum* was confirmed by the significance test. However, macro algae *Carpophyllum maschalocarpum, Sargassum* *sinclairii, Zonaria turneriana* along with herbivorous gastropods *Turbo smaragdus, Trochus viridus* Predatory whelk *Thais orbita* were verified of having positive correlation with increased sediment deposition percentages observed through SOMs.

A SOM and an ESOM (figure 6.17 a & b) were created with RICBIS using 42 species average count data along with sediment deposition rates. In the maps annual variations within the 30 beach sites could be observed. Year 2001 data points on the right of the SOM look more dissimilar than in the earlier years (1999 & 2000). In the ESOM (figure 6.17b) year 2001 points look scattered all over the map.

In the SOM (figure 6.17a) year 1999 data points look more similar, falling into one cluster at the left bottom corner, year 2000 sites could be seen breaking up into two clusters and for 2001 completely broken into two distinctive clusters. This could be interpreted that the population change in the subtidal community for year 2001 as different to that of the annual variations in the previous years.

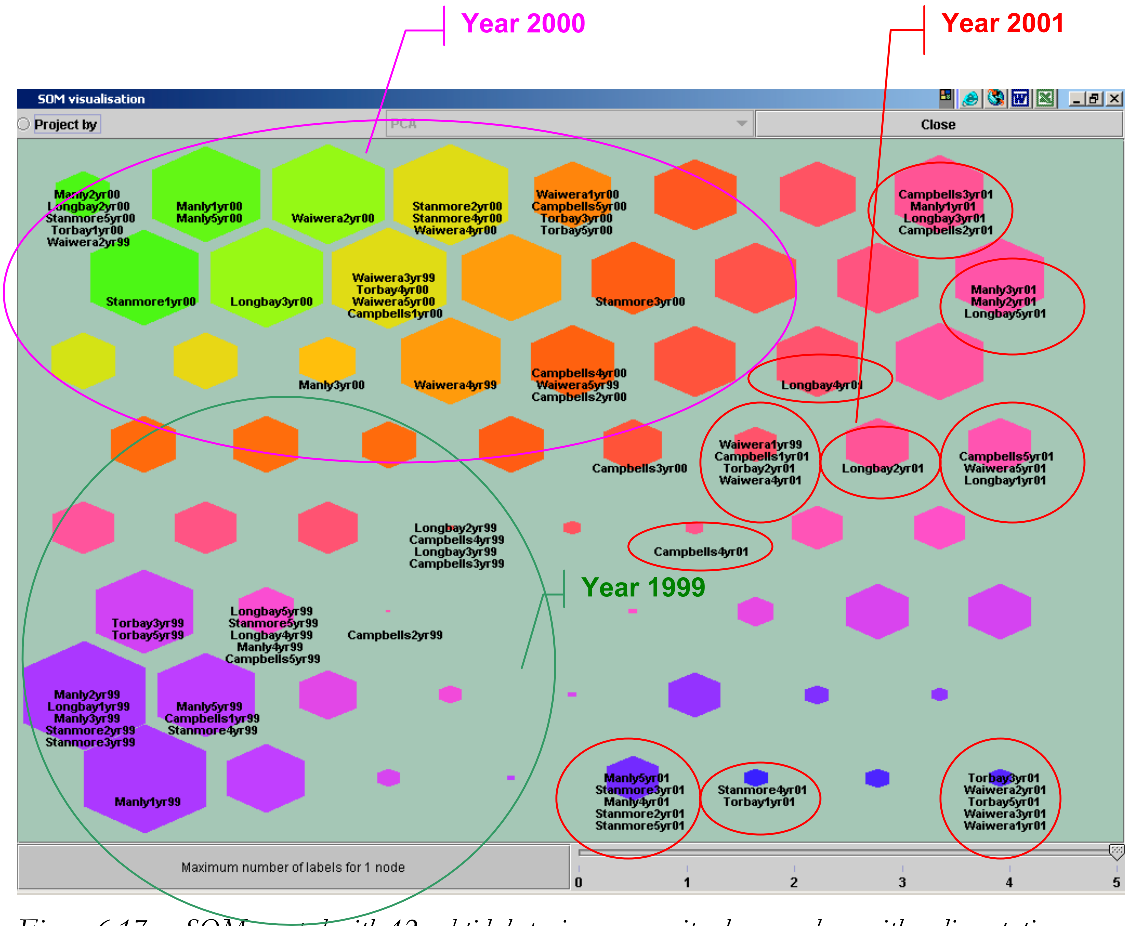


Figure 6.17 a: SOM created with 42 subtidal species community changes along with sedimentation values using RICBIS

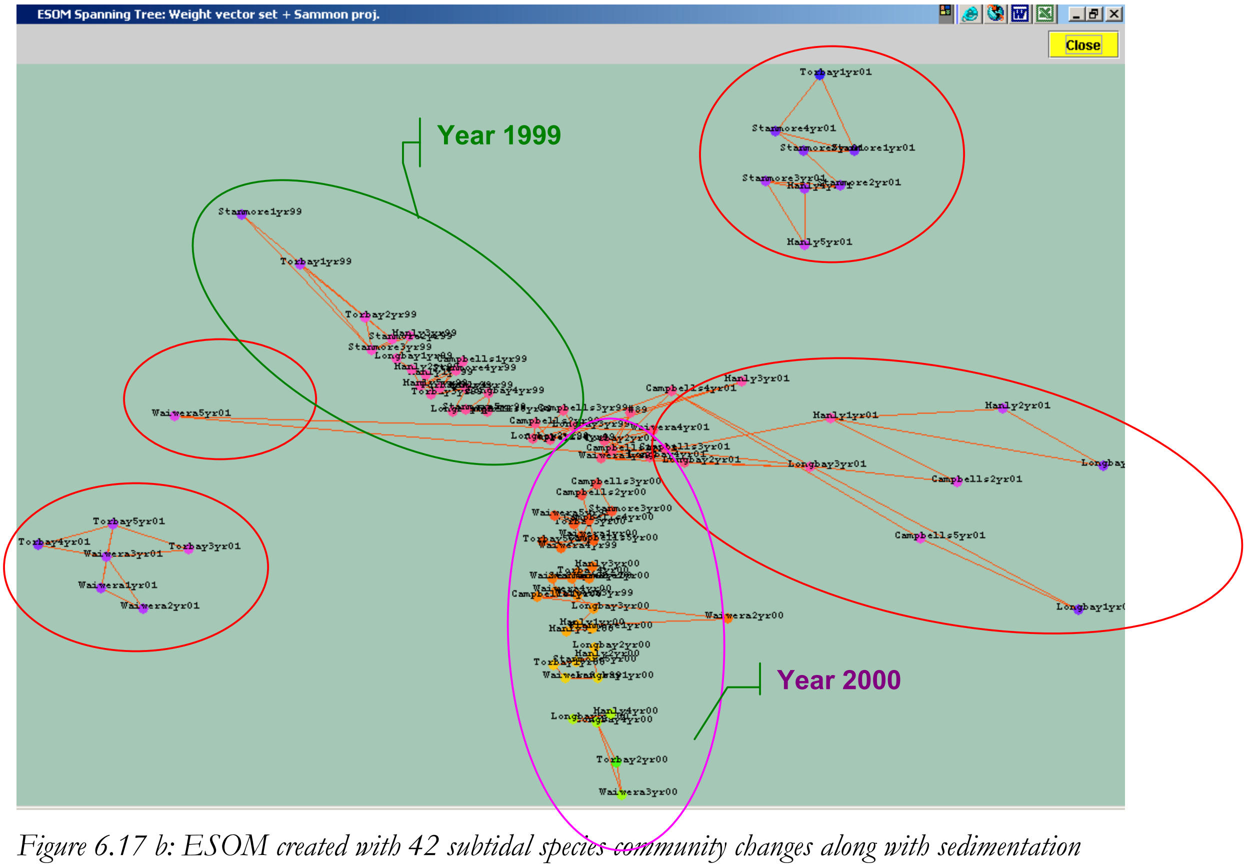


Figure 6.17 b: ESOM created with 42 subtidal species community changes along with sedimentation values using RICBIS

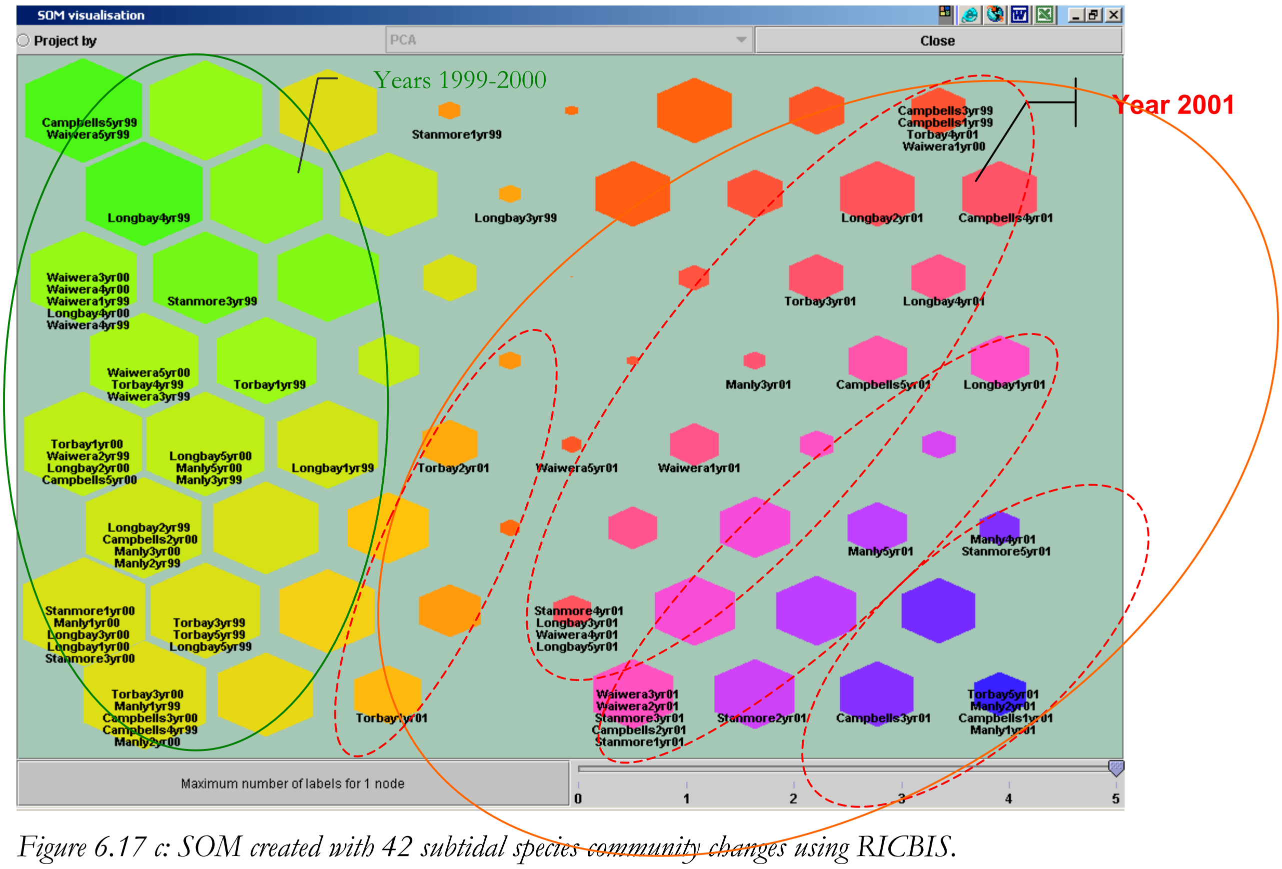


Figure 6.17 c: SOM created with 42 subtidal species community changes using RICBIS.

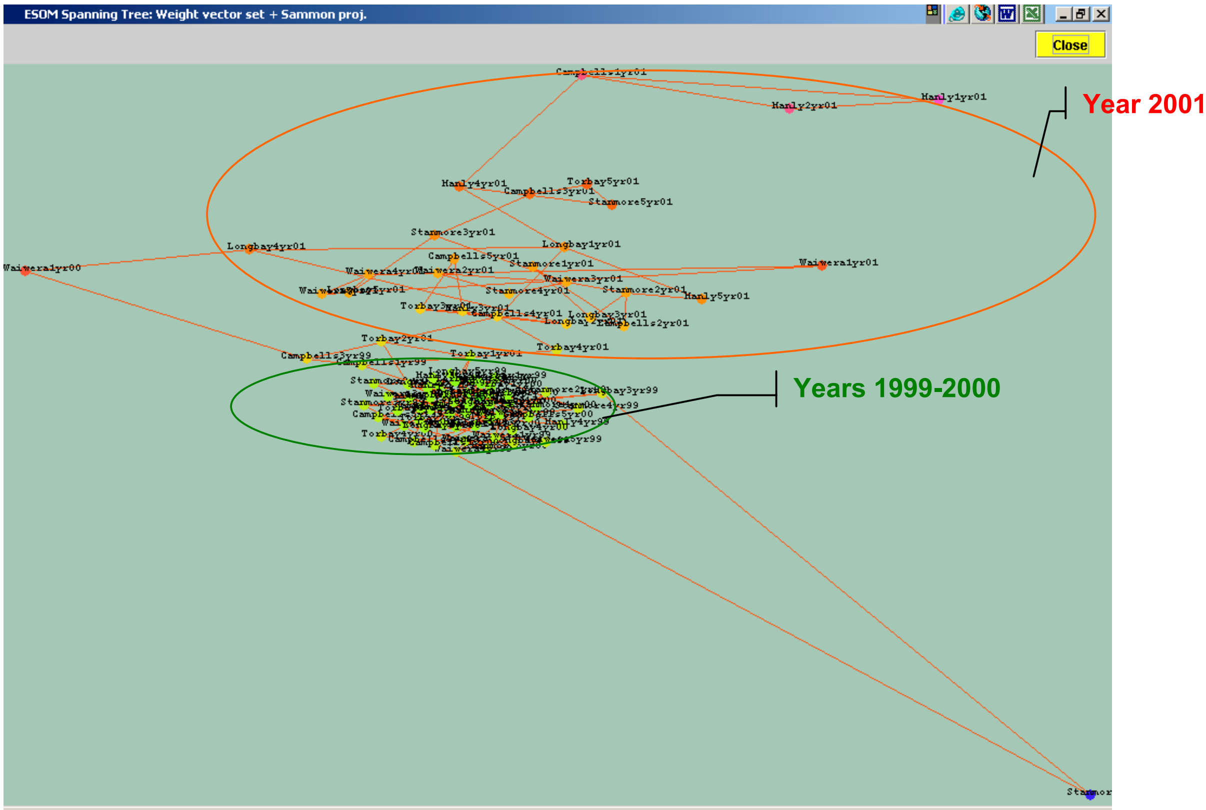


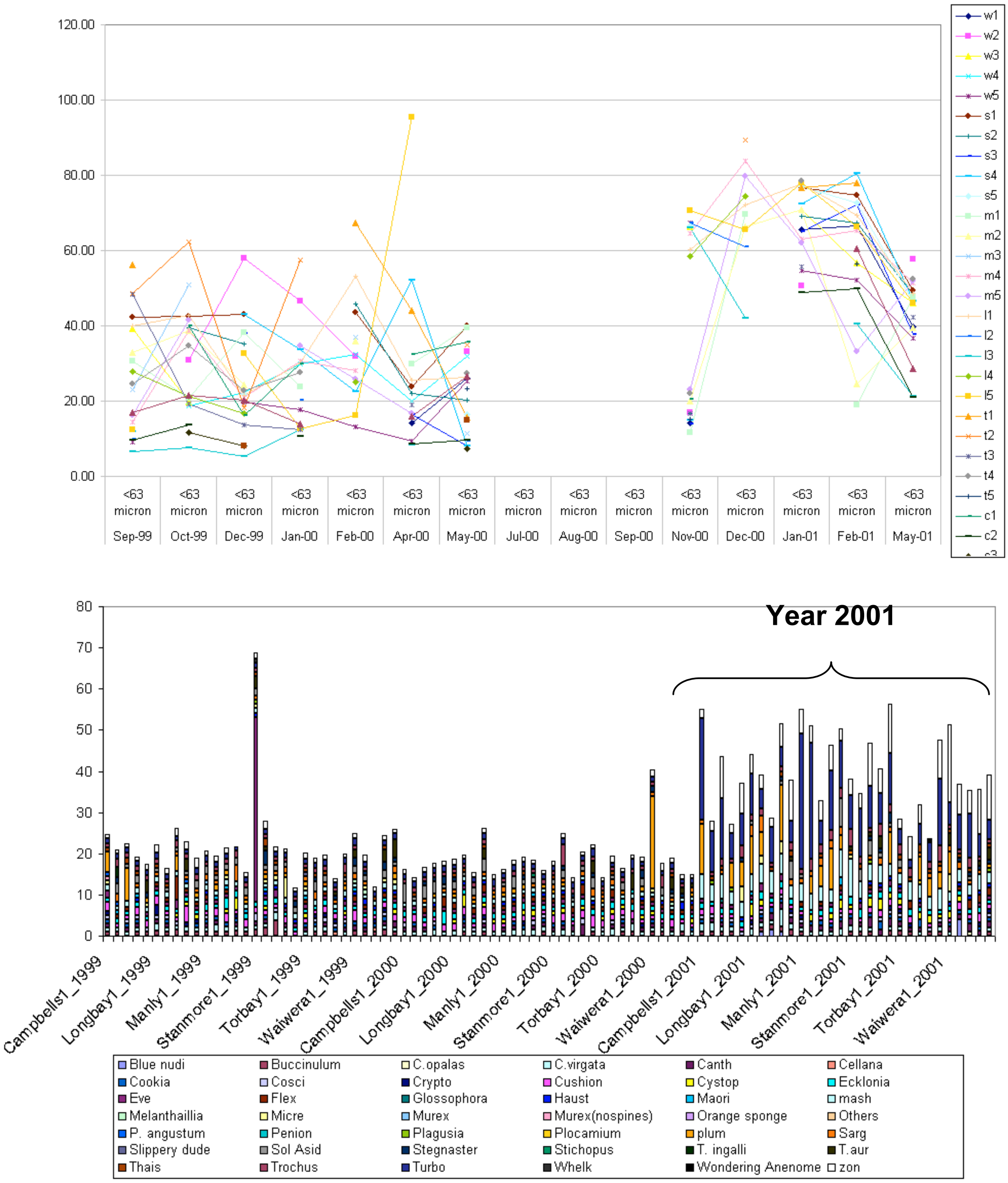
Figure 6.17 d: ESOM created with 42 subtidal species community changes using RICBIS.

On the maps, created with the species data alone without the sedimentation values (figures 6.17 c & d) years 1999 and 2000 fall into one cluster and year 2001 data fall into a different cluster. It suggests that the subtidal population dynamics for year 2001 as different from that of the previous two years.

6.5.2 Conventional data analysis

As part of the research, in addition to the SOM analyses so far discussed, the data on subtidal population dynamics and sediment deposition rates were analysed separately using conventional data analysis methods. Simple bar graphs plotted on the subtidal species average counts and sediment deposition data against time (figures 6.18 a & b) are worthy of mention. The graphs do not distinguish the annual variations within this data, however, illustrate a striking variation on species communities of all beaches for year 2001, coinciding with increased sedimentation rates. The increased percentages of <63 microns of the total sedimentation from November 2000 onwards have a good correlation to the unusual subtidal population dynamics of the beaches monitored in the ARC's programme.

Walker et al. (2000) concluded that the rigorous data analysis methods used in their design to be capable of detecting the effects increased sedimentation from that of annual variations with continued monitoring of intertidal population dynamics. Also made suggestions for using *Evechinus chloroticus* as a possible biological indicator in the future.



Figures 6.18 a & b: Graphs plotted for sedimentation and subtidal community changes in the beaches monitored off northeastern coast of Auckland.

6.6 Conclusion

The results of this case study, to apply complex industrial system modelling with SOM techniques to ecosystem modelling show potential. SOM analyses provide a means to relate and study the environmental changes with the observed biological responses such as population dynamics, in visual formats even with regional scale monitoring data. The patterns in the data can be analysed directly linking the causal process and the environmental effects on biological systems dynamics at ecosystem level even combined with economic data, within an integrated framework.

All the variables in ARC's LTB programme data on saline water quality of east coast and the Waitemata Harbour were analysed separately with 2 D graphs in (Wilcock and Stroud 2000), in the sense a graph for each and every attribute overall 21 graphs for every single beach. The SOMs created for this research with the same sets of data are able to show the patterns in the saline water quality data (21 x11 graphs); the regional scale data could be analysed in one SOM. They revealed the annual and seasonal trends among the monitored beaches and more importantly at a particular site as well as across the region, such as similarities and dissimilarities within the monitoring data were apparent in the SOMs. The example showed that SOMs are capable of modelling the spatial and temporal variations within the monitoring data as the SOM clustering classified the different beach data into different clusters, for instance Goat island into top left corner and Kaipara into right bottom corner. Goat Island data, clustered into one corner of the SOM, shows that this beach as different to other beaches analysed, considered as a reference site in (Wilcock and Stroud 2000).

The SOM created with developmental activities (such as building consent data) combined with the saline water quality data of Browns Bay showed promising results for integrated analysis of dissimilar data sets. Similarly, data from other aspects, economic and other developmental activities could be analysed within an integrated framework to analyse trade-offs at different scales and levels, even with dissimilar data sets.

Studies carried out prior to the ARC's subtidal population dynamics monitoring programme, had also stated that the sediment deposition generated by earthworks was increasing at Long Bay, since last two years (Walker et al. 2000). It was also noted that there was an observable variation in the biodiversity of the coastal environment, caused either by natural causes or induced by increased sedimentation. None of the earlier research effects was able to exactly detect the subtidal population dynamics or its cause either. The SOM maps created with the same set of data were able to differentiate the annual variations in biodiversity from those owing to increased sedimentation rates, through visual formats.

ARC's biomonitoring programme carried out by UoA is expected to continue at least for another few years as it is based on BACI design (Walker et al. 2000). Even with extensive information on the habitat species, before and after an impact occurs, it is extremely difficult to establish the correlation between the biological system changes and the environmental impact using the BACI and other more complex design methods used in the series, argued (Thrush et al. 1995; Osenberg 1996) (see chapter 3). The reasons for this are, apart from the variability seen in species threshold response and other impact reciprocating mechanisms, the extent of an environmental impact could vary significantly within the analysed ecosystem due to spatial and temporal variations.

SOMs are capable of modelling the spatial as well as temporal variations using the available subtidal community monitoring data. SOMs created with RICBIS also depicted the clustering patterns within the beach sites from the biomonitoirng data. The SOMs and ESOMs distinguished the sediment induced population dynamics from that of annual variations. Hence, the case study results prove SOMs techniques as a useful tool for ecosystem modelling with a quantitative approach using the available data, and abilities to overcome the inadequacies with conventional data analysis methods.

6.7 Future work

NIWA has produced models showing the possible sediment generation under different future development scenarios on North Shore (Green et al. 2000). Further research is intended to use SOMs and the same hypothesis to investigate possible techniques to predict the population dynamics for NIWA's predicted sedimentation data. Prediction is possible in Viscovery by associating the related variable/s in the data set with the SOM. In the case of subtidal community changes, associating it with the current sedimentation rates and then creating new SOMs for extending the correlation between the two sets of components is a possible way of predicting the environmental effects on the biodiversity.

Summary

The second case study of this research to examine how best SOM methods could be applied to modelling ecosystem processes at regional scales and to an extent along with economic data, within an integrated framework, produced promising results and showed potential for future use. The various SOM methods utilised in the case study elaborated the application of SOM feature extraction in ecological data analysis.

The saline water quality example showed how SOM clusters could be effectively used to carry out a comparative analysis on the available data. Component plane analyses showed how SOM components, visual representations of the multidimensional, disparate data sets could be applied to studying the correlations within the different variables of fused data sets. Also the example illustrated SOMs capabilities in detecting the spatial variations within the monitoring data, seen to be impossible by traditional methods.

The subtidal community biomonitoring example illustrated, SOM abilities to explore correlations among variables, in particular between environmental parameters and biological indicators, which are generally found to be extremely difficult if not impossible, even with the use of rigorous traditional statistical methods.

In the next chapter SOMs are applied to ecosystem modelling with global data to look for patterns at a further wider scale than in this chapter. The shortcomings with the contemporary ecosystem modelling methods, such as pressure-state-response and information pyramid models, and the pilot projects already implemented for this purpose by the Dutch government (also adopted by various international institutions and Ministry for Environment in New Zealand) will be discussed in detail.

Chapter 7

SOM techniques to analyse global data

The last two chapters provided details on case studies that examined the use of SOM methods, utilising large volumes of data sets with inconsistent labelling from a single site and at regional scales. Potential applications of Kohonen's SOM techniques to modelling biological and environmental systems within an integrated framework are being investigated in this research. Chapter 6 illustrated the efficient use of SOM methods with ecological monitoring (biological and environmental) and development related (building consent) data of an area, in analysing the correlations in them. In this chapter, data sets at higher levels and wider scales are analysed with SOMs to study the patterns in them. The main objective is to detect global trends in the effects of urbanisation on biodiversity based on data compiled by international institutions. Many environmentally concerned institutions (national and international, such as Organisation for Export Corporation and Development (OECD), the World Bank, World Resources Institute (WRI) and Ministry for the Environment in New Zealand), have embarked on activities to developing methods and formulating legislation to save ecosystems from massive environmental deterioration. The next section gives a background on the case study being analysed herein.

7.1 Background

The World Bank's matrix of environmental indicators was developed based on an explicit conceptual model within a pressure-state-response (PSR) model framework, to measure the human interaction with the environment (see chapter 3.). The matrix indicators (appendix 3) are categorised into four major issues. They are:

(i) source indicators
(ii) sink or pollution indicators
(iii) life support indicators and
(iv) human impact indicators.

As the World Bank's matrix of indicators is structured based on the PSR framework, data on these indicators could be directly used for quantitative analysis of ecosystems using SOMs. Data tables and reports released by the World Bank and WRI for the above stated issues are available for public access (Hammond et al. 1995c; World Bank 2002d; World Bank 2002a; World Bank 2002c; World Bank 2002b). SOMs were created with different combinations of the World Bank's data tables to analyse the global trends in developmental activities and their impact on biological diversity.

7.2 Objective

The following are the objectives of this case study using the World Bank's statistical data:

(i) to further investigate the SOM based approaches to analyse the effects of human interaction on our global ecosystem in which many living beings coexist in a dynamic equilibrium. Methods that can detect an equilibrium shift towards any undesirable direction by tracking the systems dynamics, within the available statistical data of different countries, could be of use to study the global ecosystem trends.

(ii) to explore SOM approaches to analyse disparate global data on biological and environmental changes along with economic and social aspects within an integrated framework (or with a systems approach). These approaches could be applied to economic trade-off analysis by collectively analysing different combination of data sets, to explore the relationships among the chosen variables without any physical models.

7.2 Contemporary indicators for ecosystem analysis

The PSR indicator framework, first introduced by the Dutch government to analyse diverse ecosystems has been adopted by many international institutions, such as OECD, UNDP, WRI, and the World Bank despite the hassles involved in the aggregation process. In this section, PSR model framework and the information pyramid based aggregation processes are explained using the sink or pollution indicators.

7.2.1 Sink or pollution indicators

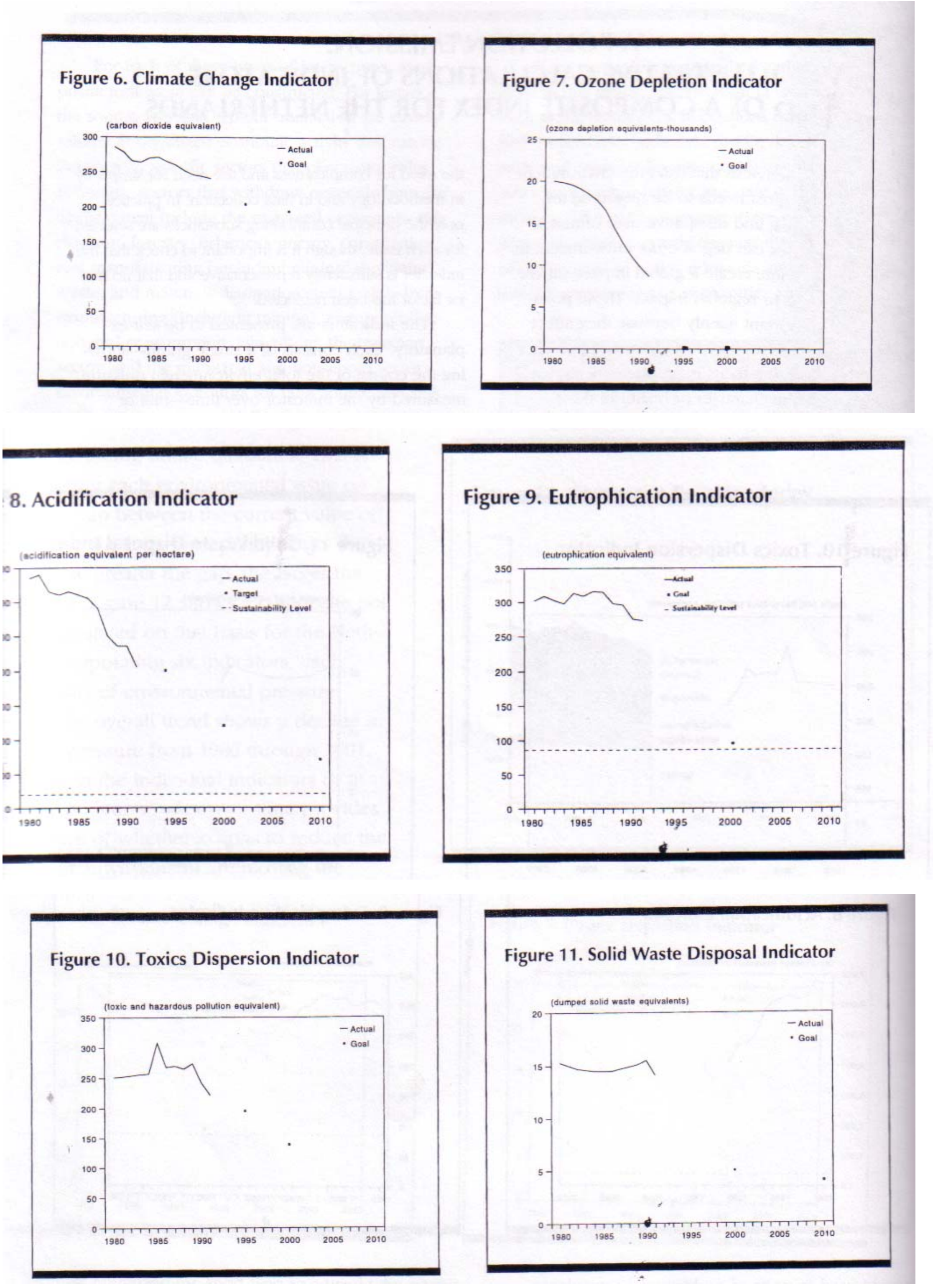
In the PSR approach, different aggregation methods are used to produce composite indices to measure the state, pressure and response of an environment (see chapter 7). The aggregation processes, based on the information pyramid model, are used to reduce a number of primary indicators/ data pertaining to an issue, into a single composite index per issue. Similarly, all primary indicators are converted into composite indices i.e. GDP.

In the Netherlands, environmental indicators, developed based on the PSR indicator framework, have been tested with considerable success despite the complex calculations involved in the aggregation process. These indicators have been applied to gauging the state, pressure, and response of the environment and natural resources with an aim to support ecosystem functioning and biodiversity for human wellbeing. The following are the indicators listed by the Dutch government as the six main human activities that primarily alter the character of the earth's physical and biological systems:

(i) Climate change indicator
(ii) Ozone depletion indicator
(iii) Acidification indicator
(iv) Eutrophication indicator
(v) Toxics dispersion indicator
(vi) Solid waste disposal indicator

Of this six listed issues, ozone depletion indicator is further described to see as how the first level indicators are aggregated to produce the final single composite index for pollution. Initially, the ozone depletion indicators are aggregated, depending on the extent of damage caused to the ozone layer, by giving a weight to the next level of aggregated indicator. For instance, the damage caused by Halon 1301 to the ozone layer is ten times more than that of the substances categorised as CFC-11. The weighted, summation of ozone depletion indicators is estimated as ozone depletion equivalents. Accordingly, the estimation for 1980 was 20,000 units and by 1991 it dropped to 8,721 units. This shows an indication of 56 % improvement on the environment's state by ozone depletion substance emission over this period (figure 7.1). The Dutch government policy target was to reach nearly complete termination of producing ozone depletion substance by the year 2000. Similarly, all other five environmental issues from such material/ substance emissions, relating to the state indicators were calculated and analysed using 2D graphs (figures 7.1-7.3).

These six already aggregated indices were then further aggregated into a single composite pollution index to obtain the overall state in the use of environment as a sink. The weighting for aggregation of these six environmental issues, in relation to their contribution was calculated based upon the gaps between the current values of indicators and their respective long-term policy target for sustainability. The longer the gap the more the value assigned to that indicator in the aggregation. Using the annual composite index values, from 1980 to 1991, the trend in the state of the environment as a sink or pollution state over this period was analysed (figure 7.4).



Figures 7.1, 2 & 3: Climate change and ozone depletion indicators, 7.2: Acidification and eutrophication indicators and 7.3: Toxic dispersion and solid waster disposal indicators (Hammond et al. 1995e).

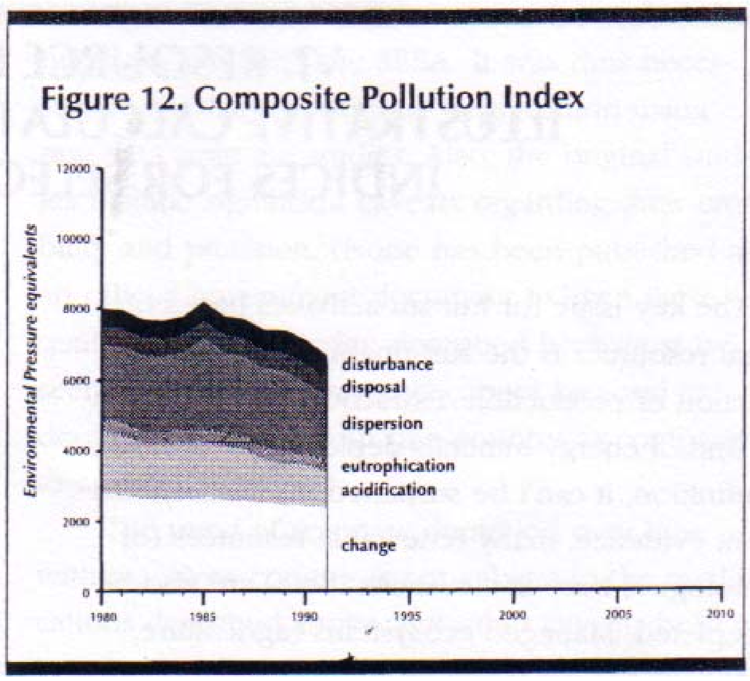


Figure 7.4: Composite pollution index (Hammond et al. 1995e).

Even though the specific problems with regard to pollution from emissions would vary from country to country, environmental indicators could be devised to suit their own country's requirements using this approach (Hammond et al. 1995e). It is noted here that the Ministry for the Environment (MfE) in New Zealand too, has introduced legislation to develop a set of indicators based on the PSR model framework for future use (Chapman 1999; Ministry for the Environment 2002a; Ministry for the Environment 2002c; Ministry for the Environment 2002b). However, there is a major constraint in this approach, which is, unless the physical process of an ecosystem issue is known there is no way of aggregating the primary data/ indicators to form the composite index.

7.2.2 Biodiversity indicators

The major issue encountered in the initial development of indicators for biodiversity is examined in this section. In the World Bank's list of matrix, there are a few undefined issues (Hammond et al. 1995d). Of those undefined issues, a particular item that is also of great interest to this research is the biodiversity. The following are the details of the matrix elements for biodiversity listed by WRI and the World Bank, based on the PSR framework:

Issues	Pressure	State	Response
Biodiversity	Land conservation, land fragmentation	Species abundance comp. to virgin area	Protected areas

Figure 7.5a: Biodiversity in WRI matrix (Hammond et al. 1995b).

Issues - iii) Life support indicators	Pressure	State	Response
1. Biodiversity	Land use changes	Habitat/ NR	Protected areas as % threatened
2. Oceans	Threatened, extinct species % total	…………………..	…………………..
3. Special lands (i.e. wetland)	…………………..	…………………..	…………………..

Figure 7.5b: Biodiversity in the World Bank's matrix (Hammond et al. 1995b)

Indicators of biodiversity can be considered as a proxy for measures of fundamental life support functions (Hammond et al. 1995a) despite the constraints encountered in developing them. Biodiversity is considered to play a significant role in ecosystem functioning, by providing support to all life on the Earth, such as oxygen production, water purification and many more. However, the spatial variations within an ecosystem pose a major constraint in devising a set of indicators for biodiversity at wider scales (see chapter 3), even the PSR framework approach has not been useful in this effect. Some suggestions made by Hammond et al. (1995a) to overcome this problem are discussed herein as they are applied to SOM analyses in this case study: The suggestions are:

i) Policies to preserve biodiversity should be directed at ecosystem or habitat level for which measure of biodiversity change could be made at species level either by counting species or listing endangered species. With the recent advances in DNA studies, genetics and ecological modelling, it possible to measure the diversity in life at various levels, such as gene, species and ecosystem. Despite the fact that all these are capable of reflecting the important elements of the Earth's biological heritage, many interactive processes, critical to all life take place at ecosystem level. Thus, measures to study the preservation of biodiversity should be carryout at this level with a systems/ holistic approach.

ii) To associate the human impact on ecosystems (that could be measured in terms of changes in human activities, such as cutting of forests), with monitoring of ecosystem changes (that could be measured in terms of biological population dynamics). As ecosystems can be roughly studied corresponding to their geographical units, the administrative responsibility for conservation and land management activities as well should be assigned to these units.

iii) To use the following widely reported factors that are considered as useful in devising biodiversity indicators within the PSR model framework,

a. list of endangered species and
b. statistics on wilderness areas, both as state indicators.

Statistics on percentage of land legislated for conservation/ protected could be considered as response indicators. None of the above could be used as pressure caused from human activities on ecosystem state. This is because the changes in biodiversity arising from human activities vary significantly even within an ecosystem due to variations in species response (often described as threshold and non-linear) and population dynamics (described in terms of species richness and evenness). Such variations can only be mapped spatially, such as digital maps, in terms of currently available methods.

iv) To extract national level summaries on the relationships among various human activities and biodiversity from Geographical Information Systems (GIS) to analyse the spatial distribution of a particular ecosystem. There are digital maps featuring the spatial distribution of vegetation types or other markers of broad ecosystem type with the basic physical data on land type and microclimate, including the location and intensity of various human activities. These primary data could be used to create concise indicators of biodiversity, as the more recent GISs allow for integration of other data sets and manipulation of the same for further analysis and extraction of information at higher levels.

v) To use the indirect measures on human induced pressure from GISs, instead of the direct measures that are difficult to obtain. There are many countries with spatial distribution of data in some indirect measures (for instance, human population distribution, and the presence of roads or infrastructure), already superimposed on digital maps. These could be used as a time series analysis to study the effects of human activities in ecosystem biodiversity. Some of the human activities that create direct pressure on ecosystems, such as clearing of forests, filling in wetlands, overharvesting (firewood or overgrazing of domestic animals), introduction of exotic species and pollution or diversion of water, cannot be measured easily. To overcome this problem, indirect measures, such as population increase, could be used to analyse the direct pressure imposed by humans.

vi) To use the latest GISs parameters as indicators of non-linear variables, such as vulnerability or the threshold of a varying pressure, within an ecosystem. The spatial variations within an ecosystem may significantly affect the extent of an impact. Hence, using an GIS, the measures of inherent sensitivity that are dependent on the distribution of geographical and biological parameters such as soil type, climate zones, slope and proximity to waterways could be analysed quantitatively along with the current degree of modification in the area, such as habitat fragmentation or soil erosion and risk pressure. By combining the pressure and vulnerability data on digital maps, the region's ecological thresholds could be established and used as an indicator of relative risk of degrading biodiversity throughout the ecosystem.

Based on the above suggestions, WRI in collaboration with the World Conventional Monitoring Centre, RIVM, Conservation International and the Institute for Sustainable Development began work on preparing maps of preliminary pressure, sensitivity and ecosystem risks for a few African countries using the already developed digital maps. Originally, these maps of population distribution and infrastructure were developed for these African countries as a development planning tool (Hammond et al. 1995a).

The recent suggestions and developments in ecosystem dynamics modelling to incorporating biodiversity were outlined in the section. Many international institutions have constantly expressed their concern over the escalating human activities, deleterious to the environment. They also have embarked on efforts to develop biodiversity indicators based on the PSR model framework despite the above discussed drawbacks.

In the next section, how SOM techniques could be best applied to studying the pressure trends in biodiversity at ecosystem level is described. The main aim is to analyse the effects on biodiversity due to human activities at global scales using the statistical data compiled by the World Bank and WRI.

7.3 Methodology

In this case study, data on the pressure and state of biodiversity are collectively analysed using SOMs, to see the patterns in them. Human activities, such as urbanisation and deforestation are used as the pressure indicators and threaten species are used as the state indicators, as suggested in the PSR framework. As the global statistical data tables compiled by the World Bank were classified based on the PSR model approach, collectively analysing them was found to be relatively easy, compared to the earlier two case studies. Visovery® SOMine lite version 4.1 by eudaptics software gmph package was used as it could create maps even with a few missing values.

7.4 Results and discussion

The SOM results of two examples with global data are explained. In the first example, atmospheric concentrations of greenhouse gases are analysed. In the second, urbanisation and biodiversity data are studied based on the PSR model framework without any aggregation processes.

7.4.1 SOM analysis on greenhouse and ozone-depleting gases

Global warming of greenhouse gases is a major factor that contributes significantly towards environmental degradation. It is also a widely recognised issue by many professionals, such as scientists, policymakers and even the general public. The total global, atmospheric concentrations of greenhouse and ozone-depleting gases, estimated for 1980 to1998 time period, consist of two main categories and are based upon (Carbon Dioxide Information Analysis Center Data 2000-2001). The two categories are:

(i) Chlorofluorocarbons: CFC-11 (CCl_3F), CFC-12 (CCl_2F_2), and CFC-113 ($C_2Cl_3F_3$).

(ii) Total gaseous chlorine: Calculated by multiplying the number of chlorine atoms in a unit of the chlorine-containing gases by their concentrations. Chlorine acts as a catalyst in the destruction of ozone.

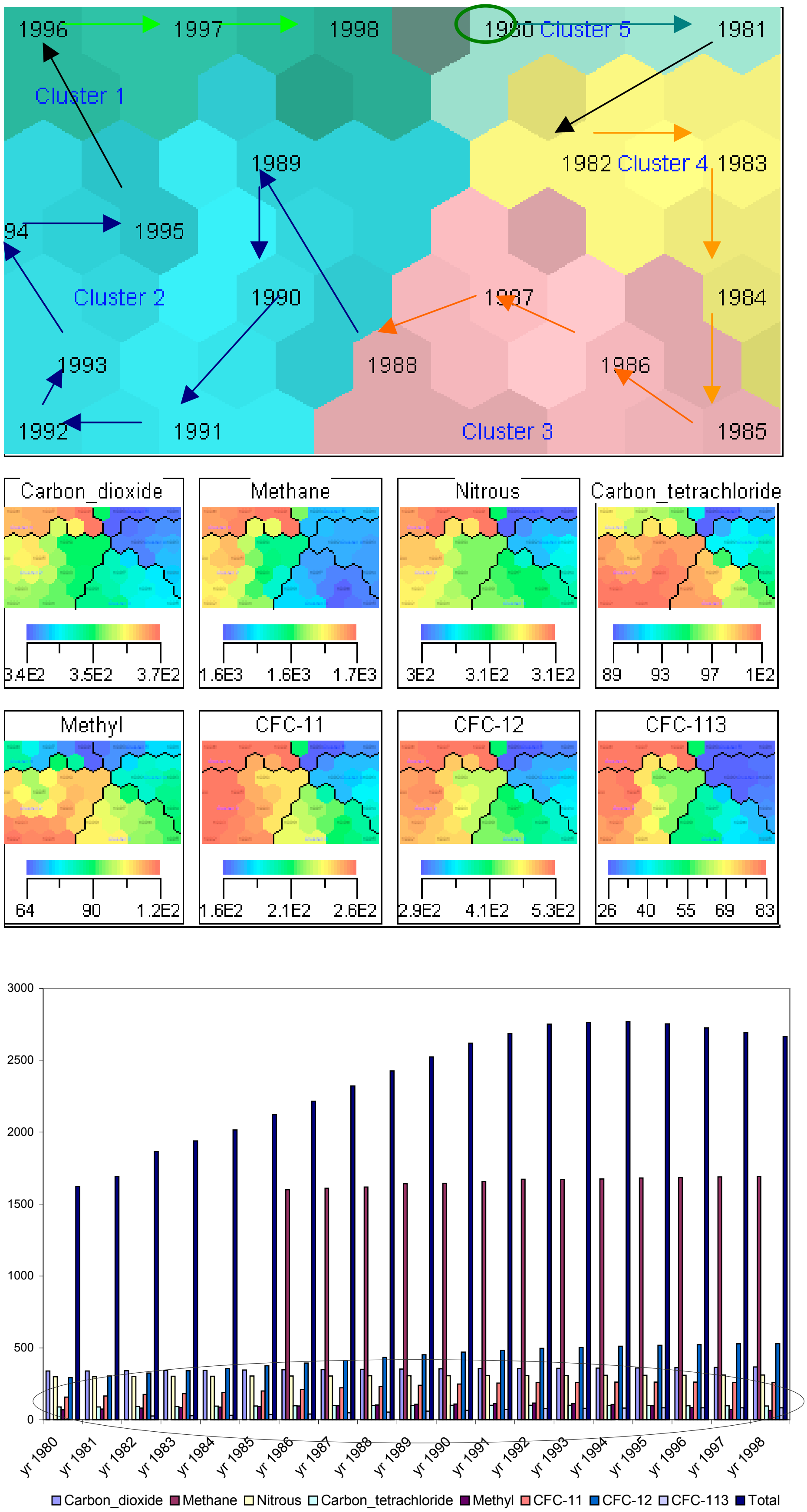


Figure 7.6 a: SOM created with atmospheric concentration data of greenhouse and ozone-depleting gases, for a period of 18 years. b: Component planes of the SOM and c: Histogram of the same data.

Using this data on atmospheric concentrations of greenhouse and ozone-depleting gases, for a period of 18 years, a SOM map (figures 7.6 a & b) was created with 100 nodes and all other map creation parameters set to default values. A histogram of the same data is shown in figure 7.6 c. The SOM cluster maps as well as the component planes (figure 7.6 b) show an effective means to analyse the multidimensional data and this is explained herein.

SOM component planes (figure 7.6 b) depict the vectors in easily understandable formats. For instance, carbon dioxide plane (figures 7.6 b and d) illustrates the atmospheric concentration of carbon dioxide over 1980-1998. The scale beneath the plane shows the range (3.4 E2 – 3.7 E2, which is 340- 370 concentration units) depicted in that space. Similarly, other planes and their corresponding scale depict the ranges covered within the same space for other gases.

In the cluster map, vectors of each and every data point may be visualised. The mean values for each cluster (such as in figure 7.6 e) are used in the interpretation of the map. On the contrary, comparative analysis of gases, using the histogram (figure 6.7 c), looks cumbersome because of the number of gases and the period analysed, especially the very small differences between the shorter bars cannot be visualised with accuracy. With SOMs, where similar data points are grouped together, patterns and trends in the gas emission data, spanning a period of 18 years can be collectively analysed; using the cluster statistics even minor details could be compared.

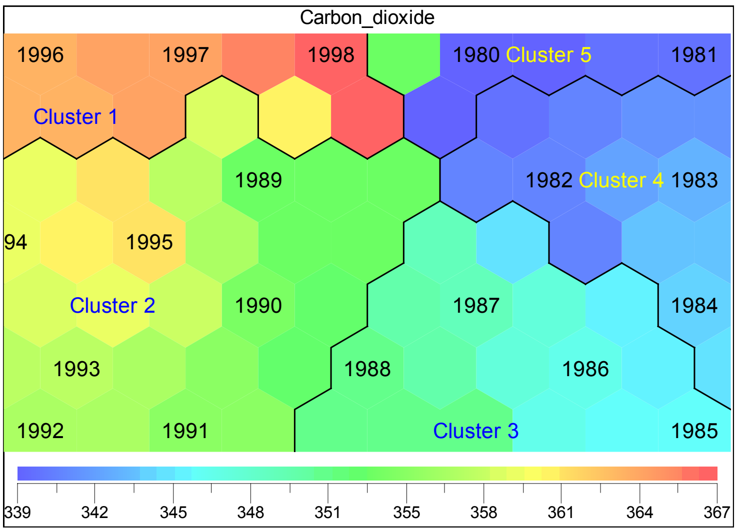
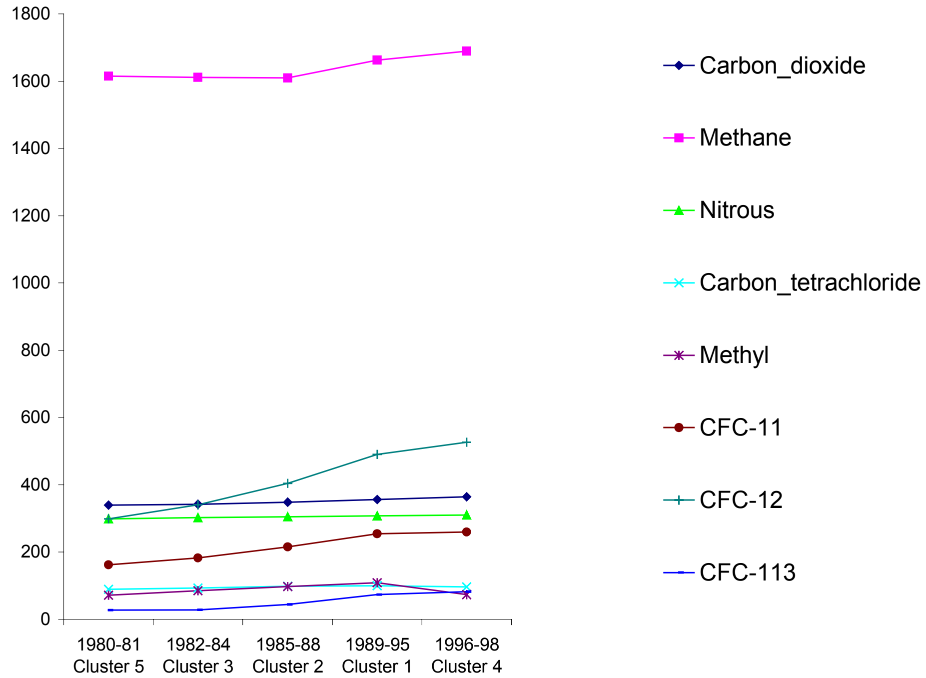


Figure 7.6 d: SOM Component plane of atmospheric carbon dioxide concentration from 1980-1998. The scale shows the atmospheric carbon dioxide range 339-367 in concentration units.



	Cluster 5	Cluster 3	Cluster 1	Cluster 2	Cluster 4
Period	1980-81	1982-84	1985-88	1989-95	1996-98
Carbon_dioxide	339.5	342.7	348.3	356.4	364.7
Methane	1615	1611	1609	1663	1689
Nitrous	299	302	305	307.6	310.7
Carbon_tetrachloride	89.5	93	98	100.4	97
Methyl	72	84.7	97.3	108.7	74
CFC-11	162	182.3	215.5	254.9	260
CFC-12	299	340.3	404	490.7	527
CFC-113	27.19	28.33	44.25	73.71	82.33

Figure 7.6 e: Graph of SOM (figure 7.6 a) cluster details and d: SOM cluster values.

The following are the interpretations arrived at from this map (figure 7.6 a and b):

(i) Cluster 5, consisting of 1980 and 1981 shows the lowest volumes for all gases except for methane and nitrous oxide.

(ii) Cluster 3, consisting of 1982 - 84, with low volume of all gases, except for CFC 11 and 12, released in reasonably high volumes.

(iii) Cluster 1, covering a period of four years from 1985 to 1988, exhibits values same as cluster 3, except for CFC 12 and 113 both of which show considerable increase than the earlier time period.

(iv) Cluster 2, consisting of 1986, 1987 and 1988 exhibits the release of very high volumes of methane, CFC- 11, 12 and CFC-113.

(v) Cluster 4, consisting of 1991 - 1995 exhibits the release of very high volumes of CFC gases, especially 113.

The atmospheric concentrations of these gases were very low during the 1980-84 period. From 1985, the concentrations of methane and CFC have shot up and steadily increased till 1998. However, there had been a reduction on methyl chloroform and carbon tetrachloride concentrations during 1996 to1998 period.

The example illustrated how SOMs could be used to analyse without having to calculate their contribution towards the issue as carried out in the PSR and information pyramid models. In fact data relating to pressure, response and economic outcome as well could be studied using SOMs and will be elaborated in the next section.

7.4.2 SOMs in global data on rural environment and land use

A SOM map (figure 7.6 a and b) was created using the following data of land use changes, to study the global trends and patterns in different countries on their rural environment and land use in years 1980 and 2000 along with their annual growth (World Bank Report 2001a):

(i) Rural population: Total percentages in 1980 and 2000 along with percentage of average annual growth between 1980-2000.

(ii) Rural population density: People per sq.km of arable land in 1999.

(iii) Land area: Thousand sq.km in 1999.

(iv) Land use: Percentages of arable land in total land area in 1980 and 1999, percentages of permanent cropland out of total land area in1980 and 1999 and percentages of other land area out of total land area in 1980 and 1999.

The total population figures are World Bank estimate. Data on land area and land use are from Food and Agriculture Organisation's (FAO's) electronic files, published in its Production Yearbook. FAO gathers these data from national agencies through annual questionnaires and national agricultural censuses (World Bank Report 2001a). A SOM was created with the pressure factors affecting the state of biodiversity from the World Bank data with 100 nodes and all other map parameters set to default values to look for any major deviations in these two years, 1980 and 2000. The following are the interpretations derived from the six cluster SOM map (figures 7.7a and b):

i) Cluster 6 consists of Australia, Brazil, Canada, Russian Federation, United States and China show the highest land area (mean) along with the lowest percentages of cropland and people/ 1000 sqk land 1999. These countries show the lowest percentages for the latter two variables because of their high land area. The only difference between China and the rest of the countries in this cluster is that China has the percentage of rural population 58.05 and 49.19 for 1980 and 2000 respectively, whereas for all the other countries it is low.

ii) Cluster 1 consists of Algeria, Argentina, Armenia, Austria, Azerbaijan, Bolivia, Bosnia & Herzegovir, Chile, Colombia, Congo Republic, Ecuador, Finland, Gabon, Georgia, Hong Kong-China, Iran Islamic Republic, Iraq, Ireland, Israel, Japan, Jordan, Kazakhstan, Korea Democratic Republic, Liberia, Libya, Mauritania, Mexico, Mongolia, Morocco, Mozambique, New Zealand, Nicaragua, Norway, Oman, Panama, Peru, Saudi Arabia, Sierra Leone, Singapore, Slovenia, Sweden, Switzerland, United Arab Emirates, Uruguay and Venezuela RB. The variations between clusters 1 and 6 are:

   a) cluster 1 countries have low total land area nonetheless have high percentage of other land area, high as cluster 6.

   b) cluster 1 is densely populated in that its people/ 1000 sqk land is as mush as three times more than that of cluster 6.

iii) Cluster 2 consists of Afghanistan, Angola, Burkina Faso, Cambodia, Chad, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Honduras, Kenya, Lao PDR, Lesotho, Madagascar, Malawi, Mali, Myanmar, Namibia, Nepal, Niger, Papua New Guinea, Senegal, Somalia, Sudan, Swaziland, Tajikistan, Tanzania, Turkmenistan, Uzbekistan, Vietnam, Yemen Republic, and Zimbabwe. The countries in this cluster show the highest rural population, 78.32 and 70.35 in 1980 and 2000 respectively and at the same time, high annual growth and because of this reason differ from cluster 1.

iv) Cluster 4 consists of Burundi, Costa Rica, CôteD'lvoire, Cuba, Dominican Republic, El Slvadoe, Ghana, Greece, Guatemala, Haiti, Indonesia, Italy, Jamaica, Lebanon, Malaysia, Philippines, Portugal, Puerto Rico, Rwanda, Spain, Sri Lanka, Trinidad And Tobago, Tunisia and Uganda. These countries show high to medium percentages of crop land (8.96 and 10.36) as well as rural population (59.14 and 49.26) for 1980 and 2000.

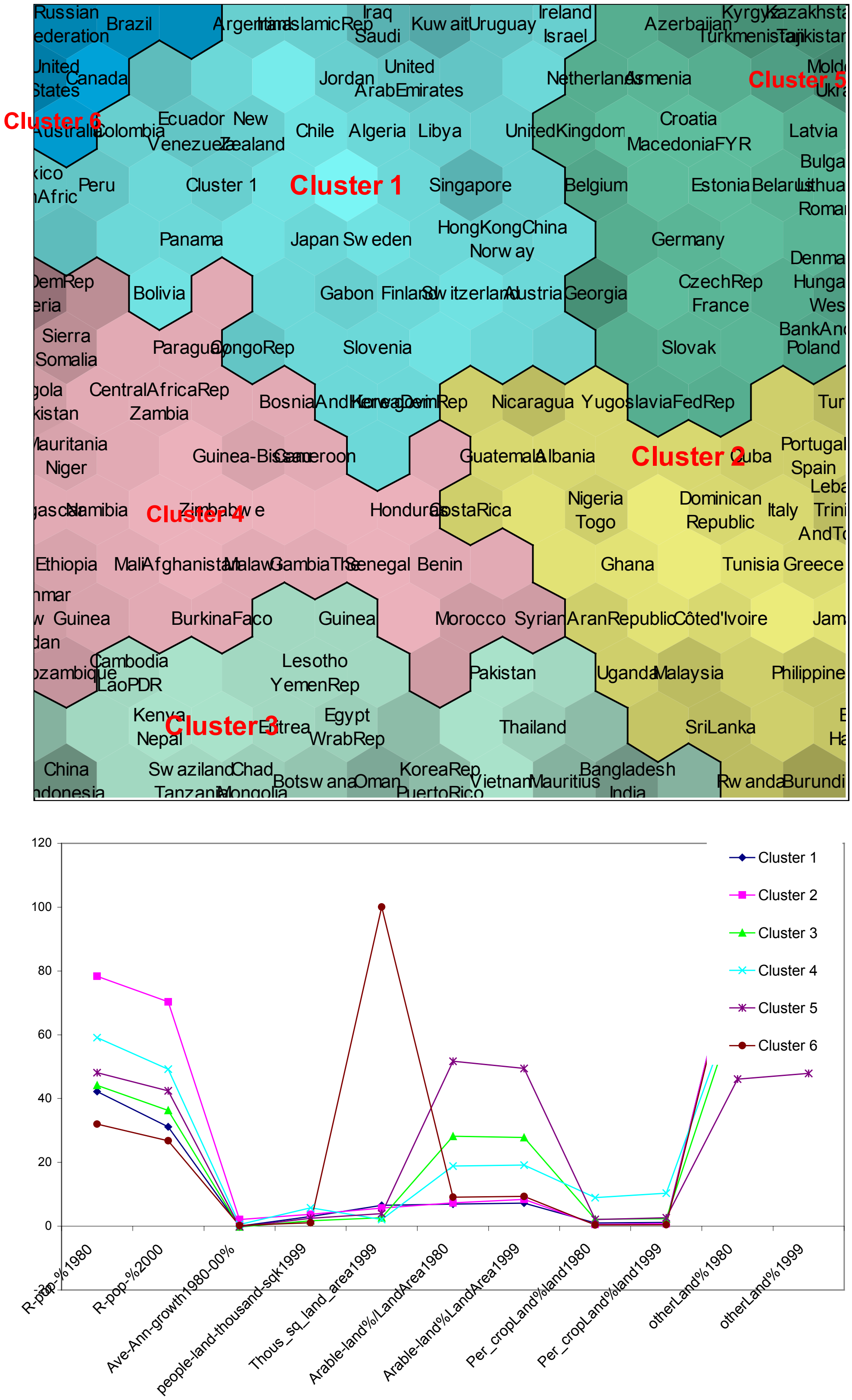
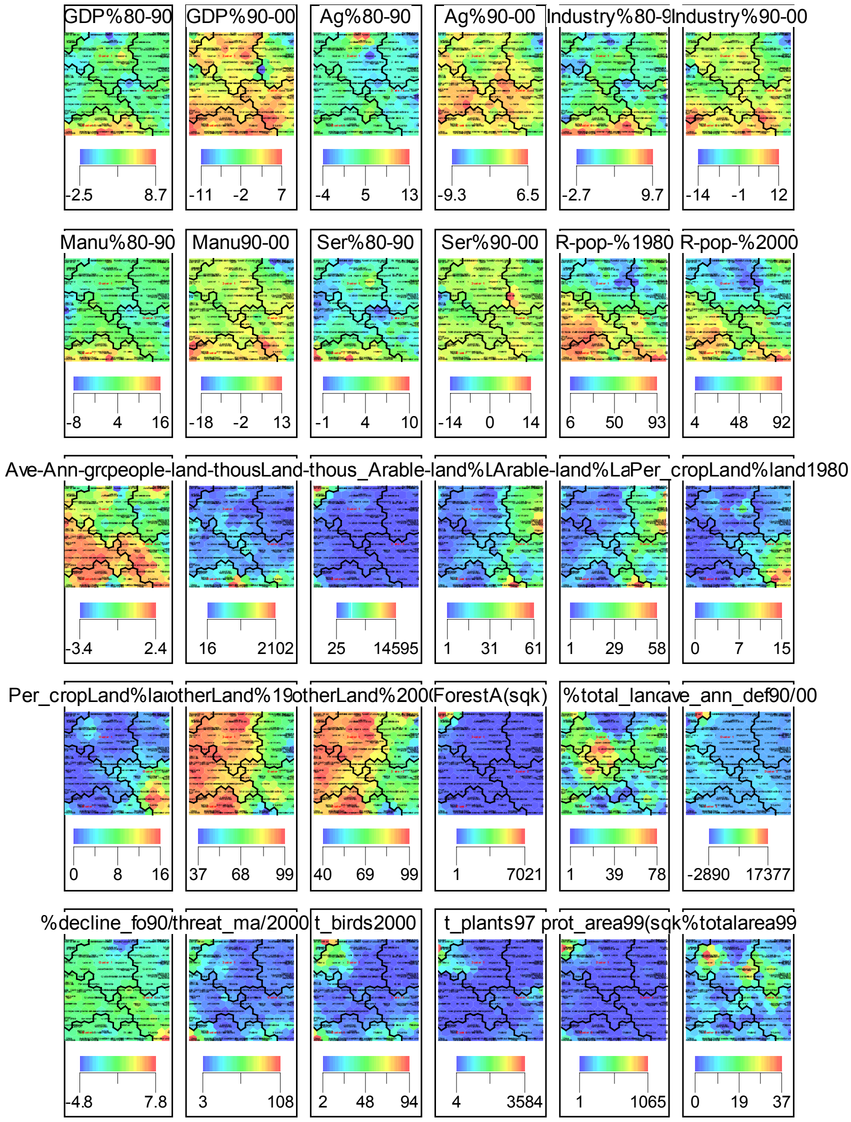


Figure 7.7 a: SOM created using land use change data that are suggested as the state indicators of biodiversity (Hammond et al. 1995a).

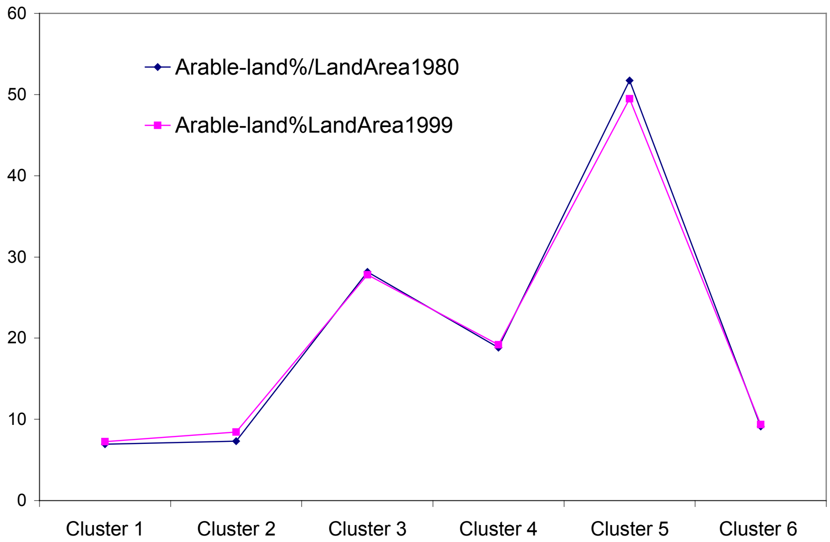


	C 1	C 2	C 3	C 4	C 5	C 6
R-pop-%1980	42.24	78.32	44.13	59.14	48.15	31.99
R-pop-%2000	31.12	70.35	36.27	49.26	42.47	26.81
Ave Annual growth 1980-00%	0.015	2.081	-0.206	0.558	-0.124	0.25
People/land-1000 sqk '99	306	375	174	579	249	109
Land area 1999 1000 sqk	658	572	262	208	396	10005
Arable-land%/ Land Area 1980	6.95	7.31	28.16	18.82	51.73	9.14
Arable land% Land Area 1999	7.25	8.44	27.79	19.2	49.48	9.37
Per_cropLand%land1980	1.06	0.53	2.2	8.96	2.09	0.37
Per_cropLand%land1999	1.19	0.77	2.31	10.36	2.66	0.45
otherLand%1980	91.9	92.1	69.6	72.2	46.1	90.5
otherLand%1999	91.5	90.8	69.9	70.4	47.9	90.1

Figure 7.7 b: Component planes of SOM created with land use change data that are suggested as the state indicators of biodiversity (Hammond et al. 1995a).

a. Cluster 3 consists of Albania, Belarus, Belgium, Croatia, Estonia, France, Germany, Korea Republic, Latvia, Macedonia FYR, Netherlands, Nigeria, Pakistan, Slovak Republic, Syrian Arab Republic, Thailand, Togo, Turkey, United Kingdom and Yugoslavia Federal Republic. These countries show low to medium of all variables. They also show the lowest average annual growth of rural population, which is in negative (-0.206)

b. Cluster 5 consists of Bangladesh, Bulgaria, Czech Republic, Denmark, Hungary, India, Lithuania, Mauritius, Moldova, Poland, Romania and Ukraine with attributes same as cluster 3, except for high percentages of arable area and low percentages of other land in 1980 and 2000.



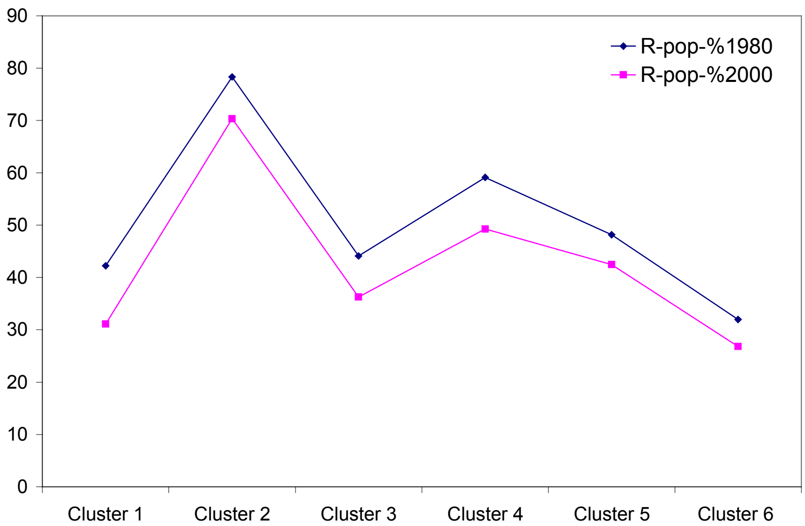


Figure 7.7 c: Graphs showing the cluster details of rural population and percentage of arable land/ total land.

The percentages of rural population are less in 2000 than in 1980 for all six clusters. However, clusters 3 and 5 show, negative average annual growth in rural population during the period between 1980-2000. As far as the land use changes is concerned, in all clusters, mean percentage of cropland/total land is high in 1999 than in 1980, however, total arable land show a different scenario, cluster 3 with less mean percentage for 1999.

7.4.3 SOM analysis on global biodiversity data

SOM analysis was carried out on the following table of indicators compiled by the World Bank to study the global trends in biodiversity for year 2000/ 2001:

(i) Forest area: Forest area in thousand square km for the year 2000, obtained from the FAO's report on the state of the World's Forests 2001[20].

(ii) Average annual deforestation: Average annual deforestation from 1990 to 2000 in square km and percentage of decline in forest area during the same time period, obtained from (World Bank Report 2001b). The report considers deforestation to be a major cause for biodiversity loss.

(iii) Biodiversity: Data on the status of threatened species, gathered from global scale surveys, carried out on certain selected groups of organisms. Such knowledge on an area's threatened species is considered to be an indicator of its biodiversity loss as well as a meaningful alternative indicator of the area's species richness. In ecological studies measures of species richness is considered to be the most straightforward approach for describing the biodiversity of an area. Sampling of plots is usually carried out to produce the estimation of small plants and animals, as these analyses are time-consuming, involving manipulation of large amounts of data. The following is a summary of the groups analysed in the global surveys for threatened species:

   a. Mammals: Based upon the estimate, 45 percent of mammal species remain to be assessed.
   b. Birds: The only group of which the status of all species has been assessed.

[20] This information was based on a survey, stated to be the latest global forest assessment and the first ever to use a uniform global definition for forests. In addition, percentage of forest area cover of the total land area in year 2000 too, was used to indicate the remaining forest area of a nation. The forest cover data included the natural forest and plantation as such the figures for deforestation might give an underestimate of the disappearance of natural forests in some countries.

c. Higher plants[21]: The first ever comprehensive listing of threatened species of plants on a world scale was produced by the World Conservation Union's (IUCN) 1997 red list of threatened plants. Based on the list, nearly 34,000 plant species, 12.5 percent of the total plants are threatened with extinction. This report is a result of more than 20 years' work by botanists from all over the world.

d. Nationally protected areas: The World Conservation Monitoring Centre (WCMC) has compiled the protected area and threatened species (within certain species groups) details in different countries. However, cross-country compatibility of the WCMC data is subject to anomalies, due to the differences in taxonomic concepts and coverage adopted by different countries.

(World Bank Report 2001b)

A SOM map (figures 7.8 a, b and c) was created using the above discussed data set with 100 nodes and all other map parameters set to default values.

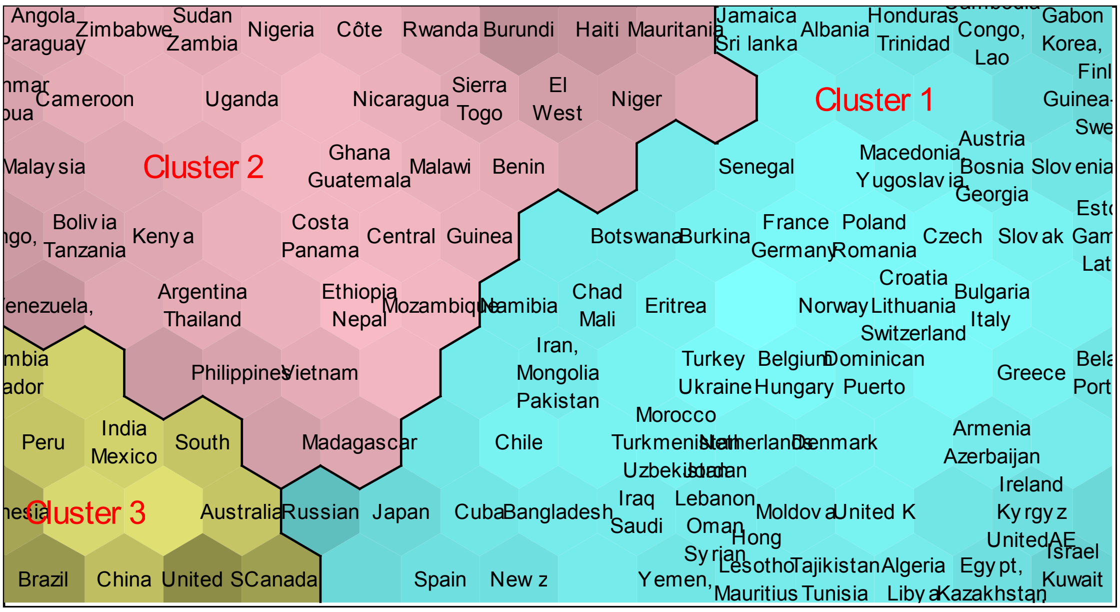
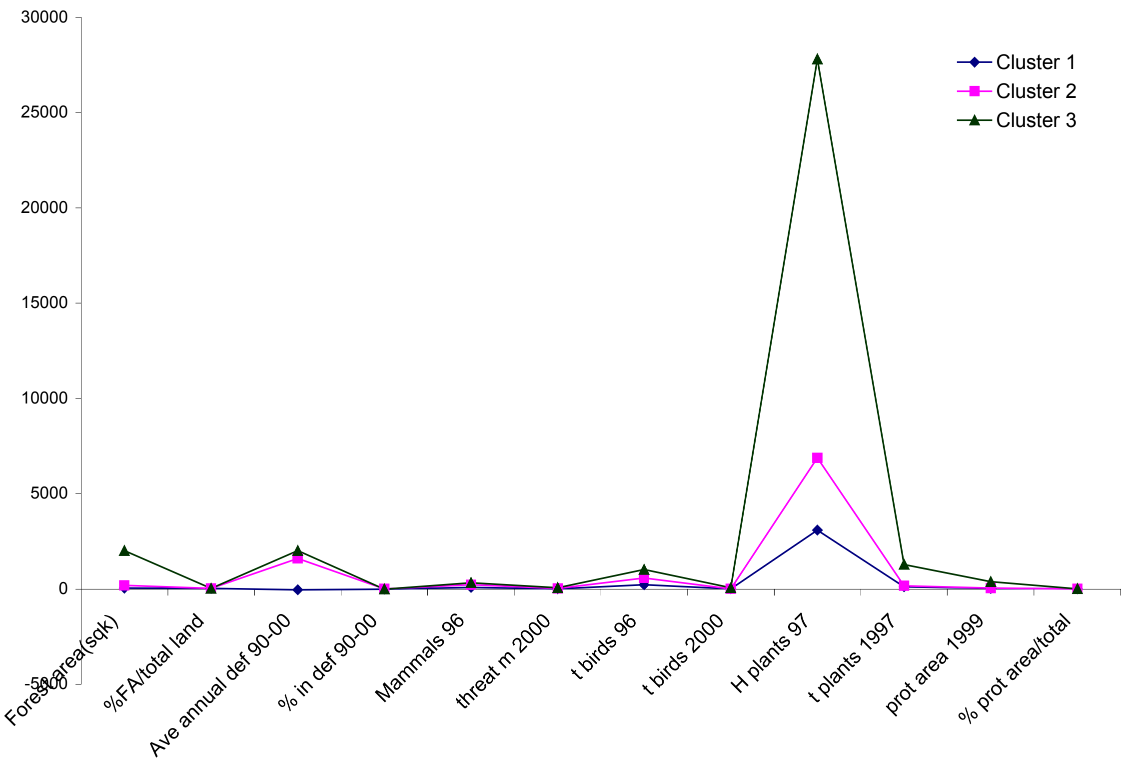


Figure 7.8 a: SOM created with deforestation and biodiversity data from (World Bank Report 2001b).

[21] Higher plants are the native vascular plant species. Source (World Bank Report 2001b)

Component	Cluster 1	Cluster 2	Cluster 3
Forest area(sqk)	50	188	2013
%FA/total land	24.85	29.56	35.39
Ave annual deforestation 90-00	-43	1599	2003
% in deforestation 90-00	-0.39	1.848	0.264
Mammals 96	82	223	336
Threatened mammals 2000	10	21.6	58
Threatened birds 96	221	572	1020
Threatened birds 2000	8	17	60
Higher plants 97	3094	6873	27799
Threatened plants 1997	118	176	1288
protected area 1999	21	39	379
% protected area/total	7.84	7.23	9.6



Figures 7.8 b & c: SOM cluster details and graph of deforestation and biodiversity data from (World Bank Report 2001b).

Initially, the map was classified into three major clusters and the interpretations arrived at are discussed below:

(i) Cluster 3 countries Colombia, Ecuador, Peru, India, Mexico, South Africa, Indonesia, Australia, Brazil, China, United States and Canada exhibit the highest world values for all variables analysed except for the percentage of deforestation (0.264) despite their average annual deforestation for 1990/2000 time period. This is because these countries have very large total land areas compared to that of other countries. The cluster also shows the world highest number of mammal, plant and bird species. Threatened plant, bird and mammal species also are high in these countries, including the world's highest values. This cluster could be future divided into two:

a. 3A countries Colombia, Ecuador, Peru, India, Mexico, South Africa, Indonesia and Brazil are in the high end of the ranges for all attributes except for forest area (1932, 2196 sqk) and threatened plants (189, 808).
b. 3 B countries, China, Australia, United states and Canada show the highest values of forest area and threatened plants. Percentage of protected/ total land for 1999 in these countries are 6.54, 7.96, 11.81 and 10.07 respectively.

(ii) Cluster 2 of this map consists of Angola, Paraguay, Zimbabwe, Sudan, Zambia, Nigeria, Côte, Rwanda, Burundi, Haiti, Mauritania, Myanmar Papua, Cameroon, Uganda, Nicaragua, Sierra Togo, El West, Niger, Malaysia, Ghana Guatemala, Malawi, Benin, Congo, Bolivia Tanzania, Kenya, Costa Panama, Central, Guinea, Venezuela, Argentina Thailand, Ethiopia Nepal, Mozambique, Philippines, Vietnam and Madagascar. It shows medium values for all attributes except for percentage of deforestation 1990-2000, which is at 1.848 the highest in the whole map. Within this cluster, Brazil shows the highest forest area. Indonesia has the highest threatened species for all categories. The cluster can be further divided into two;

a. 2A countries consist of values at the high end of the range for all attributes except for the percentage of deforestation 1990-2000.
b. 2B on the lower end of the spectrum.

(iii) Cluster 1 consists of all other countries and has the lowest values for all attributes.

(iv) The total species numbers as well as threatened species have a corresponding correlation to forest area. This could be interpreted as confirmatory of the theories adopted by WRI to use forest area data to represent species diversity and loss

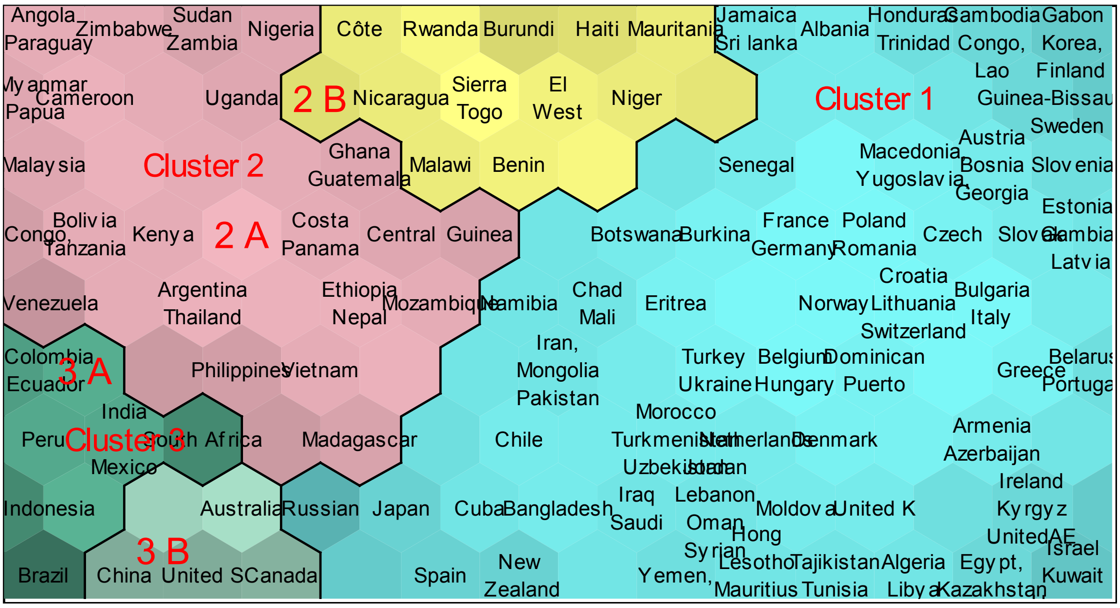


Figure 7.8 d: SOM created with deforestation and biodiversity data (World Bank Report 2001b).

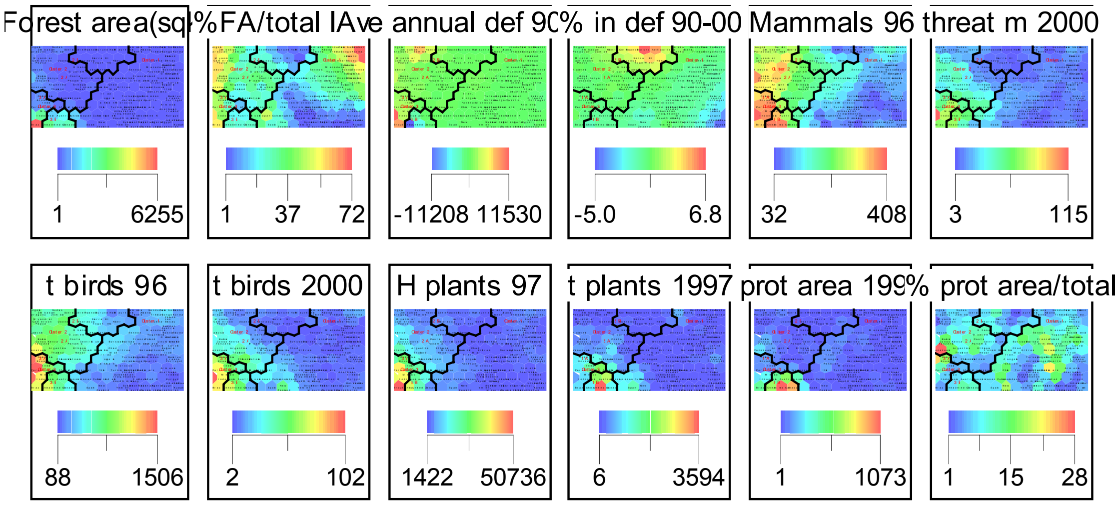
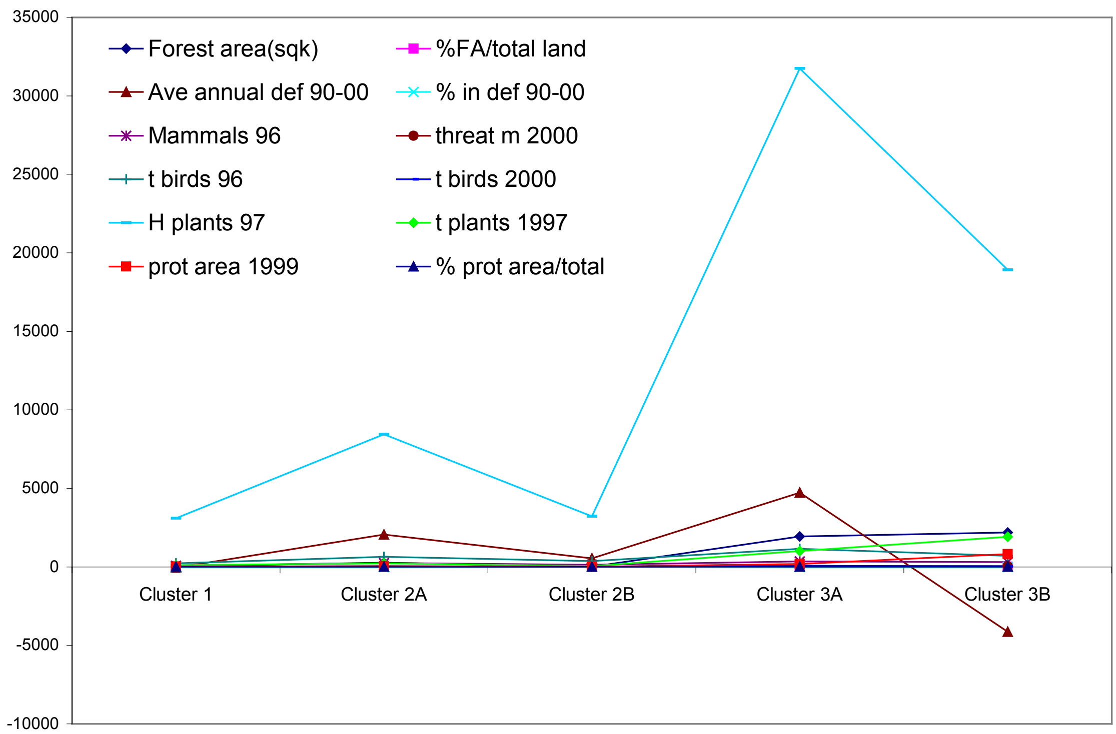


Figure 7.8 e: Component planes of SOM created using deforestation and biodiversity data from the World Bank report (World Bank Report 2001b) with five clusters.

Component	Cluster 1	Cluster 2A	Cluster 2B	Cluster 3A	Cluster 3B
Forest area(sqk)	50	261	17	1932	2196
%FA/total land	24.85	36.94	12.52	40.51	23.87
Ave annual def 90-00	-43	2059	538	4725	-4121
% in def 90-00	-0.39	0.983	3.844	0.505	-0.281
Mammals 96	82.7	256.3	146	345.9	314.4
threat m 2000	10.6	27.6	7.7	63.4	46.4
t birds 96	221	658	372	1160	704
t birds 2000	8.3	22.3	5.8	67.6	42.9
h plants 97	3094	8453	3227	31741	18928
t plants 1997	118	227	58	1010	1914
prot area 1999	21	50	13	189	808
% prot area/total	7.84	7.99	5.49	9.83	9.1



Figures 7.8 f & g: SOM cluster details and graph of deforestation and biodiversity data from (World Bank Report 2001b).

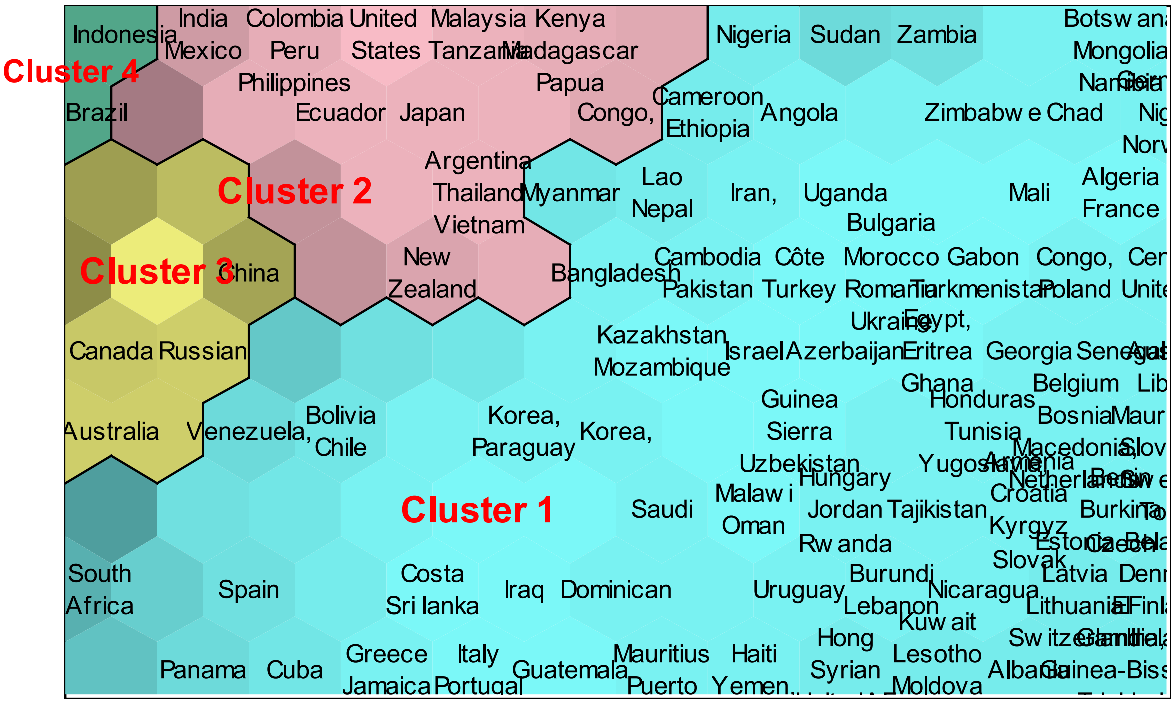


Figure 7.9 a: SOM created with 100 nodes and other map parameters set to default values using the World Bank data on deforestation and the number threatened species.

In order to study the relations between the state and response of biodiversity, only the following indicators stated in (Hammond et al. 1995a) from (World Bank Report 2001b) were used to create a SOM map with 100 nodes and all other map parameters set to default values:

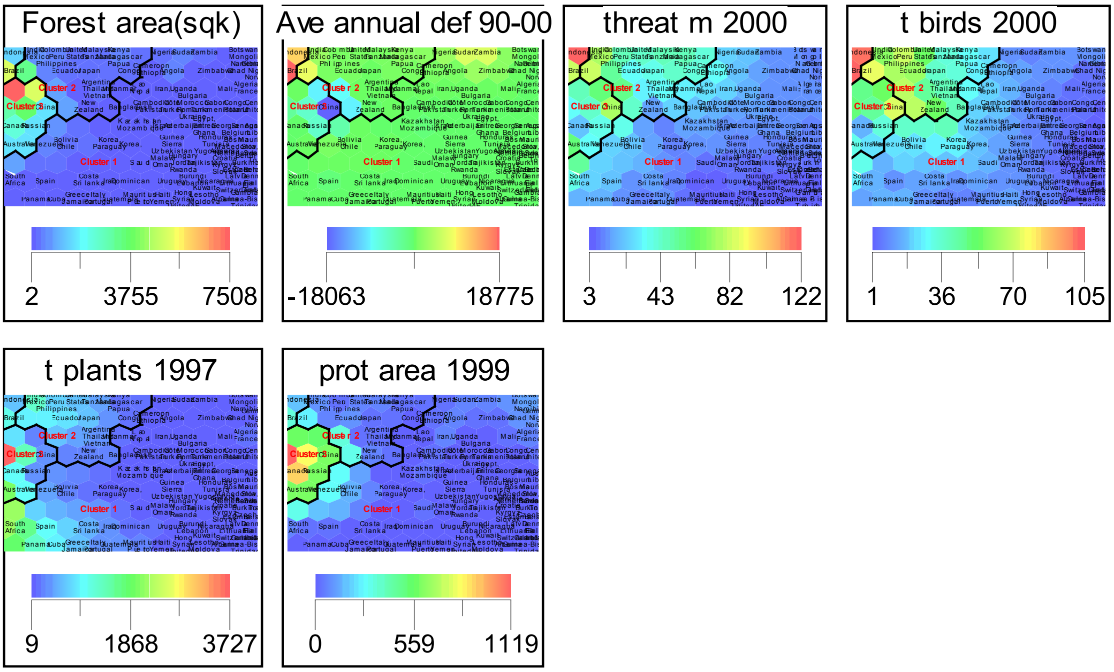
(i) List of endangered species from these group of organisms: mammals, birds, and higher plants as a state indicator for biodiversity

(ii) Statistics on wilderness, such as forest area and decline in forest area, also considered as a state indicator for biodiversity.

(iii) Statistics on percentage of protected area of total land area as a response indicator

Of the six components (figure 7.9 b), except for the average annual deforestation 1990-2000 all other variables show a corresponding correlation in that all the components leaving the former consist of similar high and low areas.

It should noted that the average annual deforestation 1990-2000 values are misleading as some countries classify plantation as reforestation (World Bank Report 2001b).



Component	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Forest area(sqk)	72	357	3221	3292
Ave annual def 90-00	329	1921	-4156	15411
threat m 2000	11.6	44.6	47.1	101.8
t birds 2000	8.5	45.3	43.3	103.4
t plants 1997	135	494	1563	855
prot area 1999	25	55	763	273

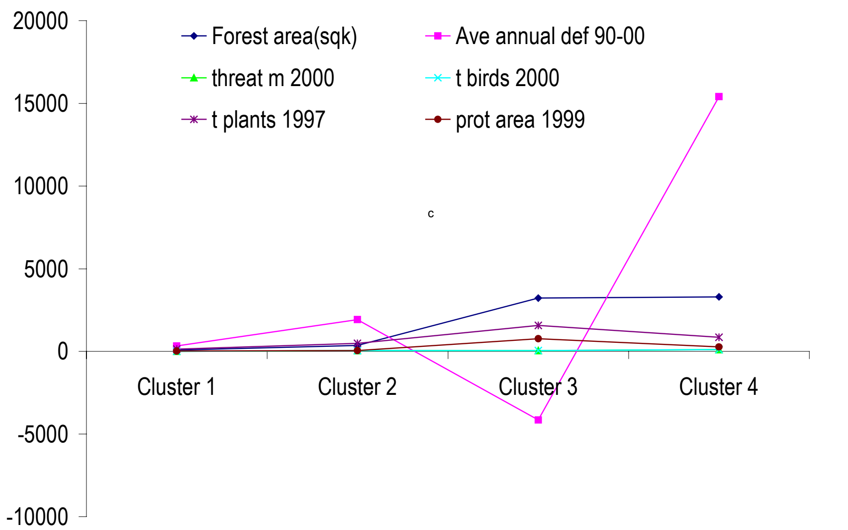


Figure 7.9 b: Component planes of SOM created with World Bank data on deforestation and number threatened species. c: SOM cluster details and d: graph showing the difference in clusters 1 to 4.

The interpretations derived from the map (7.9 a, b & c) are:

(i) Cluster 4 countries of the SOM, consisting of Indonesia and Brazil are seen with the highest values for forest area, average annual deforestation, threaten mammals and birds for year 2000. They are seen to be less responsive in terms of protected areas, as would be seen in figures 7.9 b-e.

(ii) Cluster 3 consists of China, Canada, Russian and Australia with the second highest values for forest area and threatened plants and the highest protected area for 1999, the values being 598, 916, 613 and 603 sqk respectively. These countries also show negative values for deforestation. It should be noted that forest area includes actual forest areas as well as areas under plantation. Hence, deforestation negative values could be due to areas of plantation.

(iii) Cluster 2 counties, India, Mexico, Colombia, Peru, Philippines, United States, Malaysia, Tanzania, Kenya, Madagascar, Papua New Guinea, Ecuador, Japan, Congo, Argentina, Thailand, Vietnam and New Zealand show similar attributes as cluster 4 countries, but with lesser values. Hence, exhibit the second worst state in terms of threaten species in all three categories for 2000.

(iv) All other countries fall into cluster 1, with the lowest values for all variables except for deforestation.

Cluster 3 countries consists of high threaten plants. Cluster 1 countries show low threatens mammal and bird species while cluster 4 countries show high values for these species. Clusters 2 and 3 show medium threaten species for these two categories.

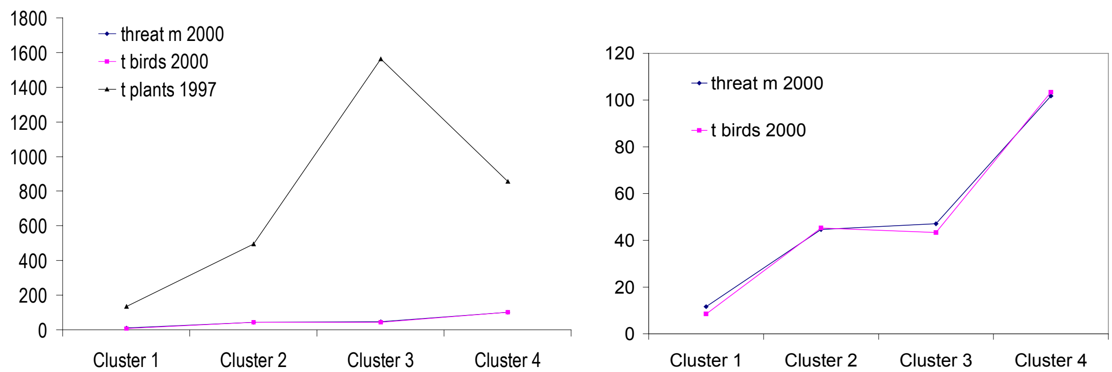


Figure 7.9 c: Graphs showing SOM cluster details of the World Bank data on deforestation and number threatened species.

7.4.3 SOMs to analyse composite global data

Data on urbanisation and biodiversity used in the above SOM analyses were combined to develop a new SOM map to see the patterns between development and biodiversity. The variables included in the SOM (figure 7.10 a and b) were developmental activities, Gross Domestic Product (GDP), agriculture, Industry and manufacturing services from (World Bank Report 2002) with biodiversity indices and rural development data, which were analysed separately earlier in this chapter.

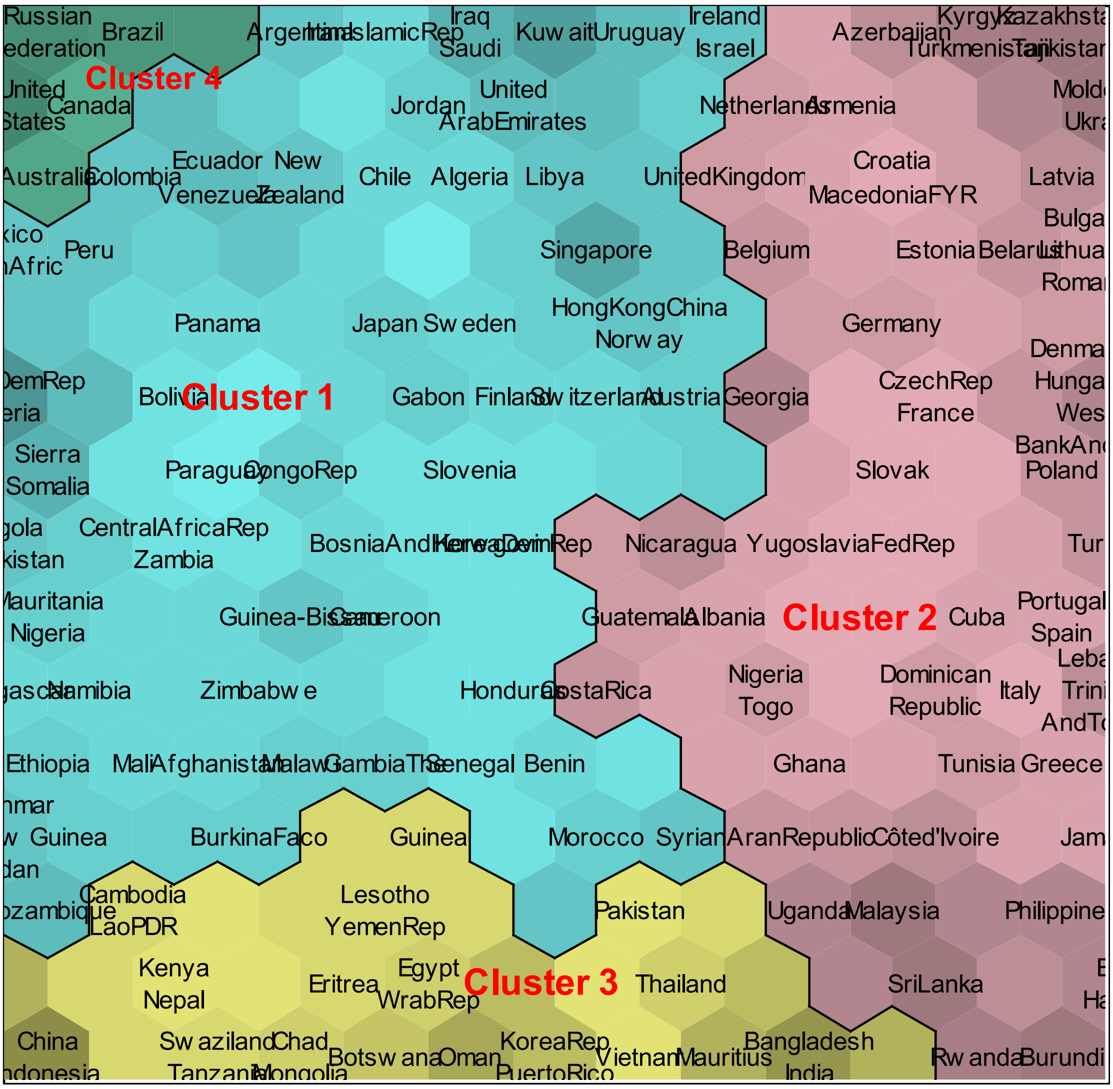


Figure 7.10 a: SOM created with development and biodiversity data compiled by the Word Bank (World Bank Report 2001a; World Bank Report 2001b; World Bank Report 2002) with 100 nodes and all other map parameters set to default values.

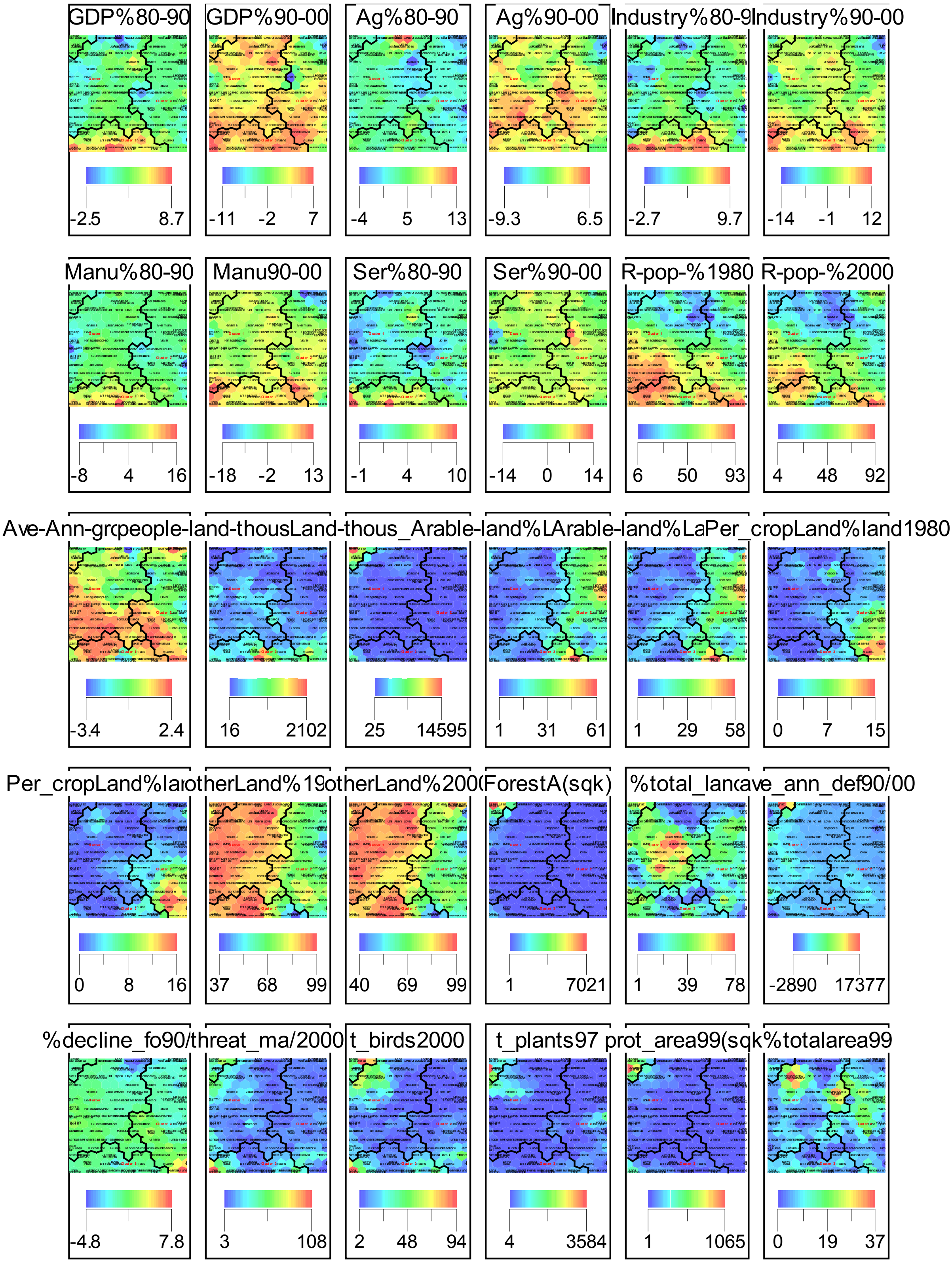


Figure 7.10 b: SOM component planes of development and biodiversity data.

Component	Cluster 1	Cluster 2	Cluster 3	Cluster 4
GDP%80-90	1.93	2.16	5.84	3.12
GDP%90-00	2.79	0.83	4.62	1.68
Ag%80-90	2.93	0.79	3.33	2.67
Ag%90-00	2.35	0.34	2.61	0.12
Industry%80-90	1.9	2.55	7.33	2.56
Industry%90-00	2.27	0.18	5.79	0.14
Manu%80-90	2.17	2.3	9.56	2.29
Manu90-00	2.06	-0.43	6.12	2.67
Ser%80-90	2.1	2.85	6.26	3.31
Ser%90-00	2.27	2.57	4.9	2.27
R-pop-%1980	53.53	48.78	75.18	27.15
R-pop-%2000	43.72	41.68	63.31	22.35
Ave-Ann-growth1980-00%	0.809	0.206	1.167	0.215
people-land-thousand-sqk1999	271	268	737	51
Land-thous_sq_km1999	687	214	944	9761
Arable-land%LandArea1980	6.88	28.16	15.08	9.08
Arable-land%LandArea1999	7.39	26.47	15.24	8.74
Per_cropLand%land1980	0.92	4.93	1.28	0.43
Per_cropLand%land2000	1.02	5.41	1.75	0.43
otherLand%1980	92.1	66.9	84	90.5
otherLand%2000	91.5	68.1	83.3	90.8
ForestA(sqk)	194	40	210	3895
%total_land	31.78	23.57	23.82	37.86
ave_ann_def90/00	1081	93	123	3471
%decline_fo90/00	0.228	0.483	0.236	0.075
threat_ma/2000	17.7	11.8	28.8	46.1
t_birds2000	15.4	8	25.6	51.1
t_plants97	195	175	178	1677
prot_area99(sqk)	44	13	73	674
%totalarea99	8.11	8.63	7.02	7.47

Figure 7.10 c: Cluster details of SOM created with urbanisation and biodiversity data

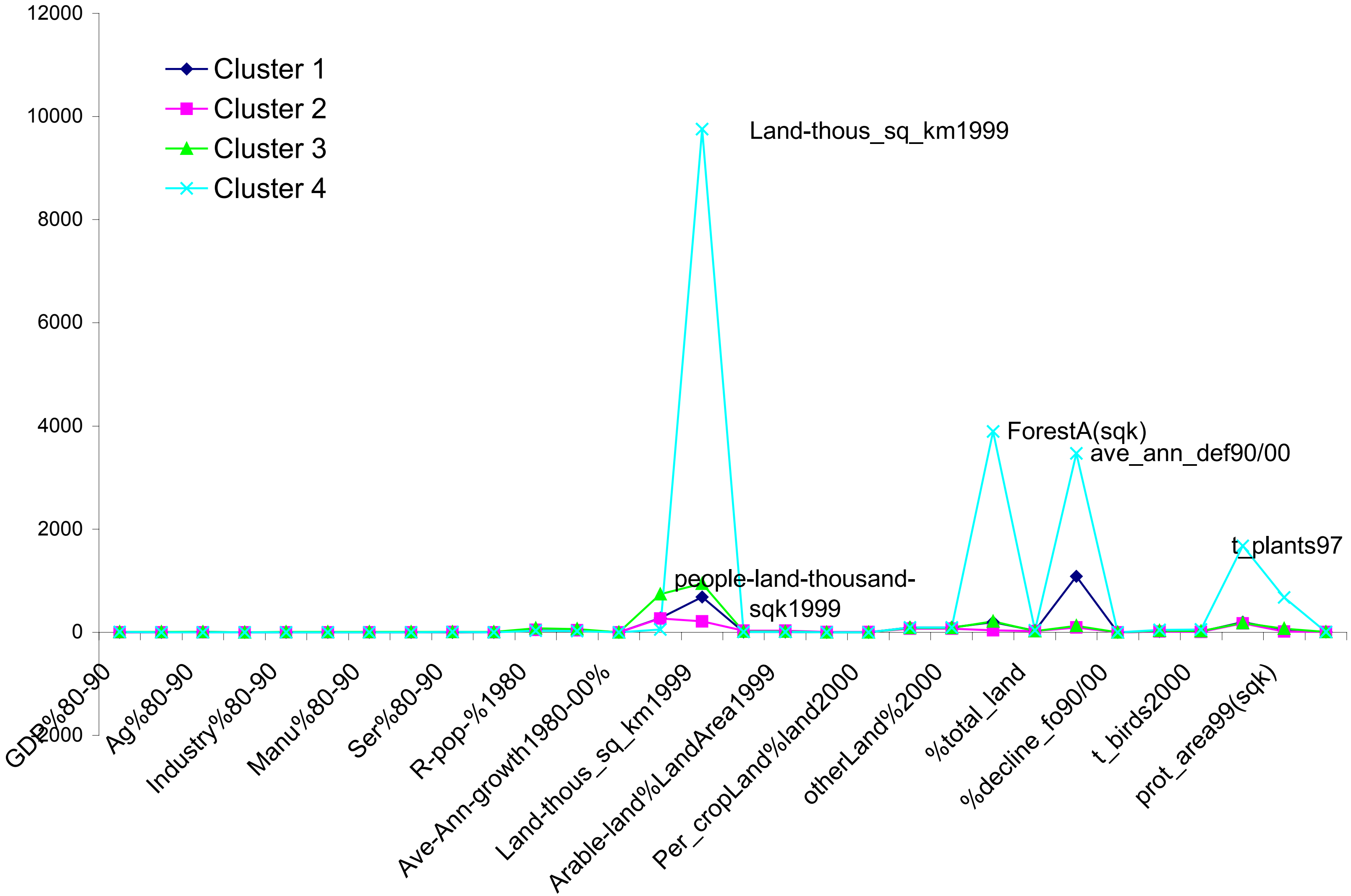


Figure 7.10 d: Graph showing the cluster details of SOM created with urbanisation and biodiversity data

The following are the interpretations arrived at from this map:

(i) Cluster 4 consisting of Russian Federation, Brazil, United States, Canada and Australia show medium GDP, average annual percentage growth of Agriculture, Industry, Manufacturing and Services in 1980 to 1990 and 1990 to 2000 time periods. These counties also show medium to high numbers of mammals, birds, higher plants and high number of threatened species of the same. It can also be seen that they have large areas of protected land, but when converted into percentage of total land area these countries have low percentages. Only variable that is low in this cluster is the percentage of decline in forestation because of their high total land area.

(ii) Cluster 3 countries, Papua New Guinea, Cambodia, LaoPDR, Lesotho, Yemen Repbulic, Pakistan, Kenya, Nepal, Eritrea, Egypt WrabRep, Thailand, China Indonesia, Swaziland Tanzania, Chad Mongolia, Botswana, Oman, Korea Republic, Puerto Rico, Vietnam, Mauritius, Bangladesh and India show high GDP (both in 1980-90 and 1990-2000), agriculture, manufacture, rural population, people/ 1000 sqk land (737 the map highest), average annual growth in rural population and percentage decline in forest (1990-2000). The cluster has high values of threaten species for all three categories.

(iii) Within this cluster 3, China and Indonesia in one node show the highest GDP, highest values of mammals and birds for 1996 and the highest rates of threatened species for both in the year 2000. These two counties also have enjoyed the highest GDP for both years with high industry, manufacturing and percentage of average annual deforestation (1990/2000). It could be stated that their GDP growth has come at the expense of biodiversity.

(iv) Cluster 2 countries, Azerbaijan, Kyrgyz, Turkmenistan, Kazakhstan Tajikistan, Netherlands, Armenia, Moldavia, Ukraine, United Kingdom, Croatia Macedonia FYR, Latvia, Belgium, Estonia, Belarus, Bulgaria Lithuania, Romania, Germany, Georgia, Czech Republic France, Denmark Hungary, West Bank And Gaza, Slovak, Poland, Nicaragua, Yugoslavia, Turkey, Guatemala, Albania, Cuba, Portugal Spain, CostaRica, Nigeria Togo, Dominican Republic, Italy, Lebanon Trinidad AndTobago, Ghana, Tunisia, Greece, Syrian Arab Republic, Côted' lvoire, Jamaica, Uganda, Malaysia, Philippines, Sri Lanka, El Haiti, Rwanda and Burundi show low to medium GDP, agriculture, industry, services, rural population, forest and total land. However, they have high (268) people/1000sqm land (1999), compared with that of cluster 1's 51.

(v) Cluster 1 counties, Argentina, Iran, Iraq Saudi, Kuwait, Uruguay, Ireland Israel, Jordan, United Arab Emirates, Colombia, Ecuador Venezuela, New Zealand, Chile, Algeria, Libya, Mexico South Africa, Peru, Singapore, Panama, Japan, Sweden, Hong Kong, Norway, Congo Liberia, Bolivia, Gabon, Finland, Switzerland, Austria, Sierra Somalia, Paraguay, CongoRep, Slovenia, Angola Uzbekistan, Central Africa, Zambia, Bosnia And Herwgovir, Korea, Mauritania Niger, Guinea-Bissau, Cameroon, Madagascar, Namibia, Zimbabwe, Honduras, Ethiopia, Mali, Afghanistan, Malawi, Gambia, Senegal, Benin, Myanmar, Papua New Guinea Sudan, Burkina Faco, Morocco and Mozambique show low to average values for all the attributes analysed.

(vi) Variables attributing to the growth of development have a corresponding, correlation in the time intervals (1980/1990 & 1990/2000) analysed.

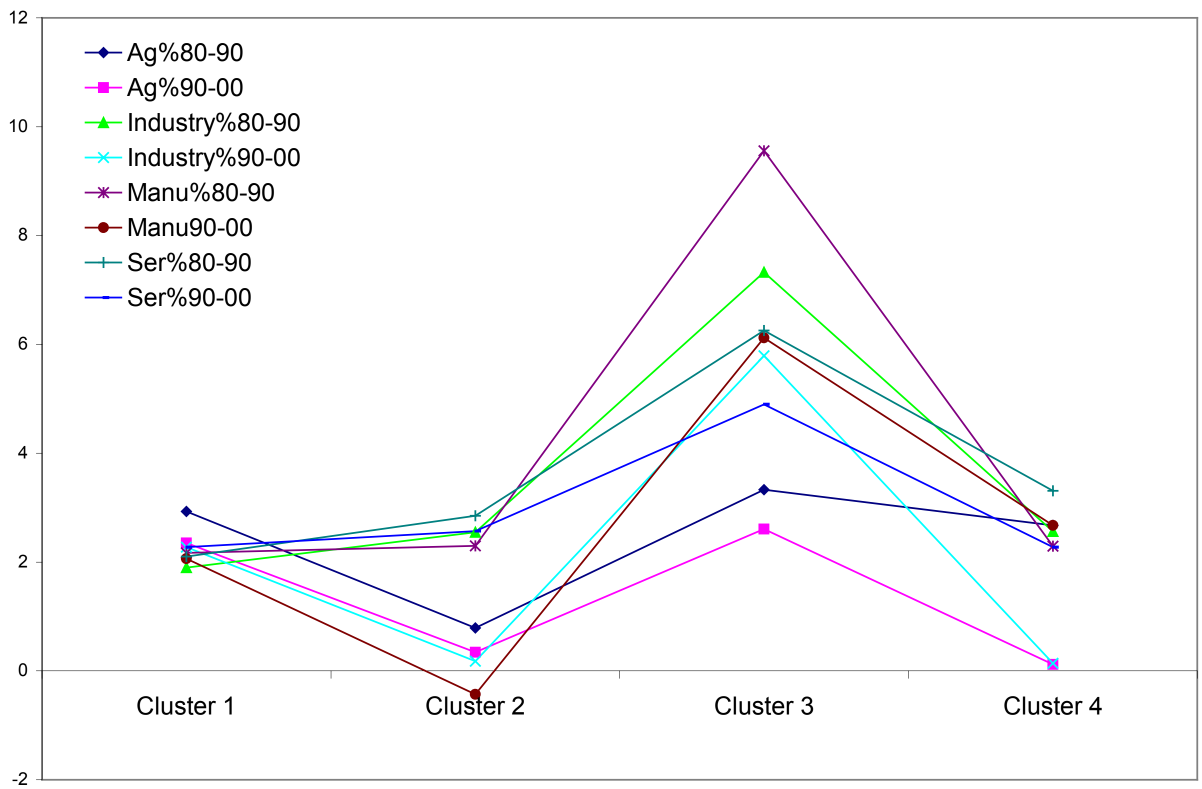


Figure 7.10 e: Graph showing the cluster details based on developmental activities.

The SOMs of combined data gave a means to look for patterns in these indicators without any aggregation. From the SOM maps, the major contributing factors in the fused data set can be distinguished and if necessary more priority could be given during the training process so that they carry more weight in the clusters map. On the other hand, indicators of the pressure, state and response of an ecosystem could be analysed to see the correlations, patterns, and trends in them. As SOM are visual analyses they can indicate whether the state of an ecosystem is improving or getting worse as well as the relationships between the pressure and state for the observed trends when used with appropriate data sets.

7.5 conclusion

The examples used in this chapter showed how SOM analyses could be applied to studying multidimensional disparate data sets at global scales, within an integrated framework with the available knowledge on the data being analysed. They even can be applied to learning more information, in particular the relationships between the pressure, state and response indicators of diverse ecosystems using data from statistical tables.

Summary

The third case study of this research attempt to experiment the use of SOM methods, as used in industrial process modelling and financial analysis produced good results. It illustrated a sensible approach for the implementation of the current approaches such as the PSR and information pyramid concepts, even with issues of limited knowledge. Data within PSR framework could be analysed with SOMs, alleviating the problems faced in the calculation of appropriate weights for indicator aggregation. The chapter illustrated the use of SOMs to analyse disparate global data of biological and developmental activities along with economic as well as social interests within an integrated framework, in particular to analyse the economic trade-offs on the present and future decisions contemplated by resource managers.

SOM methods could be used to collectively analyse multidimensional data sets with little prior knowledge, to learn about the relationships, structures and trends across scales using the available data, as required in recent environmental projects, such as the Millennium ecosystem assessment (MEA). The previous two chapters along with this third case study gave details on how SOM techniques could be best applied to tracking a systems dynamics in ecosystem modelling, using ecological monitoring data without any physical models. SOMs provided a means to analyse highly complex and diverse ecosystems incorporating their spatial and temporal variations using the abundant numerical data sets, collected by academic, state and research institutions.

The next chapter analyses the benefits of SOM applications to environmental sciences as well as the advantages and disadvantages with the different software used in this research for the implementation of SOM techniques.

Chapter 8

Results and discussion

SOM based approaches (cluster analysis, dependency analysis, decision support system and time series analyses/ trajectories) are investigated for exploratory data analysis in biological and environmental sciences, as applied in highly complex industrial system process modelling and financial data analysis. Chapter 4 gave details of the approaches adopted for investigation in the research. The last three chapters described case studies and the results arrived at, by using SOM techniques to model extremely complex, highly diverse, naturally evolving ecosystems. Modelling is also possible at different scales (regional and global) and levels (environmental, biological and ecosystem). The overall results on the use of SOM techniques to model various ecosystems are revealed in this chapter. In addition, the reasons for suggesting SOM methods to bridge the gaps in the existing ecological data analysis methods are explained in detail.

8.1 Ecological data analysis and ecosystem modelling

In general, the existing conventional ecological data analysis methods are increasingly seen to be inadequate in the sense, they are unable to inform resource management of environmental conditions or any major threats to ecosystem functioning. The highly sophisticated, professional design methods of ecology, such as BACI, BASIPS are complicated, rigorous and yet incapable of distinguishing the effects of environmental impact, whether due to human activities or natural causes. However, not only the detection of such environmental effects and their causes, but also the prediction of deleterious effects are stressed and reiterated to created a better human-environment relationship that protects natural habitats along with their biodiversity. The same issues have been emphasised over and over again since the late 1980s. A large volume of literature is reviewed in chapter 3, providing details from different scientific perspectives.

The need for reliable environmental impact analysis and prediction models has never been so great; anthropogenic degradation of the environment continues to increase and also the kind and severity of human influence causing them. The demand on ecosystems, such as natural resources and produces, continues to escalate at unprecedented proportions due to the exponential growth of the world population, imposing more pressure on the degraded and overexploited natural systems (Reid 2000). Scientists, ecologists and many national and international institutions have expressed their concerns over the issues and taken remedial measures to address them. UN's MEA, EPA's STAR programmes and WRI's efforts are a few recent attempts that have resulted from this (Environmental Protection Agency; Bierbaum et al. 2001; Ministry for the Environment 2002a; Ministry for the Environment 2002b).

A better, co-ordinated effort on decisions affecting ecosystems is seen as imperative for humanity's wellbeing in the long run. Scientists and stakeholders need to work together to predict ecosystem responses, so that natural habitats sustain human activity, was elaborated upon in chapter 3. But in the past, environmental impact assessment (EIA) of conventional statistical methods has caused wrangling between environmentalists and land developers over the approval of proposed development (Buckeridge 1999). Often arguments are based on the soundness of conventional ecological assessment methods (Mapstone 1996). Stakeholders encourage as many as possible developmental activities as these invariably improve a nation's current socio-economic status by generating employment and revenue. Blanket restriction on development could cause undue hardship by eliminating even the environmentally non-harmful ones, and *vice versa* could affect future generations with massive environment degradation and loss to biological diversity. As supported by a number of scientific papers reviewed in the thesis, it has become necessary to distinguish the human induced environmental impact from that of natural and global variations, such as spatial, temporal and climate change.

The following are the questions set out for investigation through this research, to fulfil the need for better methods in ecosystem modelling:

(i) How could SOM methods be best applied to unravel the structure of highly diverse, extremely complex and naturally evolving ecosystems, and to predict their system dynamics

(ii) Could this approach be applied for the conversion and dissemination of disparate (i.e. ecological and socio-economic) data sets to a wider community, to preserve various ecosystems along with their innate biodiversity and functioning for human wellbeing?

The three case studies experimented in this research show that SOMs could be used for ecosystem modelling in an efficient and very constructive manner. SOM methods provide an approach for the conversion of abundant monitoring data (at different levels and scales), into meaningful information in visual formats. They are useful in modelling diverse ecosystems including their biodiversity and environmental changes with spatial and temporal variations, using disparate data, such as biological, environmental as well as economic data, at ecosystem level within an integrated framework. The case study results on the investigation of how best SOM methods could be utilised for ecosystem modelling are elaborated upon.

8.2 Use of SOMs in Long Bay Okura Marine Reserve data

In the Long Bay Okura Marine Reserve analysis, the use of SOM techniques provided a means to collectively analyse the reserve's ecological monitoring data along with NSCC's *Enterococci* data. These data sets were collected and analysed separately by AUT students and NSCC staff to study the reserve's ecosystem (physical and biological) changes and the beach water quality at Long Bay respectively.

The SOMs also revealed the non-linear relationships between the input vectors. As SOM can display the input vectors on two-dimensional formats, they could be used assess the environmental impact needed by resource managers for decision making on future development and present practices.

The SOM clustering was found to be useful in analysing the Long Bay Reserve data. They showed SOM approaches to be capable of delineating intricate patterns, even in disparate data sets collected at irregular intervals. The detection of correlations among environmental and biological variables was found to be relatively easy on the SOMs created with the fused data sets, from different sources with inconsistent labelling and missing values.

The dependent component plane analysis proved SOM abilities to relate the reserve's physical and biological processes within the intertidal zone. The component planes were useful in discerning the major contributing variables for the observed physical system changes and their effects on the sciaphilic colonisation at the reserve. For instance, the SOMs portrayed the eutropic conditions observed in the reserve, from the raw data.

The difficulties in incorporating spatial and temporal variations, especially in analysing the effects of environmental impact over large areas using conventional methods could be overcome with SOM approaches. The SOM clustering that distinguished the different intertial, littoral zones (S1-S4) from the Reserve data showed its potential to model spatial variations within the monitoring data.

The time series analyses or trajectories on SOMs showed how temporal variations and their dynamics within the reserve's monitoring data could be modelled for prediction purposes. However, the trajectories of Long Bay Reserve data were not as regular as seen in complex industrial system dynamics modelling examples, as the reserve data was intermittent, not producing a smooth flow in the trajectory.

8.3 Use of SOMs in council monitoring data

The use of SOM techniques in ARC's LTB saline water quality provided an approach to comparative analysis on beaches across a region, north of Auckland using the available data. SOMs project the input vectors on two dimensional formats, which are capable of displaying the non-linear relationships and spatial variations in the data even without knowing the class memberships in them. In an earlier study by (Wilcock and Stroud 2000), each and every attribute of the data was analysed separately with several two dimensional graphs for every single beach water sampling, which made comparative analysis of the beaches very difficult.

SOM analyses gave a means to study the relationships between the sedimentation deposition rates with the subtidal marine habitat population dynamics of selected beaches between Campbells Bay and Waiwera, northeast of Auckland. The biological data set of species average count produced complex matrices of numerical values (Walker et al. 2000). Analysing such complex matrices of data is a common problem faced in multivariate analyses by ecologists (Giraudel and Lek 2001). Furthermore, the multivariate analytical methods generally used by ecologists to correlate the environmental and community patterns are data dependent and may produce misleading, confusing, unstable or incomputable results (see chapter 3). The SOM analyses (implemented with RICBIS and Viscovery, a commercial data depiction software) not only distinguished the annual variations from that of sediment induced changes on the subtidal population dynamics. They also provided a quick and simple means to establish the link, within the monitoring data only spanning a period of three years and with inconsistent labelling. The detection of such correlations, such as determining whether an impact was due to human influenced or of natural causes, has been difficult if not impossible, as the impact of an ecosystem may or may not co vary with the ecological impact within an ecosystem, due to the following reasons:

(i) the spatial variations of the ecosystem, and the effects of the extensively varying impact (such as urbanisation) being analysed (Thrush et al. 1995).

(ii) many species typically exhibit a non-linear, threshold response to a physical environmental change (Raimondi and Reed 1996).

(iii) biological system responses are slow, subtle and are often mistaken as 'no response' by many conventional data analysis methods (Clark et al. 2001).

In the subtidal community population dynamics example, the species that were found to be potential indicators by fine tuning the cluster maps; one among the identified species in (Walker et al. 2000) is, *Evechinus chloroticus*

Even though some of these results were derived using standard statistical analyses, because of the complexity in the conventional methods, stakeholders and the general public have limited the use of these findings in their decision making processes (Buckeridge 1994). But in the SOM method approach, data sets are mapped onto easily understandable, visual displays, providing a means to visualise the correlations among the various causes and their resulting effects. As multidimensional data sets can be directly projected onto two-dimensional displays, SOM maps also provide a more plausible approach than the existing conventional methods.

Using conventional MDS clustering methods (Walker et al. 2000) failed to detect any constructive patterns in the monitoring data sets. It is stated in the report that the MDS clustering methods carried out by them had failed to reveal any constructive patterns in the community structure on the species abundance data for years 1999-2000. This may be because the MDS methods used did not show any useful patterns, or the analysts were not able to interpret them. MDS clustering was projected onto two-dimensional displays using the famous Bray-Curtis similarity matrices calculated on 4^{th} root transformed species abundance data. The multivariate community analyses were used to determine the community structure patterns, undetectable by univariate methods carried out by the analysts (Walker et al. 2000). This drawback was overcome by SOMs in the case study.

SOM clustering results could be incorporated with GISs for easy viewing. The SOM clustering results incorporated into a GIS, such as Arc View, version 8.2 showed how SOM cluster results could be incorporated with GISs.

8.4 Use of SOMs in global data

The results achieved through the third case study of this research aimed at investigating the use of SOM techniques in ecosystem modelling at higher levels and scales are reviewed in this section. The main aim of the case study was to produce an approach, to detect global trends in the effects of urbanisation on the global ecosystem and its biodiversity using statistical numeric data. It has the potential to bring the three main participants (scientists, stakeholders and the general public) needed together to preserve ecosystems for future generations. Conventional methods do not create a common trust among the main participants; they neither encourage integrated, interdisciplinary environmental research nor provide a means to implement the triple bottom line (TBL) and similar model concepts (Harris, 2000). The twentieth century research efforts of gaining in-depth knowledge with a fragmented image of nature has been blamed for this (Bowler 1992).

Unavailability of predictive models of ecosystem response to human influence or inability to invent such models is seen as a major factor for the recent global environmental degradation. The historic approach of sectoral management and the resulting inevitable response management has led to the regional environmental problems becoming global, such as habitat destruction, local species extinction, emission of greenhouse and CFC gases and overfishing.

As the world population continues to escalate at unprecedented proportions, the demand on natural resources and biodiversity as well continues to increase. The need to preserve severely failing natural systems for future generations seems greater than ever. The requirement for understanding diverse ecosystems to improve human-environment relationship has led to redirection on research efforts towards introducing interdisciplinary approaches with innovative integrated data analysis and modelling techniques.

Chapters 3 and 7 discussed how we deal with environmental issues is crucial for human wellbeing in many ways i.e. economics, social status, future prospects. The socio-economic status of a region, a nation and ultimately the natural systems of this planet depends on how we deal with these issues in our daily life. This resulted in the development of the TBL, 4Es (economics, ecology, ethics and engineering) (Buckeridge 1994) and many more concepts, which set out rules for implementation of environmentally sustainable development. These concepts were considered as a useful means as they permit only the developments that would not change the natural ecosystem functioning. However, implementation of such concepts has never been easy. The current EIA practices in many countries cause wrangling between environmentalists and land developers (Mapstone 1996). The twentieth century's fragmented image of nature (Bowler 1992) has widened the gap between different professionals. The mistrust between these professionals does not encourage co-ordinated effort for sustainable environmental management (Harris 2002). The reliability of scientific findings and predictions are questioned because of the knowledge divide and the qualitative interpretations of scientific reports that are increasing criticised as overstating, such as marine dessert, or understating, such as no observable deviation, the ecosystem damage (Ambrose et al. 1996; Mapstone 1996; Buckeridge 1999; Harris 2002).

Natural processes are complicated with many subdued and slow interrelated reactions and compensating mechanisms (Clark et al. 2001). Despite the advances achieved by scientists and ecologists over the decades, the issues remain the same (see chapter 3). Indeed, the more we learn about ecosystem processes the greater the complexity in modelling them. Because of this model uncertainties found in traditional methods tend to undermine their use. "… given the variety of ways in which regions can differ, it is unlikely the model uncertainties will disappear. Indeed it is unlikely that we will ever have an exactly correct model. Thus formal inference will need to include both diagnostic checks to exclude plausible models that do not fit the data, and rough measures of model uncertainty from those not excluded…" (Stewart-Oaten 1996b:129). SOMs give a plausible quantitative approach making use of large amounts of monitoring data sets; incorporating spatial and temporal variations along with any other deviations that may exist within the ecosystem being analysed.

While the constraints for environmentally unsustainable activities continue to increase, the need for improved procedures for detection and interpretation of environmental impact becomes ever greater. The recent advances achieved in many research disciplines are capable of detecting the critical ecological changes at a range of levels such as DNA, gene and tissues. However, methods to measure the deviations at ecosystem level were recommended to analyse natural system for sustainable environmental management; as many of the important changes that affect ecosystem functioning, occur at this level (Hammond et al. 1995). In chapter 3, it was shown that new approaches adopted should be combined with logic and some form of scientific reasoning; described by the field ecologists as more academic field experiments. The perceived lack of rigor in environmental monitoring against the academic ecological experimentation has been expressed to be a difference between 'applied' and 'pure' science for a long time now (Underwood 1996; Harris 2002); considered as a reason for the lack of proper communication between scientists and others. Nevertheless, it is important to distinguish the good from bad practices instead of arguing about pure and applied. Ecologists should be more responsive in adopting new methods that could convince stakeholders and the public. If not, decisions on resolving environmental issues may continue to be more expensive and dominated by the more stochastic processes of law (Lester 1996).

The efforts made by the Canadian and the Dutch governments resulted in the establishment and use of environmental indicators based on the PSR model. The World Bank and WRI as well have devised lists of environmental indicators (see appendices 3 and 4 for details) based on this model. The MfE in New Zealand initiated programmes to develop an indicator system based on this model (). In the PSR indicator system, primary data from various sources are first aggregated to produce a set of indicators, which are then further aggregated to produce a set of indices with concise information based on the information pyramid (see chapters 3 and 7). Ultimately, the indices on the top of the pyramid are made easy for use in decision making processes, similar to the indices of other disciplines, such as GDP. However, complex calculations based on how these indicators contribute to the final index need to be worked out. Hence, unless explicit knowledge on physical processes is available the approach cannot be applied. In chapter 7, it was shown that with the use of SOM techniques, PSR indicators or even disparate raw data sets could be analysed revealing the many non-linear relationships within them. As SOMs directly map the numerical data, they provide a method of quantitative analysis on the complex ecosystem process state, pressure and response without any physical models as used to study the dynamics of industrial system processes. SOMs as well provide an approach for the implementation of indicator aggregation based on the information pyramid and PSR models even without explicit physical process knowledge. In issues, where knowledge is limited for calculating the weights of the different components to create aggregated or composite indices SOMs seems to be an excellent tool for converting the data into concise information. SOMs are useful in identifying the contributing variables in initial data analyses, with options for studying correlations among disparate data sets as there is no need to understand the physical process involved in this regard. Biological, environmental as well as economic data could be analysed at ecosystem level, within an integrated framework.

SOMs can even be useful for analysing issues concerning biodiversity for which indicators have not been developed by WRI and the World Bank (appendices 3 and 4). The analysts were working on approaches to overcome the issues by making use of the ecosystem variations, such as soil type, integrated into GISs, already developed for use in other areas, such as developmental planning in some South African countries, fire services with varying risk factors in the US. Chapter 7 provided sufficient details on the use of SOM techniques, to analyse the issues without proper indicators. For easy viewing, SOM clustering could be integrated to a GIS, as carried out in case study two, where SOM clustering results on the subtidal population dynamics and sediment deposition data were incorporated to a GIS software Arc View 8.2.

SOMs can also be used to analyse the contributing factors for index aggregation based on the information pyramid concept. In the information pyramid approach adopted by the World Bank and WRI, primary data/ indicators are converted into condensed information by adding all factors, multiplied by their respective contribution towards the issue (see chapter 7, for details on the aggregation of weighted data/ indicators). With the use of SOM components, indicators of unknown contribution towards the final effect can be studied even without explicit knowledge on the issue being analysed.

SOMs can be used to detect water quality trends on the coastal habitats through the monitoring data as in freshwater systems. The predictive models, such as RIVPACS, AUSRIVAS have been successfully used for assessing river health in Britain and Australia. Even though AUSRIVAS system was primarily developed for lotic environments, efforts are currently being made to extend it by widening its scope, focusing further into estuarine and wetland environments (National River Health Program 2002). The current level of knowledge on coastal and marine systems is significantly limited, let alone the complexity, which involves oceanography, climate and atmospheric variations. Thus SOMs could be an ideal tool as they are capable of analysing systems with little prior knowledge using measurable variables; as seen in initial financial data analysis, explained in chapter 3.

8.5 Conclusion

The results of the research attempt to investigate the use of SOM techniques in biological and environmental process modelling proved the approach as practical. The hypothetical approach examined to apply different SOM methods, similar to their use in complex industrial process dynamics and financial analysis produced promising results.

The SOM techniques experimented with case studies from biological and environmental sciences gave evidence of a quantitative analytical approach useful in many aspects. The approach bridges the existing gap, critical for the conversion of multidimensional disparate data sets into concise information for use by resource managers. The conventional methods are often suited for hypothesis testing; confirmatory statistical methods, such as null hypothesis, confidence intervals and regions. Hypothesis tests are generally more rigorous, objective way of decision making and are not suitable for exploratory analysis of environmental monitoring data (Stewart-Oaten 1996a). As our current knowledge on ecosystem behaviour is limited, mainly in the context of human activities, postulation of hypotheses and testing them seem very difficult, in fact it is a different practice altogether and will not permit to harness the knowledge embedded in the widely available environmental and biomonitoring data.

The variations observed among and within ecosystems cause major constraints to model them with the existing methods. Especially the spatial and temporal variations observed within an ecosystem, along with genetic variations exhibited within a species cause the formulation of a global ecosystem model not at all practical. The response or threshold variations exhibited by different species within a region further complicates modelling. A similar dilemma faced in industrial process system modelling was successfully overcome with the use of SOM techniques. It provided an approach for exploratory data analysis and this was applied in this research for ecosystem modelling. The results provided sufficient potential for use to analyse highly varying, equally complex ecosystems.
The use of SOM techniques in case study chapters 5, 6 and 7, illustrated how SOM based approaches could be best applied at different levels and scales i.e. sites, regions, national and global, to analyse ecosystem dynamics. In particular, SOMs were seen to be very useful in analysing the effects of urbanisation and human activities on the environment and its biodiveristy. They were also found to be very useful in converting large amounts of desperate, redundant, numeric data into meaningful information, understandable by various professionals and the public as well.

Of the software used in this research Viscovery (Eudaptics software gmbh 1998; Eudaptics software gmbh 2002) was found user friendly and enabled the researcher to fine tune and analyse much complex data sets. It outperformed other software of public access and academic use because of its ability to handle data even with some missing values. However, RICBIS clustering was found to be useful in distinguishing the annual subtidal community changes from those of stimulated by sedimentation in near shore, within the available monitoring data between 1999 and 2001.

A major disadvantage encountered in the use of SOMs as a data mining tool is that domain expertise plays a major role in making the results meaningful. Discretion of data elements depending on the aim of the analysis is vital and without domain expertise any such analysis could not be made meaningful.

SOM limitations for prediction purposes can be overcome. They can be used for predicting interpolated values within the available data without any constraints, but not for extrapolation of values as the number of nodes used in the creation of SOMs cannot be changed. The SOM limitations for extrapolation purposes could be overcome by the following methods:

(i) by adding simulated values in advance, nodes for accommodating the abnormal scenarios in the output display could be made possible or

(ii) by studying the error values, by calculating the deviations for the new values, that were not within the original SOM ranges. The new values are normally added to the nearest possible nodes, with high approximation errors.

Summary

Overall, the research results show that the use of different SOM techniques (cluster analysis, dependent components analysis, decision support systems and time series analysis/ trajectory) for biological and environmental process dynamics modelling without any physical models, to be possible and promising for future use. SOMs were found effective in ecosystem modelling using numeric data. Many constraints faced in conventional ecological analysis methods could be overcome by SOMs i.e. data reduction could be achieved very quickly, without losing much of its important information to better visualise the complex and multidimensional data, revealing hidden patterns. SOMs provide a means to analyse data without knowing its class membership. This is a very useful feature to analyse disparate data sets (ecological, social and economic data), as human knowledge on ecosystem response is limited-not sufficiently comprehensive for the 'rich picture' needed to adequately map all the valuable information that exists for any single ecosystem.

SOM methods can be classified as a useful quantitative analytical approach with a great potential for meaningful, rational interpretation of monitoring data sets. They are capable of revealing more information from the input vectors (i.e. numeric data), compared to any other, currently available conventional data analysis methods. They can be used as an effective and quick approach for exploratory data analysis, prior to analysis using conventional methods such as statistical hypothesis testing methods. SOM methods can serve as a useful tool for initial data analysis of large amounts of numerical data where prior knowledge is limited. They are particularly suitable for analysis of environmental and biomonitoring data, separately and collectively even if integrated with socio-economic data with a systems approach.

This thesis proves that SOM techniques could be applied to modelling of not only the natural systems complexity with spatial variations but also its functioning and dynamics, incorporating temporal variations as well and to overcome the constraints with conventional methods as applied in other disciplines such as industrial process modelling and financial data analysis

References

Aleksander, I. (2000). I compute therefore I am, BBC Science and Technology.

Allinson, N. M. (2001). WSOM' 01, WORKSHOP ON SELF-ORGANIZING MAPS. Proceedings of Advances in Self-Organising Maps, Neural Networks 2002 Special Issue: NEW DEVELOPMENTS IN SELF-ORGANIZING MAPS.

Amari, S.-i. (1995). Foreword. Foundations of Neural Networks, Fuzzy systems and Knowledge Engineering. N. K. Kasabov, The MIT Press: xi.

Ambrose, R. F., R. J. Schmitt, et al. (1996). Predicted and Observed Environmental Impacts - Can We Foretell Ecological Changes. Detecting ecological impacts - Concepts and Applications in Coastal Habitats. R. J. Schmitt and C. W. Osenberg: 345-368.

Anderson, D. and G. McNeill (1992). **ARTIFICIAL NEURAL NETWORKS TECHNOLOGY**. New York, Kaman Sciences Corporation,258 Genesse Street, Utica, New York 13502-4627: 87.

ANNIMAB-1 (2000). Artificial Neural Networks - In Medicine And Biology. Artificial neural networks in medicine and biology, Goteborg, Sweden.

Australian and New Zealand Environment and Conservation Council (November 1992). 7.2.2 Biological indicators. Australian water quality guidelines for fresh and marine waters: 7-7.

Bajic, V. and I. Bajic (1999). Neural Network System for Promoter Recognition. ICONIP/ANZIIS/ANNES'99 International Workshop.

Baldi, P. and S. Brunak (2000). Bioinformatics (The Machine Learning Approach). ACM Press Release, second edition.

Bierbaum, R., S. Carpenter, et al. (2001). Millennium Ecosystem Assessment Sub-Global Component: Purpose, Structure and Protocols - Millennium Ecosystem Assessment, Strengthening Capacity to Manage Ecosystems sustainably for human well-being.

Biological Monitoring (2000). Biological Monitoring. **2000**.

Bowler, P. J. (1992). The problem of perception. the fontana history of The environmental sciences. R. potter. London, Great Britain, HarperCollins Manufacturing, Glasgow: 634.

Brown, C., M. Schreibder, et al. (1999). Information Science and Bioinformatics. ICONIP/ANZIIS/ANNES'99 International Workshop.

Buckeridge, J. S. (1994). "Introducing philosophy and ethics to the engineering curriculum." Transactions of the Institution of Professional Engineers New Zealand **21 Number 1/GEN**.

Buckeridge, J. S. (1999). Stochastic Urban Accretion and Marine Reserves: Complementary or Conflicting options? 19th Annual Meeting of the International Association for Impact Assessment for a New Century, Glasgow, Scotland.

Buckeridge, J. S. (2000). Modern Zoological Education : Impediments and Opportunities. XVIII (New) International Congress on Zoology, Athens, Greece. 28 August – 2 September, 2000.

Buckeridge, J. S. and B. A. Tapp (1999). Ethics and that Ethic called Sustainability. Australasian Environmental Engineering Conference, Auckland, New Zealand.

Buckeridge, J. S. (2001). Ethics, Environment and Culture: Their significance within Engineering Education. International Ecological Engineering Conference ©International Ecological Engineering Society (IEES) and New Zealand Engineers for Social Responsibility (ESR), Lincoln University.

Buckeridge, J. S. and D. Gordon (2000). *Species 2000 New Zealand* : Outcomes of the February Symposium. XVIII (New) International Congress on Zoology, Athens, Greece. 28 August – 2 September, 2000.
Cao, Y., D. P. Larsen, et al. (2002). "Rare species in multivariate analysis for bioassessment: some considerations." BRIDGES: 144-153.
Carbon Dioxide Information Analysis Center Data (2000-2001). A Guide to the Global Environment, cited in World Resources Institute, UNEP, UNDP, and World Bank, World Resources 2000-01.
Ce´re´ghino, R., J. L. Giraudel, et al. (2001). Spatial analysis of stream invertebrates distribution in the Adour-Garonne drainage basin (France), using Kohonen self organizing maps.
Chapman, B. (1999). Indicators for Monitoring the HSNO Act. Proceedings of the Environmental Indicators Symposium, University of Otago.
Clark, J. S., S. R. Carpenter, et al. (2001). "Ecological Forecasts: An Emerging Imperative." The American Association for the Advancement of Science. **Volume 293, Number 5530**(Issue of 27 Jul 2001): 657-660.
Clark, J. S., S. R. Carpenter, et al. (2001). "Ecological Forecasts: An Emerging Imperative - Supplementary Material." The American Association for the Advancement of Science. **Volume 293, Number 5530**(Issue of 27 Jul 2001): 657-660.
Clark, R. B., C. Frid, et al. (1997). Measuring change. Marine Pollution, Oxford University Press**:** 11-22.
Concise Science Dictionary (1991). Oxford Reference.
Couriel, E. D., R. Park, et al. (2000). Project CARE. North Shore, Auckland, North Shore City Council**:** 55.
Deboeck and Kohonen (1998a). Best Practices in Data Mining using Self-Organizing Maps. Visual Explorations in Finance with Self-organizing Maps**:** 203-229.
Deboeck and Kohonen (1998b). Visual Explorations in Finance with Self-organizing Maps.
Deboeck, G. (1998a). Investment Maps of Emerging Markets. Visual Explorations in Finance with Self-organizing Maps. G. Deboeck and T. Kohonen**:** 83-105.
Deboeck, G. (1998b). Software Tools for Self-Organizing Maps. Visual Explorations in Finance with Self-organizing Maps. G. Deboeck and T. Kohonen**:** 188.
Deboeck, G. (2002). Data Mining with Self-Organzing Maps: Part II : A Practical Application. **2002**.
Deboeck, G. and T. Kohonen (1998). Introduction. Visual Explorations in Finance with Self-Organising Maps, Springer**:** xxix.
Deng, D. and N. Kasabov (1999). Evolving Self-Organizing Maps and its Application in Generating A World Macroeconomic Map. ICONIP/ANZIIS/ANNES'99 International Workshops.
Deng, D., I. Koprinska, et al. (1999). RICBIS: New Zealand Repository for Intelligent connectionist-Based Information Systems. ICONIP/ANZIIS/ANNES'99, Dunedin, New Zealand.
Department of Conservation (2002). about DOC, Department of Conservation. **2002**.
Department of Conservation (2002). DOC's Role in Resource Management, Department of Conservation. **2002**.
Department of Oceanography at Texas A & M University (2002). Ecosystem Modelling Group. **2002**.
Environmental Protection Agency Environmental Monitoring and Assessment Program (EMAP).
Estebon, M. D. (1992). Perceptorns: An Associative Learning Network. **2001**.
Eudaptics software gmbh (1998). Viscovery SOMine lite User's Manual. Austria.

Eudaptics software gmbh (2002). on line manual - Viscovery SOMine ver 4. Austria.
European Environment Agency Environmental themes (2002). Indicator: Hazardous substances in blue mussels in the north-east Atlantic. Policy issue: Have reductions in emissions led to a better environment for marine life?
Giraudel, J. L. and S. Lek (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination.
Graedel, T. E., A. Alldredge, et al. (2001). Grand Challenges in Environmental Sciences. Grand Challenges in Environmental Sciences. Washington, National Academy Press: 96.
Green, M. O., M. J. Stroud, et al. (2000). "Long Bay Sedimentation Study: Stage I (Sediment Generation) Stage II (Sediment Fate)." **NIWA Client Report NSC00209**: 98.
Gustavsson, L. (2001). On Ecological Engineering and Sustainable Development in a Swedish Context. International Ecological Engineering Society (IEEE) and Engineers for Social Responsibility (ESR), Lincoln University, New Zealand.
Hammond, A., A. Adriaanse, et al. (1995a). Background and context, Chapter II. ENVIRONMENTAL INDICATORS - A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development, World Resources Institute: 5-9.
Hammond, A., A. Adriaanse, et al. (1995b). Biodiversity: All illustrative approach to the development of composite indicators, Chapter VII. ENVIRONMETAL INDICATORS - A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development, World Resources Institute: 27-28.
Hammond, A., A. Adriaanse, et al. (1995c). Foreword. ENVIRONMETAL INDICATORS - A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development, World Resources Institute: 1-3.
Hammond, A., A. Adriaanse, et al. (1995d). Introduction, Chapter I. ENVIRONMETAL INDICATORS - A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development, World Resources Institute: 1-3.
Hammond, A., A. Adriaanse, et al. (1995e). Organizing Environmental Information: Indicator types, Environmental issues, and a proposed conceptual model to guide indicator development - Chapter IV. ENVIRONMETAL INDICATORS - A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development, World Resources Institute: 11-16.
Hammond, A., A. Adriaanse, et al. (1995f). Pollution/ Emission: Illustrative calculations of indicators and or a composite index for the Netherlands, Chapter V. ENVIRONMETAL INDICATORS - A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development, World Resources Institute: 17-21.
Harris, G. (2002). Integrated assessment and modelling: an essential way of doing science. Science Direct, Environmental Modelling and Software, Science Direct. **Volume 17:** 201-207.
Harris, T. (1993). A Kohonen SOM based machine health monitoring system which enables diagnosis of faults not seen in the training set. International Joint Conference on Neural Networks (JCNN).
Hecht, C. (2001). "Monitoring the shallow water marine environment in the Long Bay-Okura Marine Reserve (Biological Systems)." 46.

Hickey, C. W. (2000). Freshwater Mussel watch: real-time biomonitoring. Water & wastes in New Zealand (2000), NIWA**:** 25-26.
Higgs, J. D. (1999). Monitoring of Sciaphilic organisms in the Long Bay-Okura Marine Reserve, Auckland. P. s. b. J. Buckeridge.
Himberg, J., J. Ahola, et al. (2001). The Self-Organising Map as a Tool in Knowledge Engineering. Pattern Recognition in Soft Computing Paradigm. N. R, World Scientific Publishing**:** 38-65.
Hodju, P. and J. Halme (1999). Neural Networks: What are Artificial Neural Networks? **2001**.
Hollmen, J. and O. Simula (1996). Prediction models and sensitivity analysis of industrial production process parameters by using Self-organizing map. proceedings of the IEEE Nordic Signal Processing Symposium (NORSIG'96).
Johnson (1993). Biological Monitoring.
Jowett, I. (2001). "Artificial intelligence and river biomonitring." Water & Atmosphere **9(1) 2001**: 6.
Kasabov, N. (1999). Evolving Connectionist and Fuzzy Connectionist Systems: Theory and Applications for Adaptive, on-line Intelligent Systems. Studies in Fuzziness and Soft Computing. N. Kasabov and R. Kosma. Germany, Physica-Verlag**:** 111-144.
Kasabov, N. K. (1995). Foundations of Neural Networks, Fuzzy systems and Knowledge Engineering, The MIT Press.
Kasabov, N. K. (2000). "On-line learning reasoning, rule extraction and aggregation in locally optimized evolving fuzzy neural networks." Neurocomputing **41**(Issue 1-4): 25-45.
Kaski and Lagus (1996). Comparing self-organizing maps. C. von der Malsburg, W von Seelen, J C Vorbruggen and B Sendhoff. ICANN 1996, International Conference on Artificial Neural Networks, Bochum, Germany,.
Kirby, A. (2000). "Humans stress ecosystems to the limit." World Resources Institute(19 September 2000).
Kiviluoto (1996). Topology preservation in self-organizing maps. Proceeding of the International conference on Neural Networks (ICNN 1996), IEEE, New York, NY,USA.
Kohonen, T. (1982). Self-organised formation of topographically correct feature maps. *Biological Cybernetics*, 43:59-69.
Kohonen, T. (1997). Justification of Neural Modelling, Berlin; New York : Springer.
Kremer J. N. and S. W. Nixon (1978). A coastal Marine Ecosystem: Simulation and Analysis, Springer-Verlag, Berlin.
Lake Champlain Basin Program (2002). Lake Champlain Zebra Mussel Monitoring Program.
Lash, J. (1995). Foreword. ENVIRONMETAL INDICATORS - A systematic approach to measuring and reporting on environmental policy performance in the context of sustainable development. A. Hammond, A. Adriaanse, E. Rodenburg, D. Bryant and R. Woodward, World Resources Institute.
Longhurst, A. R. Vertical migration In. D. H. Cushing and J. J. Walsh. Ed. (1976). The Ecology of the Seas, Blackwell Scientific Publications, Oxford.
Mackey, B. (2003). Department of Geography, The Australian National University. **2003**.
Mann, K. H. (1982). Models and Management. Ecology of Coastal Waters - A Systems Approach, Oxford : Blackwell Scientific. **Studies in Ecology Volume 8:** 260-285.

Mapstone, B. D. (1996). Scalable Decision Criteria for Environmental Impact Assessment. Detecting ecological impacts - Concepts and Applications in Coastal Habitats. R. J. Schmitt and C. W. Osenberg**:** 67-80.
Matsumato, G. (1999). The Brain can Acquire its Algorithm in a Self-Organized Fashion. ICONIP/ANZIIS/ANNES'99 International Workshop 22-23 November 1999 p116.
Meyer, S. (2000). Monitoring the shallow water marine environment in the Long Bay-Okura Marine Reserve (Biological Systems)**:** 67.
Ministry for the Environment (2002a). About the Ministry, Ministry for the Environment. **2002**.
Ministry for the Environment (2002b). Environmental Performance Indicators, Ministry for the Environment. **2002**.
Ministry for the Environment (2002c). Environmental Performance Indicators programme, Ministry for the Environment. **2002**.
Ministry for the Environment (2002d). Marine Indicators, Ministry for the Environment. **2002**.
Mitchell, T. M. (1999). Machine learning and Data Mining. Communications of the ACM.
National Center for Environmental Research (NCER) Office of Research and Development (ORD) - US Environmental Protection Agency (EPA) (2000). Environmental Indicators in the Estuarine Environment Research Program - 2000 Science To Achieve Results (STAR) Program.
National Center for Environmental Research (NCER) Office of Research and Development attached to United States EPA (2001). ORD/NCER STAR Grants, Ecological Assessment and Indicators Research, April 2001 and ORD/NCER STAR GRANTS, Urban Sprawl Research January 2001.
National River Health Program (2002). Ausrivas.
neuscience (2000). Neural networks - Unsupervised neural networks.
Norris, R. H. (1999). Environmental Indicators: Recent Developments in Measurement and Application for Assessing Freshwaters. Proceedings of the Environmental Indicators Symposium, University of Otago.
North Shore City Council - 0800SAFESWIM "What does the beach water quality programme test for?"
North Shore City Council - Project care (2002). Project Care.
North Shore City Council - Wai Care (2002). Wai Care - Together fro healthy waterways.
O'Connor, M. and W. J. Walley (2001). A River Pollution Diagnostic System for England and Wales, School of Computing, Staffordshire University. **2001**.
O'Connor, T. (2002). "Chemical Contaminants in Oysters and Mussels", State of the Coast Report. Silver Spring, MD: NOAA. National Oceanic and Atmospheric Administration (NOAA). 1998 (on-line),. **2002**.
On-line dictionary (2002). "indicator."
On-line Medical Dictionary (2002).
Osenberg, C. W. and R. J. Schmitt, Eds. (1996). Detecting ecological impact caused by human activities. Detecting Ecological Impacts - Concepts and Applications in Coastal Habitats. London, Academic Press, Inc.
Palmer, M. (2002a). A GLOSSARY OF ORDINATION-RELATED TERMS.
Palmer, M. W. (2002b). Ordination Methods - an overview.
Parker, P., R. Letcher, et al. (March 2001). "Progress in integrated assessment and modelling." **2002**.
Patte, H. H. Hierarchy Theory In. Ed. (1973). George Braziller, New York.

Potter, J. H. Management of the tidal Thames In. A. L. H. Gameson. Ed. (1973). Mathematical and Hydraulic Modelling of Estuarine Pollution, HMSO, London.

Pregibon, D. and W. DuMouchel (2001). What Is Data Mining? (An Why Should Statistician Care?). 53rd Session of the International Statistical Institute, Seoul, Korea.

Raimondi, P. T. and D. C. Reed (1996). Determining the spatial extent of ecological impacts caused by local anthropogenic disturbances in coastal marine habitats. Detecting ecological impacts - Concepts and Applications in Coastal Habitats. R. J. Schmitt and C. W. Osenberg: 179-198.

Ravetz, J. (2000). "Integrated assessment for sustainability appraisal in cities and regions." Environmental Impact Assessment Review **Volume 20**(Issue 1): 31-64.

Reid, W. V. (2000). The Delicate Balance: Environment, Economics, Development - Ecosystem Data to Guide Hard Choices. ISSUES IN SCIENCE AND TECHNOLOGY online.

Sallis, P., L. Hill, G. Janee, K. Lovette and C. Masi (1998). A methodology for profiling users of large interactive systems incorporating Neural Networks for Data mining Techniques.

Sastry, A. N. and D. C. Miller (1980). Application of biochemical and physiological responses to water quality monitoring. Biological Monitoring of Marine Pollutants. F. J. Vernberg, A. Calabrese, F. P. Thurberg and W. B. Vernberg. New York, Toronto, London, Sydney, San Francisco, Academic press, A subsidiary of Harcourt Brace Jovanovich Publishers: 265-294.

Scharader, D. (2000). Monitoring the shallow water marine environment in the Long Bay-Okura Marine Reserve (Physical Systems): 66.

Schellnhuber, H. J. (1999). 'Earth system' analysis and the second Copernican revolution, Nature 402, C19 - C23 (1999) © Macmillan Publishers Ltd.

Schmitt, R. J., C. W. Osenberg, W. J. Douros and J. Chesson (1996). The art and science of administrative environmental impact assessment. Detecting ecological impacts. Eds. R. J. Schmitt and C. W. Osenberg, Academic Press: 281-315.

Serrano-Cinca, C. (1998). Let Financial Data Speak for Themselves. Visual Explorations in Finance with Self-organizing Maps. Eds. Deboeck and Kohonen: 3-23.

Shalvi, D. and N. Declaris (2000). An Unsupervised Neural Network Approach to Medical Data Mining Techniques. http://www.ee.umd.edu/medlab/papers/dcsThShort/thpaper1.html

Simula, O., J. Vesanto, et al. (1999). Analysis and modeling of complex systems using the self-organizing map. Studies in Fuzziness and Soft Computing Neuro-Fuzzy Techniques for Intelligent Information Systems. N. Kasabov and R. Kosma. Germany, Physica-Verlag: 3-22.

Simula, O., J. Vesanto, et al. (1998). Analysis of Industrial Systems Using the Self-Organising Map. Proceedings of the International Conference on Knowledge-based Intelligent Systems KES '98, Adelaide, Australia.

Smith, R. W., B. B. Bernstein, et al. (1988). Community-Environmental Relationships in Benthos: Applications of Multivariate Analytical Techniques. Marine Organisms as Indicators. D. F. Soule and G. S. Kleppel. New York, Springer-Verlag: 247-324.

Snowdon, R. (1999). Monitoring of Sciaphilic Organisms in the Long Bay-Okura Marine Reserve: 26.

Song, Q. and D. Hegg (1999). A methodology for Biological Data Modeling, Based on Evolving Connectionist System. Proceedings of the ICONIP/ANZIIS/ANNES'99 International Workshop.

Soule, D. F. (1988). Marine Organisms as Indicators: Reality of Wishful thinking. Marine Organisms as Indicators. D. F. Soule and G. S. Kleppel. New York, Springer-Verlag**:** 1-10.

Soule, D. F. and G. S. Kleppel (1988a). Marine Organisms as Indicators. New York, Springer-Verlag.

Soule, D. F. and G. S. Kleppel (1988b). Preface. Marine Organisms as Indicators. D. F. Soule and G. S. Kleppel. New York, Springer-Verlag**:** v-vii.

Stewart-Oaten, A. (1996a). Goals in Environmental Monitoring. Detecting Ecological Impacts - Concepts and Applications in Coastal Habitats. C. W. Osenberg and R. J. Schmitt. London, Academic Press, Inc**:** 17-27.

Stewart-Oaten, A. (1996b). Problems in the analysis of environmental monitoring data. Detecting ecological impacts - Concepts and Applications in Coastal Habitats. R. J. Schmitt and C. W. Osenberg**:** 109-131.

Thrush, S. F., R. D. Pridmore, et al. (1995). Impacts on Soft-sediment Macro fauna. Detecting Ecological Impacts - Concepts and Applications in Coastal Habitats. R. J. Schmitt and C. W. Osenberg.

Tryba, V., S. Metzen, et al. (1989). Designing basic integrated circuits by self-organising feature maps. Neuro-Nimes'89 - International Workshop. Neural Networks and their Applications.

Underwood, A. J. (1996). On beyond BACI - Sampling designs that might reliably detect environmental disturbances. Detecting ecological impacts - Concepts and Applications in Coastal Habitats. R. J. Schmitt and C. W. Osenberg**:** 151-175.

Vant, B. (1999). Analysing and Reporting SoE Monitoring Results for Rivers in the Waikato Region. Proceedings of the Environmental Indicators Symposium, University of Otago.

Vesanto, J. and E. Alhoniemi (2000). Clustering of the Self-Organising Map. IEEE Transactions on Neural Networks. **Volume 11:** 586-600.

Vitousek, P. M., J. D. Aber, et al. (1997). "Human alteration of the global nitrogen cycle: sources and consequences." Science Direct, Ecological Applications **Volume 7, Issue 3**: 737-750.

Walker, J., R. Babcock, et al. (2000). The Long Bay Monitoring Program Sampling Report - July 1999 / June 2000.

Walley, W. J., R. W. Martin, et al. (2001). Self-Organising Maps for the Classification of River Quality from Biological and Environmental Data, School of Computing, Staffordshire University. **2001**.

Walley, W. J. and M. A. O'Connor (2001). "Unsupervised pattern recognition for the interpretation of ecological data." Elesevier **Ecological modelling 146(2001) 219-230**: 219-230.

Wasserman, P. (1989). Neural Computing - Theory and Practice.

Watts, M. (1999). Evolving Connectionist Systems for Biochemical Applications. ICONIP/ ANZIIS/ ANNES '99.

Weiss, P. The living system: determination stratified In. A. Koestler and J. K. Symthies. Ed. (1969). Beyond Reductionism, Hutchinson, London.

What's Biomonitoring? (2002). What's Biomonitoring?, Merriam-Webster. **2002**.

Wilcock, R. J. and M. J. Stroud (2000). Year 2000 Summary Report, Baseline Water Quality - Stream, Lake, and Saline Waters, Auckland Regional Council**:** 219.

World Bank (2002a). 1.1 Size of the economy.

World Bank (2002b). 1.6 Key indicators for other economies.

World Bank (2002c). 5.1 Private sector development.

World Bank (2002d). Rural environment and land use.

World Bank Report (2001a). 3.1 Rural environment and land use. World Development Indicators, World Bank. **2001**.
World Bank Report (2001b). 3.4 Deforestation and biodiversity. World Development Indicators, World Bank. **2001**.
World Bank Report (2002). 4.1 Growth of output.
World Resources Institute (2000). WHAT IS THE PROBLEM? **2001**.

Glossary

Artificial neural networks (ANNs): Biologically inspired networks of neurons used for information processing to incorporate heuristics into conventional algorithmic computational processing.

Biogeocenosis: "a combination on a specific area of the Earth's surface of atmosphere, mineral strata, soil, vegetation, animal and microbial life, water - possessing its own specific type of interactions of these components and interchange of their matter and energy among themselves and other natural phenomena…"(Mackey 2003)

Biological responses: A chain of changes eventually causing an action or movement in a living organism.

Complex: The many interrelating components and mechanisms those make the final outcome of the system process, difficult to understand and to predict upon system behaviour under differing conditions.

Correlation: A causal, complementary, parallel, or reciprocal relationship, especially a structural, functional, or qualitative correspondence between two comparable entities from http://www.dictionary.com

Data mining: A term exclusively used to describe the extraction of knowledge stage in the whole knowledge discovery process. Data mining is also referred as exploratory data analysis.

Dispersion models: Dispersion models for prediction purposes are useful in the diagnosis of harmful pollutant depositions under certain conditions. They cannot be used to predict the pollutant effects as they do not utilise biological information that can be interpreted as susceptibility.

Ecosystem: "A biological community *termed as the biological system in the research* and the physical environment, *which in turn termed as the physical/ environmental system* associated with it". (Concise Science Dictionary 1991) (*Attention of italics for the clarification of this research)*.

Exploratory data analysis: same as data mining.

Higher plants: The native vascular plant species are referred to as higher plants (World Bank Report 2001).

Knowledge Engineering: A term used to refer the academic research in developing models, methods and basic technologies for representing and processing knowledge and building intelligent knowledge-based systems

Knowledge discovery: A term used to describe the whole process of the extraction of knowledge (knowledge means relationships and patters between data elements). It has recently become a multi disciplinary approach, involving machine learning, database technology, expert systems and data visualisation, all possibly contributing to the extraction of new knowledge from raw data.

Ordination techniques: Operations on community data matrix to visualise the arrangement of species and/or samples along gradients, considering it as a synonym for multivariate gradient analysis (Palmer 2002; Palmer 2002)

Projection methods: A data visualisation method for representing the input data in a chosen low dimensional space, where certain properties of the structure of original data are preserved as faithfully as possible.

Rotations: "(Mathematics) A transformation of a coordinate system in which the new axes have a specified angular displacement from their original position while the origin remains fixed" from http://www.dictionary.com

Self-organising map (SOM): SOM is a connectionist paradigm of feed forward artificial neural networks with an unsupervised algorithmic training. They are capable of projecting multidimensional input vectors on a low dimensional, topology preserving output display of self-organising neurons.

Species: A taxonomic category, subordinate to a genus (or subgenus) and superior to a subspecies or variety. It is composed of individuals possessing common characters distinguishing them from other categories of individuals of the same taxonomic level. In taxonomic nomenclature, species are designated by the genus name followed by a Latin or Latinised adjective or noun. (On-line Medical Dictionary 2002)

Translation: "(Kinematics) Motion in which all the points of the moving body have at any instant the same velocity and direction of motion; -- opposed to rotation" from http://www.dictionary.com

Trophic levels: The various levels used to define the different organisms in an ecosystem based on their positions in food chains, by which nutrients and energy move round the ecosystem in loops and cycles.

Appendix 1

What are neural networks used for?

Their applications are almost limitless but they fall into several main categories.

Classification

Business

· Credit rating and risk assessment, Insurance risk evaluation, Fraud detection

· Insider dealing detection, Marketing analysis, Mail shot profiling

· Signature verification, Inventory control

Engineering

· Machinery defect diagnosis, Signal processing, Character recognition

· Process supervision, Process fault analysis, Speech recognition

· Machine vision, Speech recognition, Radar signal classification

Security

· Face recognition, Speaker verification, Fingerprint analysis

Medicine

· General diagnosis, Detection of heart defects

Science

· Recognising genes, Botanical classification, Bacteria identification

· Modelling

Business

· Prediction of share and commodity prices,. Prediction of economic indicators

Engineering

· Transducer linearisation, Colour discrimination, Robot control and navigation

· Process control, Aircraft landing control, Car active suspension control

· Printed Circuit auto routing, Integrated circuit layout, Image compression

Science

· Prediction of the performance of drugs from the molecular structure.

· Weather prediction, Sunspot prediction

Medicine

· Medical imaging and image processing

Forecasting

· Future Sales, Production Requirements, Market Performance

· Economic Indicators, Energy Requirements, Time Based Variables.

Novelty Detection

· Fault Monitoring, Performance Monitoring, Fraud Detection,

· Detecting Rare Features

· Different Cases.

Web address: http://www.ncs.co.uk/nn_intro.ht

Appendix 2

Matrix of environmental indicators adopted by WRI

Issues	Pressure	State	Response
Climate change	(GHG) emissions	Concentrations	Energy intensity; env measures
Ozone depletion	(Halocarbon) emissions; production	(Chlorine) concentrations; O_3 column	Protocol sign.;CFC recovery; Fund contribution
Eutrophication	(N, P water, soil) emissions	(N, P, BOD) concentrations	Treatment connection; investments/ costs
Acidification	(SO_2, NO_2, NH_3) emissions	Deposition; concentrations	Investments; sign agreements
Toxic contamination	(POC, heavy meal) emissions	(POC, heavy meal) concentrations	Recovery hazardous waste; investment/ costs
Urban environment quality	(VOC, NO_x, SO_x) emissions	(VOC, NO_x, SO_x) concentrations	Expenditures; transport policy
Biodiversity	Land conversion; land fragmentation	Species abundance composition to virgin area	Protected areas
Waste	Waster generation municipal, industrial, agricultural	Soil/ groundwater quality	Collection rate; recycling investments/ cost
Water resources	Demand/ use intensity residential/ industrial/ agricultural	Demand/ supply ratio	Expenditures; water pricing; savings policy
Forest resources	Use intensity	Area degr. Forest; use/sustain. Growth ratio	Protected area forest, sustain. Logging
Fish resources	Fish catches	Sustainable stocks	Quotas
Soil degradation	Land use changes	Top soil quality	Rehabilitation/ protection
Oceans/ coastal zones	Emissions; oil spills; depositions	Water quality	Coastal zone management; Ocean protection
Environmental index	Pressure index	State index	Response index

Matrix of environmental indictors, Source: Environment Indicators. (Hammond, Adriaanse et al. 1995).

Appendix 3

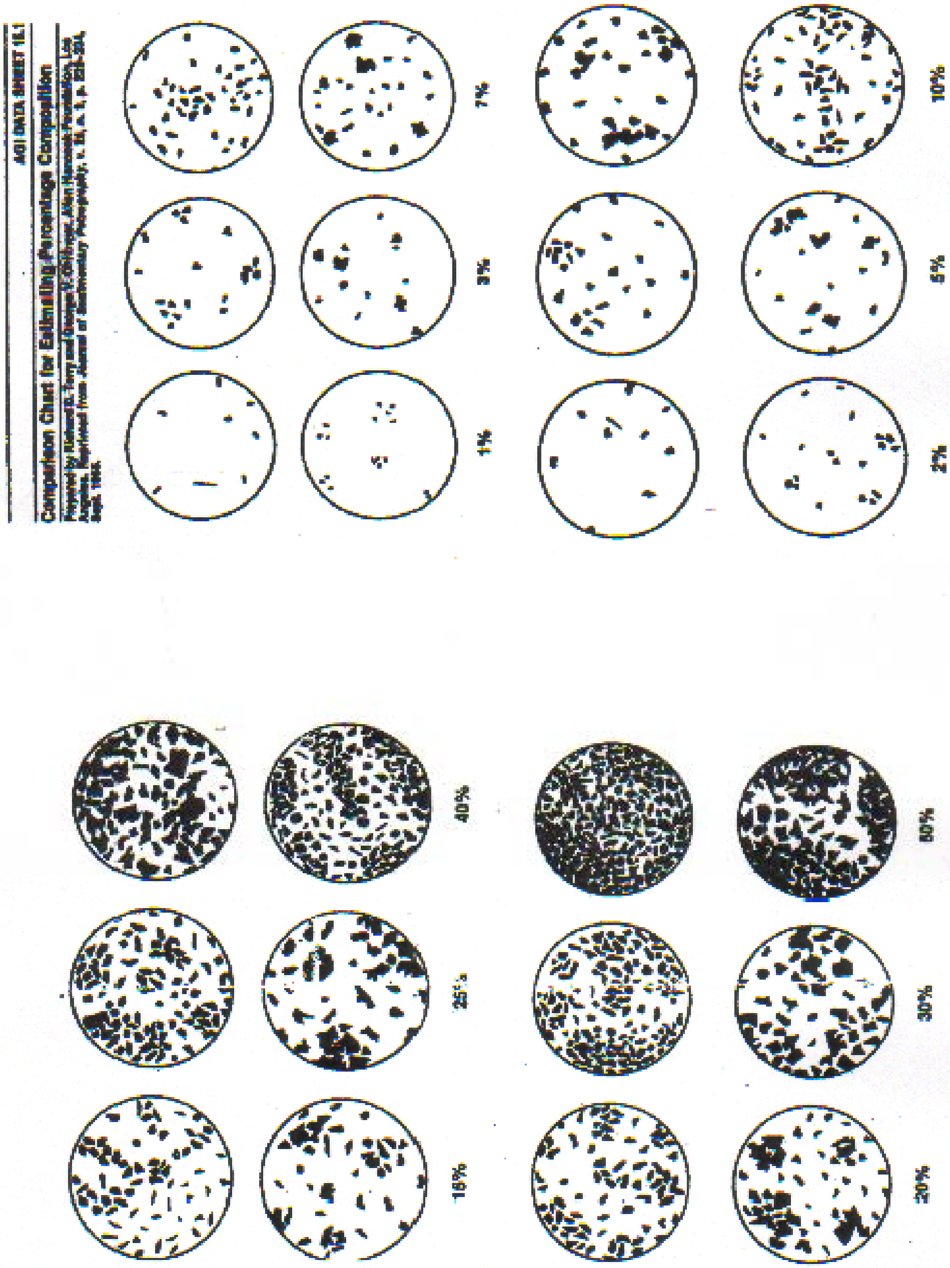
Matrix of environmental indicators adopted by the World Bank

Figure 4. Matrix of Environmental Indicators

Issues	Pressure	State	Response
I. Source Indicators			
1. Agriculture	Value Added/Gross Output	Cropland as % of wealth	Rural/Urban Terms of Trade
a. Land Quality	Human-Induced Soil Degrad.	Climatic Classes & Soil constraints	...
b. Other	...	...	...
2. Forest	Land Use Changes, Inputs for EDP	Area, volumes, distribution; value of forest	In/Output ratio, main users; recyc. rates
3. Marine Resources	Contaminants, Demand for Fish as Food	Stock of Marine Species	% Coverage of Int'l Protocols/Conv.
4. Water	Intensity of Use	Accessibility to Pop. (weighted % of total)	Water efficiency measures
5. Subsoil Assets	Extraction Rate(s)	Subsoil assets % wealth	Material balances/NNP
a. Fossil Fuels	Extraction Rate(s)	Proven Reserves	Reverse Energy Subsidies
b. Metals & Minerals	Extraction Rate(s)	Proven Reserves	In/Output ratio, main users; recyc. rates
II. Sink or Pollution Indicators			
1. Climate Change	...	...	...
a. Greenhouse Gases	Emissions of CO^2	Atmosph. Concentr. of Greenhouse Gases	Energy Efficiency of NNP
b. Stratospheric Ozone	Apparent Consumption of CFCs	Atmosph. Concentr. of CFCs	% Coverage of Int'l Protocols/Conv.
2. Acidification	Emissions of SO_x, NO_x	Concentr. of pH, SO_x, NO_x in precipitation	Expenditures on Pollution Abatement
3. Eutrophication	Use of Phosphates(P), Nitrates(N)	Biological Oxygen Demand, P, N in rivers	% Pop. w/waste treatment
4. Toxification	Generation of hazardous waste/load	Concentr. of lead, cadmium, etc. in rivers	% Petrol unleaded
III. Life Support Indicators			
1. Biodiversity	Land Use Changes	Habitat/NR	Protected Areas as % Threatened
2. Oceans	Threatened, Extinct species % total	...	...
3. Special Lands(e.g.,wetland)	...	...	...
IV. Human Impact Indicators			
1. Health	Burden of Disease (DALYs/persons)	Life Expectancy at birth	% NNP spent on Health, vaccination
a. Water Quality	...	Dissolved Oxygen, faecal coliform	Access to safe water
b. Air Quality	Energy Demand	Concentr. of particulates, SO_2, etc.	...
c. Occupat'l Exposures,etc.	...	...	...
2. Food Security & Quality	...	...	...
3. Housing/Urban	Population Density (persons/km^2)	...	% NNP spent on Housing
4. Waste	Generation of industrial, municipal waste	Accumulation to date	Exp. on collect. & treatmt., recyc. rates
5. Natural Disaster	...	...	...

Source: The World Bank

Appendix 4


Comparison Chart for Estimating Percentage Composition
1%
3%
7%
2%
5%
10%
15%
25%
40%
20%
30%
50%

Appendix 5

Data mining with self-organising maps: Main steps

Electronically reformatted version

DATA MINING WITH SELF-ORGANIZING MAPS: PART I : MAIN STEPS **By Guido Deboeck, Ph.D***

Many articles and courses outline principles and the details of algorithms that can be used for data mining e.g. neural networks, genetic algorithms, fuzzy logic. They emphasis technique rather than practice This article summarizes "best practices" in data mining, clustering and visualization of large multi-dimensional data sets in finance, economics or marketing. These best practices are based on the lessons learned from many applications presented in *Visual Explorations in Finance with self-organizing maps* (Springer-Verlag,1998), lessons extracted from many papers presented at neural net conferences and the expertise from people who have several years of hands-on experience in applying neural networks in finance and economics. The process described here for data analysis, clustering, visualization, and evaluation can be applied to many applications. As an illustration we will use self-organizing maps(SOM), which is a technique based on unsupervised neural networks that uses competitive learning in order to create a reduced two dimensional representation of a large multi-dimensional data set. Part II of this article will apply the steps outlined here to the problem of assessing country credit risks based on economic, financial and stock market data.

Main steps

The financial, economic and marketing applications of self-organizing maps outlined in *Visual Explorations in Finance with self-organizing maps* show that there are no specific procedures or optimal methods for applying SOM that are valid for all applications. Similar to the design of other neural network models, to create a self-organizing map, is still an art more than a science. Like for many other approaches the "engineering" aspects of SOM , e.g. for selection of a SOM array; the scaling of the input variables; the initialization of the algorithm; the selection of the neighborhood size, the learning rate; the interpretation and color coding of the map, are easy to obtain. However these are not sufficient for the entire process of data analysis. Hence, we present here a series of steps that details the process of data mining rather than the application of any specific algorithm or technique.

Box 1: Main steps in clustering and visualization of data
Step 1. Define the purpose of the analysis;
Step 2. Select the data source and -quality;
Step 3. Select the data scope and variables;
Step 4. Decide how each of the variables will be preprocessed;
Step 5. Use relevant sample data that are representative for your system;
Step 6. Select the clustering and visualization method(s);
consider the use of hybrid methods;
Step 7. Determine parameters : in case of SOM the desired display size, map ratio,
the required degree of detail;
Step 8. Tune the output or map for optimal clustering and visualization;

Step 9. Interpret the results, check the values of individual nodes and clusters;
Step 10.Define or paste appropriate map labels;
Step 11.Produce summary results that highlight the differences between clusters;
Step 12.Document and evaluate the results.

1 Define the purpose of the analysis

Without proper definition of the goals and objectives for the design of a neural network model, supervised or unsupervised, it will be difficult to assess the effectiveness of the outcome. Neural net models can be designed for many different objectives. As shown in several applications the main objectives for design of a neural network or self-organizing map in our case can be for

(i) classification, clustering, and/or data reduction;
(ii) visualization of the data;
(iii) decision-support;
(iv) hypothesis testing;
(v) monitoring system performance;
(vi) lookup of (missing) values;
(vii) forecasting.

If clustering and visualization are the main objectives, various alternative visualization and clustering methods should be considered. Several traditional statistical methods for clustering and data visualization exist. Combining traditional statistical methods with neural network techniques like SOM may generate better results than the use of one technique by itself. It may also be useful to determine a priori how much data reduction is desired. If decision-support is the main objective then it is essential to define precisely what decisions need to be supported, what is the scope of these decisions, and what is their time frame. For example, predicting the direction of a market is quite different from predicting future price levels.
If hypothesis testing is the main objective one needs to define a priori what hypotheses will be tested and what will be the standard for acceptance or rejection. For example, when applying neural networks to banking data the hypothesis may be is there a significant difference between the various banking institutions in various markets around the world. If monitoring systems performance is the objective the goals of the monitoring process need to be defined e.g. monitoring for quality purposes, fault detection, standard compliance. If forecasting is the objective, it is important to spell out what is the forecasting window, the desired accuracy, how will the performance be evaluated. For example, for what time window are the predictions for, should the predictions be accurate in terms of the level or just the direction; will the price predictions be evaluated in terms of the percentage of correct predictions or in terms of the cumulative profit or loss achieved by implementing the predictions in a given period.

2. Select the data source and data quality

Thee importance of using high quality data can not be underestimated. It is important that the data comes from reputable sources. Good sources of high quality financial and economic data are - national and international agencies (e.g. government statistical offices, the United Nations, specialized agencies of the UN, World Bank, IMF, and the like), - well-established information services (e.g. Bloomberg, Reuters, Telerate, Knight-Ridder, Standard & Poor), or - data base providers (e.g. Value Line, Morningstar, DRI, Moody's, American OnLine, CompuServe etc.) and many others.

Data that is freely available on the web may or may not be of high quality. It is therefore advisable to be skeptical about what is freely offered on the web.

3. Select the data scope and variables

To define the data scope in relation to the objectives of the study is important for any kind of analysis. Neural network techniques based on learning techniques and competitive learning in particular may cause laziness and or attempts to "through in the kitchen sink", i.e. use all the available data on a particular subject rather than a selective set relevant to the objectives of the study. Furthermore it is important to use domain expertise, or to collaborate with those that have such expertise. For example, when studying structures in investment data, credit risk data, poverty data, proper analyses cannot be done without domain knowledge. One should also be careful in the selection of the appropriate indicators. Once the data scope has been properly defined, some important tips to remember in selecting variables to be included in the analysis are

• do not get wed to your data, learn to discriminate, discard and delete
• select only those variables that are meaningful in relation to the objectives
• select the variables that are most likely to influence the results
• consider to use combinations of variables, such as ratios, time-invariants etc.
• use domain expertise or involve in the analysis people who have domain expertise
• do not assume that the data is normally distributed
• adding of one or more irrelevant variables can dramatically interfere with the cluster recovery
• omission of one or more important variables may also affect the results.

4. Decide how each of the variables will be preprocessed

Pre-processing of data is important particularly in neural network design. When preprocessing data for clustering the pre-processing may specifically involve *data standardization*, -*transformations*, and *setting of priorities.* The main reason for *data standardization* is to scale all data to the same level. Often the data range of each variable varies from column to column. If no preprocessing is applied this may influence the clustering and the ultimate shape of the output map. There are many ways in which data can be standardized. The most are to standardize all data based on the standard deviation. Other methods are to standardize on the basis of the range e.g. $z= [x - \min(x)] / [\max(x)-\min(x)]$. Some studies have shown that standardizing the data based on the range can be superior in certain cases, in particular if the variance is much smaller than the range. *Data transformations* can be applied to any or all variables to influence the importance and/or influence of each variable on the final outcome. Transformations may also be used to „equalize“ the histograms. Two typical data transformations are *logarithmic* and *sigmoid*. The former squeezes the scale for large values, the latter takes care of outliners. Applying data transformations redefines the internal representation of each variable and should be applied with caution. *Setting the priority* of a variable to a value greater or lower than one has the same effect as changing the standardization explicitly. By giving a priority to a variable you can provide a weighting of the variables in the mapping process. For example, if in the selection of investment managers, the 'launching date' of a mutual fund is considered less important, then this variable can be given a low priority.

5. Use relevant sample data representative for your system
Training a neural network on a set of sample data will yield better results when using random initial input vectors. By selecting representative input vectors for the training of a SOM map one reduces noise and can obtain a sharper map. This map then can be used for testing on all the remainder input data sets. Furthermore, depending on the applications, the use of input vectors that represent outliners may be of crucial importance for training a SOM. Outliners provide contrasts and can sharpen the differences between clusters. However, this can be to the cost of sensibility for the other parts of the map. If outliners are not representative, they should of course be eliminated.

6. Select the clustering and visualization method(s); consider the use of hybrid methods
In this article we focus on SOM however combining SOM with other methods can yield better results. For example, a hybrid system of SOM and genetic algorithms can improve the performance of trading models; overlaying the results from SOM on top of principal component analysis can improve visualization; combining SOM with a Geographic Information System can improve interpretation. In financial, economic and marketing applications combining SOM with other statistical methods is common practice. A SOM map by itself provides a topological representation of the data which needs to be translated in operational or actionable outcomes. Financial analysts, economists and certainly marketing professionals will want to know what are the main features of the clusters, how they differ from each other, and how to use the newly found structures or patterns for forecasting or decision-support. Thus a SOM map by itself can not be a final outcome.

7. Determine the desired display size, shape, and the required degree of detail
Bigger maps produce more detail; input vectors are spread out on a larger number of nodes. Smaller maps can contain bigger clusters or more input vectors can cluster on a smaller set of nodes. Which is better ? This will depend on the application and the usage of the map. Smaller is not necessarily better. More detail may be desirable in some cases. In general, smaller numbers of nodes stand for higher generalization, and this may also be useful if the data contains much noise. Higher numbers of nodes normally yield nicer map images but must not be over-interpreted in later use. The key in determining the size of the map will be how the map will be used. A simple analogue would be to compare the use of a country atlas with that of highway or street maps. If using SOM for lookup of information a larger SOM map may be more esirable; however when using SOM to select investment opportunities or investment managers a smaller map that clusters managers and investment opportunities in five to seven categories may be more optimal.

8. Tune the output or map for optimal clustering and visualization
Once a SOM has been trained you can inspect the map by looking at the number of nodes that contain input vectors, the mean values of the nodes and clusters, the number of clusters that were created, and the number of matching input vectors for each cluster. Fine-tuning a map can be done by increasing or reducing the cluster threshold and/or the minimum cluster size. A larger cluster threshold or higher minimum cluster size will reduce the number of clusters, it will increase the coarse-ness of the clustering. Lowering the cluster threshold will show more details of the map.

9. Interpret the results, check values of individual nodes and clusters,
Once a topological representation of the data is created, it is important to check the validity of the map. This can be done in several ways. Again domain expertise will be a key ingredient. A simple check may consist in printing a list of the input vectors sorted by node or cluster of the map. Another one may be to calculate some simple summary statistics on each cluster. Depending on which software tool is used the mean values of the clusters may be even displayed on the screen. In this case the user can interactively check each cluster and judge whether the summary values make sense. Comparisons of values among nodes and clusters will then allow the user to decide on how more detailed the map needs to be, which data transformation could be needed, how to fine-tune the priority of some components, or what the generalization capability of the map eventually may be. In other words, an interactive capability to check the values for nodes and clusters is important in order to allow the process to be dynamic and to incorporate the user's domain expertise and knowledge about the data.

10. Define or pasting appropriate map labels
The importance and difficulty of defining appropriate labels has been discussed in many articles. When using SOM to classify countries, states or cities, or when using SOM to cluster investment opportunities, companies, or banks the labels to be used are obvious: each input vector can be extended with an appropriate or abbreviate name of the country, state, city, security, company or bank it represents. When using SOM for process control labeling may be restricted to a few input vectors, picking on those that represent failures, or idle states. When using SOM to classify wines or whiskeys, multiple labels may be necessary to identify the country, region, vineyard or distillery. In sum, flexibility in automatic labeling of nodes or clusters from the input data vectors is of crucial importance. This automatic labeling capability is of particular importance for finance, economic and marketing applications.

11. Produce a summary of the map results that highlight the differences between clusters
The production of summary statistics may be automatic or manual depending on which software tool is used for SOM. Newer software packages have built-in capabilities for automatic production of summary statistics. This has advantages over software tools that do not provide any post-processing capability. In finance, economics and marketing, post-processing of SOM results, information extraction of value added, and how SOM results can be used is very important. A post-processing capability that allows to create summary statistics for each node and each cluster showing at the minimum the mean, standard deviation, minimum, maximum value, and the sum of the input vectors is a great advantage.

12. Document and evaluate results
For SOM to be useful in finance, economics and marketing, it is essential to demonstrate its value added. "Look Mom what a nice picture I made" will just not fly in boardrooms, management meetings or strategic marketing sessions. When we applied supervised neural networks to create financial models, we measured the value added by measuring the performance (return), the risks, and the portfolio turnover of the models; we compared results with those of benchmarks (e.g. performance of human traders, or models based on more traditional methods). Return is usually compared to risk to obtain the risk-adjusted return. This risk adjusted return can be compared to a benchmark (e.g. the risk-adjusted return of the Standard and Poor

500). By adding portfolio turnover one can take into account the costs of trading. The higher the turnover the higher the transaction costs. The tradeoff between risk-adjusted return and costs then provides a measure of effectiveness of trading models. The quality of an unsupervised neural net model can and should be measured on the basis of (i) the number of clusters; ii) the quality of clustering; (iii) the stability of clustering (as measured by the similarity or lack of similarity obtained by varying the testing data set). If we assess unsupervised neural net models in this way we are likely to find that there are many tradeoffs between quantity, quality, and stability of the clusters. It is then be up to the user to determine what is the best combination in the light of the objectives of the study. Some applications may demand maximum data reduction (minimum number of clusters), and can live with coarse map quality and low stability; other applications may demand refined maps (i.e. sharp differences between clusters), good stability, but do not require a lot of data reduction. For example, in macro-economic analyses, analyses of world development indicators, environmental conditions, analyses of global poverty and the like, maximum data reduction may be most desired because the maps would be mainly used for policy formulation and macro decision-support. In other applications such as mapping opportunities for options and future trading, fund manager selection, client segmentation, product differentiation, or market analyses, much finer differentiation between clusters may be desired.

There is a vast domain of research and innovation to be done in this area, in particular in developing standards and a standard method for measuring the value added of clustering using self-organizing maps in financial, economic and marketing applications. * Guido Deboeck is an expert on advanced technology and its applications for financial engineering and management In the past twenty years he has been a leading innovator and advisor on technology to the World Bank in Washington. He holds an MA and Ph.D. degrees in Economics from Clark University. E-mail:gdeboeck@erols.com

References

Guido Deboeck & Teuvo Kohonen : Visual Explorations in Finance with self organizing maps, Springer-Verlag, 1998, 250 pp.
Guido Deboeck : Trading on the Edge: Neural, Genetic and Fuzzy Systems for Chaotic Financial Markets, John Wiley and Sons, New York, April 1994, 377 pp.
Teuvo Kohonen: Self-Organizing Map, Springer Verlag. 2nd edition, 1997, 426 pp.

Appendix 6

SOM maps of ARC's saline water quality data

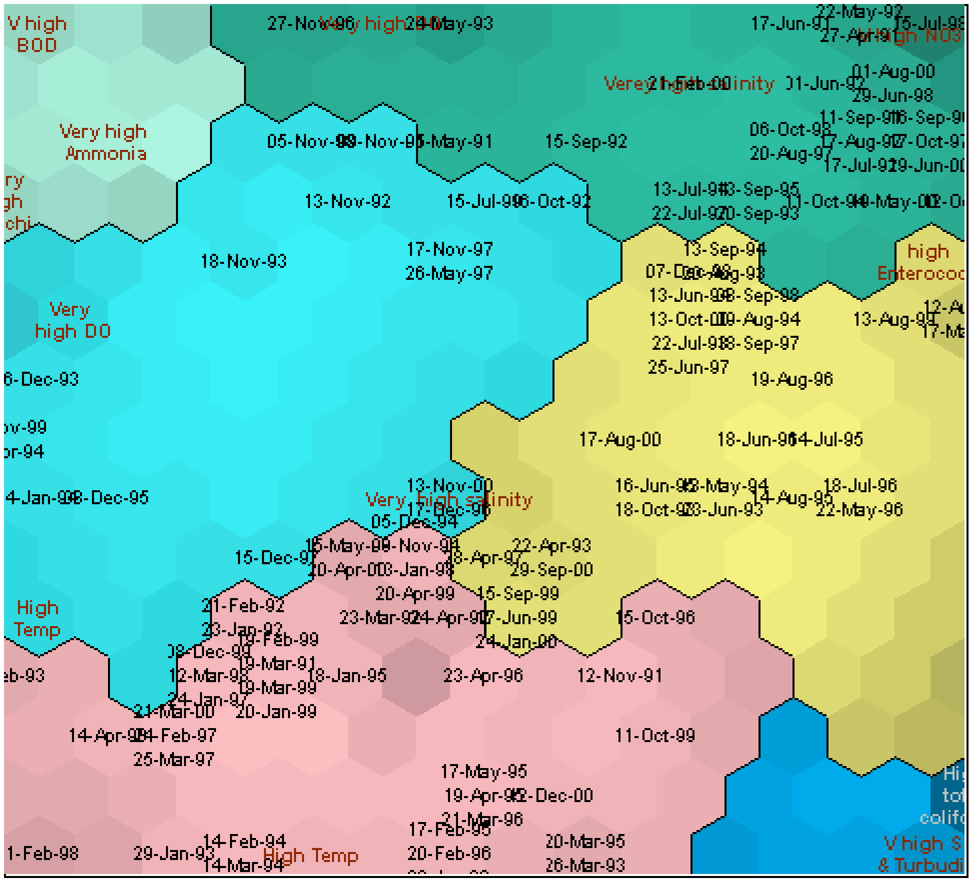


Figure A 6.1: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Chelsea data.

Chelsea: Most of the data points lie in the right bottom diagonal half of the map, an area where pH is between 7.9-8.25, chloride ranging from high to medium 15,056-20,261, DO with high to medium values, DRP with a high range from 0.012 - 0.042 and BOD around 1.5. But on 12 August 1998 and 17 May 1999 the site has experienced the highest *Enterococci* count and on 24 May 1993, the highest chloride levels at 0.013.

On 15 July 1998, the site has recoded the highest DRP value in the map and is seen in the high NO_3NO_2 area. However no data is found in very high BOD, Ammonia and secchi disk area.

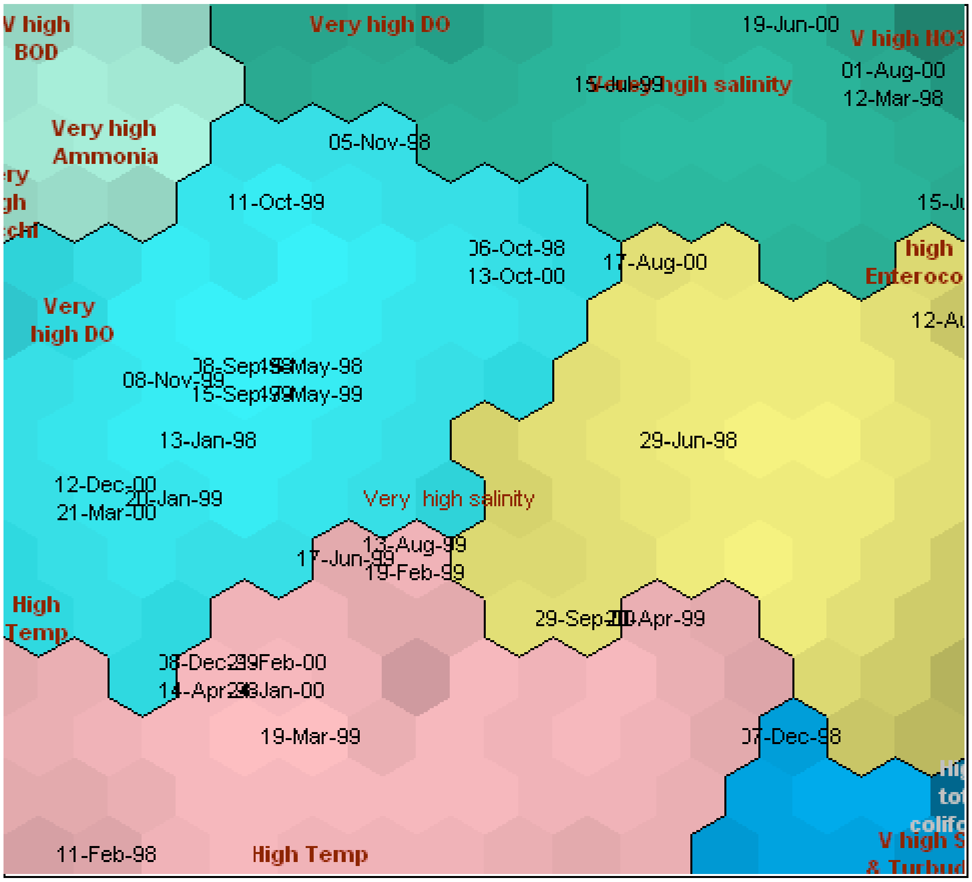


Figure A 6.2: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Depth data.

Depth: The Depth data represents bottom water columns (5-10 meters from the surface) of seawater samplings collected from different locations. Depth data was available only for a period of three years from 13 January 1998 to 12 December 2000. Within this period on two occasions (1 August 2000 and 12 March 1998) the site has experienced very high NO_3NO_2 levels.

15 July 1998 and 12 August 1998 are seen closer to the high *Enterococci* area on the right top corner of the map.

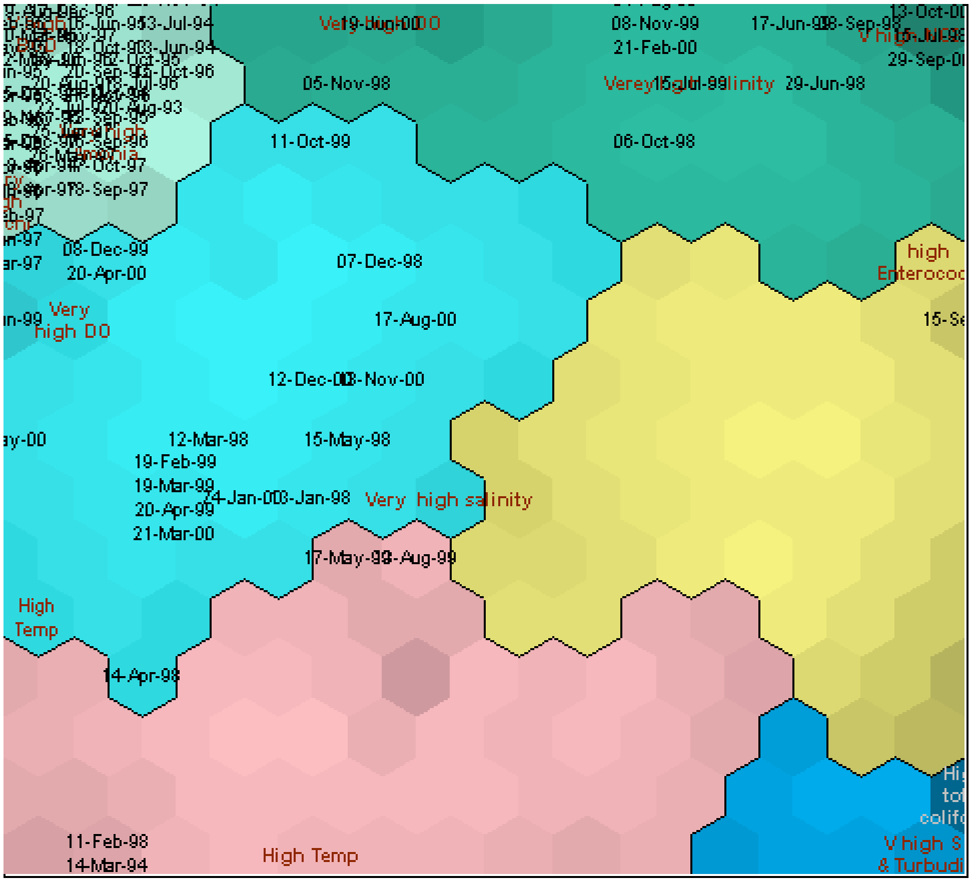


Figure A 6.3: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Goat Island data.

Goat Island: Goad Island is described to be the effective reference site with the best water quality and least variability of the 11 saline sites included in ARC's LTB monitoring programme (Wilcock and Stroud 2000b). In the SOM map most of the Goat Island data fall in the top left corner showing the highest BOD, Ammonia and secchi disk values. Further, it could be noticed that 12 August 1998, 13 October 2000, 15 July 1998 and 29 September 2000 data points fall in the top right corner, with the highest values of NO_3NO_2. The highest value for *Enterococci* count (148) could be observed on the 15 September 1999.

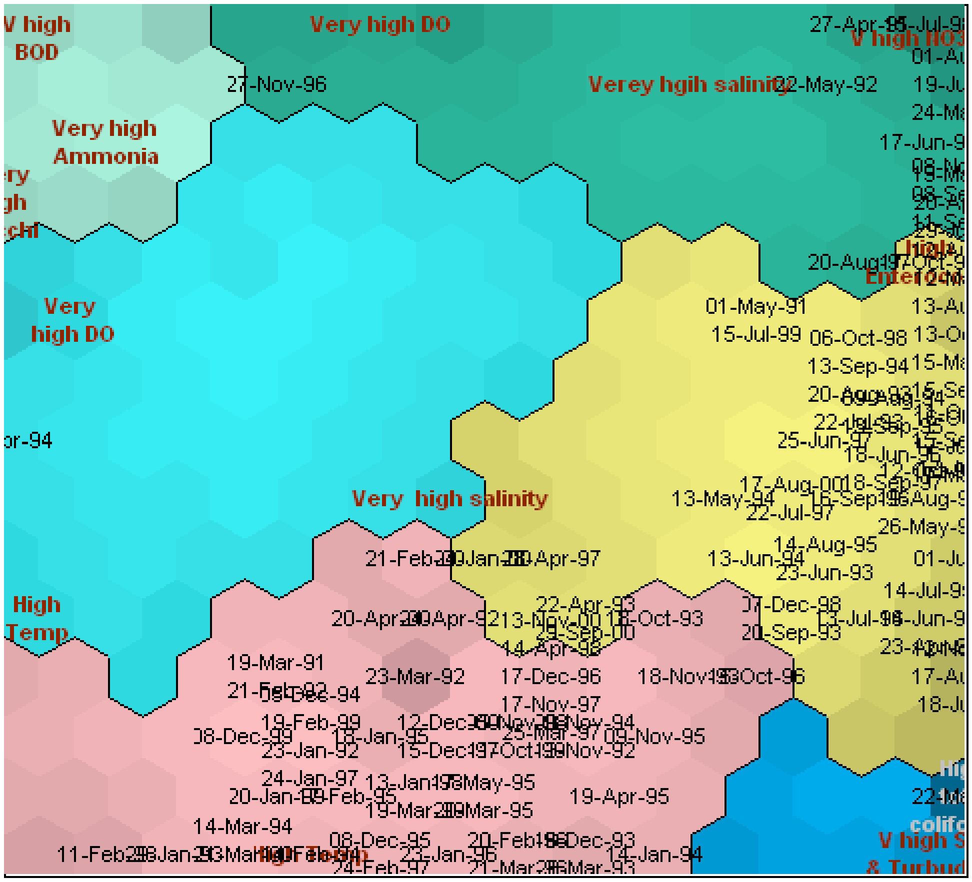


Figure A 6.4: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Henderson data.

Henderson: All the data points fall in the right bottom diagonal half of the map, most of them falling in very high NO_3NO_2, high *Entercocci* count area and on 22 May 1996, the site has experienced the highest turbidity and high total coliform values of the map. 27 November 1997 fall in the top left corner in the high BOD, high ammonia and high secchi disk value patch. 15 July 1998, 27 April 1991, 22 May 1992, 01 August 2000, 19 June 2000 are 24 May1993 are seen in the high NO_3NO_2 area.

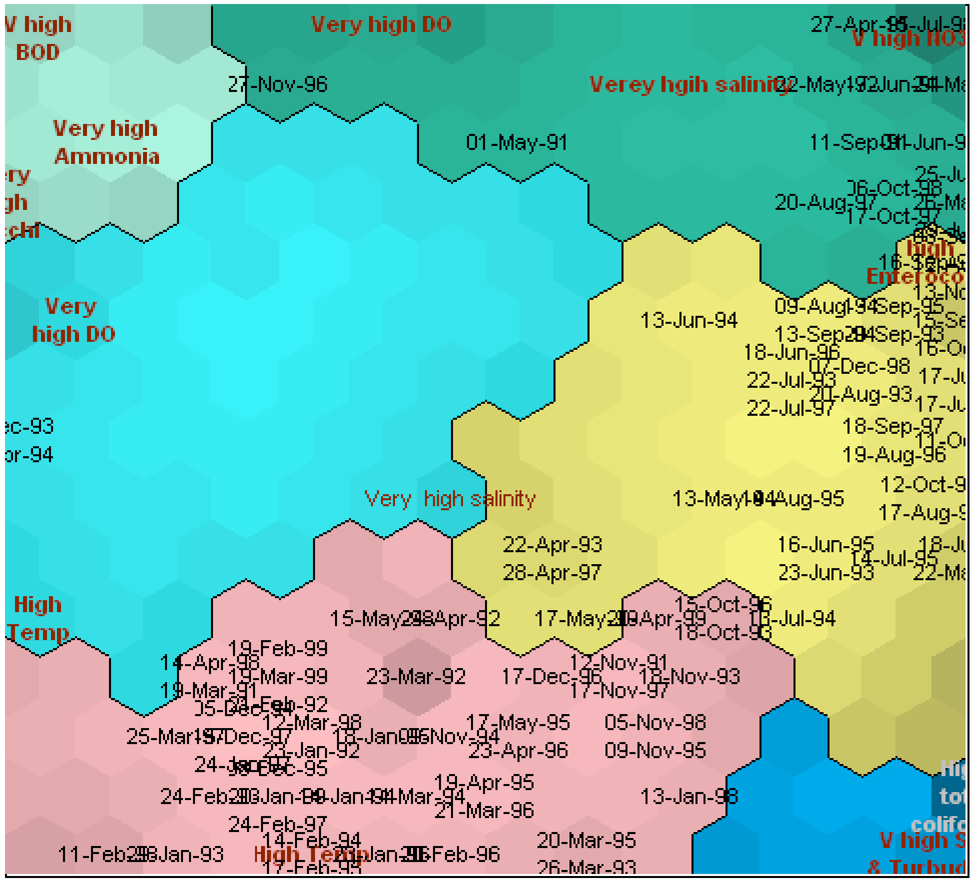


Figure A 6.5: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Hobsonville data.

Hobsonville: Data spread shows a pattern similar to that of Henderson. All data points fall in the left bottom triangle. However, no data point could be found in the high total coliform and turbidity area in the left bottom corner. On 1 May 1991 and 27 October 1996 site values have deviated from the general trend. 27 April 1991 and 15 July 1998 are found in the very high NO_3NO_2 area.

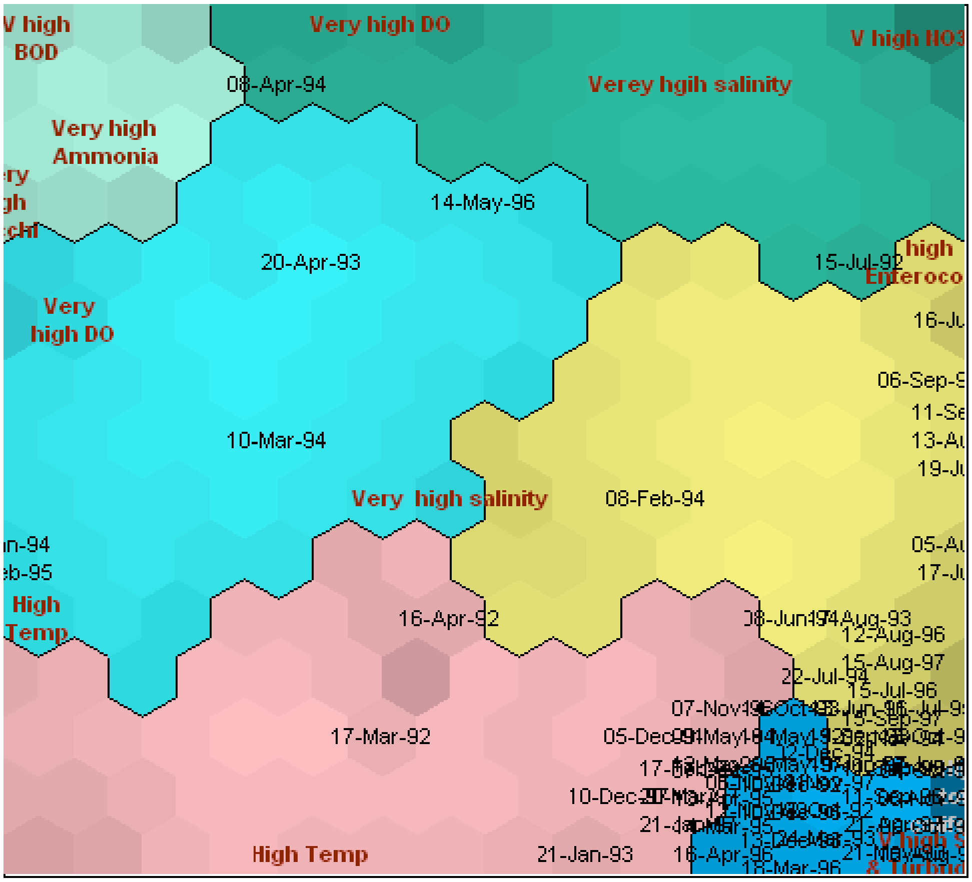


Figure A 6.6: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Kaipara (Shelly Beach) data.

Kaipara (Shelly Beach): Most of the data points fall in the bottom right corner of the map, where total coliform and turbidity show the map's highest values. However, certain months have shown deviations from the general trend: 10 January 1994, 14 February 1995, 10 March 1994, 20 April 1993, 14 May 1996, 08 April 1994 and 08 February 1994

Some months are seen in close to high *Enterococci* count area: 15 July 1992, 16 June 1992, 11 September 1992, 13 August 1992, 19 July 1993, and 06 September 1994

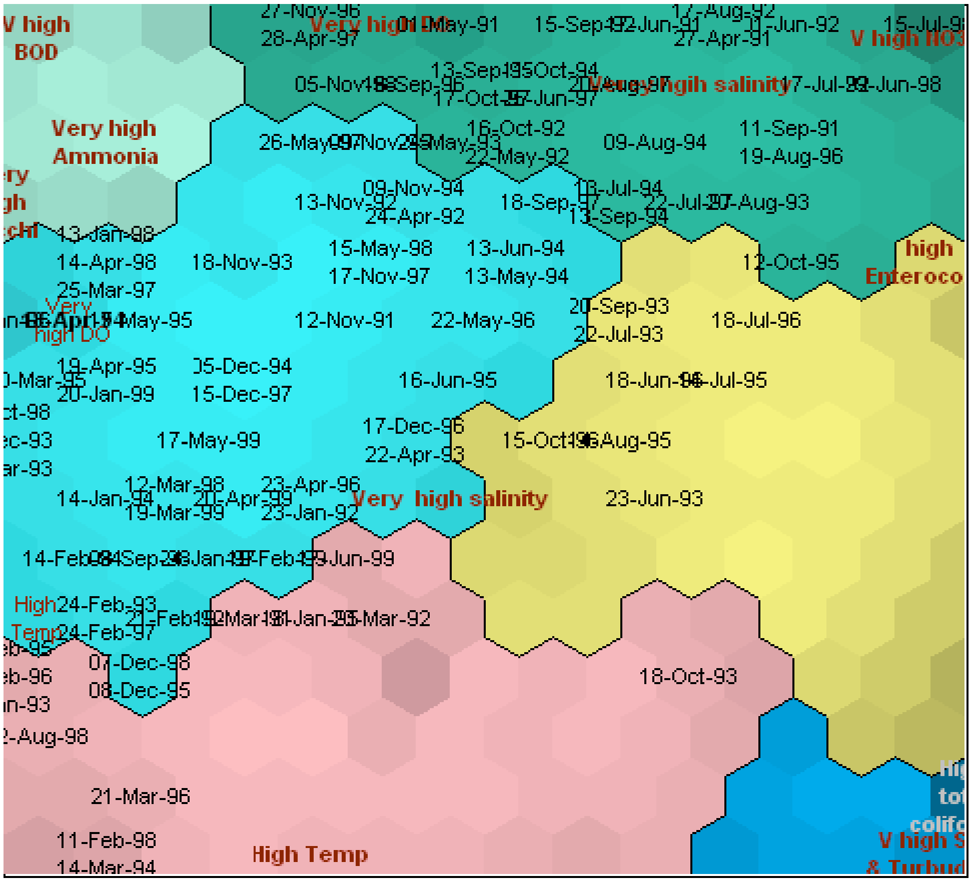


Figure A 6.7: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Kawau Bay data.

Kawau Bay: Most of the data points fall in the top right diagonal half of the map with some points in the very high DO and salinity patch, except for 15 July 1998 with very high NO_3NO_2. No data could be found in the top left corner where ammonia, BOD and secchi disk values show the map's highest values. Some data fall into areas outside the general trend; 21 march 1996, 11 February 1998 and 14 March 1994.

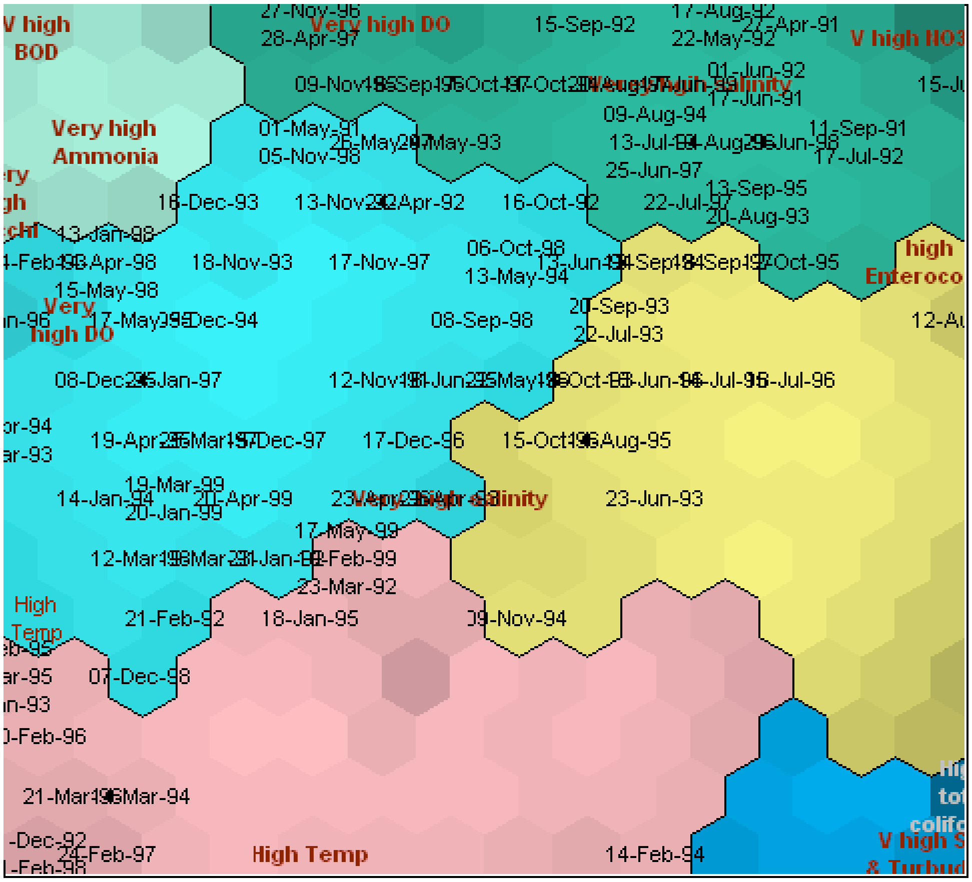


Figure A 6.8: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Mahurangi data.

Mahurangi: All the data points for Mahurangi fall in the top right diagonal half on the map. No data points fall in the top left or bottom right corners. 15 July 1998 is seen in the very high NO_3NO_2 area.

14 March 1994, 21 March 1996, 24 February 1997, 01December 1992,11 February 1998, 20 February 1996, 17 February 1995, 20 March 1995 and 29 Jan 1993 are seen deviating from the normal trend.

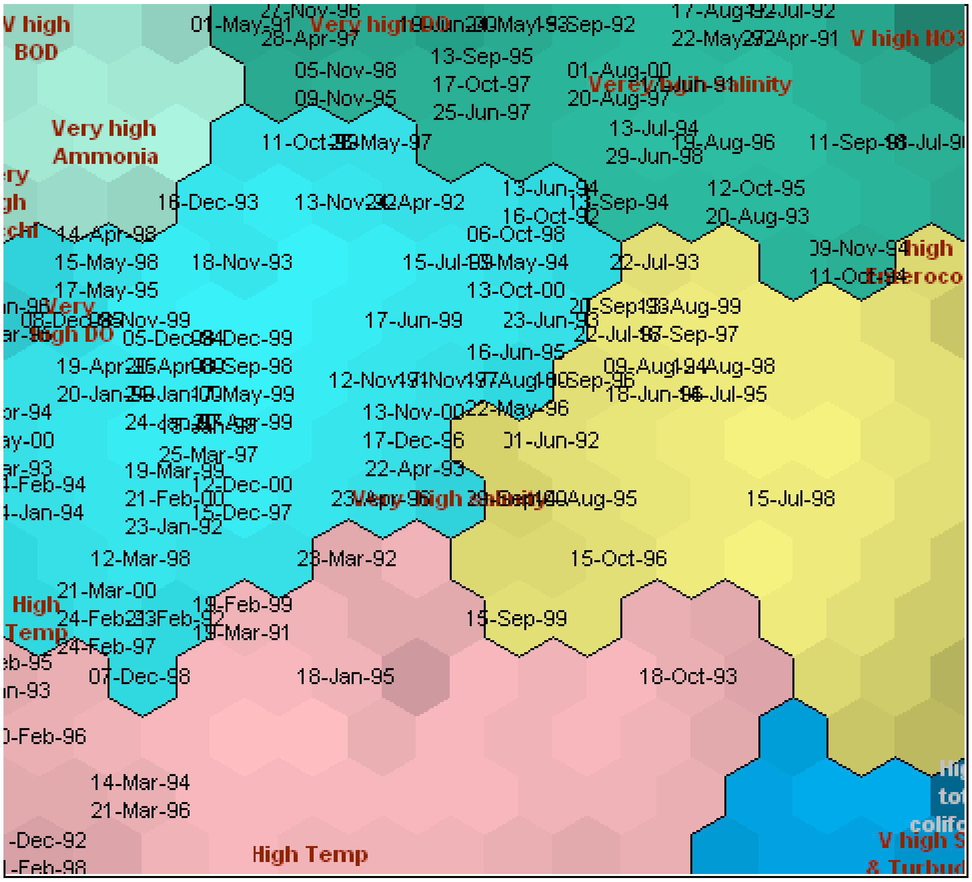


Figure A 6.9: SOM map of the saline water quality data from the 11 beach water sampling locations; included in ARC's programme with Orewa data.

Orewa: Most of the data points fall in the top left diagonal half of the map. Still no points are seen in the top left corner. 15 July 1998 is seen in a high NO_3NO_2 area.

Here again a few data points fall outside the general trend; 17 February 1995, 29 January 1993, 20 February 1996, 01 December 1992, 11 February 1998, 14 March 1994, 21 March 1996, 18 January 1995, 18 January 1995, 18 October 1993, 18 January 1995, 15 October 1996 and15 July 1998

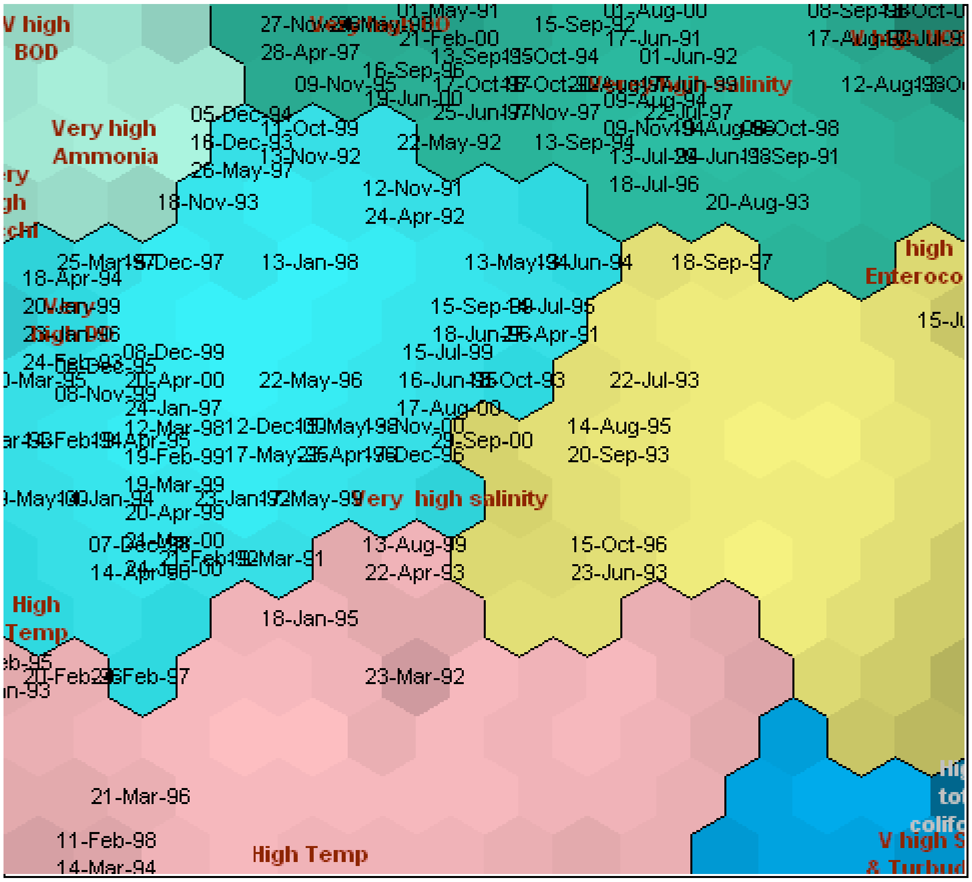


Figure A 6.10: SOM map of the saline water quality data on the 11 beach water sampling locations; included in ARC's programme with Ti Point data.

Ti Point: Most of the data fall in the top left diagonal half except for 15 July 1998 falling in the high *Enterococci* count area. Also 08 September 1998 and 17 August 1992 are seen in the high NO_3NO_2 area.

Some data points fall outside the general trend; 17 February 1995, 29 January 1993, 21 March 1996, 20 February 1996, 11 February 1998, 14 March 1994, 18 January 1995, 23 March 1992, 13 August 1999, 22 April 1993, 15 October 1996, 23 June 1993, 20 September 1993, 14 August 1995, 22 July 1993 and 18 September 1997.

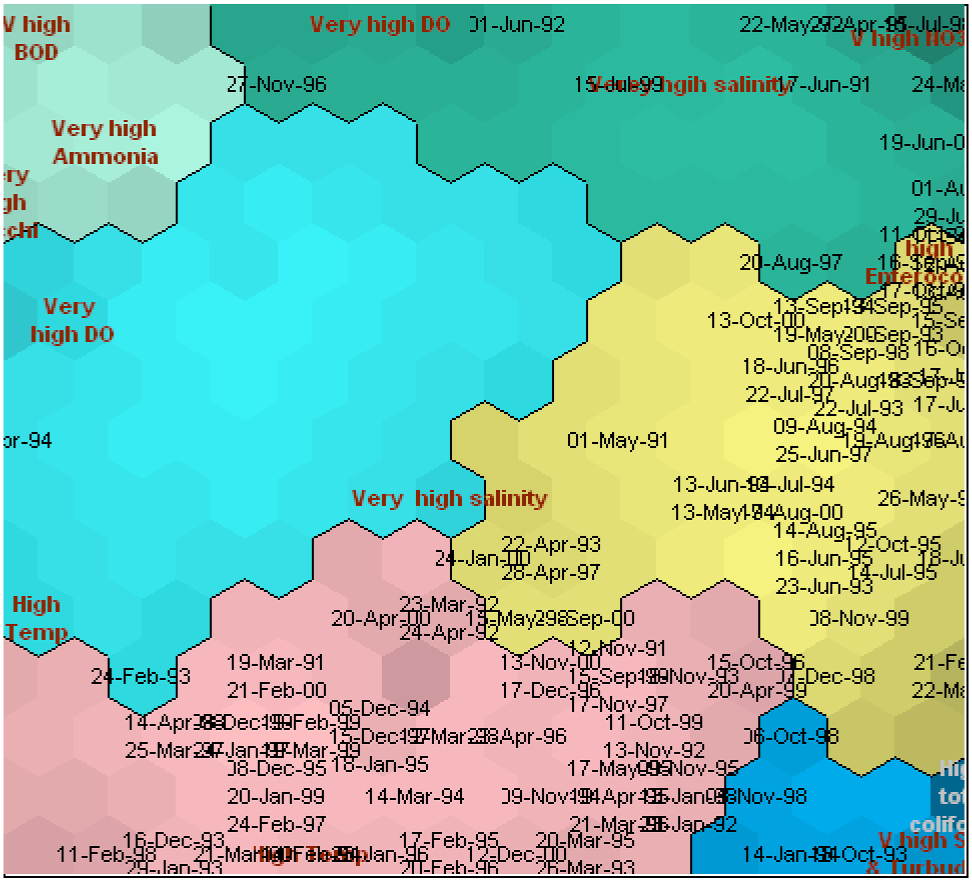


Figure A 6.11: SOM map of the saline water quality data on the 11 beach water sampling locations; included in ARC's programme with Whau data.

Whau: All data fall in the right bottom diagonal half of the map, except for 27 November 1996 and 01 June 1992 falling in the very high DO area. 14 January 1994 and 18 October 1993 are seen in the very high suspended solids, turbidity and high total coliform area. 22 May 1992, 27 April 1991, 15 July 1998 and 24 May 1993 are seen in the high NO3NO2 area. Three points are seen outside the general trend: 01 June 1992, 27 November 1996 and 18 April 1994.

It is interesting to note that on 15 July 1998 all the beaches have experienced high NO_3NO_2 values except for Ti Point and Depth, which show high *Enterococci* count on this day.

Appendix 7

Significance tests on % <63 micron sediment rates of total sedimentation

These are all the correlations with their p-values, between all variables i.e. sedimentation and subtital community changes. Correlations with p-values less that 0.05 are considered as significant.

Correlations: % ave<63(mio, tot ave(g/d), Blue nudi, Buccinulum, C.opalas, C.vir

	% ave<63	tot ave(	Blue nud	Buccinul	C.opalas	C.virgat	Canth	Cellana
tot ave(	-0.288							
	0.010							
Blue nud	0.093	0.083						
	0.412	0.463						
Buccinul	-0.052	-0.036	-0.027					
	0.647	0.754	0.804					
C.opalas	0.137	0.085	-0.019	-0.108				
	0.226	0.454	0.860	0.312				
C.virgat	-0.011	0.054	-0.029	0.198	-0.178			
	0.920	0.636	0.785	0.062	0.093			
Canth	0.258	-0.052	0.147	-0.031	0.180	-0.017		
	0.021	0.647	0.166	0.771	0.090	0.872		
Cellana	-0.077	-0.004	-0.027	-0.038	-0.016	-0.254	-0.202	
	0.499	0.969	0.801	0.723	0.881	0.016	0.057	
Cookia	0.079	0.000	0.165	-0.057	0.079	-0.044	0.278	0.112
	0.488	0.997	0.120	0.592	0.460	0.682	0.008	0.293
Cosci	0.077	0.068	-0.070	0.196	-0.095	0.134	0.012	0.007
	0.500	0.548	0.512	0.064	0.376	0.209	0.914	0.949
Crypto	0.044	-0.171	-0.096	0.190	-0.057	-0.053	-0.028	0.230
	0.698	0.129	0.367	0.073	0.593	0.619	0.792	0.029
Cushion	-0.084	-0.209	0.059	0.090	-0.011	-0.045	-0.009	-0.049
	0.458	0.063	0.580	0.401	0.916	0.674	0.934	0.647
Cystop	0.404	-0.096	0.078	-0.189	-0.083	-0.034	0.167	-0.118
	0.000	0.398	0.466	0.074	0.438	0.751	0.116	0.269
Ecklonia	0.111	-0.055	-0.029	-0.018	-0.039	0.107	-0.020	-0.102
	0.329	0.627	0.786	0.868	0.717	0.314	0.854	0.338

Eve	0.029	-0.102	-0.022	0.152	-0.029	0.019	0.038	-0.041
	0.796	0.368	0.839	0.153	0.787	0.858	0.725	0.701
Flex	-0.250	0.159	-0.045	-0.157	0.172	-0.114	0.026	-0.130
	0.025	0.158	0.674	0.140	0.104	0.284	0.805	0.220
Glossoph	-0.000	0.000	-0.019	0.055	-0.011	0.035	0.028	-0.016
	1.000	1.000	0.860	0.609	0.916	0.745	0.791	0.881
Haust	0.107	-0.259	0.032	0.096	0.114	-0.251	0.180	0.018
	0.345	0.020	0.762	0.371	0.286	0.017	0.089	0.868
Maori	0.042	-0.104	-0.064	0.017	-0.038	0.103	-0.028	-0.054
	0.709	0.357	0.548	0.875	0.722	0.332	0.796	0.612
mash	0.535	0.120	0.208	0.057	0.039	0.149	0.322	-0.084
	0.000	0.289	0.049	0.593	0.717	0.162	0.002	0.433
Melantha	0.155	-0.106	-0.033	0.038	-0.019	-0.221	0.069	-0.028
	0.169	0.348	0.759	0.720	0.855	0.037	0.519	0.795
Micre	0.038	-0.013	0.065	0.138	-0.040	0.099	0.033	-0.057
	0.738	0.911	0.544	0.194	0.709	0.355	0.758	0.595
Murex	-0.044	-0.080	-0.019	0.055	-0.011	-0.008	0.028	-0.016
	0.698	0.480	0.860	0.609	0.916	0.942	0.791	0.881
Murex(no	-0.056	-0.007	-0.019	0.055	-0.011	0.002	0.028	-0.016
	0.619	0.949	0.860	0.609	0.916	0.988	0.791	0.881
Orange s	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
Others	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
P. angus	-0.184	0.142	-0.033	0.096	-0.020	0.010	-0.149	-0.028
	0.103	0.210	0.757	0.370	0.854	0.926	0.161	0.793
Penion	-0.072	-0.029	-0.019	0.055	-0.011	0.163	0.028	-0.016
	0.528	0.800	0.860	0.609	0.916	0.125	0.791	0.881
Plagusia	0.138	0.147	-0.043	0.162	-0.026	0.099	0.135	-0.037
	0.223	0.193	0.685	0.127	0.810	0.353	0.204	0.732
Plocamiu	0.080	0.219	-0.019	-0.108	-0.011	-0.008	0.028	-0.016
	0.483	0.051	0.860	0.312	0.916	0.942	0.791	0.881
plum	0.072	0.089	0.035	-0.029	-0.055	0.281	-0.021	-0.079
	0.526	0.434	0.741	0.784	0.604	0.007	0.846	0.460

Sarg	0.301	-0.120	0.118	-0.008	0.010	-0.062	0.084	-0.081
	0.007	0.287	0.269	0.944	0.929	0.559	0.431	0.449
Slippery	0.126	-0.115	-0.019	-0.108	-0.011	-0.008	0.028	-0.016
	0.266	0.310	0.860	0.312	0.916	0.942	0.791	0.881
Sol Asid	-0.031	-0.172	-0.154	0.012	-0.018	0.149	0.129	-0.146
	0.782	0.128	0.147	0.913	0.867	0.162	0.226	0.169
Stegnast	-0.114	-0.148	0.055	0.025	0.083	-0.191	0.005	0.010
	0.313	0.192	0.610	0.812	0.439	0.071	0.963	0.923
Stichopu	0.097	-0.039	-0.019	-0.108	-0.011	0.097	0.028	-0.016
	0.391	0.734	0.860	0.312	0.916	0.362	0.791	0.881
T. ingal	-0.162	0.235	-0.019	0.055	-0.011	-0.008	-0.142	-0.016
	0.152	0.036	0.860	0.609	0.916	0.942	0.182	0.881
T.aur	0.062	-0.092	-0.087	0.139	-0.051	-0.085	0.076	-0.006
	0.585	0.419	0.417	0.191	0.630	0.427	0.477	0.959
Thais	0.253	0.156	0.132	0.175	0.095	0.136	0.150	0.003
	0.023	0.167	0.214	0.098	0.372	0.201	0.159	0.977
Trochus	0.312	-0.032	-0.055	-0.094	0.137	-0.031	0.227	-0.158
	0.005	0.778	0.608	0.381	0.197	0.773	0.031	0.137
Turbo	0.395	0.099	0.100	-0.014	0.115	0.086	0.224	-0.079
	0.000	0.382	0.351	0.897	0.281	0.422	0.034	0.457
Whelk	0.038	-0.094	-0.019	0.136	-0.011	-0.008	0.028	-0.016
	0.736	0.406	0.860	0.202	0.916	0.942	0.791	0.881
Wonderin	0.186	-0.003	-0.019	0.055	-0.011	-0.178	-0.142	-0.016
	0.099	0.981	0.860	0.609	0.916	0.093	0.182	0.881
zon	0.502	0.092	0.148	-0.036	0.097	-0.052	0.361	-0.080
	0.000	0.416	0.163	0.733	0.364	0.623	0.000	0.454

	Cookia	Cosci	Crypto	Cushion	Cystop	Ecklonia	Eve	Flex
Cosci	-0.112							
	0.294							
Crypto	-0.023	0.157						
	0.832	0.140						
Cushion	-0.166	-0.080	0.160					
	0.118	0.451	0.132					
Cystop	0.221	-0.065	0.062	-0.119				
	0.037	0.540	0.561	0.265				

Ecklonia	-0.045	-0.106	-0.046	0.133	0.015			
	0.676	0.322	0.664	0.212	0.887			
Eve	-0.100	0.303	-0.058	0.038	-0.062	-0.065		
	0.347	0.004	0.585	0.721	0.563	0.543		
Flex	-0.259	-0.054	-0.005	0.127	-0.072	0.064	-0.070	
	0.014	0.610	0.965	0.234	0.499	0.546	0.511	
Glossoph	0.079	-0.095	-0.057	-0.069	-0.083	0.054	-0.004	-0.092
	0.460	0.376	0.593	0.517	0.438	0.612	0.971	0.390
Haust	0.137	-0.082	-0.048	0.012	0.002	-0.075	0.113	0.005
	0.198	0.440	0.652	0.911	0.984	0.484	0.290	0.966
Maori	0.056	-0.040	-0.132	0.077	0.167	-0.027	-0.038	-0.023
	0.602	0.709	0.215	0.469	0.116	0.802	0.723	0.826
mash	0.240	0.070	-0.043	-0.236	0.164	0.168	-0.073	-0.265
	0.022	0.513	0.685	0.025	0.123	0.112	0.496	0.012
Melantha	0.037	-0.164	-0.099	0.175	-0.144	-0.091	-0.022	0.075
	0.732	0.123	0.353	0.098	0.177	0.394	0.834	0.485
Micre	-0.008	0.008	-0.152	-0.012	-0.036	0.063	0.085	-0.232
	0.939	0.942	0.153	0.914	0.738	0.553	0.424	0.027
Murex	0.079	0.104	-0.057	-0.069	-0.083	0.129	-0.006	0.049
	0.460	0.329	0.593	0.517	0.438	0.226	0.954	0.644
Murex(no	0.079	0.104	0.162	0.139	-0.083	-0.100	0.011	-0.092
	0.460	0.329	0.128	0.190	0.438	0.348	0.919	0.390
Orange s	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
Others	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
P. angus	-0.087	-0.050	0.155	0.305	-0.070	-0.103	-0.014	-0.018
	0.417	0.642	0.143	0.003	0.510	0.332	0.896	0.864
Penion	-0.114	-0.095	-0.057	0.105	-0.083	0.080	-0.029	0.076
	0.286	0.376	0.593	0.326	0.438	0.452	0.787	0.477
Plagusia	0.004	0.238	-0.031	0.035	-0.065	-0.165	0.433	-0.050
	0.969	0.024	0.775	0.741	0.543	0.120	0.000	0.643
Plocamiu	0.271	-0.095	-0.057	-0.243	0.131	0.023	-0.006	-0.092
	0.010	0.376	0.593	0.021	0.220	0.832	0.954	0.390

plum	0.212	0.095	-0.115	-0.228	0.172	-0.025	-0.034	-0.136
	0.045	0.371	0.282	0.031	0.104	0.816	0.752	0.200
Sarg	0.048	0.022	0.022	-0.054	0.492	0.015	0.015	-0.053
	0.652	0.836	0.834	0.615	0.000	0.888	0.886	0.623
Slippery	0.079	-0.095	-0.057	-0.026	0.065	-0.100	0.001	0.030
	0.460	0.376	0.593	0.810	0.543	0.348	0.989	0.778
Sol Asid	-0.129	0.106	0.038	0.030	0.033	0.091	0.040	0.080
	0.224	0.318	0.724	0.782	0.755	0.392	0.710	0.455
Stegnast	0.057	-0.193	0.029	0.121	-0.077	0.060	-0.072	0.087
	0.594	0.068	0.788	0.256	0.472	0.574	0.502	0.414
Stichopu	-0.114	0.104	-0.057	-0.069	0.124	-0.100	-0.006	0.071
	0.286	0.329	0.593	0.517	0.245	0.348	0.954	0.507
T. ingal	-0.114	-0.095	-0.057	-0.069	-0.083	-0.064	-0.029	0.030
	0.286	0.376	0.593	0.517	0.438	0.548	0.787	0.778
T.aur	-0.086	-0.012	0.037	0.368	-0.010	-0.039	0.210	0.053
	0.422	0.908	0.730	0.000	0.926	0.714	0.047	0.622
Thais	0.190	0.113	0.155	-0.061	-0.078	-0.055	0.085	-0.203
	0.074	0.291	0.144	0.568	0.466	0.604	0.426	0.055
Trochus	0.024	-0.006	-0.078	-0.164	0.185	0.103	-0.060	-0.053
	0.822	0.953	0.465	0.124	0.081	0.335	0.577	0.619
Turbo	0.118	0.065	-0.015	-0.280	0.126	-0.010	-0.074	-0.204
	0.270	0.543	0.885	0.008	0.235	0.928	0.488	0.053
Whelk	-0.114	0.303	-0.057	0.035	-0.083	-0.082	0.993	-0.092
	0.286	0.004	0.593	0.742	0.438	0.442	0.000	0.390
Wonderin	-0.114	0.104	-0.057	-0.069	0.173	-0.005	-0.029	-0.092
	0.286	0.329	0.593	0.517	0.102	0.966	0.787	0.390
zon	0.202	-0.025	0.101	-0.160	0.150	0.057	-0.061	-0.096
	0.056	0.818	0.344	0.133	0.159	0.594	0.566	0.366

	Glossoph	Haust	Maori	mash	Melantha	Micre	Murex	Murex(no
Haust	-0.089							
	0.406							
Maori	-0.038	-0.129						
	0.722	0.225						
mash	0.252	-0.166	0.042					

	0.017	0.118	0.694					
Melantha	-0.019	0.127	0.083	-0.126				
	0.855	0.233	0.439	0.236				
Micre	-0.040	0.155	-0.036	0.090	0.012			
	0.709	0.144	0.739	0.398	0.908			
Murex	-0.011	0.114	-0.038	-0.046	-0.019	-0.040		
	0.916	0.286	0.722	0.665	0.855	0.709		
Murex(no	-0.011	-0.089	-0.038	-0.051	-0.019	0.078	-0.011	
	0.916	0.406	0.722	0.634	0.855	0.467	0.916	
Orange s	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
Others	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
P. angus	-0.020	-0.155	0.100	-0.109	-0.034	-0.001	-0.020	0.571
	0.854	0.144	0.348	0.306	0.749	0.991	0.854	0.000
Penion	-0.011	-0.089	-0.038	-0.074	-0.019	-0.040	-0.011	-0.011
	0.916	0.406	0.722	0.490	0.855	0.709	0.916	0.916
Plagusia	-0.026	0.075	-0.087	0.134	-0.045	-0.037	-0.026	-0.026
	0.810	0.483	0.414	0.207	0.677	0.726	0.810	0.810
Plocamiu	-0.011	0.114	-0.038	0.072	-0.019	-0.040	-0.011	-0.011
	0.916	0.286	0.722	0.501	0.855	0.709	0.916	0.916
plum	-0.055	-0.212	-0.004	0.253	-0.096	0.274	-0.055	0.092
	0.604	0.044	0.967	0.016	0.368	0.009	0.604	0.388
Sarg	0.010	-0.134	0.194	0.161	0.046	0.026	0.076	-0.123
	0.929	0.210	0.067	0.129	0.665	0.811	0.477	0.247
Slippery	-0.011	-0.089	0.247	-0.058	-0.019	0.078	-0.011	-0.011
	0.916	0.406	0.019	0.584	0.855	0.467	0.916	0.916
Sol Asid	-0.103	0.069	0.338	-0.130	0.032	-0.167	0.055	-0.103
	0.335	0.519	0.001	0.221	0.768	0.115	0.608	0.335
Stegnast	-0.068	0.306	0.060	-0.164	0.222	0.059	-0.068	-0.068
	0.524	0.003	0.574	0.122	0.036	0.580	0.524	0.524
Stichopu	-0.011	-0.089	0.247	-0.070	-0.019	-0.040	-0.011	-0.011
	0.916	0.406	0.019	0.512	0.855	0.709	0.916	0.916
T. ingal	-0.011	-0.089	0.247	-0.065	-0.019	-0.040	-0.011	-0.011

	0.916	0.406	0.019	0.540	0.855	0.709	0.916	0.916
T.aur	-0.051	0.255	-0.094	-0.095	0.158	0.152	-0.051	-0.051
	0.630	0.015	0.380	0.371	0.137	0.154	0.630	0.630
Thais	0.095	-0.017	-0.045	0.070	-0.061	0.013	-0.091	0.095
	0.372	0.876	0.672	0.514	0.570	0.900	0.395	0.372
Trochus	0.021	-0.084	0.029	0.413	-0.086	-0.004	-0.023	-0.050
	0.841	0.429	0.785	0.000	0.419	0.972	0.827	0.642
Turbo	0.190	-0.143	-0.037	0.696	-0.091	-0.054	-0.050	-0.054
	0.072	0.179	0.730	0.000	0.395	0.614	0.639	0.610
Whelk	-0.011	0.114	-0.038	-0.063	-0.019	0.078	-0.011	-0.011
	0.916	0.286	0.722	0.554	0.855	0.467	0.916	0.916
Wonderin	-0.011	-0.089	-0.038	0.064	-0.019	-0.040	-0.011	-0.011
	0.916	0.406	0.722	0.547	0.855	0.709	0.916	0.916
zon	0.236	0.129	-0.113	0.527	-0.085	0.265	-0.055	-0.062
	0.025	0.226	0.287	0.000	0.424	0.012	0.607	0.563

	Orange s	Others	P. angus	Penion	Plagusia	Plocamiu	plum	Sarg
Others	*							
	*							
P. angus	*	*						
	*	*						
Penion	*	*	-0.020					
	*	*	0.854					
Plagusia	*	*	-0.045	-0.026				
	*	*	0.673	0.810				
Plocamiu	*	*	-0.020	-0.011	-0.026			
	*	*	0.854	0.916	0.810			
plum	*	*	-0.011	-0.055	-0.032	0.113		
	*	*	0.919	0.604	0.763	0.287		
Sarg	*	*	0.011	-0.123	-0.160	0.010	0.099	
	*	*	0.921	0.247	0.131	0.929	0.355	
Slippery	*	*	-0.020	-0.011	-0.026	-0.011	-0.021	0.032
	*	*	0.854	0.916	0.810	0.916	0.847	0.767
Sol Asid	*	*	-0.015	-0.103	-0.027	-0.103	-0.261	0.089
	*	*	0.890	0.335	0.798	0.335	0.013	0.406

Stegnast	*	*	0.013	-0.068	-0.018	-0.068	0.049	-0.079
	*	*	0.905	0.524	0.867	0.524	0.648	0.461
Stichopu	*	*	-0.020	-0.011	-0.026	-0.011	-0.055	0.010
	*	*	0.854	0.916	0.810	0.916	0.604	0.929
T. ingal	*	*	0.571	-0.011	-0.026	-0.011	-0.055	0.132
	*	*	0.000	0.916	0.810	0.916	0.604	0.216
T.aur	*	*	0.076	0.329	0.187	-0.051	-0.166	0.009
	*	*	0.474	0.002	0.078	0.630	0.118	0.933
Thais	*	*	-0.051	-0.091	0.048	0.281	-0.031	-0.143
	*	*	0.636	0.395	0.657	0.007	0.775	0.179
Trochus	*	*	0.018	-0.032	-0.106	-0.174	0.030	0.163
	*	*	0.868	0.762	0.322	0.100	0.782	0.125
Turbo	*	*	-0.099	-0.034	0.091	0.013	0.266	0.157
	*	*	0.351	0.751	0.393	0.902	0.011	0.140
Whelk	*	*	-0.020	-0.011	0.437	-0.011	-0.055	0.010
	*	*	0.854	0.916	0.000	0.916	0.604	0.929
Wonderin	*	*	-0.020	-0.011	-0.026	-0.011	0.078	0.043
	*	*	0.854	0.916	0.810	0.916	0.467	0.689
zon	*	*	-0.089	-0.063	0.188	0.059	0.203	0.122
	*	*	0.404	0.555	0.076	0.580	0.055	0.253

	Slippery	Sol Asid	Stegnast	Stichopu	T. ingal	T.aur	Thais	Trochus
Sol Asid	-0.103							
	0.335							
Stegnast	-0.068	0.064						
	0.524	0.551						
Stichopu	-0.011	0.025	-0.068					
	0.916	0.818	0.524					
T. ingal	-0.011	0.095	0.158	-0.011				
	0.916	0.372	0.137	0.916				
T.aur	-0.051	-0.042	0.234	-0.051	-0.051			
	0.630	0.698	0.027	0.630	0.630			
Thais	0.095	-0.100	-0.079	-0.091	-0.091	-0.048		
	0.372	0.349	0.458	0.395	0.395	0.651		
Trochus	-0.050	0.079	-0.066	-0.050	0.091	-0.177	-0.133	
	0.642	0.458	0.535	0.642	0.391	0.096	0.212	

```
Turbo    -0.055   -0.143   -0.119   -0.053   -0.057   -0.124   -0.063    0.274
          0.604    0.179    0.264    0.621    0.596    0.246    0.552    0.009

Whelk    -0.011    0.033   -0.068   -0.011   -0.011    0.234    0.095   -0.050
          0.916    0.757    0.524    0.916    0.916    0.027    0.372    0.642

Wonderin -0.011   -0.103   -0.068   -0.011   -0.011   -0.051    0.095    0.013
          0.916    0.335    0.524    0.916    0.916    0.630    0.372    0.906

zon      -0.047   -0.145    0.075   -0.063   -0.057    0.139    0.140    0.247
          0.661    0.172    0.484    0.555    0.593    0.191    0.189    0.019

          Turbo    Whelk Wonderin
Whelk    -0.054
          0.611

Wonderin  0.011   -0.011
          0.919    0.916

zon       0.474   -0.048   -0.095
          0.000    0.650    0.372

Cell Contents: Pearson correlation
               P-Value
```

These are the correlations and p-values from the second Leigh data set.

Correlations: ave tot_sed, less63sedi, % less 63 se, C.opalas, C.virgata, Canth,

```
       ave tot_ less63se % less 6 C.opalas C.virgat    Canth  Cellana   Cookia
less63se -0.383
          0.117

% less 6 -0.242    0.687
          0.334    0.002

C.opalas  0.239    0.187    0.279
          0.340    0.457    0.262

C.virgat -0.174   -0.029    0.352   -0.192
          0.489    0.908    0.152    0.446

Canth    -0.285    0.699    0.650    0.299    0.234
          0.252    0.001    0.004    0.228    0.350

Cellana  -0.115   -0.079   -0.198   -0.086   -0.205   -0.087
          0.650    0.755    0.431    0.735    0.416    0.733

Cookia   -0.023    0.183    0.575   -0.115    0.100    0.239   -0.043
```

	0.926	0.466	0.013	0.649	0.693	0.339	0.865	
Cosci	-0.070	0.117	0.060	0.010	0.081	0.104	0.108	0.019
	0.784	0.643	0.814	0.970	0.750	0.681	0.670	0.941
Crypto	0.365	0.200	0.185	0.114	-0.103	0.121	0.166	0.223
	0.136	0.427	0.462	0.653	0.683	0.632	0.511	0.375
Cushion	0.367	-0.066	-0.440	0.154	-0.533	-0.131	0.041	-0.267
	0.134	0.795	0.068	0.543	0.023	0.605	0.871	0.285
Cystop	-0.446	0.646	0.622	0.174	0.112	0.572	0.026	0.403
	0.063	0.004	0.006	0.491	0.657	0.013	0.920	0.098
Ecklonia	-0.121	0.622	0.715	-0.108	0.330	0.544	-0.246	0.389
	0.634	0.006	0.001	0.670	0.181	0.020	0.325	0.111
Eve	-0.231	0.071	-0.105	-0.109	0.104	-0.119	-0.102	-0.158
	0.356	0.778	0.680	0.668	0.681	0.639	0.687	0.531
Flex	0.360	0.061	-0.055	0.371	-0.737	-0.059	-0.022	0.209
	0.143	0.809	0.828	0.130	0.000	0.815	0.932	0.405
Glossoph	0.165	-0.302	0.079	-0.059	0.356	-0.166	-0.086	-0.115
	0.512	0.223	0.755	0.817	0.147	0.509	0.735	0.649
Haust	-0.182	0.070	-0.183	0.088	-0.484	-0.258	0.178	-0.126
	0.471	0.781	0.468	0.729	0.042	0.302	0.479	0.619
Maori	-0.274	-0.098	-0.066	-0.202	0.228	0.004	0.294	-0.012
	0.271	0.699	0.796	0.422	0.363	0.989	0.236	0.964
mash	-0.009	0.374	0.817	0.138	0.487	0.653	-0.215	0.551
	0.973	0.126	0.000	0.584	0.041	0.003	0.392	0.018
Melantha	-0.181	-0.158	-0.265	-0.086	-0.233	-0.018	0.470	-0.036
	0.473	0.532	0.287	0.736	0.352	0.944	0.049	0.888
Micre	0.143	-0.011	0.312	0.638	0.380	0.276	0.006	-0.127
	0.571	0.966	0.207	0.004	0.119	0.268	0.981	0.615
Murex	-0.123	-0.090	-0.144	-0.059	-0.012	-0.166	-0.086	0.005
	0.627	0.721	0.568	0.817	0.961	0.509	0.735	0.983
Murex(no	0.424	-0.328	-0.380	-0.059	-0.181	-0.166	-0.086	-0.115
	0.079	0.184	0.120	0.817	0.473	0.509	0.735	0.649
Orange s	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
Others	*	*	*	*	*	*	*	*

	*	*	*	*	*	*	*	*
P. angus	0.652	-0.421	-0.495	-0.086	-0.187	-0.242	-0.125	-0.168
	0.003	0.082	0.037	0.735	0.457	0.332	0.621	0.506
Penion	-0.222	-0.004	0.097	-0.059	-0.080	-0.166	-0.086	-0.115
	0.377	0.988	0.702	0.817	0.752	0.509	0.735	0.649
Plagusia	-0.061	0.372	0.404	0.454	0.130	0.355	-0.189	-0.128
	0.810	0.129	0.096	0.059	0.607	0.148	0.453	0.612
Plocamiu	0.100	0.032	0.398	-0.059	-0.136	-0.055	-0.086	0.910
	0.693	0.898	0.102	0.817	0.590	0.827	0.735	0.000
plum	0.014	0.345	0.777	0.193	0.553	0.681	-0.233	0.415
	0.956	0.161	0.000	0.443	0.017	0.002	0.352	0.087
Sarg	-0.082	0.078	0.372	0.090	0.610	0.522	0.002	0.351
	0.745	0.757	0.128	0.723	0.007	0.026	0.992	0.153
Slippery	-0.063	-0.003	-0.077	-0.059	-0.089	-0.166	0.686	-0.115
	0.803	0.991	0.763	0.817	0.726	0.509	0.002	0.649
Sol Asid	-0.090	0.185	0.064	-0.254	-0.092	0.226	-0.041	0.178
	0.721	0.461	0.799	0.310	0.718	0.366	0.870	0.481
Stegnast	-0.339	0.203	-0.055	-0.013	0.025	0.054	0.109	-0.381
	0.169	0.418	0.828	0.960	0.921	0.830	0.665	0.118
Stichopu	-0.162	0.098	-0.042	-0.059	0.126	-0.166	-0.086	-0.115
	0.521	0.699	0.867	0.817	0.620	0.509	0.735	0.649
T. ingal	0.424	-0.328	-0.380	-0.059	-0.181	-0.166	-0.086	-0.115
	0.079	0.184	0.120	0.817	0.473	0.509	0.735	0.649
T.aur	0.060	0.017	-0.276	0.113	-0.351	0.006	0.141	-0.351
	0.813	0.948	0.268	0.655	0.153	0.982	0.576	0.153
Thais	-0.173	0.364	0.532	-0.141	0.414	0.391	-0.206	0.619
	0.492	0.138	0.023	0.576	0.088	0.109	0.412	0.006
Trochus	0.102	0.336	0.704	0.098	0.088	0.492	-0.287	0.678
	0.686	0.173	0.001	0.698	0.728	0.038	0.248	0.002
Turbo	-0.051	0.240	0.636	0.144	0.326	0.519	-0.215	0.297
	0.842	0.338	0.005	0.570	0.187	0.027	0.392	0.232
Whelk	-0.313	0.041	-0.059	-0.059	0.106	-0.166	-0.086	-0.115
	0.206	0.873	0.815	0.817	0.676	0.509	0.735	0.649
Wonderin	0.100	0.032	0.398	-0.059	-0.136	-0.055	-0.086	0.910

	0.693	0.898	0.102	0.817	0.590	0.827	0.735	0.000
zon	0.140	0.425	0.777	0.668	0.195	0.669	-0.233	0.377
	0.579	0.079	0.000	0.002	0.439	0.002	0.352	0.123

	Cosci	Crypto	Cushion	Cystop	Ecklonia	Eve	Flex	Glossoph
Crypto	0.119							
	0.639							
Cushion	0.236	0.218						
	0.346	0.384						
Cystop	-0.049	0.460	-0.275					
	0.846	0.055	0.270					
Ecklonia	0.100	0.185	-0.318	0.347				
	0.694	0.463	0.198	0.158				
Eve	0.209	-0.286	-0.073	-0.132	-0.184			
	0.405	0.251	0.774	0.601	0.464			
Flex	-0.075	0.367	0.372	0.128	0.033	-0.496		
	0.768	0.134	0.128	0.613	0.896	0.036		
Glossoph	0.010	-0.368	-0.170	-0.462	-0.034	-0.070	-0.458	
	0.970	0.133	0.499	0.054	0.895	0.782	0.056	
Haust	-0.191	-0.475	0.273	-0.341	-0.273	0.059	0.132	0.088
	0.449	0.046	0.273	0.166	0.272	0.815	0.600	0.729
Maori	0.139	0.207	-0.358	0.326	-0.270	0.207	-0.361	-0.202
	0.581	0.411	0.144	0.187	0.279	0.411	0.141	0.422
mash	0.045	0.100	-0.469	0.360	0.595	-0.131	-0.233	0.376
	0.859	0.692	0.050	0.142	0.009	0.606	0.351	0.124
Melantha	-0.605	-0.164	0.042	0.054	-0.384	-0.158	0.015	-0.086
	0.008	0.514	0.870	0.831	0.116	0.531	0.953	0.736
Micre	-0.067	-0.170	-0.090	-0.069	-0.107	0.083	-0.367	0.510
	0.793	0.500	0.723	0.787	0.672	0.743	0.135	0.030
Murex	0.010	0.114	0.342	0.188	-0.214	-0.109	0.019	-0.059
	0.970	0.653	0.165	0.455	0.395	0.668	0.939	0.817
Murex(no	0.010	0.114	0.307	-0.462	-0.390	0.123	-0.044	-0.059
	0.970	0.653	0.215	0.054	0.109	0.628	0.863	0.817
Orange s	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*

Others	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
P. angus	0.014	0.429	0.421	-0.365	-0.323	0.099	0.129	-0.086
	0.956	0.075	0.082	0.136	0.191	0.696	0.611	0.735
Penion	0.010	0.114	0.056	-0.049	0.233	-0.109	0.105	-0.059
	0.970	0.653	0.825	0.847	0.352	0.668	0.679	0.817
Plagusia	0.104	-0.280	-0.061	-0.046	0.155	0.468	-0.325	0.454
	0.681	0.260	0.811	0.856	0.539	0.050	0.188	0.059
Plocamiu	0.010	0.114	-0.170	0.170	0.211	-0.109	0.371	-0.059
	0.970	0.653	0.499	0.500	0.400	0.668	0.130	0.817
plum	0.008	0.077	-0.448	0.325	0.571	-0.157	-0.291	0.429
	0.975	0.761	0.062	0.189	0.013	0.534	0.241	0.076
Sarg	0.070	0.276	-0.207	0.429	0.168	-0.091	-0.317	-0.199
	0.782	0.268	0.409	0.076	0.506	0.720	0.200	0.428
Slippery	0.139	0.114	-0.178	-0.023	-0.040	-0.032	-0.044	-0.059
	0.583	0.653	0.481	0.928	0.875	0.901	0.863	0.817
Sol Asid	0.012	0.086	-0.218	0.234	0.361	-0.158	0.236	-0.531
	0.964	0.735	0.384	0.350	0.141	0.530	0.345	0.024
Stegnast	0.022	-0.415	0.016	-0.301	0.018	0.579	-0.412	-0.013
	0.931	0.087	0.950	0.225	0.943	0.012	0.089	0.960
Stichopu	0.010	0.114	-0.299	0.145	-0.047	-0.025	-0.008	-0.059
	0.970	0.653	0.229	0.566	0.854	0.923	0.974	0.817
T. ingal	0.010	0.114	0.307	-0.462	-0.390	0.123	-0.044	-0.059
	0.970	0.653	0.215	0.054	0.109	0.628	0.863	0.817
T.aur	0.043	0.163	0.580	-0.148	-0.228	0.138	0.112	-0.288
	0.865	0.519	0.012	0.557	0.363	0.586	0.658	0.247
Thais	0.109	-0.003	-0.298	0.246	0.438	0.514	-0.343	-0.141
	0.666	0.991	0.229	0.325	0.069	0.029	0.163	0.576
Trochus	0.010	0.162	-0.322	0.307	0.502	-0.197	0.114	0.264
	0.968	0.520	0.192	0.215	0.034	0.433	0.651	0.290
Turbo	0.033	0.025	-0.513	0.328	0.392	-0.172	-0.185	0.419
	0.898	0.921	0.030	0.184	0.108	0.495	0.461	0.084
Whelk	0.160	-0.368	-0.141	-0.072	-0.196	0.964	-0.458	-0.059
	0.525	0.133	0.576	0.777	0.436	0.000	0.056	0.817

Wonderin	0.010	0.114	-0.170	0.170	0.211	-0.109	0.371	-0.059
	0.970	0.653	0.499	0.500	0.400	0.668	0.130	0.817
zon	0.005	0.153	-0.230	0.385	0.413	-0.169	0.101	0.225
	0.984	0.544	0.359	0.115	0.089	0.503	0.691	0.370

	Haust	Maori	mash	Melantha	Micre	Murex	Murex(no	Orange s
Maori	-0.380							
	0.119							
mash	-0.375	0.054						
	0.125	0.831						
Melantha	0.322	0.017	-0.249					
	0.192	0.946	0.319					
Micre	0.059	-0.069	0.432	0.006				
	0.817	0.785	0.073	0.980				
Murex	0.088	0.202	-0.152	-0.086	-0.001			
	0.729	0.422	0.548	0.736	0.996			
Murex(no	0.088	0.202	-0.137	-0.086	-0.001	-0.059		
	0.729	0.422	0.587	0.736	0.996	0.817		
Orange s	*	*	*	*	*	*	*	
	*	*	*	*	*	*	*	
Others	*	*	*	*	*	*	*	*
	*	*	*	*	*	*	*	*
P. angus	-0.323	0.000	-0.216	-0.125	-0.142	-0.086	0.686	*
	0.191	1.000	0.390	0.622	0.575	0.735	0.002	*
Penion	0.088	-0.202	-0.186	-0.086	-0.193	-0.059	-0.059	*
	0.729	0.422	0.461	0.736	0.443	0.817	0.817	*
Plagusia	0.194	-0.222	0.447	-0.189	0.690	-0.130	-0.130	*
	0.441	0.375	0.063	0.454	0.002	0.608	0.608	*
Plocamiu	0.088	-0.202	0.325	-0.086	-0.193	-0.059	-0.059	*
	0.729	0.422	0.189	0.736	0.443	0.817	0.817	*
plum	-0.411	0.024	0.982	-0.221	0.525	-0.167	-0.149	*
	0.090	0.924	0.000	0.379	0.025	0.509	0.554	*
Sarg	-0.514	0.469	0.473	-0.014	0.318	0.146	0.150	*
	0.029	0.050	0.047	0.957	0.199	0.563	0.553	*
Slippery	0.088	0.202	-0.143	-0.086	-0.001	-0.059	-0.059	*
	0.729	0.422	0.572	0.736	0.996	0.817	0.817	*

Sol Asid	-0.302	0.372	0.074	-0.109	-0.530	-0.054	0.058	*
	0.224	0.129	0.769	0.667	0.024	0.830	0.819	*
Stegnast	0.455	-0.110	-0.210	0.165	0.173	-0.364	0.128	*
	0.058	0.663	0.402	0.513	0.494	0.138	0.613	*
Stichopu	0.088	0.202	-0.180	-0.086	-0.193	-0.059	-0.059	*
	0.729	0.422	0.475	0.736	0.443	0.817	0.817	*
T. ingal	0.088	0.202	-0.137	-0.086	-0.001	-0.059	1.000	*
	0.729	0.422	0.587	0.736	0.996	0.817	*	*
T.aur	0.113	-0.277	-0.447	0.283	-0.084	-0.288	0.259	*
	0.655	0.265	0.063	0.255	0.739	0.247	0.299	*
Thais	-0.144	0.055	0.519	-0.206	0.113	-0.141	0.061	*
	0.569	0.828	0.027	0.413	0.654	0.576	0.810	*
Trochus	-0.229	-0.062	0.871	-0.272	0.149	-0.134	-0.007	*
	0.360	0.806	0.000	0.275	0.555	0.597	0.980	*
Turbo	-0.446	0.252	0.865	-0.219	0.326	-0.151	-0.147	*
	0.063	0.314	0.000	0.382	0.186	0.550	0.560	*
Whelk	0.088	0.202	-0.143	-0.086	0.105	-0.059	-0.059	*
	0.729	0.422	0.571	0.736	0.677	0.817	0.817	*
Wonderin	0.088	-0.202	0.325	-0.086	-0.193	-0.059	-0.059	*
	0.729	0.422	0.189	0.736	0.443	0.817	0.817	*
zon	-0.203	-0.152	0.820	-0.214	0.643	-0.153	-0.168	*
	0.418	0.547	0.000	0.393	0.004	0.544	0.506	*

	Others	P. angus	Penion	Plagusia	Plocamiu	plum	Sarg	Slippery
P. angus	*							
	*							
Penion	*	-0.086						
	*	0.735						
Plagusia	*	-0.189	-0.130					
	*	0.453	0.608					
Plocamiu	*	-0.086	-0.059	-0.130				
	*	0.735	0.817	0.608				
plum	*	-0.193	-0.180	0.484	0.174			
	*	0.443	0.474	0.042	0.490			
Sarg	*	-0.019	-0.154	-0.117	0.012	0.495		

	*	0.939	0.543	0.643	0.961	0.037		
Slippery	*	-0.086	-0.059	-0.130	-0.059	-0.166	-0.117	
	*	0.735	0.817	0.608	0.817	0.510	0.643	
Sol Asid	*	-0.124	-0.023	-0.438	0.024	0.013	0.241	0.024
	*	0.624	0.929	0.069	0.926	0.958	0.335	0.926
Stegnast	*	-0.172	0.338	0.427	-0.364	-0.177	-0.142	-0.013
	*	0.495	0.170	0.077	0.138	0.482	0.574	0.960
Stichopu	*	-0.086	-0.059	-0.130	-0.059	-0.181	-0.153	-0.059
	*	0.735	0.817	0.608	0.817	0.473	0.546	0.817
T. ingal	*	0.686	-0.059	-0.130	-0.059	-0.149	0.150	-0.059
	*	0.002	0.817	0.608	0.817	0.554	0.553	0.817
T.aur	*	0.378	0.441	-0.062	-0.288	-0.393	-0.109	-0.106
	*	0.122	0.067	0.807	0.247	0.107	0.666	0.677
Thais	*	-0.059	-0.141	0.310	0.466	0.440	0.459	-0.141
	*	0.817	0.576	0.211	0.051	0.067	0.055	0.576
Trochus	*	-0.054	-0.175	0.317	0.584	0.794	0.187	-0.166
	*	0.832	0.489	0.200	0.011	0.000	0.458	0.510
Turbo	*	-0.220	-0.155	0.372	0.114	0.861	0.298	-0.146
	*	0.381	0.538	0.128	0.652	0.000	0.230	0.563
Whelk	*	-0.086	-0.059	0.454	-0.059	-0.174	-0.106	-0.059
	*	0.735	0.817	0.059	0.817	0.490	0.675	0.817
Wonderin	*	-0.086	-0.059	-0.130	1.000	0.174	0.012	-0.059
	*	0.735	0.817	0.608	*	0.490	0.961	0.817
zon	*	-0.195	-0.177	0.601	0.255	0.831	0.343	-0.167
	*	0.437	0.483	0.008	0.308	0.000	0.163	0.507

	Sol Asid	Stegnast	Stichopu	T. ingal	T.aur	Thais	Trochus	Turbo
Stegnast	-0.137							
	0.588							
Stichopu	-0.115	-0.013						
	0.650	0.960						
T. ingal	0.058	0.128	-0.059					
	0.819	0.613	0.817					
T.aur	-0.222	0.454	-0.288	0.259				
	0.376	0.058	0.247	0.299				

Thais	0.084	0.247	-0.141	0.061	-0.139			
	0.740	0.324	0.576	0.810	0.582			
Trochus	0.128	-0.379	-0.164	-0.007	-0.427	0.427		
	0.614	0.121	0.516	0.980	0.077	0.077		
Turbo	0.207	-0.299	-0.149	-0.147	-0.462	0.165	0.789	
	0.411	0.228	0.556	0.560	0.054	0.513	0.000	
Whelk	-0.146	0.549	-0.059	-0.059	0.077	0.466	-0.228	-0.149
	0.564	0.018	0.817	0.817	0.762	0.051	0.363	0.555
Wonderin	0.024	-0.364	-0.059	-0.059	-0.288	0.466	0.584	0.114
	0.926	0.138	0.817	0.817	0.247	0.051	0.011	0.652
zon	-0.104	-0.181	-0.171	-0.168	-0.252	0.316	0.737	0.706
	0.680	0.473	0.498	0.506	0.313	0.202	0.000	0.001

	Whelk	Wonderin
Wonderin	-0.059	
	0.817	
zon	-0.152	0.255
	0.546	0.308

Cell Contents: Pearson correlation
P-Value

* NOTE * All values in column are identical.