Self-organising maps to study the effects of urbanisation at Long Bay in New Zealand

Subana Shanmuganathan[1], Philip Sallis[2] and John S Buckeridge[3]

[1&2]Auckland University of Technology, New Zealand

[3]Earth & Oceanic Sciences Research Centre, Auckland University of Technology

www.aut.ac.nz

(Email: [1]subana.shanmuganathan@aut.ac.nz, [2]philip.sallis@aut.ac.nz, [3]john.buckeridge@aut.ac.nz)

Abstract

Biologically inspired Artificial Neural Network (ANN) modelling methods provide a means of problem solving that incorporates heuristics with conventional algorithmic processing. Over the last few decades, new techniques for neuron relationship modelling and network architecture algorithms have been introduced to solve a wide range of problems across many fields. This paper looks into the aspects of using SOMs to a biological example to permit understanding of a complex environmental process. The preliminary research results to study the effects of urbanisation on marine life at the Long Bay-Okura Marine Reserve, situated in northern New Zealand is discussed in detail. This was the country's first urban, marine reserve to be established (1995), and resulted from growing concern of environmental groups and general public of the area. Since then many institutions have conducted research to find the cause for the observed environmental change. All these data sets are fused and analysed collectively to study the patterns in them. The use of SOMs to industrial process monitoring has been very successful in many areas and is applied here to study the process dynamics in environmental process modelling with SOM trajectories. The analyses show the relationships found in the data sets from different sources in easily perceivable formats without having to model the complex physical process.

Keywords: Self-organising maps, Urbanisation, Data mining, Environmental process monitoring

1.Introduction

Biologically inspired Artificial Neural Networks (ANNs) provide a new approach to computational modelling. They are considered as simulations for human intelligence. McCulloch and Pitts (1943) were the first to introduce a mathematical model of a neuron. Most of their work involved a simple neuron model and these single layer network systems were generally referred to as perceptions. This led to the innovations and novel approaches in providing a means to bridge the gap found in conventional computing, in anticipation that this would solve problems that needed heuristics. Restrictions found in these networks almost ended the ANN research when Minsky (1969) proved that single layer perceptrons could not learn and represent linearly inseparable solutions in solving problems. He also expressed his concerns over learning algorithms for multi-layer networks. This turned the interest of many researchers to Artificial Intelligence and Expert Systems. The connectionist network models that were first introduced in 1977 led to a considerable progress during the last two and a half decades. Researchers from a wide range of fields have been working on new ANN paradigms suitable for their own applications; of particular interest here is the Kohonen's Self-Organising Map (SOM) (1982). Its application in knowledge discovery is very significant. The SOM displays the input vectors on a two dimensional grid while preserving the topology of the original data. In recent times it is described as the best tool for data depiction enabling detection of hidden patterns. Its ability to discover implicit knowledge from numerical data is yet to be matched by any other computational modelling methods that are available today.

Many factors, such as ever increasing computing power, networking capability, communication facilities and storage capacity at constantly decreasing prices, continue to increase the capture of data. This trend and the desire to reap the potential of these stored data paved the way for the increased interest in data mining techniques [6]. Two decades on, the use of SOMs in data mining for knowledge discovery is an approach that is found to be very successful in a wide range of applications in a variety of disciplines i.e. analysing medical outcomes, predicting customer purchase behaviour, predicting the personal interests of web users, optimising manufacturing processes, predicting the stock market, financial analysis and sales in real estate investment appraisal of land properties [7-9]. This paper describes an example where SOM techniques are used to study the effects of

urbanisation on marine life at the Long Bay-Okura Marine Reserve in northern New Zealand.

The application of SOM in this analysis is fusion of data sets in an integrated format to further investigate the data, collected by various research groups. It is achieved by carrying out cluster analysis, system dependency analysis and system process modelling i.e. trajectories on the biological example, explained here. The data used in this study covers a period of two years from 1999 until recently.

2. Long Bay-Okura Marine Reserve

The Long Bay-Okura Marine Reserve was established in 1995. The impetus for establishment came from the North Shore's environmental groups and general public, who were concerned about the massive environmental degradation. Buckeridge (1999) described the situation as "stochastic urban accretion (SUA)", in reference to the apparent lack of planning within city development. The Authorities in the past failed to give required attention to the impact caused upon the surrounding environment, when granting resource consent for proposed development. Poor monitoring of environmental change and improper impact assessment on proposed development have contributed to this situation. As a result, an ever increasing load on the existing public utility systems is argued as 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 more wastewater pumping station overflows, storm water leaks into wastewater systems, storm water infiltration into wastewater systems and wastewater leak into ground water. [4]. All of this result in increased amount of untreated water entering the sea causing degradation in marine biodiversity.

As the Okura Marine Reserve was New Zealand's first urban marine reserve, the establishment of the Reserve itself was subjected to severe criticisms from the land developers of the North Shore area. Even though it was controversial, Local Authorities, Academic and Research Institutions took advantage of an opportunity to evaluate the effectiveness of the maturing systems (as part of the Resource Management Process). Research involving biomonitoring, environmental monitoring carried out by staff and students of the Auckland University of Technology and other institutions provide evidence of environmental changes, the causes for biodegradation. Their research and monitoring programmes gave a formal approach to assess the effects of urban development on marine life at the Okura Marine

Reserve. The research findings continue to emphasise that the current urban development in North Shore is unsustainable by the nearby coastal and marine environment.

3. Data from North Shore and the Reserve

3.1. Biochemical

The North Shore City Council (NSCC) carries out *Entrococci* tests on water samples taken from mid sea at Long Bay. The NSCC and the Ministry of Environment (1999) consider the bacteria as the indicator most closely correlated with health effects in New Zealand marine waters. *Enterococci* counts up to 40 are considered harmless for human bathing in seawaters. *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 i.e. sewage and control, storm water discharges, land use, such as sediment control, works in watercourse. The data from 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 and measures are being made to obtain these data from the relevant local and regional government authorities.

Biomonitoring is the use of biological responses to assess changes in the environment; generally those are due to anthropogenic causes. Biomonitoring mainly involves indicator species or indicator communities that accumulate pollutants in their tissues from the surrounding environment and thus reflect the environmental conditions. Even though it is very expensive to develop such indicator systems, the use of environmental effects of ecosystem (i.e. individuals, species, communities) biomonitroing provides information on the magnitude and ecological effects of the pollutants on the system [1].

The National Institute of Water and Atmospheric Research (NIWA) has produced models to predict the current and future sediment runoff during current and two different future development scenarios at Long Bay. The two scenarios discussed in this report are: (i) mixed development: which includes 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 lot 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 predicts the daily sediment generation for different return periods (i.e. 0 to 30 years) and for different development stages (i.e. 10 and 20 years) and is based upon [5].

The Auckland Regional Council (ARC) conducts subtidal and intertidal monitoring programmes to study the rates of sediment deposition and the impact

this has on species abundance and composition in the coastal habitats from Waiwera to Campbells Bay. It shows that Long Bay and Campbells Bay have the highest sediment depositions among the monitored sites. Further, the ARC report argues that there exists continuity in the lack of recruitment of bivalves (pipi and tuaua) to Long Bay and low numbers of these species recorded across the monitoring sites [10].

Auckland University of Technology staff and students collected the following data to measure the environmental changes caused by urban development activities in the Long Bay-Okura Marine Reserve:

1. Date
2. Place (Data collection monitoring stations)
3. Time
4. Temperature [° C]
5. Specific Conductivity (Sp Cond) [mS/cm]
6. Dissolved Oxygen [mg/l] / DO [%]
7. pH
8. Ammonia [mg/l]
9. Nitrate [mg/l]
10. Turbidity [NTU] / Turbidity [cm]
11. High Tide
12. BOD-5 [mg/l] (biological Oxygen in demand)
13. Phosphate [mg/l]
14. Air temp. [° C]
15. Special remarks

3.2. Geological

The Long Bay-Okura cliffs and shore are part of the Waitemata group of rocks, and were deposited as sediments derived from underwater avalanche deposits from the lower Miocene age [2].

The following are the different tidal zones (figure 1) with slate tiles positions in a chosen transect covering from uppermost to upper sub tidal zones:

1. Lower supralittoral
2. Upper littoral
3. Mid littoral
4. Lower littoral
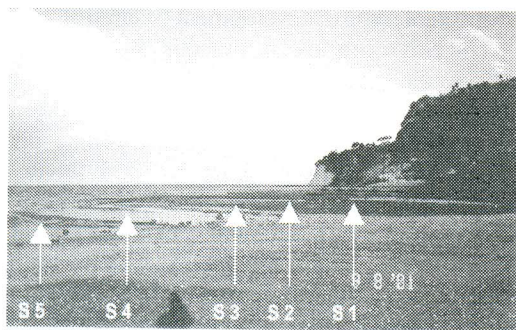5. Upper sublittoral

Figure 1: Photograph showing the data collecting monitoring stations at the Long Bay-Okura Marine Reserve

The five intertidal divisions have characteristic zoning of sciaphilic organisms (those organisms that encrust the rock surfaces). Observations are made on a monthly basis by taking photographs in order to monitor the colonisation patterns of these sciaphilic organisms [3].

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

4. Methodology

In this analysis, the data collected by Auckland University of Technology research groups to analyse the biodegradation and its causes are used together with NSCC *Enterococci* test results. The physical systems sampling of these different data collecting monitoring stations (S1, S2, S3 and S4, each station lying within the lower supralittoral upper littoral, mid littoral and lower littoral respectively) were carried out 5-6 times a month. Photographing of these plates to study the bio cover was carried out once a month. This creates wide gaps in the data sets, as the physical system data and photographing of the sciaphilic colonisation may not necessarily coincide. In order to overcome this problem the following measures are being taken:

1. Missing values in bio cover of sciaphilic colonisation data are filled by interpolation after studying their sequence. i.e. Bio cover on a day in week 2 is determined as 10% if the bio cover during week 1 and week 3 are 10%, assuming that no organisms grew and died in between.
2. Visovery® SOMine lite version 2.1 by eudaptics software gmph package is used as it can create maps even with a few missing values.

5. Results and Discussion

The data sets from the two different studies i.e. physical systems and sciaphilic colonisation research are analysed along with NSCC *Enterococci* test data. Initially a cluster analysis is carried out to see the

clustering patterns in the S1-S4 slate tile data. A SOM (figure 2) is created using Viscovery with 200 nodes, priority for all components set to 1 and other map creation parameters set to default values.

The sub divisions (S1-S4) found within the intertidal zone (figure 1) have different topography and microclimate and thus serve as ideal habitats for different species. Interestingly, these variations have been picked up in this cluster analysis. Their contribution has resulted in the categorisation of the SOM (figure 2) into two distinguished areas discussed below. The SOM (figure 2) also consists of component windows. Interpretations arrived from this map are:

1. The data collecting monitoring stations S1 and S2 are separated from S3 and S4 by a diagonal line running from the upper right to the lower left of the map except for six nodes in the right bottom. No data has been included about their location description, yet the self-organisation of the results shows them separated into two categories. Please note that S1 is the closest station to the landside and as the number increases exposure time to seawater increases.

2. The S3 and S4 data, grouped in the S1 and S2 cluster of the map are, S4 in March 1999, S3 in October 1999, S3 and S4 in November - December 2000 and S3 and S4 during the months of January to April 2001. Incidentally, this feature (S3 and S4 falling in the S1 and S2 cluster) is more in summer. It gives an indication that during the summer on occasions S3 and S4 show similar attributes (i.e. more dried) to S1 and S2 monitoring stations. This was more common in 2001 than in 2000.

3. Component planes of Sp Cond, pH, Nitrate, bio cover and Entrococci count show a good correlation to the clusters map. This means Sp Cond and pH are higher in S1 and S2 monitoring stations than in S3 and S4. Under normal conditions it should be the other way around, pH and Sp Cond of S1 and S2 should be lesser than that of S3 and S4. This may be due to evaporation resulting from high temperatures experienced in S1 and S2, which in turn, result in higher concentrations of salts showing increased pH and Sp cond values. Bio cover is higher in S3 and S4 than in S1 and S2.

4. Bio cover is high in the left top corner in S4 and is decreasing towards the right side. Enterococci counts are high in the top left with the highest at the centre top and decreasing towards the right bottom corner. Enterococci counts are high during May to September 1999, November and January 2000 and in January 2001. Interestingly, during this period bio cover is quite low and the temperature values is found to be quite high. Please note that Enterococci tests are carried out on water samples from mid sea and not on water
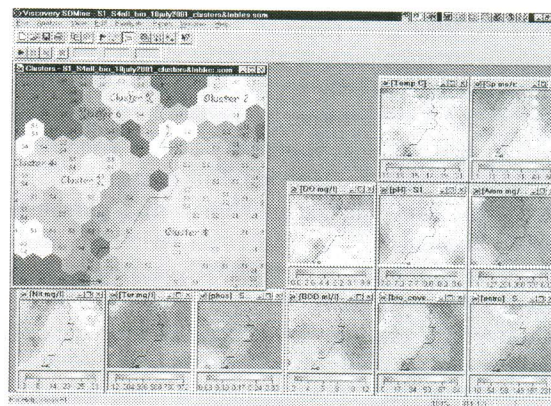

Figure 2: SOM of the Long Bay - Okura Marine Reserve physical and biological system data with all components

samples near the monitoring stations. Still considered as an indicator for pollution in seawaters and this study.

Figure 4 shows above details superimposed on the clusters map where the relationship between the environmental change and bio cover can be easily visualised. Apart from the component analysis the following are derived from the SOM clusters (figure 2):

1. In the map, S1 and S2 monitoring station area can be distinguished into two major clusters (i.e. cluster1 and cluster 2) and they directly correspond to the Enterococci counts and nitrate concentrations in the area.

2. S3 and S4, the left diagonal part seems to have four clusters: cluster 3, cluster 4, cluster 5 and cluster 6.

3. The bottom left corner has a patch with the highest values for BOD, Ammonia, DO, and phosphate along with high temperature.

4. Bio cover is high at the top left corner where temperature is between 11-17°C, low Enterococci counts, quite high DO and slightly high Ammonia and nitrate values.

The SOMs of dependency analyses of individual monitoring stations are also interesting where the effects of urban development on these organisms can be studied (figure 3). Environmentally unsustainable anthropogenic activities lead to certain conditions i.e. high BOD, high nutrient levels (i.e. nitrate, ammonia) that affect the growth of marine organisms causing alternation in ecosystem structure. Buckeridge (1999) pointed out that in the past, on two different occasions between the period 1996-1999, during warmer periods, high nutrient levels produced eutrophic conditions, which affected the presence of E. plicata and the normally prolific intertidal algae Corallina officinalis both resulting in cyanobactria colonies dominating the ecosystem at this Reserve. Hence, the

135

SOMs created with physical and biological system data together should reveal the relationship in them.

S1: The SOM of S1 includes a patch of 15% bio cover, at the top left corner. *Enterococci* count is very high on the right top corner at 254 and getting decreased towards the left side. Nitrate, has its highest value at the right side and ammonia with its highest on the left bottom corner. Bio cover, with the highest value 15% corresponds to high DO values on the left side. S1 plate was fixed on this position quite recently and therefore, as usual, it took a while for the initial settlement of sciaphilic organisms.

S2: High BOD, Ammonia, Nitrate, phosphate values and *Enterococci* counts do not seem to cause any harm to the bio cover according to the map components. But usually these conditions do harm sciaphilic organisms. The reason for this is that the bio cover in S2 ranges from 2-12%. The highest value is observed in the lower left area, where low values of any of the above could not have caused a negative influence on these organisms and 12% cover is quite low in terms of growth.

S3: Bio cover is very high at the top left corner with 54% where ammonia, nitrate, phosphate and *Enterococci* counts are low, shows a good correlation. BOD is very high at the bottom right and getting reduced towards the upper edge. *Enterococci* count is very high at the top right corner and very low at the left corner.

S4: Bio cover has its highest value at the bottom left corner with 90% where ammonia, nitrate and *Enterococci* values are reasonably low compared to other areas. *Enterococci* have the highest value at the top right corner and the lowest at the left corner and other areas ranging from 1 to 127 ml/l. Ammonia has its highest value at the bottom right corner and nitrate at the right lower edge. Both values are getting decreased towards the top left corner. DO has a corresponding and coordinating pattern to bio cover. Phosphate has the highest value at top right corner and getting decreased towards lower edge, but has a peak at the lower left edge. BOD has its highest value of 18.63 at the bottom left corner where bio cover also has its highest value of 89.99%, which means BOD is not effecting the bio cover in the S4 area.

It is also possible to observe the process dynamics with the use of trajectory maps. Figure 4 shows areas of different conditions, the periods and the monitoring station in which these conditions are experienced. The animation of trajectories in these maps can be used to study a particular monitoring station's progress in time under various conditions. Shown in the figure 5 are the progress dynamics of monitoring stations S1 and S3 in overall SOM of S1-S4. It also could be used with simulated values to predict the bio cover under different sediment depositions or development scenarios. For this, data on sediment depositions has to be incorporated as a component in the creation of SOMs. This technique can be applied
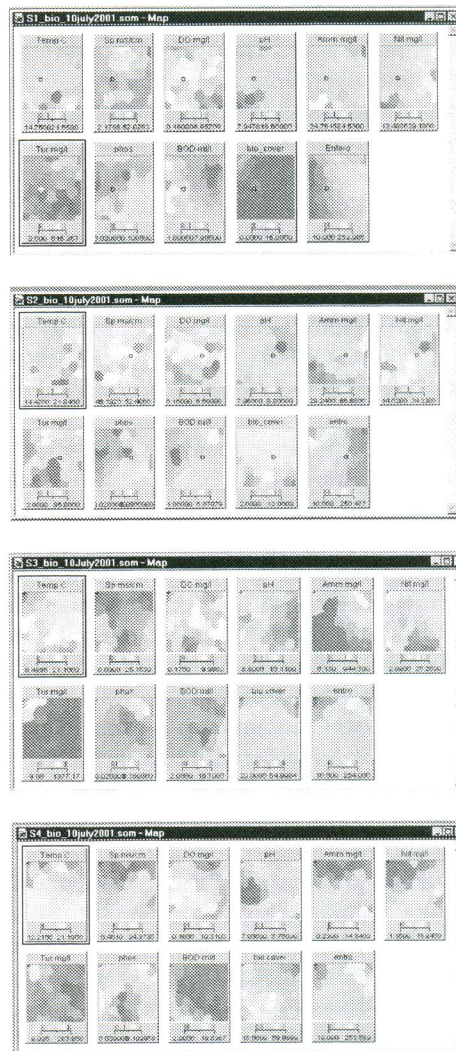

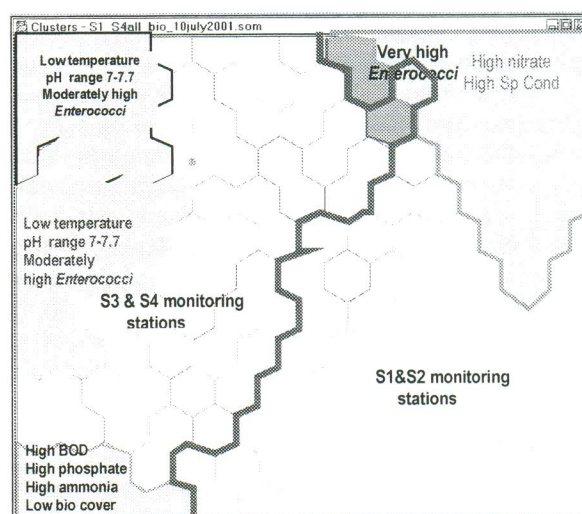Figure 3: Top to bottom self-organising maps (SOMs) of S1, S2, S3, and S4

Figure 4: Superimposition of the component map details on cluster map of S1 - S4 data collecting monitoring stations
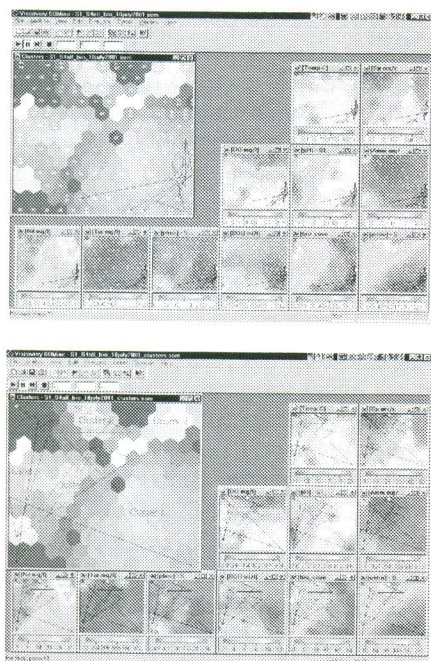
Figure 5: Trajectories of S1 and S3 on S1- S4 SOM

to analysis of other coastal data individually and collectively.

6.Conclusion

In summary, it could be stated that with SOM analyses it is possible to relate and analyse the environmental changes with biological responses in formats that are highly visual. The patterns in the data can be analysed directly linking the causes and responses in environmental process modelling.

The results support the hypothesis to apply complex industrial process modelling to environmental process modelling.

The fused data of physical systems and biological systems used in this study had missing values at critical points as different research groups have collected them on different days. Not withstanding, the SOMs created with extrapolated values give a good correlation in the study. It should be noted that we could not carryout any constructive analysis using conventional statistical methods on these fused data sets. Previous research groups carried out their own analyses of data collected and studied the physical and biological systems separately.

7.Future work

Previous studies by the University of Auckland for the Auckland Regional Council show that the sediment deposition, generated by earthworks are on the increase at Long Bay since last two years. [10]. It

also argues that there is an observable variation in the bio diversity of the coastal environment, caused due to natural causes or induced by increased sedimentation. The National Institute of Water and Atmospheric Research has produced models showing the possible sediment generation under different future development scenarios [5].

The North Shore City Council has developed models to predict the storm water based pollution for the North Shore area [4]. We intend to apply the same hypothesis and to search for any possible evidence, correlating the data on bio diversity to the predicted model data sets on environmental change, using self-organising maps.

Acknowledgements

References

[1] BiologicalMonitoring, 2001. http://h2osparc.wq.ncsu.edu/info/biomon.html

[2] J.S. Buckeridge. Land stability in Urban Sites of the North Shore, Auckland, NZ. in *In Landslides. Glissements de terrain. Bell, D H (Ed). Proceedings of the Sixth International Symposium on Landslides.* A A Balkema Publishers, Rotterdam, 1995.

[3] J.S. Buckeridge and B.A. Tapp. Ethics and that Ethic called Sustainability. in *Australasian Environmental Engineering Conference.* Auckland, New Zealand, 1999.

[4] E.D. Couriel, et al., Project CARE, North Shore City Council: North Shore, Auckland. pp 55, 2000.

[5] M.O. Green, et al., Long Bay Sedimentation Study: Stage I (Sediment Generation) Stage II (Sediment Fate). NIWA Client Report NSC00209: pp 1-98, 2000.

[6] T.M. Mitchell. *Machine learning and Data Mining.* in *Communications of the ACM*, 1999.

[7] P. Sallis, et al., A methodology for profiling users of large interactive systems incorporating Neural Networks for Data mining Techniques, 1998.

[8] D. Shalvi and N. Declaris, An Unsupervised Neural Network Approach to Medical Data Mining Techniques, 2000.

[9] O. Simula, et al., Analysis and Modelling of Complex Systems Using the Self-Organizing Maps, in *Neuro-Fuzzy Techniques for Intelligent Information Systems*, N. Kasabov and R. Kosma, Editors. pp 3 - 22, 1999.

[10] J. Walker, et al., The Long Bay Monitoring Program Sampling Report - July 1999 / June 2000, pp.1-72, 2000.