

Neural Networks and Fuzzy Logic Analysis in Archaeology

Claire Reeler

Abstract

Neural networks are designed to simulate the way a human brain learns and interprets patterns. Fuzzy logic is a way of quantifying classifications that are not easily restricted to opposites. These are both techniques borrowed from the field of artificial intelligence. Used together these two techniques can give us a different perspective on archaeological analysis and can allow us to model the world in a way that is closer to the way that humans think and behave. This process allows us to attempt to understand the processes of human cognition that lie behind the archaeological evidence with which we work. By building human cognition into the analysis, we make the data easier to work with for ourselves, and potentially get closer to the people whom we are trying to understand through the archaeological record. Both techniques can be applied to almost any kind of archaeological analysis, but provide a particularly useful adjunct to conventional statistical methods and GIS. This paper uses an example from New Zealand to show the potential of these forms of analysis in archaeology.

The use of artificial intelligence techniques in archaeology is a new and exciting field. These techniques build on from the application of other quantitative techniques, and especially multi-variate statistics. The use of neural networks, particularly as part of hybrid systems with other artificial intelligence techniques, has attracted interest in archaeology in recent years (Gibson 1993, Claxton 1995). Neural networks are also particularly adaptable for use with spatial data and Geographic Information Systems (GIS). This paper reports on the application of a neural network / fuzzy logic hybrid system to spatial and other excavation data from Maori pa sites in New Zealand. The pa sites are defended sites, often situated on raised, easily defensible landforms where the natural defences provided by the landscape were enhanced by the addition of artificial earthwork defences. These sites were constructed primarily during the period from about 1500 AD to the early 1800s (Schmidt 1996). In the early 1800s the European colonisation of New Zealand and the increasing availability of firearms led to a change in site form and by the start of the 20th century, pa sites were no longer constructed.

Pa sites have been excavated in New Zealand since the 1940s and there is now a wealth of information concerning the sites. As part of a PhD project at the University of Auckland, New Zealand, this information has been gathered into a database. The maps of the sites have been digitised and linked to the database. Some sites have been re-surveyed in order to provide data for 3-dimensional maps and additional, unexcavated pa have been surveyed in order to examine the relationship between surface and subsurface features. This data is currently being

analysed in an attempt to understand the factors involved in the construction and layout of these sites, particularly with respect to choice of landform, defensive device and internal organisation of features and structures. The role that the sites played in the society that built them is also being addressed from the perspective of the factors mentioned.

The use of neural networks and fuzzy logic allows a new perspective on pa sites to be obtained. By using these two techniques in combination the patterning within the sites can both be extracted and explained. The extraction of patterning resulting from human behaviour in spatial data is one of the most important goals of spatial analysis in archaeology (Holl 1993a). The interpretation of this patterning has been achieved using many different techniques over the past few decades (see for example Clark 1954, Whallon 1973a, Whallon 1973b, Whallon 1974, Hodder and Orton 1976, Ferring 1984, Whallon 1984, Parkington, Nilssen, Reeler and Henshilwood 1992, Roebroeks, De Loecker, Hennekens and van Ieperen 1992, Agorsah 1993, Holl 1993b, Levy 1993). The introduction of artificial intelligence methods allows interpretation to be aided by information extracted from the data itself. Neural networks when used with fuzzy logic, can explain the patterning purely in terms of the numbers used to produce the patterning, thereby adding to and refining the archaeological interpretations themselves.

Cluster analysis provides a perfect example of this. Cluster analysis has been used to extract information from spatial data about sites (see for example Whallon 1984). However, cluster analysis has never allowed us to actually extract the rules for the

clustering. We can interpret the clustering in terms of what we know about the data and the sites, but we have not been able to know how well this interpretation actually fits the patterning within the data itself. The use of neural networks and fuzzy logic allow one to extract the rules for the clustering. These rules can then be compared to the archaeological interpretations in order to add to, or modify, these interpretations from a different perspective.

In order to extract the rules from the neural network a 'fuzzy' neural network must be created. The use of fuzzy logic provides another important aspect to the analysis. Fuzzy logic provides a consistent way to deal with vague or imprecise categories. Neural networks trained on 'fuzzy' data tend to train more efficiently and give more meaningful results. The use of fuzzy logic can also be used to achieve a greater subtlety in results compared with conventional quantitative methods. The process of 'fuzzification' of data is an important one and precedes the training of neural networks. The introduction of fuzzy techniques is an important theoretical, as well as methodological step, in the analysis.

Conventional quantitative techniques are vital prerequisites in order to bring trends within the data to light. The fuzzification of data allows one to build *a priori* assumptions about the data into the analysis. Some of these assumptions can be built in on the basis of knowledge that the researcher has concerning the data, for example as the result of experience in a particular field. Other assumptions should be made on the basis of analysis of the data using simple descriptive statistics. All fuzzy numbers have memberships to which they belong. The memberships describe the fuzzy states of the data (Brulè 1995). For example, sites may be divided into three memberships on the basis of size - small, medium and large. The latter are fuzzy concepts describing the data and these concepts form the memberships. A site with an area of tens of square metres would be classed as small, and a site with an area of tens of thousands of square metres would be classed as large. Membership terms such as small, medium and large have defined numerical meanings when used in fuzzy logic, removing ambiguity when used in this way.

The choice of the number of memberships into which to divide fuzzy data should be based on detailed analysis of the data with descriptive quantitative techniques. It is important that the number of memberships reflect the patterns within the data. For

example, plotting sizes of sites on graphs with different scales and intervals along the axes might show that sizes of sites are best divided into four memberships and that there is a case for including the category 'very large' for some sites. The calculations of the ranges of memberships also need to be substantiated. Fuzzy logic provides a tool whereby 'small' sites grade gradually into 'medium' ones and 'medium' into 'large'. In other words a site of 100 m² may be 'small', but a site of 600 m² may be partly 'small' and partly 'medium' (see Figure 1). One is therefore not restricted to choosing ranges that are wholly one thing or another.

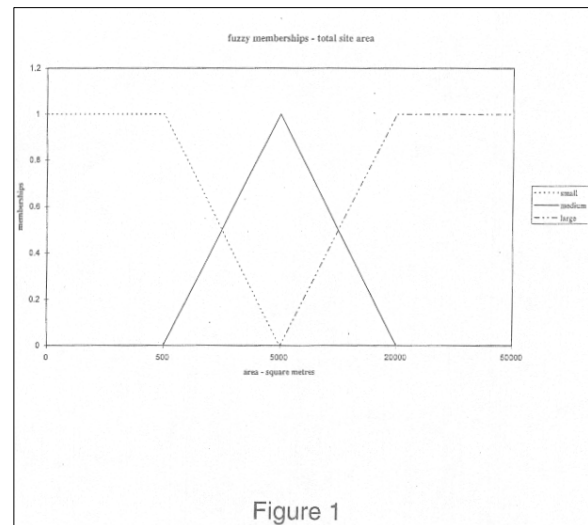


Figure 1.

Even so, the choices of how to subdivide the ranges of values must be grounded in a clear understanding of the basic trends within the data. It is no good dividing the size of sites into 3 or 4 equal divisions to be used with a sample that is very clearly skewed, unless one specifically wants to underrate some of the divisions for the purposes of analysis. Fuzzy logic allows one to have unequal divisions and does not restrict one to having the memberships sum to 1 (Brulè 1995). In some cases, but clearly not all cases, it may be useful to allow the memberships to sum to more than one.

It may be useful to see the extent to which examples belong to more than one membership or classification if one is working with cultural factors. For instance, a cultural influence on the style of an artefact may be related to the degree to which the people who made the artefact subscribed to a certain cultural group. A simplistic example may be made using a New Zealand backdrop, although the picture presented is not meant to reflect the archaeological reality of New

Zealand, which is far more complex than this example. This example is meant to be purely hypothetical in order to present the point about fuzzy memberships.

It is possible that people may have constructed pa sites partly to reflect cultural identity and that cultural identity was related to belonging to certain cultural groupings. People wholly within one cultural group (group A) might have made pa sites that expressed their cultural identity by favouring a certain type of landform, whilst people wholly within another, distinct cultural group (group B) might have constructed pa that would reflect the other cultural identity by having a certain pattern to the defences (see Figure 2).

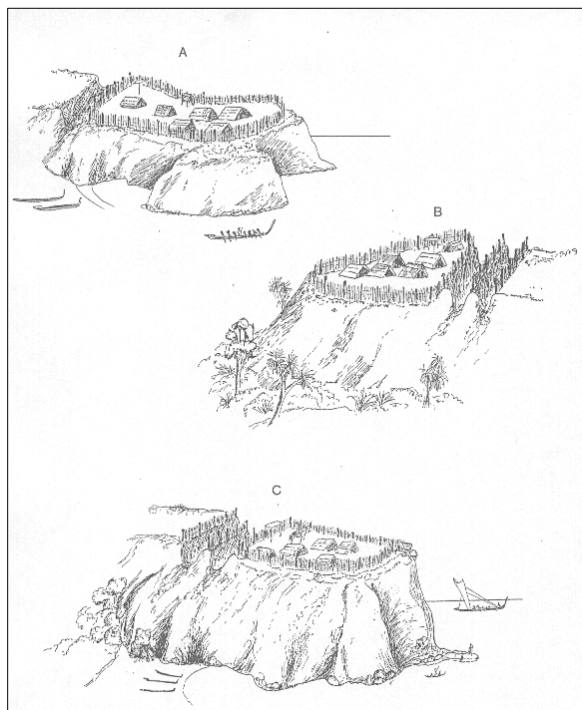


Figure 2.

People who may have inhabited an area between the two groups and who might have had allegiance or family ties to both groups (group C) might have constructed pa sites that contained aspects of both cultural identities. For example, the latter might be situated on the landforms favoured by the first group, but have the defensive system favoured by the second group (see figure Figure 2). If the sites were classified on the basis of cultural identity, the third type of site would be given membership in both cultural classifications. The third type of site might rate highly in factors defining both of the first and second groups. It would therefore be given high

memberships in both groups - a good case for allowing memberships to sum to more than 1. It should be stressed that this example is included purely for illustration and should not be taken as a reliable interpretation of the factors influencing the construction of Maori pa. Although this project does not address the issues of the way in which specific cultural identity was expressed in pa site architecture, the method allows for the possible inclusion of information of this nature.

The ability of neural networks to interpret fuzzy data allows them to be sensitive to subtle patterning such as that reflected in the above example. This has several implications. Firstly, the neural networks are able to pick up patterns within the data that might be missed by conventional quantitative methods that do not utilise fuzzy logic. The neural networks are therefore able to point out patterning that the researcher may have missed. It remains for the researcher to determine whether or not the patterning described by the neural network is archaeologically important or not, but since the neural networks also give the conditions for the patterning in great detail, this is made easier. Once the neural networks have been trained, they are also able to classify sites themselves, and this may produce interesting and potentially very useful results.

The neural networks used in this study are primarily Backpropagation networks, which learn to classify data according to classifications of training data on which the neural networks are trained (Clarkson 1990, Carpenter 1991, Kasabov 1993). The Backpropagation neural networks are made up of three levels (see Figure 3) (Aleksander and Morton 1990). The input level is the level at which the variables to be used for the analysis are introduced. The variables are the inputs. The output level is the level where the classifications of the variables are made. The classifications are the outputs. In between these two layers is a single layer of hidden units. Other neural network architectures may allow for more than one hidden layer, but all the neural networks used in this study use a single hidden layer. The hidden units allow the neural network to interpret complexity within the data. The more hidden units are used, the greater the number of concepts the neural network can learn. In general a number of hidden units somewhere between the number of inputs and the number of outputs is used. The connections between the nodes in the network are where the weights are stored. These weights are generated randomly when the network is created, but are altered during the training process until they

reflect the patterning which causes the classification of the data (Aleksander and Morton 1990, Carpenter 1991).

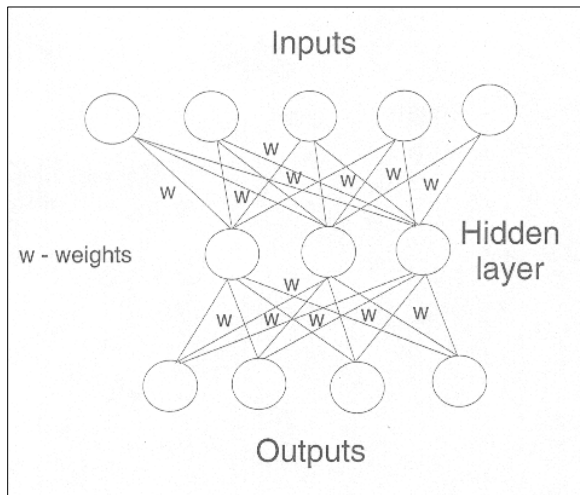


Figure 3.

During the training process, the training data and correct classifications are repeatedly fed through the network. After each training run or 'iteration' the network alters its weights until it is able to match the data to the classifications. It is during this process that the network learns the patterns within the data that relate to the classifications of the data. Once the neural network has been trained, it is tested on new data, which it has not seen before, but for which the correct classifications are known. The neural network is then asked to classify this data and is assessed on its ability to do so. If sufficient training has been done and the neural network is unable to correctly classify the test data, then there is a problem with either the classifications or the data, or the relationship between the two. For example, neural networks may be unable to recognise classifications that are not clearly defined. The latter property makes them a good tool for determining a useful number of clusters for analysis in cluster analysis.

If the neural network is able to correctly classify the data, other data can be classified using the neural network, which will now perform well, as long as the data continues to conform to the patterns of the data on which the neural network was trained. Furthermore, as long as the neural network was trained and used with fuzzy data, the rules for the classifications can be extracted from the neural network. This can allow one to extract the rules for the clusters produced by a cluster analysis, for example.

Output from the neural networks produced by the software used in this study are always fuzzy, even when the network was trained on non-fuzzy data. This is because a hybrid system is used, incorporating both fuzzy logic and neural network architecture. In other words, the neural networks assign a fuzzy possibility that the data belongs to each of the possible classifications. It is an important mathematical distinction that these are possibilities and not probabilities since that frees them from the constraint of having to sum to 1 (Brulè 1995).

The possibilities reflect the degree of certainty with which the neural network makes the classification. The possibilities also reflect the degree to which the examples in the data may belong to more than one classification. Neural networks allow for the data to belong to more than one classification, and record this in the classifications that they assign to the data. This feature makes neural networks potentially very useful within archaeology. As has been shown in the example given in Figure 2, it is possible that some archaeological information may need to belong to more than one classification.

The ability of the neural networks to express the degree to which each example (in this case, sites), belongs to each classification, also makes them very sensitive to the nature of individual sites. This enables one to draw general conclusions about the sites without losing the critical dimension of the individual nature of all archaeological sites. Some aspects of analysis thrive on the general conclusions about sites that one can draw for the sites as a whole, but other aspects of analysis need the perspective provided by the individual sites. Whilst the current project of pa site analysis emphasises the first type of analysis, i.e. general conclusions, the individual nature of sites must not be forgotten. Individual sites may differ from the models explaining the sites, and it is useful to know how they differ from the models. It is also very important in the models we build of human behaviour with respect to pa site construction, to identify the sites that do not seem to fit the model and build them into some form of explanation of the model, or modify the model if necessary.

Neural networks and fuzzy logic have been used very successfully in the analysis of Maori pa sites. The application of fuzzy logic to the identification of variables describing the sites has illustrated many subtle trends in the data. Twenty variables were defined describing each site, on the basis of data collected from the excavation reports and digitised maps of the sites. Some of these variables included

data provided by the GIS used in the project, such as total site area and area of features, which is automatically calculated by ARC/INFO when the maps are digitised.

Using fuzzy logic, these 20 variables were assigned varying numbers of memberships, so that the final number of fuzzy variables was 61. The fuzzy analysis revealed the same basic trends as identified by the descriptive statistical analysis, but provided additional information about some of the relationships between variables, such as area of defences and topographic type. The process of fuzzifying the data immediately adds to the conclusions reached by the application of descriptive statistics. This process should be seen as a dialectical one, with descriptive statistics informing the fuzzification process, and the latter possibly suggesting additional dimensions to be explored by conventional quantitative techniques.

The 20 non-fuzzy variables were then analysed using Ward's method cluster analysis. Ward's method cluster analysis, whilst producing potentially useful results in itself, also works very well with the neural network architecture used in this project. The two techniques use closely related algorithms and estimates of error to interpret patterning within the data and one can therefore be reasonably confident that the results of both techniques will be comparable to each other.

The results of the cluster analysis were interpreted in order to determine the most useful number of clusters to be used in the analysis. The optimal number of clusters was chosen by visual inspection of the distances between clusters. The maximum number of useful, reasonably coherent clusters was taken to be 5 clusters. The sites composing these clusters were then examined in order to make a preliminary interpretation of the archaeological factors that might explain the clustering. This interpretation was done before the neural network analysis, in order to ensure that the interpretation was not influenced by the results of the neural network analysis. The interpretation could therefore be compared to the results of the neural network analysis once that analysis had been done in order to provide an independent check.

Once the interpretation of the clusters was complete, neural networks were trained and the results of this training indicated that whilst 6 clusters was the point at which the neural networks ceased to be able to clearly recognise the clusters, there was a good case

for using 5 clusters for interpretation. The neural networks were able to clearly distinguish up to 6 clusters, using a cutoff point of about 70% as an acceptable measure of correct classification by the neural networks. But the inclusion of the sixth cluster introduces a cluster containing only one site, thereby skewing the results slightly at this point. It was therefore decided that 5 clusters are indeed an optimum number for interpretation.

The ability of the neural networks to guide one as to the optimum number of clusters is an additional advantage of this method. Neural networks are tested once they have been trained, so that their performance can be assessed. The networks are tested by being set the task of classifying data for which the correct classification is already known. The classifications made by the neural network are then compared to the known correct classifications and the performance of the neural network is assessed. In terms of percentages of correct classifications, neural networks are regarded as performing well at anything above 70 - 75% (Singh pers. comm. 1996).

For the purposes of identifying the optimum number of clusters, neural networks were trained using the classifications provided by the cluster analysis, starting with two clusters and incrementing by one cluster each time, through to 11 clusters. The neural networks were able to recognise two clusters very easily, giving results of 90%. As the data was subdivided into more clusters the performance of the neural networks decreased until by 11 clusters the neural networks were achieving only 48% correct classifications. It would be possible to extract the rules for all levels of clustering, and thus track the patterns in the data as more groupings are made, but for the purposes of this study, rules were extracted only for 5 and 6 clusters.

Not only did the percentage of correct classifications decrease, but the neural network's 'confidence' in the classifications also decreased. The fuzzy nature of the classifications given by the neural network software indicates the certainty with which the neural network assigns the classifications. For example, if the neural network finds that the data fits the pattern for the classification of cluster 1 very well, it will assign that site to cluster 1 with a 'possibility' of about 0.9 (a value of 1 indicates a perfect correspondence and a value of 0 indicates a complete lack of correspondence with the classification). However, if the data does not fit the patterning for cluster 1 very well, the neural network might assign that site to cluster 1 with a 'possibility' of only about 0.5. This

might still be the highest 'possibility' assigned to that site, the other possibilities for the other classifications being between 0.3 and 0.001, and it may represent a correct classification, if it is known that the site does in fact belong to cluster 1. However, there is a difference in what might be termed 'confidence' between a classification of 0.9 and a classification of 0.5.

The rating of the classifications in this way can be useful for determining the degree to which sites fit into the patterns used. The results emphasise the individual nature of the sites. The rating of the classifications can also be used to identify sites that could potentially belong to more than one classification. For example, if the sites referred to in the example given in figure 2 were classified as belonging to different classifications, then the third site would have a relatively high possibility of belonging to both classifications.

The neural networks were initially trained on a selection from the total sample. This selection comprised the training data for the neural networks. The selection of sites to be included in the training data was made randomly since it was not known which sites were good indicators of the clusters and which were not. In other applications of this technique, training data could be selected according to criteria to ensure that they adequately reflect the classification. The selection of training data in this case seems to have been adequate, since the neural networks were able to identify the clusters which were more clearly differentiated.

However, before the rules for clustering were extracted, the neural networks were retrained on the entire sample. This was done to ensure that any potential biases introduced by the selection of training data were removed. The rules were then extracted and analysed. The nature of the software used (FuzzyCOPE, developed by the Knowledge Engineering Laboratory at the University of Otago), gives rules for both inclusion in and exclusion from each cluster. In this instance the rules for inclusion in each cluster were deemed to be the most useful for analysis.

Once the rules were analysed, they were found to contribute significantly to the analysis. Generally, the patterns distinguished from a visual examination of the clusters were also present in the rules for clustering. The rules used by the neural networks fitted well with the earlier interpretation made of the clustering. This result shows that the neural networks

can be used to add to conventional archaeological interpretations. Furthermore, the rules were also able to explain clustering that was not easy to interpret. Trends in the data within these clusters could be determined. The rules were also able to show which variables were influential in the clustering and which patterns within the data were most significant.

The analysis of the rules extracted from the fuzzy neural networks has suggested ways in which sites might be grouped into analytical units based on the interplay of a number of different variables. It has emphasised the importance of a number of different variables in describing the variability of the sites. The fuzzy neural network analysis therefore forms an integral part of the analytical process. Analysis of the rules in turn leads to new ways of grouping and examining the sites.

The influential variables identified within the rules extracted from the neural networks suggest avenues for further research. These variables suggest important patterns within the variables extracted from the sites. Several of the influential variables suggest the choices made by the prehistoric people who built the sites. However, the analysis also informs about the archaeological process itself. Some aspects of recording bias have been detailed and this information can be used to suggest changes in the way in which similar sites are surveyed and recorded in future. The rules have revealed both patterning within the clusters distinguished by cluster analysis, as well as overall patterning with respect to the sample as a whole.

These results are being used to reassess our understanding of these sites in terms of their variability as well as the choices governing their construction. Refinements in recording technique can also be suggested. The use of fuzzy neural networks forms a valuable addition to the other analytical techniques used in this study, such as multivariate statistics and GIS. The results of the analysis of the rules extracted from the neural networks are being used to inform on further analytical procedures and have suggested new directions for investigation. Furthermore, one is able to identify influences made at the recording level and distinguish these from influences that may reflect the way the sites were used and positioned in the landscape.

One of the main contributions of the ability to extract rules from the fuzzy neural networks is that all criteria for groupings of sites are made explicit. Once the criteria are identified, they can be assessed and

further research builds on from a firm base. It is possible to state exactly what patterns in the data form the groupings identified by the Ward's method cluster analysis, and therefore it is possible to identify how much the cluster analysis contributes to the understanding of the sites. The more influential and interesting patterns can be further explored, whilst potential directions in the research that can be shown to be less significant can be avoided.

Unfortunately, we can never be sure that we have incorporated all the important aspects of sites into an analysis. In terms of neural network analysis it is therefore important to include as many aspects as possible. Although we can never be certain of reaching a full understanding of the way that prehistoric sites were built and used, fuzzy neural networks nevertheless provide an additional

perspective on the data. This methodology also integrates very well with other quantitative techniques to form part of the analytical process. The use of fuzzy logic and neural networks therefore forms a useful addition to the techniques available for use in spatial analysis as well as other forms of quantitative analysis.

Acknowledgements

Thanks are due to the Knowledge Engineering Laboratory of the University of Otago for providing beta versions of the FuzzyCOPE software and to Joan Lawrence of the Department of Anthropology, the University of Auckland, for turning my ideas about fuzzy memberships and pa sites into the wonderful illustrations in figure 2 of this paper.

Bibliography

- Agorsah K E, 1993 Archaeological Considerations on Social Dynamics and Spatial Pattern Development of Traditional Settlements in A Holl and T E Levy (eds) *Spatial Boundaries and Social Dynamics. Case Studies from Food-Producing Societies*, 7-24. . International Monographs in Prehistory, Michigan,
- Aleksander, I, and Morton, H, 1990 *An Introduction to Neural Computing*. Chapman and Hall, London.
- Brulé, J, 1995 *Fuzzy Systems - A Tutorial*. comp.ai, May, 1995.
- Carpenter, G, 1991 Neural network models for pattern recognition and associative memory in G. Carpenter and S. Grossberg (eds) *Pattern-Recognition by Self-Organising Neural Networks*,1-33. MIT Press, Cambridge, Massachusetts.
- Clark, J G D, 1954 *Excavations at Star Carr*. Cambridge University Press, Cambridge.
- Clarkson, D, 1990 Living Computers *Electronics World* March, 186-190.
- Claxton, J B 1995 Future enhancements to GIS, implications for archaeological theory in G. Lock, G, and Stancic, Z,(eds) *Archaeology and Geographic Information Systems, A European perspective*, 335-348. Taylor and Francis, London.
- Ferring, C R, 1984 Intrasite spatial patterning: its role in settlement-subsistence systems analysis in H. Hietala (ed) *Intrasite Spatial Analysis in Archaeology*,116-126. Cambridge University Press, Cambridge.
- Gibson, P M, 1993 The potentials of hybrid neural network models for archaeofaunal ageing and interpretation in J. Andreson, T. Madson and I Scollar (eds) *Computing The Past: Computer Applications and Quantitative Methods in Archaeology. CAA92*, ,263,271. Aarhus University Press, Aarhus.
- Hodder, I, and Orton, C, 1976 *Spatial Analysis in Archaeology*. Cambridge University Press, Cambridge.
- Holl, A, 1993a Introductory Note to Spatial Boundaries and Social Dynamics: Rounds, Recycling, and the Build-up of an Upward Spiral in A Holl and TE Levy (eds) *Spatial Boundaries and Social Dynamics. Case Studies from Food-Producing Societies*,1-5. International Monographs in Prehistory, Michigan.
- Holl, A, 1993b Late Neolithic Cultural Landscape in southeastern Mauritania: An Essay in Spatiometrics in A Holl and TE Levy (eds) *Spatial Boundaries and Social Dynamics. Case Studies from Food-Producing Societies*, ,95-133International Monographs in Prehistory, Michigan.
- Kasabov, N, 1993 Artificial Neural Networks and Fuzzy Systems for Knowledge Engineering *Proceedings of the 13th New Zealand Computer Society Conference* 1,338- 352
- Levy, T E, 1993 Production, Space and Social Change in Protohistoric Palestine in A Holl and T E Levy (eds) *Spatial Boundaries and Social Dynamics. Case Studies from Food-Producing Societies*,63-81. International Monographs in Prehistory, Michigan.
- Parkington, J E, Nilssen, P, Reeley, C, and Henshilwood, C, 1992 Making sense of space at Dunefield Midden campsite, western Cape, South Africa *Southern African Field Archaeology* 1(2),63-71

- Roebroeks, W, De Loecker, D, Hennekens, P, and van Ieperen, M, 1992 'A veil of stones': on the interpretation of an early Middle Paleolithic low density scatter at Maastricht-Belvédère (The Netherlands) *Analecta Praehistorica Leidensia* 25,1-16
- Schmidt, M, 1996 The Commencement of Pa Construction in New Zealand Prehistory *Journal of the Polynesian Society* 105(4),441,460
- Whallon, R, 1973a Spatial Analysis of Occupation Floors I: Application of Dimensional Analysis of Variance *American Antiquity* 38,266-278
- Whallon, R, 1973b Spatial Analysis of Paleolithic Occupation Areas in C Renfrew (ed) *The Explanation of Cultural Change*,115-130. Duckworth, London.
- Whallon, R, 1974 Spatial Analysis of Occupation Floors II: the Application of Nearest Neighbour Analysis *American Antiquity* 39,16-34
- Whallon, R, 1984 Unconstrained clustering for the analysis of spatial distributions in archaeology in H. Hietala (ed) *Intrasite Spatial Analysis in Archaeology*,242-277. Cambridge University Press, Cambridge.

Contact details

Claire Reeler
Dept of Anthropology
University of Auckland
NEW ZEALAND