Statistical Methods in Ecological Dynamics Modelling

Shanmuganathan, S. [1] Philip Sallis [1] and W Claster [2]

[1] Auckland University of Technology, New Zealand
[2] Ritsumeikan Asia Pacific University, Beppu, Oita, Japan
Email: subana.shanmuganathan@aut.ac.nz

Keywords: *Ecological dynamics modelling; Natural habitats; coastal systems.*

EXTENDED ABSTRACT

Using stochastic models of discrete-time data and continuous-time series graphs of average values (e.g., annual or monthly) modellers are able to analyse many natural systems and phenomena. These models provide decision and policymaking management with information on the system and/or phenomenon being studied. Nonetheless, since the middle of last century modelling needs have changed significantly. The focus is towards analysing the ecological dynamics of natural habitats during extreme events (i.e. heavy flooding) that could no longer be modelled using discrete-time data on normal conditions or average values. For example, information relating to the extent of the detrimental effects on a coastal habitat biota due to infrastructure failures resulting from storm water overflows, the causal factors (i.e., local and/or global) or on how these factors influence the system, is required to resolve resource and infrastructure management and land development issues.

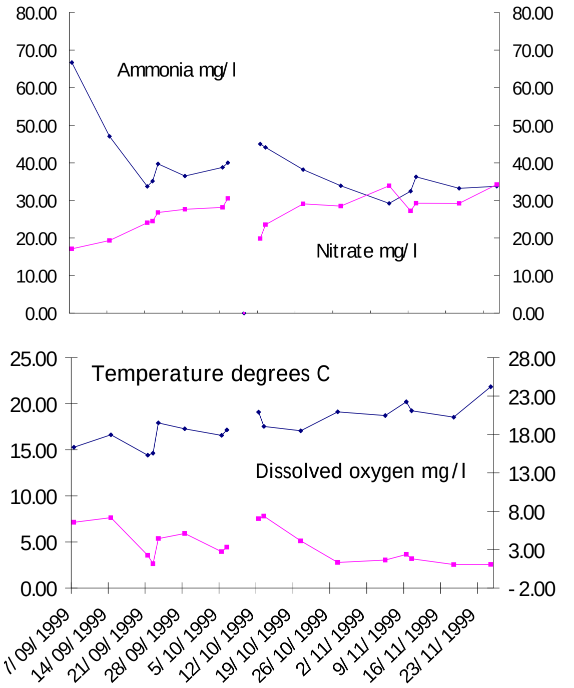

Figure 1 a & b; Graphs showing the variations in a: ammonia and nitrate, b: temperature and dissolved oxygen (DO), along the Long Bay-Okura Marine Reserve in northern Auckland, New Zealand

In the case of Long Bay Okura-Marine Reserve in northern New Zealand, ecological data available is inconsistent (in different formats) and this makes ecological dynamics modelling of the coastal habitat extremely difficult. The state institutions, such as Auckland Regional and North Shore City Councils, monitor beach water quality with many sampling locations along the northern coast of Auckland. Academic institutions as well carry out ad-hoc monitoring programmes in the Reserve (established in 1995) for scientifically validating the anecdotal evidence on the effects of urbanisation along this coastal habitat (figure 1). However, collectively analysing these data sets to model the ecological dynamics of this complex coastal system remains a difficult task. The city council efforts to study the effects of urbanisation along the beach with conventional methods show the need for better tools and data at frequent intervals on extreme conditions i.e., heavy rains.

Previous research into ecological dynamics modelling using Kohonen's self-organising map (SOM) techniques as applied to the monitoring and control of highly complex and diverse systems in industrial engineering, and their limitations (figure 2a) are explained. Finally, the paper explores some simple and complex statistical methods for resolving the issues encountered in ecological dynamics modelling of a coastal habitat using SOM techniques and discrete-time data from the Okura-Marine Reserve (figure 2b).

Figure 2 a & b: Kohonen's Self-organising map (SOM) of environmental and biological system data collected at irregular intervals. a: SOM trajectory of original data with huge gaps - difficult to study the ecological dynamics of Long Bay-Okura Marine Reserve. b trajectory - run with interpolated data calculated using simple statistical methods, shows promising results.

1. INTRODUCTION

Ecological dynamics modelling that continues to evolve is currently in need of novel approaches to meet modern day's demands of resource and infrastructure management. The use of stochastic modelling techniques with discrete-time data and continuous-time series graphs of average values is increasingly becoming inadequate in solving contemporary resource and utility service management as well as development issues. Most of the issues involve natural system behaviour during extreme conditions, such as flooding due to heavy rains. Since the middle of last century, more so in the last decade analysing data on anthropogenic factors, local as well as global, integrated with their environmental effects, within an ecosystem framework has become a vital issue in making decisions relating to a city's infrastructure maintenance and upgrade. For instance, information on the effects of global warming and climate change on a city's public utilities i.e., water supply and quality when a month's rain fall is experienced on a day, is required to strengthen the infrastructure systems and services (Jollands, et al. 2007).

2. NEW ZEALAND EXAMPLES

The paper initially looks at two recent examples from New Zealand that best illustrate the urgent need for models that provide information on the ecological dynamics/ systems behaviour of complex natural habitats under conditions never modelled before (i.e., extreme climatic conditions or extensive land development). The second part of the paper explores the use of statistical methods to enhance modelling natural system changes within an ecosystem framework by further developing Kohonen's self-organising map (SOM) based techniques using discrete-time data from multiple sources on a coastal habitat from northern New Zealand, the second example.

2.1. Hamilton infrastructure systems and services (ISS)

Jollands, et al. (2007) elaborated upon the data needed to model infrastructure systems and services (ISS) in Hamilton, New Zealand, to resist the adverse effects of climate change. The authors pointed out the need to use of daily levels instead of monthly time step when modelling the effects of climate change, its environmental impacts on the infrastructure for future research. The models of shorter time intervals could provide more information on the peak events, their frequency, impact on the infrastructure and specifically, the connections between the factors analysed for the determination of the services and the systems concerned. The important aspect discussed in the Hamilton study, also relevant to this research, is the lack of data on extreme conditions. This is a major impeding factor in modelling any public service infrastructure strength adequate to withstand the conditions or preserving natural habitats. It is significantly critical in modelling an extreme scenario relating to a natural habitat, as the environmental effects caused by a change are non-linear, highly complex and more importantly inherent to the habitat being analysed.

2.2. North Shore utility services and infrastructure

In northern New Zealand, North Shore's rapid population growth led the City Council to study the impacts of immediate growth (up until 2005 and beyond) on its utility service infrastructure. The impetus for this was "To establish and meet the community's beach water quality expectations" using sampled data from sewer, storm water systems and the beaches, including an Wastewater Treatment Plant ocean outfall (Heijs, 1999a: 3). Among the council's efforts initiated, project CARE was set up in response to pure pressure from the public to analyse the water pollution levels within the city's wastewater network caused by operational problems experienced due to flows over design capacity in its fast growing East Coast Bays and Oteha Valley Area regions (Heijs, 1999 a & b). The City Council began its water quality monitoring along the city's east coast with an aim of finding the sources of bacterial pollution and its impacts on the beaches. It did not include any other factors, such as sediments, nutrients, heavy metals and petroleum hydrocarbons, also identified as major contributing factors for the prevalent biodegradation in the coastal habitat. Heijs (1999b) having identified the use of average values of any relevant parameters as insufficient under typical wet weather conditions observed, used total data, such as number of hours that exceeded certain conditions i.e., average recurrence interval (ARI), i.e., ARI of 6 months or *Enterococci* count exceeded 100Ec/100ml, in his study.

Given that background on New Zealand's North Shore City Council efforts and the data issues in modelling the ecological dynamics of the city's costal habitats, the next section looks at the academic research conducted in a Reserve located along the same coast.

3. ECOLOGICAL SYSTEM MODELLING

Research leading up to the investigation into the use of SOMs to model the ecological dynamics of the

Long Bay-Okura Marine Reserve illustrate the compelling requirement for better information on the cause and effects of this system. "Stochastic urban accretion (SUA)" is the term Buckeridge (1999) used to describe the environmental situation in reference to the apparent lack of planning within city development on North Shore. The city council authorities failed to anticipate the significant impact caused upon the surrounding environment, especially on the coastal environment due to rapid urbanisation since the early 1990s and continued with granting permissions for new land development projects. The lack of any monitoring of environmental change and improper impact assessment on proposed development increased the load on the existing public utility infrastructure beyond its capacity, Heijs (1999a) as well pointed out this (see section 2.2 for details). As a result of the increased load on the services silt runoff and sewage infiltration continued to cause degradation to biodiversity at Long Bay until the late 1990s, without any measures being taken to improve the services such as ageing sewage, storm water systems and roading. Couriel et al. (2000) identified the following from the sewage and storm water systems that contributed to the degradation in coastal and marine biodiversity:

- continuously increasing wastewater pumping station overflows,
- storm water leaks into wastewater systems,
- storm water infiltration into wastewater systems and
- wastewater leak into ground water.

In view of the above stated need for better methods to analyse the highly complex coastal habitat, Shanmuganathan (2004) investigated into the use of self-organising map (SOM) based techniques to model the ecological dynamics of this Reserve's intertidal zone using multi-sourced discrete-time data as applied to industrial system process modelling. The next section gives an outline of this approach and the issues encountered in this regard. Consequent sections discuss simple statistical methods that show promising results to overcome the major constraint faced in modelling the ecological dynamics of complex habitats.

3.1. SOMs in ecological system modelling

In previous work, Shanmuganathan et al. (2001) illustrated how SOM methods could be best applied to analysing often 'cryptic' ecosystems in a manner similar to that applied in modelling highly complex and diverse industrial system processes for product quality/ cost optimisation purposes by Simula et al. (1999). A SOM is a single layered artificial neural network with an unsupervised training algorithm. Unlike supervised neutral nets, SOM techniques enable analysts to model multi dimensional data even without knowing their class membership. At the end of the SOM training process, similar data points (e.g., with similar vectors) within the complex data set being analysed are projected onto a low dimensional output layer with details in the raw data preserved. Hence, SOMs provide an excellent tool for visualising multidimensional data sets on low (usually 1 or 2) dimensional displays otherwise found to be difficult with standard statistical methods. Standard statistical methods are good at studying simple statistics (i.e., mean, standard deviation and so forth) of low dimensional data sets. On the other hand, rigorous statistical methods are not useful in analysing disparate data sets retrospectively and integrated.

In the original research on the Long Bay-Okura Marine Reserve, the SOM based data analysis produced promising results. A SOM created with reserve's physical and biological system data (multi sourced and discrete in time) distinguished the various littoral layers (lower supra, upper, mid and lower littoral) within the intertidal zone and their changes along with the underlying reasons for the different changes observed (figures 3 a-d) even though the data set had many missing values. The following are the three main clusters of the SOM created with 2000 nodes (figure 3 a);

C1: lower supra & upper littoral

C2: mid & lower littoral

C3: data from the above two with undesirable conditions (ecosystem state) with low DO, high nitrate, high temperature and high *Enterococci* count values.

The SOM components (figures 3c) show the factors for the SOM clustering observed (figures 3 a & b). SOM cluster (C1-C3) profiles (figure 3c) show the major difference between them.

3.2. SOMs in ecological dynamics modelling

In addition, Simula et al. (1999) illustrated the use of SOMs to track industrial system dynamics and it is possible to apply the same to modelling the ecological dynamics of natural habitats using continuous-time series data. In a SOM, as it preserves even minor details, a small change in one of the attributes of a data point would move the data to a different position from its current one in the SOM map. This is useful in tracking the ecological dynamics of complex natural systems.

With this feature, Simula et al. (1999) predicted complex industrial system process failures by running the systems vital parameter readings recorded online (i.e., running a set of time series data, on a SOM created with the systems process data previously recorded on/off line). This is a successful approach applied to modelling complex process dynamics, such as, electric/ manufacturing plant entry towards any failure in advance, provided there is sufficient reliable digital data on the process. Hollmen and Simula (1996) as well as Simula et al. (1999) predicted the output quality of a manufacturing system process using the measurements of incoming raw material characteristics and process parameter settings. However, when applied to running the Long Bay time series data sets of this coastal habitat the trajectories failed to produce a smooth progress due to huge gaps in the data sets (figures 1a & 3a). The reasons for the gaps in the data set analysed herein are, firstly, it had water quality monitoring data sampled 3-4 days consecutively and then with a 5-6 day gap before next water sampling. Secondly, biologists who observed the growth/ death of sciaphilic organisms, took photographs of four m^2 sized monitoring stations (figure 4 a-d) on a monthly basis and this makes collective analysis of the two sets of data impossible with conventional methodologies (see figures 1 a, b & 2/3 a).

4. STATISTICAL METHODS WITH SOM BASED ECOLOGICAL DYNAMICS MODELLING

The paper explores the use of two common statistical methods, they are: 1) interpolation of extra data points using a simple formula (1) and 2) the use of multiple regression analysis to study the trends in nitrate and ammonia data collected.

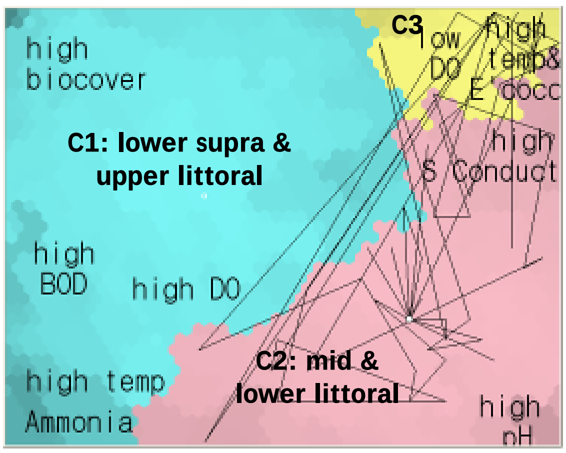

Figure 3 a. trajectory of upper littoral data with huge gaps - difficult to track the ecological dynamics of the Long Bay-Okura Marine Reserve on a SOM map (2000 nodes) using the reserve's physical and biological system data.

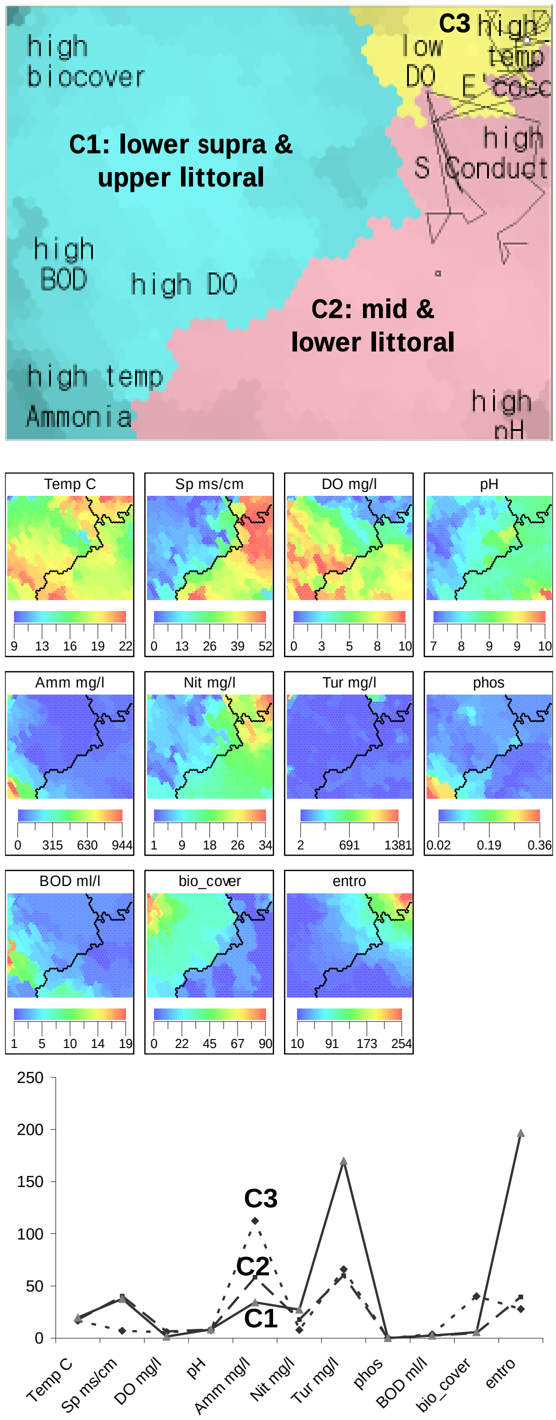

Figure 3 b, c & d; b: trajectory - run with interpolated data calculated using formula (1). c: SOM components showing the different cluster profiles. d: cluster profiles showing the variations in ammonia, turbidity and *Enterococci* count among the different littoral layers within the intertidal zone at Long Bay's Okura Marine Reserve in northern New Zealand. Please note C3 is an undesirable area (ecosystem state) with low DO, high nitrate, high temperature and high *Enterococci* count values. C1 consists of lower supra & upper littoral near land whereas C2 has mid & lower littoral mostly.

Figure 4 a-d: photographs of a: lower supra, b: upper, c: mid & d: lower littoral within the intertidal zone showing their characteristic zoning of sciaphilic organisms (that encrust the rock surfaces). Sediments and pollutants from urbanisation cause extensive degradation in these organisms by smothering & blocking their filter feeding systems, such as in barnacles.

4.1. New points added

New data points added to the original time series data set, using a simple formula (1) enhance the ecological dynamics modelling of this complex coastal habitat. The new sets of points between observations were added beginning with the missing i+1st value for the jth variable, where a gap in the recording of the values exists (as in Table 1). The trajectory ran using the new data set (with original and added), looks more smooth and useful in studying the upper littoral (within the intertidal zone) ecological dynamics of the Long Bay-Okura Marine Reserve (figures 2/3 b).

$$a_{i+1,j} = a_{i,j} + (a_{M,j} - a_{F,j})/(M-F) \quad F<i<M \qquad \textbf{(1)}$$

where the variable 'a' is the value and 'M' is the row number of the observation prior to the gap and F is the row following the gap.

4.2. Prediction of dependent variable values

In this section, multiple regression analysis is performed to predict a dependent variable values, firstly nitrate (Nit mg/l) and then ammonia (AMM mg/l) using the other determinant variables, temperature (Temp ^{o}C), pH, specific conductivity (Sp ms/cm), turbidity (mg/l), dissolved oxygen (DO mg/l) and ammonia/ nitrate. The nitrate and ammonia values predicted are then plotted against real values (figures 5 a & b) in which, nitrate values look good with daily data (in the centre of the

Table 1: original water sampling data of upper littoral of the Okura Marine Reserve.

Date	Temp C	Sp ms/cm	DO mg/l	pH	Amm mg/l	Nit mg/l	Tur mg/l	bio_cover
7-Sep-99	15.29	50.897	6.56	8.12	66.68	17.11	23.8	2
14-Sep-99	16.63	46.632	7.16	8.18	47.05	19.35	40.8	2
6-Oct-99	17.16	50.966	3.32	8.23	40.01	30.55	10.5	10
12-Oct-99	19.09	49.326	7.04	8.16	44.99	19.83	29.7	10
13-Oct-99	17.55	51.109	7.36	8.21	44.13	23.55	48.6	10
20-Oct-99	17.06	51.530	4.13	8.24	38.20	29.08	4.0	10
27-Oct-99	19.12	51.064	1.33	8.33	33.89	28.47	21.1	10
5-Nov-99	18.72	47.596	1.64	8.30	29.24	33.89	78.6	10
9-Nov-99	20.21	45.192	2.38	8.31	32.46	27.21	31.8	10
10-Nov-99	19.24	49.068	1.82	8.26	36.28	29.25	88.3	11
18-Nov-99	18.54	50.478	1.05	8.27	33.20	29.23	13.4	11
25-Nov-99	21.84	48.954	1.07	8.29	33.75	34.23	13.4	11
1-Dec-99	19.05	50.726	0.88	8.36	31.66	30.64	3.5	12

Table 2: multiple regression details of nitrate & ammonia predicted from temperature, specific conductivity, dissolved oxygen, pH, turbidity and ammonia/ nitrate of the upper littoral zone

Regression Statistics	Nitrate	Ammonia
Multiple R	0.922149	0.948294259
R Square	0.85036	0.899262002
Adjusted R Square	0.722096	0.812915146
Standard Error	2.9148	4.389336007

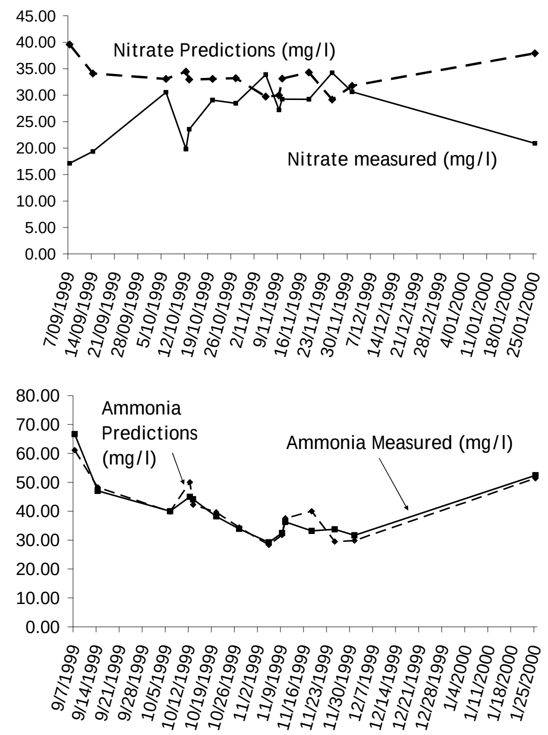

Figure 5 a & b: nitrate / ammonia predicted using multiple regression of temperature, pH, specific conductivity, turbidity, dissolved oxygen and ammonia / nitrate measured (mg/l)

graph) and the weekly ones are not (at both ends of the graph). However, the ammonia values show good results throughout (see Table 2 for stats).

Finally, nitrate and ammonia trends are analysed with graphs plotted using the new data set (measured and added points calculated using formula (1)) as well as multiple regression prediction values (see figures 6 a & b). Of the two graphs, nitrate values calculated are more close to that of the real and so is the trend however in ammonia it is the opposite. During biological decomposition, ammonia is transformed into nitrate utilising dissolved oxygen, and this could be calculated in freshwater systems but not in coastal waters due to complexity issues arriving from oceanographic factors (Wilcock and Stroud 2000) and furthermore, nitrate being converted into ammonia (McElroy, 2002). Based on the results nitrate prediction is possible using the variables analysed herein nonetheless ammonia is not.

Along the beaches, nitrogen and phosphorous that reach the coastal environment cause for the lowering of DO or hypoxia. Such eutrophic conditions affect the growth of marine organisms at times leading to major alternations in the ecosystem structure. Buckeridge (1999) illustrated the major changes observed in species composition of the Reserve's coastal habitat, which led to extinction of a species *Balanus trigonus* locally in the late 1990s.

5. DISCUSSION

Natural systems are highly complex and extremely diverse; they are open systems with dynamic equilibration, the more we know about them the greater the complexity revealed and difficulty in modelling them (Shanmuganathan 2004). Despite this dilemma, scientist and engineers are required to transform data into useful information no matter how discrete the data sets are, for critical decision-making relating to infrastructure development and maintenance. Failure to do so has resulted in significant biodegradation, in the case of Long Bay-Okura Marine Reserve. Furthermore, the poor beach water quality observed has led to health hazards along the beaches of North Shore in New Zealand. The use of SOM based approach to modelling the ecological dynamics of this complex coastal habitat as applied to that of industrial processes has its limitations because of the gaps in the data (see section 3.2). Nonetheless, the SOM trajectory of the Long Bay upper littoral intertidal zone data set with new interpolated values generated using a simple statistical technique, produced good results. Multiple regressions as well showed satisfactory results with potential for modelling discrete data of complex natural habitats.

6. CONCLUSION

In summary, the two examples outlined in the paper

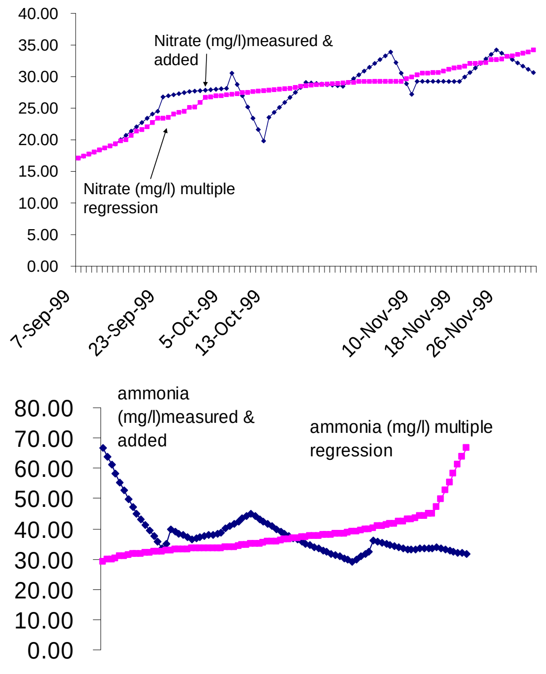

Figure 6 a & b: graphs of a: nitrate and b: ammonia trend analysis plotted with measured and added data using formula (1), and multiple regression prediction values calculated using temperature, pH, specific conductivity, turbidity, dissolved oxygen. X-axis shows only dates sampled (in table 1), the rest are interpolated data.

clearly pointed out the growing demand for models on natural systems as well as phenomena under extreme conditions that require data at shorter intervals and in the case Hamilton case study conditions never experienced before, to increase the strength of infrastructure to maintain utility services uninterrupted. This involves people's daily life as seen all over the world, and in the last decade, such interruptions caused by heavy flooding are on the increase. It is the same with pollution along the beaches. Unless proper mitigation measures are, taken biodegradation could cause unprecedented loss that could affect future generations. From the two examples, the need to model natural system behaviour under conditions never modelled before is evident and this needs to be performed using discrete data available. This is of utmost importance for appropriate decision-making relating to infrastructure upgrade and maintenance. The SOM based trajectories approach with added points to a set of inconsistent and incomplete data set using simple statistical techniques shows potential.

Natural systems are so complex they could neither be modelled using linear parameters (Clark et al. 2001) nor uncertainties relating to their behaviour resolved using simple models (Stewart-Oaten 1996). Hence, ecologists and modellers are required to develop models and techniques to study the ecological dynamics of natural habitats identifying the factors that give an indication of the prevalent trends within the whole system or with a systems approach using the existing discrete-time data sets. The paper provided a useful means for this using SOM based approaches with simple statistical methods, the former as applied to industrial process dynamics modelling with proven success. The multiple regression technique explained in the paper as well produced satisfactory results in analysing the trends within natural habitat studied.

References

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.

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.

Couriel, E. D., R. Park, et al. (2000). Project CARE. North Shore, Auckland, North Shore City Council 8pp.

Heijs, J. (1999a) Project CARE: Accelerated Options in East Coast bays and Albany/Oteha Valley FILE REF 7930-R01 POLICY REF 30 November 1999. 1-9.

Heijs, J. (1999b) Project CARE: Sources of Bacteriological Pollution on North Shore Beaches, Project CARE report FILE REF 7930-R01 POLICY REF 1-21.

Hollmen, J. and O. Simula (1996). Prediction models and sensitivity analysis of industrial production process parameters by using Self-organizing map in proceedings of the IEEE Nordic Signal Processing Symposium (NORSIG'96) 79-82.

Jollands, N., M. Ruthc, C. Bernierb, N. Golubiewski (2007), The climate's long-term impact on New Zealand infrastructure (CLINZI) project—A case study of Hamilton City, New Zealand in Journal of Environmental Management 83 462 (2007) 460–477.

McElroy Rachelle (2002) Development of a Water and Sediment Sampling & Monitoring Program for the Esquimalt Lagoon 50pp).

Shanmuganathan, S. Sallis P, and Buckeridge J (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 (ANNES2001), Dunedin, New Zealand, 22-24 Nov 2001, pp132-137.

Shanmuganathan, S. (2004) Soft systems analysis of ecosystems. PhD thesis.

http://repositoryaut.lconz.ac.nz/theses/071

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.

Sorensen Jan (1978) Capacity for Denitrification and Reduction of Nitrate to Ammonia in a Coastal Marine Sediment in Applied and environmental microbiology, Feb. 1978, Vol. 35, No. 2p. 301-305

Stewart-Oaten, A. (1996). 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.

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