Ecological modelling with self-organising maps

S Shanmuganathan[a], P Sallis[a] and J Buckeridge[b]

[b] Environmental and Oceanic Sciences Centre,
[a] Auckland University of Technology, New Zealand. subana.shanmuganathan@aut.ac.nz

Abstract: Old and new ecological models can be classified into two basic categories: Those aimed at (i) gaining more insight into ecological systems and (ii) producing predictive models of ecosystem behaviour. Many of the models successfully applied to ecological modelling are borrowed from other disciplines such as engineering, mathematics and in recent times from intelligent information processing systems motivated by neuro-physiological understandings i.e. [1]artificial neural networks (ANNs). The use of ANNs in ecological modelling is becoming a popular method with considerable success in elucidating the complexity in ecosystem processes. We critically analyse some ecological modelling applications with self-organising maps (SOMs), within the connectionist neural computing paradigms. These are used to unravel the non-linear relationships in highly complex and often cryptic ecosystems from northern New Zealand. A need to accurately predict an ecosystems response to daily increasing human influences on the environment and its biodiversity is considered to be absolutely vital to preserve natural systems. The example illustrated shows SOM abilities to extract more knowledge from the ecological monitoring data of complex matrices with numeric values of environmental and biological indicators, compared to the conventional data analysis methods. Conventional methods are seen as of little use in exploring the non-linear relationships within the data.

Keywords; *Ecological modelling; Self-organising maps; Ecological data*

1. INTRODUCTION

Over the last few decades ecological modelling techniques borrowed from other disciplines provided a useful means to analyse natural systems with considerable success. Such models of ecology, based on engineering, statistical and mathematical concepts permitted ecologists to gain more insight into ecosystem structure and functioning. However, all of them demonstrated limitations, such as inability to produce conclusive results in environmental impact analysis (i.e. whether an impact was caused either by human influence or natural causes) and to predict long-term environmental effects for management purposes. The old and the recent ecological models, i.e. Before-After-Control-Impact (BACI), Before-After-Control-Impact Paired Series (BACIPS), etc., with highly complex mathematical and statistical techniques are described to be ineffective due to the above-mentioned reasons. The shortcomings of traditional methods led ecologists to experiment with innovative approaches using the recent intelligent systems (ISs) of information processing methodologies i.e. artificial neural networks (ANNs), Fuzzy logic, etc. The recent use of ANNs in ecological modelling is seen as a popular method, successfully applied to unravel ecosystem complexity using the widely available ecological monitoring data alone. The use of different SOM methods for modelling complex ecosystems, north of Auckland in New Zealand is elaborated upon.

2. CONVENTIONAL MODELS

The following two conventional models of ecology are discussed herein:

(i) The River Thames models

(ii) BACI series models.

2.1. Simulation models: River Thames

A class of simulation models (defined by partial differential equations) designed and implemented produced considerable success in cleaning up the

[1]An ANN is a biologically inspired computational model, which consists of processing elements (called neurons), and connections between them with coefficients (weights) bound to the connections. These connections constitute the neuronal structure and attached to this structure are the training and recall algorithms. (Kasabov 1995). The recent ANN models are referred to as 'connectionist neural computing paradigms'.

tidal portion of the River Thames, below London and is discussed here, based on Mann (1982).

Objectives of the River Thames models

The objectives of the River Thames 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.

Purpose and results of the River Thames models

The River Thames models, built using some general concepts were used to predict the circumstances that would in turn, return the Thames to a well-oxygenated condition. 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)

It appears that the river was returned to a well-oxygenated condition because of recommendations from the modelling. A biological survey carried out in 1957, had showed that there were no fishes in the tidal reaches for many km below London. Following this, appropriate sewage treatment facilities were designed, constructed and brought into use in the early 1960s. As a result, by 1965 fish species were seen returning and by 1970, over 50 species had returned, mainly marine species in the lower half of the estuary and freshwater species near London. This led the 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, stated Potter (1973) cited in Mann (1982).

Later many researchers elaborated upon the success of the River Thames models. In Longhurst (1976) cited in Mann (1982), it is argued that despite the oversimplification of biological oxygen in demand (BOD) and dissolved oxygen (DO) interactions on a validated physical time-dependent model and the integrated equations for conservation of volume and materials, the models had been successful in predicting the real situations. They singled out the 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 to develop models, these could be successfully used for predicting long-term effects of an ecosystems functioning for sustainable environmental management, stated Mann (1982). However, none of the available methods (old or even the very new ones) could provide a means to detect such important properties for ecosystem modelling.

2.2. BACI series models

The Before-After-Control-Impact design proposed by Bernstein and Zalinski (1983) and Stewart-Oaten *et al* (1986) provides a means for assessing impacts. Measures of species abundance at two sites (the site of the putative impact and a similar, control, site where no impact was expected), taken on several occasions both before and after the onset of the impact are used to determine the impact. Repeated measurements are used to ascertain whether observed changes at the site of the putative impact are part of the pre-existing cycle of change. The control site is included for detecting any observed change as part of a wider effect not due to the putative impact. The use of a single control site led to criticism and eventually the inclusion of several control sites (or the paired series) in the design to overcome the issue, Underwood (1992)

"…A natural change at the control site which was coincidentally similar to that caused by the impact at the other site could lead to the impact going undetected. Alternatively, a change at the control site from before to after the onset of the putative impact while the other site remained unchanged, could result in an impact being diagnosed where there was none. The solution, proposed by Underwood (1992), is to use several randomly-selected control sites. Thus the effects of global trends can be separated from those of natural fluctuations within individual sites". Underwood (1991; 1992) based on Monitoring to detect impacts (2003).

However, it is argued in many later studies that the BACI design to be incapable of differentiating the impacts from that of global variations and natural causes. BACI, BACIPS designs, introduced to analyse an impact at a particular site before and after an activity, including control sites (and paired sampling) may not provide a good assessment for decision making stated Stewart-Oaten (1996) and identified the main reason for this as model uncertainty. BACI methods are incapable of depicting the formal biological results within a single general parameter (i.e. mean abundance) succinctly and unambiguously, noted Thrush et al. (1995). Further, stated when considering the diffuse and complex impact operating over large spatial and temporal scales, any experimental approach with these tools as impractical; often impossible as information on environmental conditions before the activity occurs (pre impact) would be required from several sites with varying exposure to the activity for this purpose.

Owing to the above limitations of conventional models used to analyse ecosystems and to be abreast of the technological advances, ecologists made attempts to use intelligent systems of information processing methodologies for ecological modelling. The use of Kohonen's self-organising maps (SOMs), based on neuro-physiological understandings of the cortex cells of the human brain has showed potential for modelling a variety of highly dynamic ecosystems i.e., freshwater and marine ecosystems.

3. RECENT ECOLOGICAL MODELS

During the last few years, the use of recent intelligent systems and data processing methodologies to analyse cryptic natural systems, has become a very popular technique in ecological modelling. An urgency to make use of the widely available digital ecological data, associated with a need to incorporate ecosystem process complexity with their spatial and temporal variations led to the exploration of novel approaches. Of those approaches, ANN applications in ecological modelling have produced promising results in revealing the non-linear relationships within the numerical ecological monitoring data.

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.

ANN 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 Deboeck and Kohonen (1998).

3.1. SOMs in ecosystem modelling

Self-organising maps are two layered, feed forward ANNs based on unsupervised algorithmic training. They are capable of projecting multi dimensional data sets onto low dimensional display of neurons, i.e. grids (hexagonal or square), while preserving the topology of the data. SOMs are unique in providing a powerful tool for visualisation of multidimensional data sets, enhancing the extraction of implicit knowledge, (i.e. in the form of structures, patterns and relationships) embedded in the input vectors. Standard statistical methods are good at analysing simple summaries on low dimensional numeric data sets, i.e. mean value, smallest value, highest value, range, etc., in studying linear relationships. However, they are seen as ineffective for use with multidimensional data sets, Deboeck and Kohonen (1998).

In ecological data analysis, SOMs are capable of preserving the spatial and temporal variations. Giraudel and Lek (2001) compared a few, widely used conventional methods of ordination i.e. Polar ordination, Correspondence analysis (CoA), Principle component analysis (PCA) and Non metric multidimensional (NMDS), with SOM analyses using data from upland forest in Wisconsin, in USA. The limitations, observed with the conventional methods were; strong distortions with non-linear species abundance relations, PCA's horseshoe effect due to unimodal species response curves, CoA's arch effect outliers, missing data, and disjoined data matrix. On the other hand, SOM algorithm was stated to be fully usable in exploratory data analysis to study community ordination, complimenting the classical techniques.

The correlations between subtidal community dynamics and sediment deposition rates were analysed with SOMs in Shanmuganathan et al. (2002). In the analysis SOM analyses segregated the annual variations in the population dynamics from those of induced by sedimentation on selected beaches of northeastern coast of Auckland in New Zealand. Furthermore, the intertidal ecosystem of the Long Bay-Okura Marine Reserve in northern New Zealand was successfully modelled with SOMs using biological and physical system monitoring data with inconsistent labelling in Shanmuganathan et al. (2001)

Modelling Patterns in Environmental Data (MOPED), a piece of software developed by NIWA based on SOM methods is successfully used to map patterns in environmental data i.e. species distribution, and elevation of freshwater bodies. MOPED could be used to predict the biological assemblages for streams with different elevation (i.e. the species that should be present in certain streams), from the available habitat data, Jowett (2001).

3.2. SOM in coastal system modelling

A SOM application to analyse saline water quality data from beaches, north of Auckland in New Zealand in illustrated here onwards.

Background

The Auckland Regional Council (ARC) and its predecessors have been monitoring the water quality as part of their Long-Term Baseline (LTB) programme since the 1980s, with an aim to study the trends and effects of human activities on freshwater streams and saline harbour sites. Of the reports released on the data, Wilcock and Stroud (2000) analysed the monitoring of 16 streams, 18 saline water sites in Manukau, Waitemata and Kaipara Harbours, and seven lakes. From this ARC's data, only the saline water quality data from the 11 beaches 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, north of Auckland (figure 1): Browns Bay, Chelsea, Goat Island, Henderson, Hobsonville, Kaipara (Shelly Beach), Kawau Bay, Mahurangi, Orewa, Ti Point and Wha.

The following are the numeric data elements from the ARC's LTB saline water quality tests: Site, Site#, pH, Temperature (deg C), Suspended Solids (SS) (mg/l), Turbidity (NTU), Chloride (mg/l), Salinity (ppt), Total Phosphorus (mgP/l), Dissolved reactive phosphorus DRP (mgP/l), Nitrite (mgN/l), Ammonia (mgN/l), BOD (mgO/l), Total Coliforms (mpn/100ml), Faecal Coliforms (mpn/100ml), Dissolved oxygen (DO %), DO (DO ppm), Secchi disk depth, Chloride, Entercocci ME, NO_2 and NO_3 NO_2.

3.3. ARC saline water data analysis

The above stated ARC data are collectively analysed with SOMs. The SOM analyses carried out on this data to study the patterns in them are first explained. The results of some conventional data analysis methods on the same data are then compared with those of SOM methods.

SOM analyses

SOM based data clustering and trajectories are used in this study. It then could be classified as an initial exploratory data analysis.

Results and Discussion: Cluster and dependent component analyses

Initially, a SOM (figure 2a and b) was created with 200 nodes, priority of all components and all other map parameters set to default values. The six different clusters of this map show the beach water quality dynamics based on the different attributes used in the programme. The SOM data clustering patterns show the monthly trends in the environmental parameters over this period as water sampling has been carried out once a month. Appropriate labels in the map provide details on the attributes analysed. The SOM was able to differentiate the spatial and temporal variations within the ARC's LTB monitoring data. The SOM clustering (figure 2) has separated the data into four major groups (listed here) that coincided with their geographical locations (figure 1). However, no details in this regard were added in the data. The four major SOM clustering groups of figure 2a, also superimposed on figure 1 are:

(i) Goat Island in the top left corner with high BOD and ammonia values, among the 11 beaches analysed.

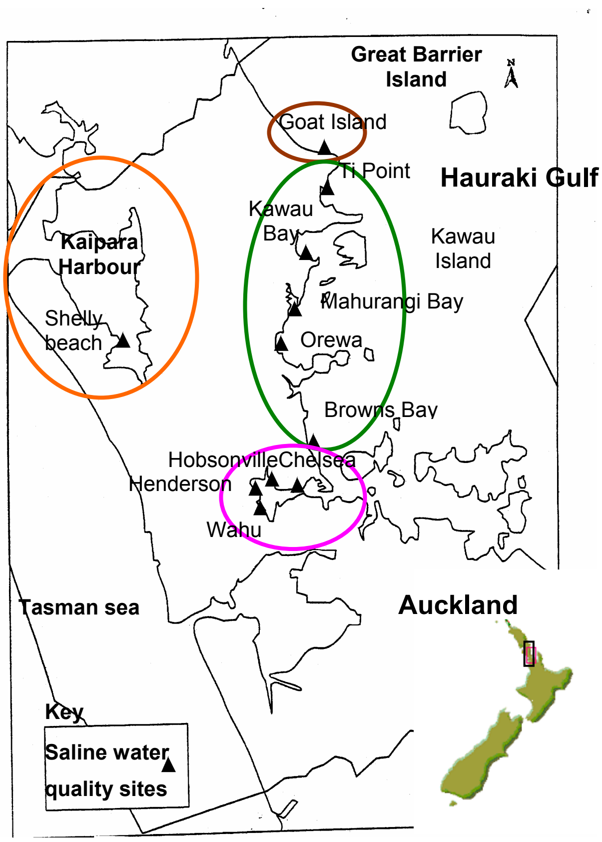
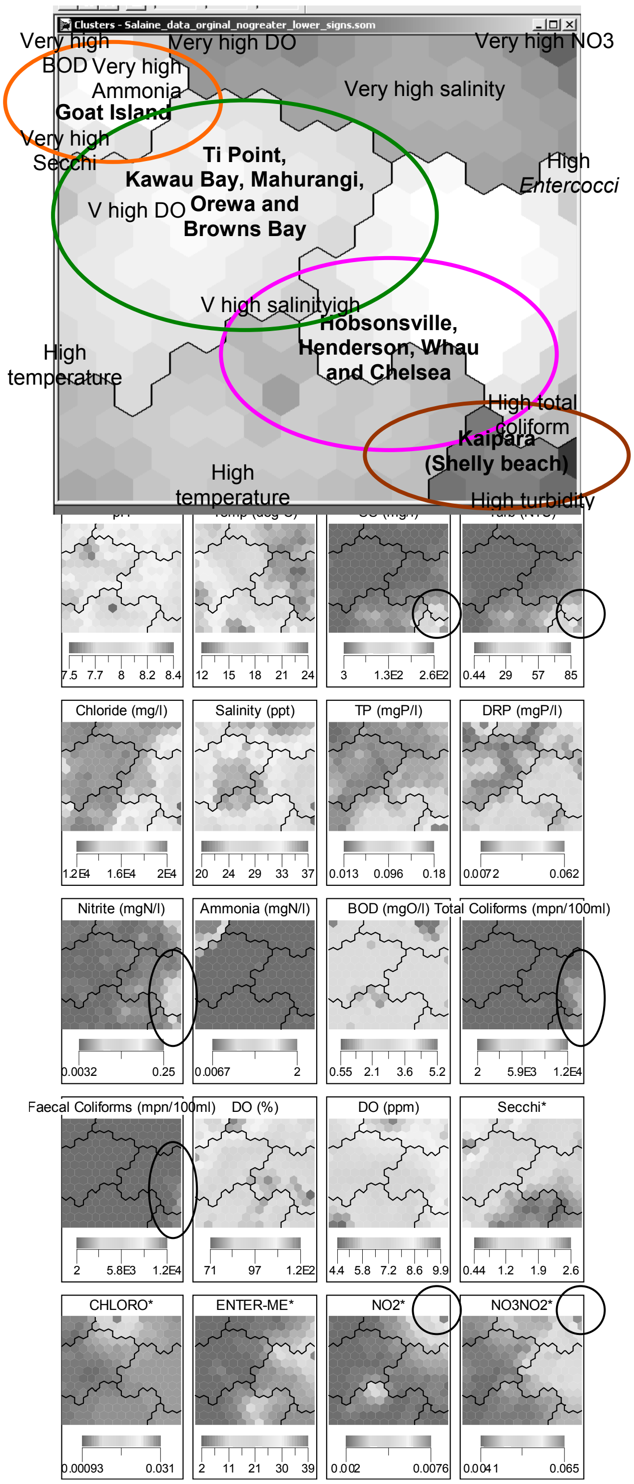


Figure 1: Saline water quality test sites, north of Auckland. Source: Wilcock and Stroud (2000:3). The circles show the SOM clustering results (figure 2) of these sites' monitoring data, coincided with their geographical locations.



Figures 2 a & b: a: SOM of LTB data from the 11 beaches with labels showing the clustering patterns in the data b: SOM components showing the spread of different data attributes.

(ii) Ti Point, Kawau Bay, Mahurangi, Orewa and Browns Bay, beaches of north eastern coast of Auckland in the top right half of the SOM map.

(iii) Hobsonsville, Henderson, Whau and Chelsea in the Waitemata Harbour in the left bottom half of the map.

(iv) Kaipara (Shelly beach) in the bottom right corner of the map with high turbidity, SS, total coliform values, among the beaches analysed.

The above SOM interpretations (figures 2 a) are superimposed on a geographical map (figure 1). On the 15 July 1998, all beach data are seen in the top right corner where *Entercocci* are high.

Time series (trajectories) analysis

A SOM map was created with the ARC's LTB data with 5000 nodes (in order to increase the solution space). By following the animation of Browns Bay data in the trajectory (figure 3), water quality dynamics at this beach could be analysed.

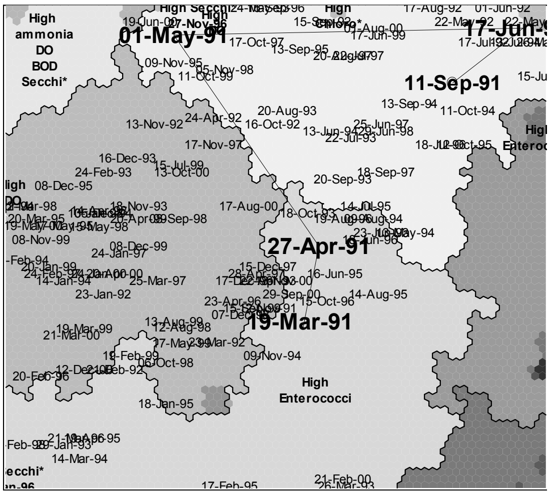


Figure 3: Trajectory of Browns Bay data.

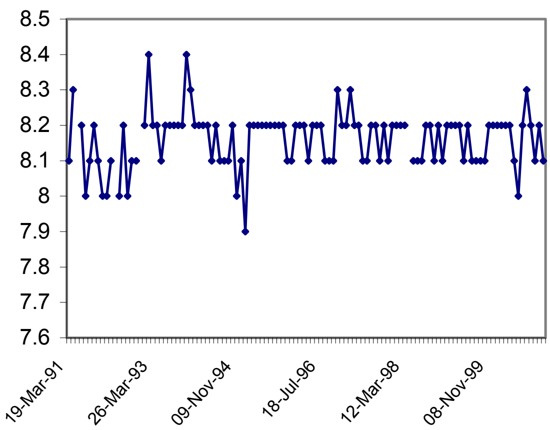


Figure 4: Graph showing the pH trend at Browns Bay (19 Mar 1991 - 15 Oct 2000)

SOM analysis verses conventional analyses

The different attributes of individual beaches, monitored through LTB programme were analysed separately with 2D graphs as in figure 4, in Wilcock and Stroud (2000). These graphs do not provide a means for comparative analysis of the beaches across the region.

3.4. SOM limitations

SOMs can be successfully used to predict values within ranges covered by the available data and not outside these ranges. This could be overcome either (i) by adding simulated values to accommodate nodes for extrapolation or (ii) by calculating the equalisation error of the nodes, the greater the error the more the deviation from the assigned node or (iii) with the use of evolving SOMs that are capable of increasing the number of nodes depending on the requirement of the input data whereas in SOMs the number of nodes is fixed.

4. SUMMARY

The perceived inadequacies with the conventional methods used over the years led to the innovative use of latest intelligent information processing techniques for ecological modelling. Of the new ANN methods, the use of SOMs is found to be capable of overcoming most of the limitations encountered in conventional ecological monitoring data analysis methods.

5. ACKNOWLEDGEMENTS

The authors wish to thank the staff of ARC and NSCC for permission to use their data in this research. The contents of this paper are from the main author's research work leading to a doctoral degree at the Auckland University of Technology, New Zealand.

6. REFERENCES

Bernstein, and Zalinsky (1983)

Deboeck and Kohonen (1998). Visual Explorations in Finance with Self-organizing Maps.

Giraudel, J. L. and S. Lek (2001). A comparison of self-organizing map algorithm and some conventional statistical methods for ecological community ordination.

Jowett, I. (2001). "Artificial intelligence and river biomonitoring." *Water & Atmosphere* **9(1) 2001**: 6.

Longhurst, A. R., Ed. (1976). *Vertical migration.* The Ecology of the Seas, Blackwell Scientific Publications, Oxford.

Mann, K. H. (1982). Models and Management. *Ecology of Coastal Waters - A System Approach*, Oxford : Blackwell Scientific. **Studies in Ecology Volume 8:** 260-285.

Monitoring to detect impacts (2003) Before/After, Control/ Impact sampling design.http://www.ukmarinesac.org.uk/communities/intertidal-reef/ir6_4.htm

Potter, J. H., Ed. (1973). *Management of the tidal Thames.* Mathematical and Hydraulic Modelling of Estuarine Pollution, HMSO, London.

Shanmuganathan, S., P. Sallis and J. S. Buckeridge (2001). *Self-organising maps to study the effects of urbanisation at Long Bay in New Zealand*,. Proceedings of the Fifth Biannual Conference on Artificial Neural Networks and Expert Systems, Dunedin, New Zealand.

Shanmuganathan, S., P. Sallis and J. S. Buckeridge (2002). *Data mining and visualisation in biological and environmental processes*. 3rd Conference of the International Society for Ecological Informatics (ISEI), Villa Grazioli Park Hotel, Grottaferrata, Rome, Italy, August 26-30, '02.

Stewart-Oaten, A. (1996). 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., Murdoch, W. W. and Parker, K. R. (1986). Environmental impact assessment: "pseudoreplication" in time? Ecology, 67, 929-940.

Thrush, S. F., R. D. Pridmore and J. E. Hewitt (1995). Impacts on Soft-sediment Macro fauna. *Detecting Ecological Impacts - Concepts and Applications in Coastal Habitats*. R. J. Schmitt and C. W. Osenberg.

Underwood, A. J. (1991) Beyond BACI: Experimental designs for detecting human environmental impacts on temporal variations in natural populations. Australian Journal of Marine and Freshwater Research, 42, 569-588

Underwood, A. J. (1992) Competition and marine plant-animal interactions. In, Plant-Animal Interactions in the Marine Benthos (eds D. John, S. J. Hawkins and J. Price) (Systematics Association Special Volume No. 46), Clarendon Press, Oxford, pp 443-75.

Wilcock, R. J. and M. J. Stroud (2000). Year 2000 Summary Report, Baseline Water Quality - Stream, Lake, and Saline Waters, Auckland Regional Council**:** 219.