Spatial ecology and artificial neural networks: modeling the habitat preference of the sea cucumber *(Holothuria leucospilota)* on Rarotonga, Cook Islands

Darrin Drumm [1], Martin Purvis[2] & Qingqing Zhou[2]

[1]Department of Marine Science
University of Otago. Dunedin, New Zealand
Phone: +64 3 479-8306 Fax: +64 3 479-8336
Email: druda019@student.otago.ac.nz

[2]Department of Information Science
University of Otago. Dunedin, New Zealand
Phone: +64 3 479-8318 Fax: +64 3 479-8311
Email: mpurvis@infoscience.otago.ac.nz

Presented at SIRC 99 – The 11th Annual Colloquium of the Spatial Information Research Centre
University of Otago, Dunedin, New Zealand
December 13-15th 1999

ABSTRACT

In this paper we discuss the ability of artificial neural networks to predict the habitat preferences of the tropical sea cucumber *(Holothuria leucospilota)* in the reef-top ecosystem of Rarotonga, Cook Islands. We suggest that ANN's combined with geographic information systems (GIS) may provide an effective method for modeling spatial patterns in ecological data. The model that we have developed is applicable to similar investigations of *H. leucospilota* from other geographic regions.

***Keywords and phrases*:** tropical marine ecology, artificial neural network, spatial modeling, habitat, sea cucumber, Cook Islands

1 INTRODUCTION

In the Pacific Islands, invertebrates including sea cucumbers are among the most valuable inshore fisheries resources (Dalzell *et al.*, 1996). The multi-species tropical sea cucumber fishery throughout the Pacific Ocean has existed for thousands of years. However, unsustainable harvesting rates can contribute to local species depletions and/or extinctions. As human activities continue to force substantial impacts on coral reef ecosystems, the management of inshore fisheries in the Pacific Islands is becoming an increasingly important priority. Effective management plans must be developed for these fisheries.

The study area includes the entire shallow-water ecosystem of the island of Rarotonga, Cook Islands (Figure 1). The sea cucumber *(Holothuria leucospilota)* forms an important part of the traditional subsistence fishery on Rarotonga, yet little is known of this species present spatial distribution and abundance around the island. Sperduto and Congalton (1996) recognize that knowledge of the distribution, biology and habitat requirements of a species are important elements necessary for its conservation.

Because environmental and ecological variables are non-linear in nature, artificial neural networks (ANN's) offer an advantage over traditional linear regression (LR) and multiple linear regression (MLR) analysis techniques, because they do not introduce any prior assumptions about the relationships between the variables. Artificial neural networks have been used to advantage in the solution of non-parametric problems (Mann and Benwell, 1996) due to their ability to discover patterns in the data that are not possible to detect using LR and MLR models. This ability has made neural networks an increasingly popular tool for modeling and prediction of ecological datasets (Baran *et al.*,1996; Lek *et al.*, 1996; Guan *et al.*, 1997; Mastrorillo *et al.*, 1997; Recknagel *et al.*, 1997, Özesmi and Özesmi, 1999).

Figure 1: Location of the Cook Islands

2 NEURAL NETWORK MODELING

Neural network models are inspired by natural physiology and mimic the neurons and synaptic connections of the brain. Figure 2 shows the architecture of a typical feed-forward neural network model. It has three layers: a layer of input nodes, an intermediate layer of nodes and a layer of output nodes. Connections are from each input node to each intermediate layer node, and from each intermediate layer node to each output node in a feed-forward manner. These nodes of different layers are the processing units of the network. A node has an output value, which is determined by a nonlinear activation function of the sum of its inputs, which are the weighted values of the outputs from other nodes that feed into it. Each internode connection has a weight (positive or negative), which is multiplied by the source node output to produce an input for the destination node.

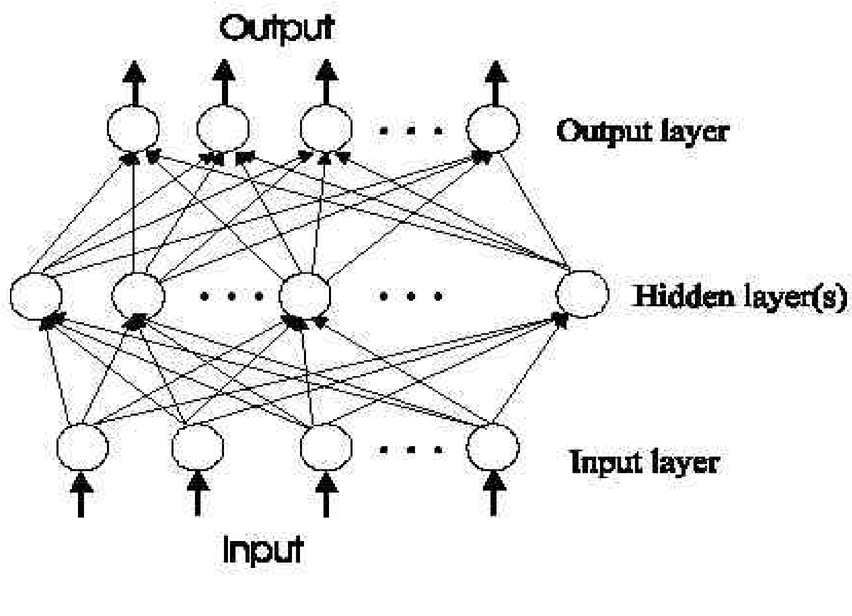

Figure 2: A feed-forward neural network model

Although simple in structure, the systemic behaviour of a network in which many such nodes as connected together can be very complicated. A feed-forward neural network works in the following way. The input values are presented to the input nodes, which feed into and activate the hidden layer(s). The values produced by the activation functions at the hidden nodes, in turn, feed forward into the output nodes, and then determine values of the output nodes. The training of neural networks employed in this paper is accomplished by means of "supervised" learning, wherein the model outputs are first compared directly with a set of known, correct values and an error function is computed. The weights associated with the internode connections are then systematically adjusted using a gradient descent procedure so as to minimise the error function. Then the next "correct" training example is presented to the network, and the error minimisation calculation is repeated. After many passes through the whole training set, the error usually converges to some value and, provided that this is error value is small enough, the network can be said to have learned a set of relationships.

3 KNOWLEDGE EXTRACTION FROM NEURAL NETWORKS

In the past few years there has been a growing interest in neural networks. Although there are many forms of neural network, they generally have two important factors in common. They can learn from examples and they can generalize, that is, they can respond suitably both to the data taught and to novel, or "test", data. Once trained for a given task, a network can then be used in the application for which it has been trained, by providing suitable data on the network inputs. However, the network remains a black box. Appropriate outputs are generated for any set of inputs, but the user usually has no direct way to understand why or how particular results are obtained. Over the past several years, there have been several efforts made to find effective algorithms to extract rules from trained neural networks in the past years (Towell *et al*, 1993; Fu, 1994) and thereby derive knowledge that is more readily understandable concerning the mapping of inputs to outputs.

"NeuroLinear" approach (Setiono *et al*, 1997) is one of the rule extraction algorithms, and it can be used to extract oblique decision rules from trained feed-forward neural networks. Rule Extraction is performed in following steps:

1. Use the *Chi2* algorithm (Liu *et al*, 1995a) to discretize the activation values of the hidden units.

2. Extract a set of rules that describe the network outputs in terms of the discretized network activation values. The *X2R* approach (Liu *et al*, 1995b) generates a set of rules which cover the most frequently occurring pattern with an error-rate not exceeding the inconsistency rate present in the patterns.

3. For each discretized hidden-unit activation value, generate a second set of rules in terms of the network's inputs.

4. Merge the two levels of generated rules into a single rule set.

4 SPATIAL DATA PREPARATION

A geographic Information system (GIS) was used to facilitate the sampling design and the spatial analysis of the physical habitat and biological attributes. The habitats of the shallow-water (backreef and lagoon) ecosystem were classified into broad zones exhibiting morphological homogeneity. The boundaries of these zones were digitised in the GIS and were delineated by the spectral signature of aerial photographs and were later confirmed by ground truthing. Habitat maps were produced using the broad classifications: reef rim, rubble/rock, sand/coral matrix, sand, mudflat and passage/harbour. Sample sites were randomly located and their allocation was weighted according to the relative percentage of each habitat zone as determined from the GIS spatial model (Figure 3). A total of 128 sites were sampled for environmental and biological variables, using 2m x 50m (100 m^2) strip transects. This size sample unit was selected to account for the patchy distribution of the animals. At each sample site, 10 environmental variables were recorded including the exposure (windward or leeward side of the island), and the following microhabitat variables were recorded as a percentage of the total area sampled: %sand, %rubble, %consolidated rubble, %boulder, %reef rock/pavement, %live coral, %dead coral, %mud/silt, and %gravel. In addition to the environmental variables, the abundance of the sea cucumber *H. leucospilota* encountered along each transect was recorded. The location of each sample site was mapped using differentially corrected global positioning system (GPS) position data to provide an accurate spatial model, integrating the environmental and ecological conditions present at each site. The spatial dataset containing these variables was then prepared for integration with a neural network.

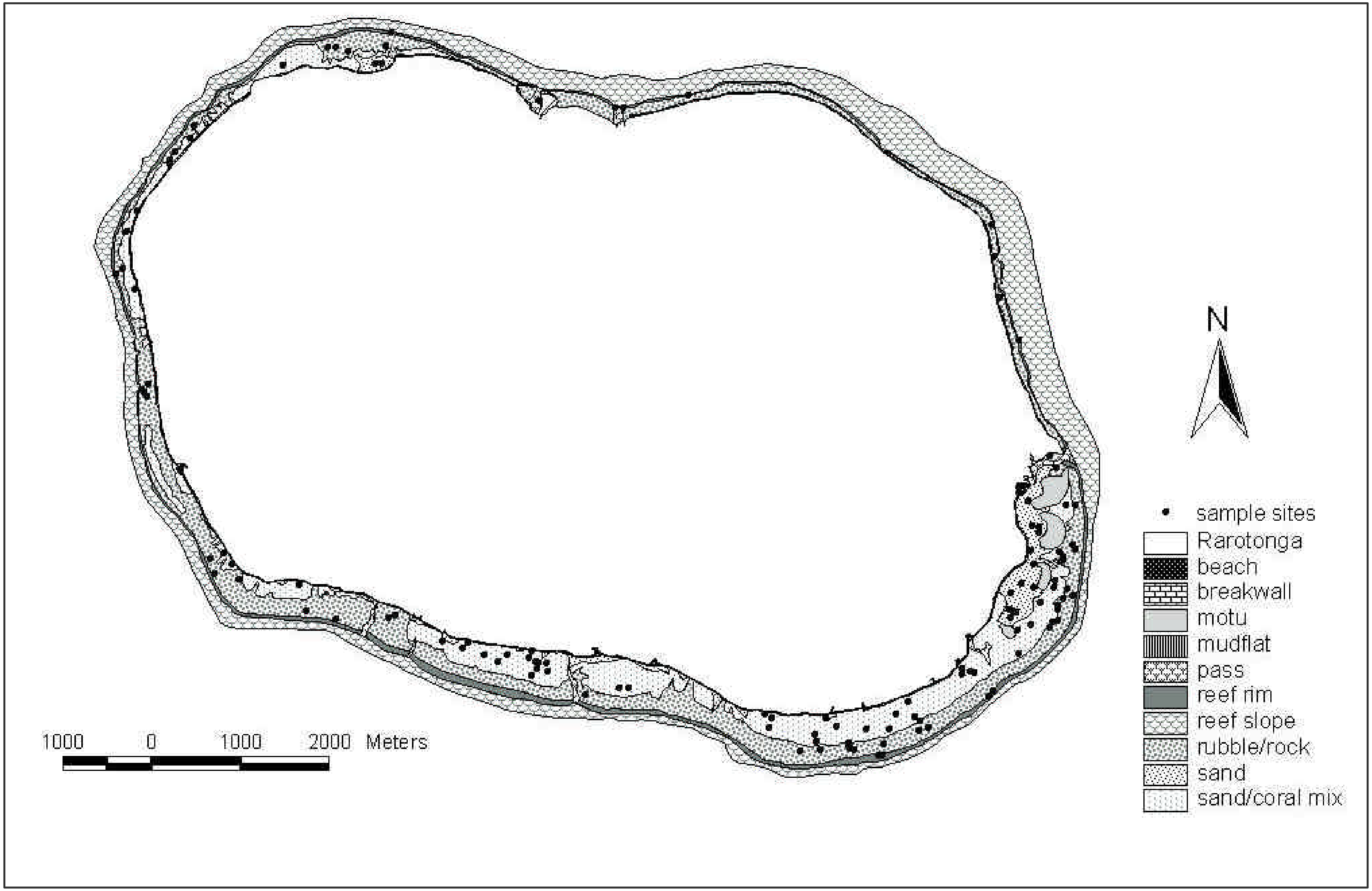

FIGURE 3: BROAD REEF-TOP HABITAT ZONES ON RAROTONGA, COOK ISLANDS SHOWING SAMPLE SITES

Ten environmental attributes which are expected to affect the habitat preference of sea cucumber are shown in Table 1.

Attribute	Description	Possible values/Range
I0	Exposure	0(Windward) or 1(Leeward)
I1	Percentage of sand	0%-100%
I2	Percentage of rubble	0%-100%
I3	Percentage of consolidated rubble	0%-100%
I4	Percentage of boulder	0%-100%
I5	Percentage of rock/pavement	0%-100%
I6	Percentage of live coral	0%-100%
I7	Percentage of dead coral	0%-100%
I8	Percentage of mud/silt	0%-100%
I9	Percentage of Gravel	0%-100%

Table 1: The related attributes of habitat preference of sea cucumber

All observed sites were divided into two distinct classes according to the frequency of animals at each location within the $100m^2$ sample unit. We named the first class as "average condition class" and was characterized by having < 1 animal/m^2. The second class of habitat, named "good condition class" was attributed to sites with > 1 animal/m^2. 'Good' condition sites were given a target value of 1, while 'average' ones had a target value of 0. The details are shown in Table2.

	Number of Sites (Patterns)	Animal Frequency Range
Class 0 (average condition class)	104	0 – 93
Class 1 (good condition class)	24	114 - 918

Table 2: Site condition class ranges

A standard re-sampling method, called *cross-validation,* was used to divide the available data set. In this procedure the available data is divided into 24 different blocks. In each block, 23 data patterns from class 1 were selected as training samples, with the remaining one used as a test sample. Similarly, 23 training data patterns and 1 test pattern were randomly selected from class 0. . For each of the 24 blocks, a different, unique test sample was selected from class 1, thereby ensuring that each of the 24 patterns of class 1 had chance to be included in the training set. Therefore, 24 different training data sets and test sets were ready for neural networks training. Thus there was a total of 46 patterns in each training set; each test set included only two patterns that categorized to class 0 and class 1 respectively.

5 NEURAL NETWORKS TRAINING

A standard three layer feed-forward neural network architecture with a single hidden layer was used. The number of input units was 10, and the number of nodes in the output layer was 2. Back propagation was used as neural network training method. In order to extract useful knowledge from networks, a pruning algorithm (Setiono, 1997) was also used to reduce the complexity of the networks by removing the redundant connections of the units in the network.

Twenty-four different neural networks were trained by using the different training sets and test sets. The number of hidden units among these 24 neural networks was either 3 or 5. The training accuracies of all 24 neural networks were over 85%, with the highest training accuracy at 100%, and the average of training accuracy at 90%.

The average accuracy of the 24 neural networks on the test set was 83.3%. Sixteen of these neural networks produced 100% test accuracy, and the other 8 neural networks had 50% test accuracy (*i.e.* 1 pattern wrong out of 2).

5.1 Knowledge Extraction from Trained Neural Networks

Twenty-four different rule sets were successfully extracted from neural networks using *NeuroLinear* knowledge extraction approach. The average of inference performance on the testing data was 83%. "If-then" types of rules consisting of linear equations were extracted, similar to the following example:

```
If (1.79784 * I2 + 5.60213 * I3) – (4.78878 * I4 + 5.49839 * I5) > 0.37,
Then condition class = good.
Else condition class = average.
```

`I2` is the percentage of rubble; `I3` is the percentage of consolidated rubble; `I4` is the percentage of boulders; and `I5` is the percentage of rock/pavement.

The coefficients and operators of the attributes in the rules provide useful information. An answer to a query such as "what is the maximum decrease/increase that can be made to attribute I2 of a pattern, such that the classification of the pattern remains the same?" is easy to obtain from the rules generated from the network. The rules also reveal that under what changing direction of associated attributes, the patterns could achieve a certain type of classification. According to the above rule, we could generally say that the 'good' condition for the sea cucumber requires a relatively higher percentage of rubble and consolidated rubble, and a relatively lower percentage of boulder and rock/pavement.

Because of the different structures and initialisations of the neural networks, different sets of rules were generated from the 24 separate neural networks. That actually allows us to reveal more information about the data and give more insight into the problem. Instead of listing the entire 24 rules set, the information induced from these 24 rule sets is summarized in Table 3.

Rule Set No.	Related Attributes	Habitat Trend with Good Condition Class
1	I0 I1 I2	I1- ; I2+
2	I2 I3 I4 I5	I2+; I3+; I4-; I5-;
3	I1 I2 I3 I4	I1-; I2+; I3+; I4-;
4	I0 I1 I2 I4 I5 I7	I1-; I2+; I4-; I5-; I7+;
5	I0 I2	I2+;
6	I2 I3 I5 I6 I7	I2+; I3+; I5-; I6+; I7-;
7	I1 I2	I1-; I2+;
8	I2 I3 I4 I5	I2+; I3+; I4-; I5-;
9	I1 I2 I3 I5	I1-; I2+; I3+; I5-;
10	I0 I1 I2 I3 I5	I1-; I2+; I3+; I5-;
11	I0 I1 I2 I3 I4 I5	I1-; I2+; I3+; I4-; I5-;
12	I0 I1 I2 I3	I1-; I2+; I3+;
13	I0 I1 I2	I1-; I2+;
14	I1 I2 I3 I4 I5	I1-; I2+; I3+; I4-; I5-;
15	I0 I1 I2 I3 I5 I6 I7 I8	I1-; I2+; I3+; I5-; I6+; I7-; I8-;
16	I0 I1 I2	I1-; I2+;
17	I1 I2 I3 I5	I1-; I2+; I3+; I5-;
18	I1 I2 I4	I1-; I2+; I4-;
19	I0 I1 I2 I3 I5 I7	I1-; I2+; I3+; I5-; I7+;
20	I2 I3 I5	I2+; I3+; I5-;
21	I1 I2	I1-; I2+;
22	I2 I3 I5 I6	I2+; I3+; I5-; I6+;
23	I1 I2 I3 I4 I5 I6 I7	I1-; I2+; I3+; I4-; I5-; I6-; I7+;
24	I0 I1 I2 I3 I4 I7 I8	I1-; I2+; I3+; I4-; I7+; I8-;

Table 3: Summarized knowledge from rule sets(-:tend to be low; +:tend to be high)

5.2 Rules Analysis

After the information shown on Table 3 was extracted from the neural networks, the general analysis was as follows:

1. Habitat type “gravel” did not appear in any of 24 rules sets, therefore it could be considered as an unnecessary feature for the problem domain.
2. Habitat types “live coral” and “dead coral” appeared in 7 of the rule sets. When one of them appears in the rules, it always tends to be high. But if both habitat types appear together in the same rule, generally one of the variables will be high and the other variable will be low.
3. The variable “exposure” has a strong influence on the output class. Since there are only two possible values (0 or 1), it affects the degree of the difference of habitat types. For example, we have a rule set shown as follows:

```
If(7.25649 * I1 - 3.91697 * I2)+ 1.03156 * I0 < -0.78,
Then condition class = good
Else condition class = average
```

```
When I0(Exposure)=0, this rule will be:
```

```
If(7.25649 * I1 - 3.91697 * I2)< -0.78,
Then condition class = good
Else condition class = average
```

```
When I0(Exposure)=1; the rule will be:
```

```
If(7.25649 * I1 - 3.91697 * I2)< -1.8156,
Then condition class = good
Else condition class = average
```

4. Finally, the general conclusion is that the “good” condition habitat class for the sea cucumber *(H. leucospilota)* typically exhibits a lower percentage of sand, boulder, rock/pavement, mud/silt and higher percentage of rubble and consolidated rubble.

6 DISCUSSION AND CONCLUSIONS

Neural networks appear to provide an attractive approach to modeling ecological applications due to their ability to “learn” and predict patterns in non-linear datasets. The rules that were extracted from the neural networks yielded predictable results for the habitat preferences of this sea cucumber when compared to the field observations. The sea cucumber, *H. leucospilota* occupies a distinct ecological niche. This animal is almost exclusively observed with its posterior end securely anchored under a rock or interstitial spaces between the rocks, and its highly mobile and extensible anterior end stretched outwards for feeding. Due to this strict habitat partitioning, one would expect to find *H. leucospilota* highly associated with substrates offering interstices such as rubble and consolidated rubble.

The live coral and dead coral substrate types appear in only 7 of the 24 rule sets. Although the suggestion from the rule set analysis is that these variables may be important to the habitat selection of the sea cucumber, we would suggest that due to their limited presence the rule sets, it is more likely that these variables offer only moderate conditions for *H. leucospilota* occupation. This is further supported by field observations. Since *H. leucospilota* is a benthic invertebrate and the corals typically have a three dimensional morphology, the only available habitat provided by the corals that meets the animal’s requirements for shelter and protection occurs at the coral/sediment interface. Of the rule sets where live and dead coral occurred together, one of the variables tended to be high and the other low. One possible explanation may be that the environmental conditions at a group of randomly selected sites used in the neural network training may favour either a living coral colony or a stressed environment (high nutrient runoff, chemical contamination, high water temperatures) not able to support a living community. Further, spatial analysis within the GIS will identify distribution patterns in living and dead coral colonies and hopefully offer further explanations.

Barren substrates offering little protection (sand, mud/silt, gravel, rock/pavement) support very few *H. leucospilota*. The suggestion that *H. leucospilota* did not favour habitats with a high percentage of boulder substrate was somewhat unexpected as the boulder habitat would appear to offer the protection and shelter required by the sea cucumber. Only 9 of the 24 rule sets attributed an influence to boulders. In each instance boulders tended to be low and rubble tended to be high. Closer examination of the data confirms the extracted rules concerning the apparent inverse relationship between boulders and rubble. Of the 38 out of 128 sample sites that contained boulders, those that had a high percentage of boulder substrate tended to have a low percentage of rubble, and consequently fewer numbers of *H. leucospilota*. In contrast, those with a lower percentage of boulder substrate tended to exhibit more rubble, and hence much higher sea cucumber frequencies. This suggests that rubble is the primary variable determining sea cucumber frequency.

The variable “exposure” appears in 11 of the 24 rule sets, thereby suggesting that it is of moderate importance for influencing the frequency of sea cucumbers. However because there are only 2 possible values for this variable, the evidence to suggest that *H. leucospilota* prefers windward or leeward sites is inconclusive.

This paper has produced evidence to suggest that spatial analysis when integrated with neural networks may offer a promising model that is suitable for predicting the habitat preferences for *Holothuria leucospilota*. The neural network and extracted rules were successful at recognizing the influence of the rubble, consolidated rubble and sand variables on the habitat preference of the sea cucumber, *H. leucospilota* on Rarotonga. Lack of representation of the other habitat variables in the available data posed limitations on the learning ability of the neural network. Additional data collection could resolve this problem. The model that we have developed can be extended to similar investigations of *H. leucospilota* from other geographic locations. Future experiments will investigate the applicability of this integrative modeling technique to additional macroinvertebrate species-habitat relationships, and more complex multi-species-habitat relationships.

ACKNOWLEDGEMENTS

The authors thank Tamsyn Dearlove, Grant Hopkins, and Naomi Sugimura for their assistance with the field data collection. We also would like to gratefully acknowledge the financial support provided by the Pacific Development and Conservation Trust, World Wide Fund for Nature (WWF), and the New Zealand Official Development Assistance (NZODA) programme. Generous donations of equipment were made by Pacific Kayak Ltd., Auckland and Divers World, Wellington, New Zealand, and Sokkia New Zealand. Tom Taranto at CSIRO Marine Research, Cleveland, Australia provided invaluable guidance and instruction with the GIS component of the research. Thanks also to Vaipo Mataora and Bruce Manuela, Cook Islands Land Information Services, Government of the Cook Islands. Thanks are extended to Arona Ngari, Cook Islands Meteorological Service for allowing me to establish my GPS base station on their premises.

REFERENCES

Baran, P., Lek, S., Delacoste, M., and Belaud, A. (1996) Stochastic models that predict trout population density or biomass on a mesohabitat scale. *Hydrobiologia,* Vol. 337, pp. 1-9.

Dalzell, P., Adams, T. G. H., and Polunin, N. V. C. (1996) Coastal fisheries in the Pacific Islands. *Oceanography and Marine Biology: an Annual Review, Vol. 34*: pp.395-531.

Fu, M. (1994) Rule Generation from Neural Networks, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 28, No. 8, 1114-1124.

Guan, B. T., Gertner, G. Z., and Parysow, P. (1997) A framework for uncertainty assessment of mechanistic forest growth models: a neural network example. *Ecological Modelling*, Vol. 98, pp.47-58.

Liu, H. and Setiono, R. (1995a) Chi2: Feature Selection and Discretization of Numeric Attributes, Proceedings of the 7th IEEE International Conference on Tools with Artificial Intelligence, pp. 388-391.

Liu, H. and Tan, S. (1995b) X2R: A fast Rule Generator, Proceeding of IEEE International Conference on Systems, Man and Cybernetics (IEEE Press).

Lek, S., Delacoste, M., Baran, P., Dimopoulos, I., Lauga, J., and Aulagnier, S. (1996) Application of neural networks to modelling nonlinear relationships in ecology. *Ecological Modelling,* Vol. 90, pp.39-52.

Mann, S., and Benwell, G. L. (1996) The integration of ecological, neural, and spatial modelling for monitoring and prediction of semi-arid landscapes. *Computers and Geosciences*, Vol. 22, No. 9, pp.1003-1012.

Mastrorillo, S., Lek, S., and Dauba, F. (1997) Predicting the abundance of minnow *Phoxinus phoxinus* (Cyprinadae) in the River Ariège (France) using artificial neural networks. *Aquatic Living Resources*, Vol. 10, pp.169-176.

Minsky, M.L. and Papert, S.A. (1969) Perceptrons. Cambridge:MIT press.

Özesmi, S. L., and Özesmi, U. (1999) An artificial neural network approach to spatial habitat modelling with interspecific interaction. *Ecological Modelling*, Vol. 116, pp.15-31.

Recknagel, F., French, M., Harkonen, P., and Yabunaka, K. (1997) Artificial neural network approach for modelling and prediction of algal blooms. *Ecological Modelling,* Vol. 96, pp.11-28.

Setiono, R. (1997) A Penalty-function Approach for Pruning Feedforward Neural Networks. *Neural Computation,* Vol. 9, No. 1, pp.185-204.

Setiono, R. and Liu, H. (1997) NeuroLinear: From Neural Networks to Oblique Decision Rules, Neurocomputing, Elsevier, Vol. 17, No.1, pp.1-25.

Sperduto, M. B., and Congalton, R. G. (1996) Predicting rare orchid (small whirled pogonia) habitat using GIS. *Photgrammetric Engineering and Remote Sensing*. Vol. 62, No. 11, pp.1269-1272.

Towell, G. and Shavlik, J. (1993) The Extraction of Refined Rules from Knowledge Based Neural Networks, Machine Learning, Vol. 131, 71-101.