



Universiteit  
Leiden

The Netherlands

**Analecta Praehistorica Leidensia 28 / Interfacing the past : computer applications and quantitative methods in archaeology CAA95 Vol. I**

Kamermans, Hans; Fennema, Kelly; et al., ; Kamermans, Hans; Fennema, Kelly

**Citation**

Kamermans, H., Fennema, K., & Et al.,. (1996). Analecta Praehistorica Leidensia 28 / Interfacing the past : computer applications and quantitative methods in archaeology CAA95 Vol. I, 272. Retrieved from <https://hdl.handle.net/1887/32752>

Version: Not Applicable (or Unknown)  
License: [Leiden University Non-exclusive license](#)  
Downloaded from: <https://hdl.handle.net/1887/32752>

**Note:** To cite this publication please use the final published version (if applicable).

ANALECTA  
PRAEHISTORICA  
LEIDENSIA

28

---

PUBLICATIONS OF THE INSTITUTE OF PREHISTORY  
UNIVERSITY OF LEIDEN

INTERFACING THE PAST

COMPUTER APPLICATIONS AND QUANTITATIVE  
METHODS IN ARCHAEOLOGY CAA95 VOL. I

EDITED BY  
HANS KAMERMANS AND KELLY FENNEMA



UNIVERSITY OF LEIDEN 1996

# contents

## VOLUME I

Hans Kamermans Kelly Fennema	Preface
<b>Data Management</b>	
Jens Andresen Torsten Madsen	IDEA – the Integrated Database for Excavation Analysis 3
Peter Hinge	The Other Computer Interface 15
Thanasis Hadzilacos Polyxeni Myladié Stoumbou	Conceptual Data Modelling for Prehistoric Excavation Documentation 21
E. Agresti A. Maggiolo-Schettini R. Saccoccio M. Pierobon R. Pierobon-Benoit	Handling Excavation Maps in SYSAND 31
Alaine Lamprell Anthea Salisbury Alan Chalmers Simon Stoddart	An Integrated Information System for Archaeological Evidence 37
Jon Holmen Espen Uleberg	The National Documentation Project of Norway – the Archaeological sub-project 43
Irina Oberländer-Tárnoveanu	Statistical view of the Archaeological Sites Database 47
Nigel D. Clubb Neil A.R. Lang	A Strategic Appraisal of Information Systems for Archaeology and Architecture in England – Past, Present and Future 51
Nigel D. Clubb Neil A.R. Lang	Learning from the achievements of Information Systems – the role of the Post-Implementation Review in medium to large scale systems 73
Neil Beagrie	Excavations and Archives: Alternative Aspects of Cultural Resource Management 81
Mark Bell Nicola King	The MARS Project – an interface with England's past 87

**Archaeometry**

- M.J. Baxter  
H.E.M. Cool  
M.P. Heyworth  
Detecting Unusual Multivariate Data: An Archaeometric Example 95
- Jon Bradley  
Mike Fletcher  
Extraction and visualisation of information from ground penetrating radar surveys 103
- Gayle T. Allum  
Robert G. Aykroyd  
John G.B. Haigh  
Restoration of magnetometry data using inverse-data methods 111
- W. Neubauer  
P. Melichar  
A. Eder-Hinterleitner  
Collection, visualization and simulation of magnetic prospection data 121
- A. Eder-Hinterleitner  
W. Neubauer  
P. Melichar  
Reconstruction of archaeological structures using magnetic prospection 131
- Phil Perkins  
An image processing technique for the suppression of traces of modern agricultural activity in aerial photographs 139
- Statistics and Classification**
- Clive Orton  
Markov models for museums 149
- Juan A. Barceló  
Heuristic classification and fuzzy sets. New tools for archaeological typologies 155
- Kris Lockyear  
Dmax based cluster analysis and the supply of coinage to Iron Age Dacia 165
- Christian C. Beardah  
Mike J. Baxter  
MATLAB Routines for Kernel Density Estimation and the Graphical Representation of Archaeological Data 179
- John W.M. Peterson  
A computer model of Roman landscape in South Limburg 185
- Sabine Reinhold  
Time versus Ritual – Typological Structures and Mortuary Practices in Late Bronze/Early Iron Age Cemeteries of North-East Caucasia ('Koban Culture') 195
- Leonardo García Sanjuán  
Jesús Rodríguez López  
Predicting the ritual? A suggested solution in archaeological forecasting through qualitative response models 203
- Johannes Müller  
The use of correspondence analysis for different kinds of data categories: Domestic and ritual Globular Amphorae sites in Central Germany 217
- J. Steele  
T.J. Sluckin  
D.R. Denholm  
C.S. Gamble  
Simulating hunter-gatherer colonization of the Americas 223

- Paul M. Gibson An Archaeofaunal Ageing Comparative Study into the Performance of Human Analysis Versus Hybrid Neural Network Analysis 229
- Peter Durham Image Processing Strategies for Artefact Classification 235  
Paul Lewis  
Stephen J. Shennan
- Gijsbert R. Boekschoten A new tool for spatial analysis: "Rings & Sectors plus Density Analysis and Trace lines" 241  
Dick Stapert
- Susan Holstrom Loving Estimating the age of stone artifacts using probabilities 251
- Oleg Missikoff Application of an object-oriented approach to the formalization of qualitative (and quantitative) data 263

## VOLUME II

### Geographic Information Systems I

- David Wheatley Between the lines: the role of GIS-based predictive modelling in the interpretation of extensive survey data 275
- Roger Martlew The contribution of GIS to the study of landscape evolution in the Yorkshire Dales, UK 293
- Vincent Gaffney Extending GIS Methods for Regional Archaeology: the Wroxeter Hinterland Project 297  
Martijn van Leusen
- Trevor M. Harris Multi-dimensional GIS: exploratory approaches to spatial and temporal relationships within archaeological stratigraphy 307  
Gary R. Lock
- Philip Verhagen The use of GIS as a tool for modelling ecological change and human occupation in the Middle Aguas Valley (S.E. Spain) 317
- Federica Massagrande The Romans in southwestern Spain: total conquest or partial assimilation? Can GIS answer? 325
- Shen Eric Lim Recent examples of geographical analysis of archaeological evidence from central Italy 331  
Simon Stoddart  
Andrew Harrison  
Alan Chalmers
- Vincent Gaffney Satellite Imagery and GIS applications in Mediterranean Landscapes 337  
Krištof Oštir  
Tomaž Podobnikar  
Zoran Staničič
- Yvette Bommeljé The long and winding road: land routes in Aetolia (Greece) since Byzantine times 343  
Peter Doorn

- Javier Baena Preysler  
Concepción Blasco      Application of GIS to images and their processing: the Chiribiquete Mountains Project 353

### **Geographic Information Systems II: The York Applications**

- Julian D. Richards      From Site to Landscape: multi-level GIS applications in archaeology 361
- Harold Mytum      Intrasite Patterning and the Temporal Dimension using GIS: the example of Kellington Churchyard 363
- A. Paul Miller      Digging deep: GIS in the city 369
- Julian D. Richards      Putting the site in its setting: GIS and the search for Anglo-Saxon settlements in Northumbria 379
- Jeffrey A. Chartrand      Archaeological Resource Visibility and GIS: A case study in Yorkshire 389

### **Visualisation**

- John Wilcock      A description of the display software for Stafford Castle Visitor Centre, UK 405
- Christian Menard  
Robert Sablatnig      Pictorial, Three-dimensional Acquisition of Archaeological Finds as Basis for an Automatic Classification 419
- Katalin T. Biró      Simple fun – Interactive computer demonstration program on the exhibition of the Szentgál-Tűzköveshegy prehistoric industrial area 433
- György Csáki  
Ferenc Redő      Documentation and modelling of a Roman imperial villa in Central Italy 437
- Maurizio Forte  
Antonella Guidazzoli      Archaeology, GIS and desktop virtual reality: the ARCTOS project 443
- Germà Wünsch  
Elisabet Arasa  
Marta Pérez      Dissecting the palimpsest: an easy computer-graphic approach to the stratigraphic sequence of Túnel VII site (Tierra del Fuego, Argentina) 457
- David Gilman Romano  
Osama Tolba      Remote Sensing and GIS in the Study of Roman Centuriation in the Corinthia, Greece 461
- F.J. Baena  
F. Quesada  
M.C. Blasco      An application of GIS intra-site analysis to Museum Display 469

### **Education and Publication**

- Robin B. Boast  
Sam J. Lucy      Teaching with objects 479

- Martin Belcher  
Alan Chalmers  
Andrew Harrison  
Simon Stoddart  
Teaching the Visualisation of Landscapes – Approaches in Computer based learning for Archaeologists 487
- Anja C. Wolle  
Stephen J. Shennan  
A Tool for Multimedia Excavation Reports – a prototype 493
- G. Gyftodimos  
D. Rigopoulos  
M. Spiliopoulou  
Exploring Archaeological Information through an Open Hypermedia System 501
- Martijn van Leusen  
Sara Champion  
Jonathan Lizee  
Thomas Plunkett  
Toward a European Archaeological Heritage Web 511
- Mike Heyworth  
Seamus Ross  
Julian Richards  
Internet archaeology: an international electronic journal for archaeology 521
- Virgil Mihailescu-Bîrliba  
Vasile Chirica  
A Survey of the Development of Computer Applications in Romanian Archaeology 529
- Kris Lockyear  
Computer-aided publication in practice 535

# An Archaeofaunal Ageing Comparative Study into the Performance of Human Analysis Versus Hybrid Neural Network Analysis

## 1 Introduction

This paper briefly reports on the completion of the first phase of a project that began in 1991 to develop a prototype computer system that could perform archaeofaunal ageing from a set of sheep mandibles. The computer system uses artificial intelligence models known as neural networks to analyse images of mandibles to assess their degree of wear and relative age (Gibson 1992a, 1992b, 1993).

In order to assess the performance of the computer system in relation to its human counterpart a comparative study that involved the analysis of a sample set of sheep mandibles by archaeologists, non-archaeologists and the computer system was undertaken. An overview of the results is presented here.

## 2 Overview of the comparative study

Age at death data of common domestic ungulates can be used to formulate an interpretation of the economy and exploitation of the livestock on a site. A number of approaches to age estimation are based on the analysis of teeth attrition. The age of an animal can be estimated by grouping its teeth into a set of wear stages based on the amount of attrition. In general, older animals have a greater degree of wear.

There are two commonly used methods of age estimation using attrition, namely Payne (1973) and Grant (1982). Payne has studied the wear stages of Anatolian sheep and goats and as a result has devised a methodology for age estimation. A more widespread study, that includes the common ungulates of pig, sheep/goat and cattle, has been undertaken by Grant.

Both methodologies concentrate on the third premolar (m3 or dP4), the fourth premolar (P4) and the three permanent molars (M1, M2, and M3). An archaeological sample can be aged by comparing the wear pattern of each tooth with the wear stages, in the form of ideograms, outlined by either methodology. This analysis results in a *tooth wear stage value* for each tooth. These values are then used to produce a *mandible wear stage value* that represents the relative age of the sample. Statistical analysis is then carried out to group all mandibles on the site into relative age stages which can be interpreted by examining the

kill-off patterns to suggest the method of animal husbandry (Payne 1973).

It is the above process of analysing the mandibles to determine age at death that has been implemented on a PC using both traditional artificial intelligence techniques and hybrid neural network models (see Gibson 1992a, 1992b, 1993). Neural networks are computing paradigms that attempt to model the cognitive phenomena of the human brain so that complex problems can be solved. In doing so, they exhibit a number of intuitive characteristics such as learning, generalisation and abstraction (see Wasserman 1989 for an explanation).

In order to establish the performance of the system a number of willing participants have been asked to attribute age to a set of mandibles using both Grant and Payne methodologies. These results have then been compared with each other and against the computer application. The aim has been to study how different the results are between each participant and between the computer system and the participants in order to determine the degree of subjectivity and accuracy. The analysis has been divided into two parts, namely

- Human vs. Human Comparison  
*Establishes the inter-observer performance*
- Human vs. Computer Comparison  
*Establishes the computer's performance in relation to the humans' performance*

Measuring the performance of humans provides a guide to establishing the reliability of the computer system. In the course of identifying the performance of human analysis a number of interesting points have been highlighted regarding the methodologies involved and the human's use of the methodologies.

The participants were a cross section of people with varying degrees of archaeological experience. The set of people also had a spread of experience in terms of the two archaeofaunal ageing methodologies. A number of the group are acknowledged experts in the use of the archaeofaunal techniques under study. In contrast, a number of the group had never before used these techniques to age animal remains. In all eleven participants undertook the study.



The sheep mandibles used for the comparative study were taken from two sources. The first set was supplied by the Environmental Archaeology Unit (EAU) at the University of York with the assistance of Dr Keith Dobney. The second set was kindly lent by Prof. Don Brothwell from his own collection.

The aim of the selection of mandibles was to provide as wide a range of wear stages, and teeth morphologies as possible without overburdening the participants. Consequently, a sample set of 22 mandibles was selected that had a combination of missing teeth, unerupted teeth, teeth in early stages of wear and teeth in moderate stages of wear. A group of mandibles was selected that appeared to be in the same state of wear. Finally, some teeth had been subject to disease.

### 3 The human vs. human comparison

In order to carry out the comparative study a questionnaire was designed to record the results of the participants' analysis and to determine facts about the participant that would be useful in the analysis, such as archaeological experience, number of years using each methodology and the preferred method. The participants were asked to age each mandible in any order using both methodologies and record the results on the questionnaire sheet. For teeth that the participant could not record they were asked to use a ? for unsure wear stages, X for present but unrecordable and a - for missing teeth.

A computer database system using DATAEASE was devised to record the results of the analysis. The source of the data entered into the database was used to produce a data file that could be analysed by another computer program, written in QBASIC, that presented the results in a manner that helped to answer the underlying objectives of the study.

### 4 Devising a method for analysing the results

Before analysing the data the main objectives of the analysis had to be made clear. To determine the performance of the human participants a number of questions needed to be addressed as part of the analysis, for example,

- Are some mandibles easier than others to age?
- Which are the most difficult mandibles and why?
- Which is the most difficult tooth in the set and in general and why?
- Which is the wear stage that causes the most disagreement and why?
- Are the experienced participants of the methodology more consistent in their interpretation than those with less experience?
- What are the factors that determine ease of observation?

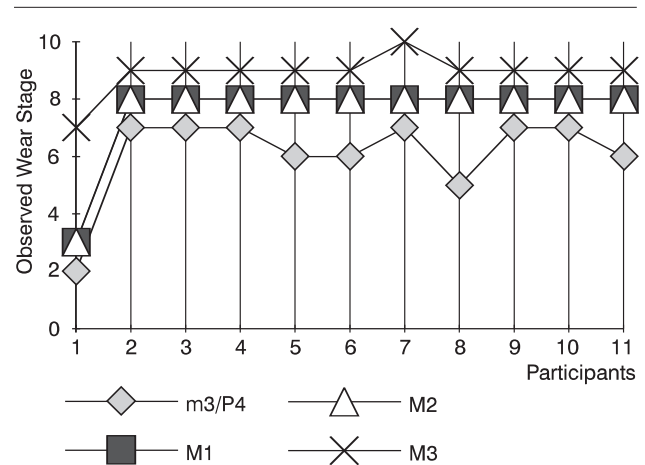


Figure 1. The Participant's Observed Wear Stage Values for Sample 2, where true wear stage values for m3 is 7, M1 is 8, M2 is 8 and M3 is 9.

- Is there any relationship between the analysis of left and right mandibles to ease of observation and general agreement?
- Which of the methodologies provides the most consistent results and why?

To answer the above questions it was necessary to establish a means of objectively analysing all the results from the participants. Keeping the objective of inter-observer comparison in mind, it was obvious that a true wear stage value for each measurement had to be used as a basis for comparison. In other words, for each of the results recorded there must be a real value by which to compare the actual observed wear stages. This 'true' wear stage value can be calculated using the mode of the participants results where 'true' effectively means 'expected in this study'. By taking the absolute value of the difference between the true wear stage and the observed wear stage it was possible to determine the amount of discrepancies between observers. Figure 1 shows a subset of the actual results using the Payne methodology illustrating discrepancies between observers.

Using the calculated discrepancy it was possible to determine the percentage agreement of tooth wear stages and mandible wear stages across participants. In addition, it was possible to rank the performance of the participants. This formed the basis for determining the reliability of the methodologies.

### 5 Analysing the data

At first glance the ranked results would appear to indicate a range of difficulties in the analysis of teeth and mandibles. On the whole both methodologies seem to

perform quite well in some areas and badly in others. It is hard to objectively state what causes such difficulties. Are they related to structure, colour, orientation of the mandible; degree of experience of the participant, speed of recording, lighting in the room at the time of analysis or to the sequence in which the mandible was examined? These questions may demand exact answers but only speculative reasons can be given through examination of the teeth and mandibles.

Firstly, there seems to be no real problem in identifying the teeth types since there were no values in the teeth columns that were invalid. Also, there is no evidence to suggest a correlation between the percentage agreement and a left or right mandible. In addition, the teeth that resulted in the most agreement were those that were missing or unerupted. All these facts suggest that the human is good at recognising simple shapes and manipulating them in order to achieve the requirements of the analysis. This may seem an obvious statement to make but such tasks are very complicated to implement using a computer. Therefore, a machine must match this performance if it is to be of any practical use.

The presence of calculus on the tooth does not appear to affect observations provided it does not obscure any important tooth structures that would differentiate wear stages. Humans have the ability to ignore such 'noise' in the analysis of surface patterns and structures, something that a computer finds more of a problem. The only time that it may affect results is when the calculus appears at a transition point from one wear stage to another.

A general observation for the overall percentage agreement graphs is that there is less accuracy in the earlier wear stages. Perhaps this is because

1. there are more features to match,
2. less distinction between wear stages since there are more wear stages in the early years, or
3. the enamel/dentine distinction is often harder to determine.

In contrast, there appears to be more accuracy in the later wear stages. Well-worn stages seem to be easier to identify perhaps because the features on the surface of the tooth are simple. The smaller distinctions between wear stages are more difficult to pick up. When the break is only partially worn a discrepancy can occur. It appears that both methodologies suffer from this problem.

In general when the tooth does not fit a single wear stage then the percentage agreement drops. The smaller the transition between wear stages, the greater the disagreement. Therefore, visual clues based on distinct structures that are evident in the ideograms are an important element in identifying wear stages. The clarity of the enamel/

dentine border is also important in the identification of wear stages.

It appears that any disfigurement of the occlusal surface of the tooth caused by a disease may affect the estimation of a wear stage depending on the degree of deformation. For example, one of the sample's M3 teeth was slightly deformed and the structure of the cusps was not as represented in the ideograms. This made establishing a wear stage rather difficult and was reflected in the percentage agreement for both methods.

To establish the overall estimate of the agreement for mandibles the average of agreements for the teeth of each mandible was considered for both methodologies. On the whole both the Payne and Grant faired similarly, with Grant gaining better agreement than Payne on some occasions and vice versa. However, the overall average agreement for Payne was 70.4% and for Grant 69.1%. Figure 2 shows a comparison of percentage agreement between Payne and Grant approaches.

To determine whether experience had any bearing on the analysis the participants were ranked for the results of mandibles, and each tooth. Again, it can be stated that experience has no real influence on the establishment of age. This conclusion is gained by examining the experience of individuals and noting where they rank in the group for each tooth. The top five were not always the most experienced.

Although this study has aimed to cover all aspects that may lead to misinterpretation of results it obviously has not been able to address all of them. There has been no consideration to the sequence in which mandibles were analysed to see if this had any influence on the results. The influence of broken and partial teeth has not been fully addressed, although they were considered in part. A larger group of participants would perhaps provide a more general and global view. Also, the effects of speed were not analysed.

This study has shown that there is not always 100% agreement in the results of observers. In addition, it has suggested why there may be discrepancies in the data. However, the main purpose of the study has been to provide a set of data that can be compared to the computer system to measure its performance.

## 6 The human vs. computer comparison

The key to the success of a neural network based system is the reliability of the data that is presented to it during the training stage of the system's development. The testing of the system is an integral part of its development and requires data that contains representative examples of all general cases that the neural network would be expected to cope with during its active operational running. Therefore, the system was trained using a series of images of

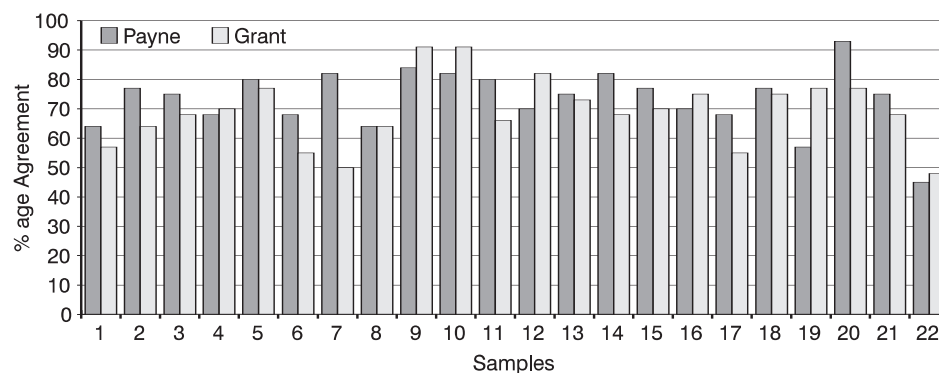


Figure 2. Comparison of percentage agreement of Payne and Grant.

mandibles with various degrees of wear. Once the system had been trained the mandibles given to the participants were presented to the computer and the results compared.

The system was measured against the participants and found to have an overall performance of 65.2% agreement. Of those results that did not match the system was only one or two wear stages out from the participants agreement. Like the human participants the system appeared to have difficulty in assessing early wear stages whilst having greater success with later stages. Again, this could be linked to the smaller transitions in some of the early wear stages. The performance of the system when faced with disfigurement of the occlusal surface was equal to that of the participants. It appeared to produce acceptable answers on the basis of what it saw and what it had learnt.

By comparing the system's performance to the overall percentage agreement of the Payne and Grant methods the result above is encouraging. However, we must be cautious not to overestimate the success of the system. It is important to note that in preparing the images for analysis, effort was made to ensure that the mandibles were presented in such a way that the system would not get confused. The success of the system relies heavily on the quality of the image. Giving the system images that had mandibles in a bad orientation or obscured by calculus deposits deteriorated the acceptability of the results. The human overcomes this problem by moving the mandible into the best position for analysis; something which is difficult to implement using a computer.

It has been seen that the system does not perform well when it is trained using a small number of examples, typically ten. By increasing the number of training examples the system shows greater tolerance to situations that it has not seen before. In one session the system was approaching a rate of 70% success in comparison to the results expected with a training data set consisting of

50 images. However, too many training examples saturate the system and its performance drops. Therefore, it is questionable whether the system will perform much better than currently measured without restructuring its basic architecture.

## 7 Conclusion and future

The human vs. human comparative study has illustrated that there is a degree of subjectivity in the analysis of age estimation using both methodologies. It has identified some areas where the subjectivity originates. Furthermore, it has shown that although the comparative study was carried out rigorously there are areas that the study has not been able to address. Although, the study has taken a *small* number of participants it has still been useful as a means of comparing the computer system.

The human vs. computer comparative study shows that given the correct conditions the system can perform acceptably in relation to the human participant. However, the human participants are still better adapted to undertaking this type of subjective analysis.

The next stage of the project is well on the way to implementing a system capable of interpreting kill-off patterns of sheep in order to ascertain their exploitation. Again, it will be necessary to undertake a comparative study to determine the performance of the computer system in terms of its human counterpart.

## Acknowledgements

The author would like to thank all the participants who took part in the comparative study. In addition, the loan of the mandibles by Dr Keith Dobney and Prof. Don Brothwell is appreciated. The support by Dr Julian Richards, University of York and Dr Terry O'Connor is gratefully acknowledged. Finally, the author is thankful for his parents' support and help.

## references

- Gibson, P.M. 1992a The potentials of hybrid neural network models for archaeofaunal ageing and interpretation. In: J. Andresen/T. Madsen/I. Scollar (eds), *Computing the Past; Computer applications and quantitative methods in archaeology CAA92*, 263-271, Aarhus: Aarhus University Press.
- 1992b An application using hybrid neural network models to perform archaeofaunal ageing, *Archaeological Computing Newsletter* 32, 1-6.
- 1993 The application of hybrid neural network models to estimate age of domestic ungulates, *International Journal of Osteoarchaeology* 3, 45-48.
- Grant, A. 1982 The use of tooth wear as a guide to the age of domestic ungulates. In: B. Wilson/C. Grigson/S. Payne (eds), *Ageing and Sexing Animal Bones from Archaeological Sites*, 91-108, BAR British Series 109, Oxford: British Archaeological Reports.
- Payne, S. 1973 Kill-off patterns in sheep and goats: the mandibles from Asvan Kale, *Anatolian Studies* 23, 281-303.
- Wasserman, P.D. 1989 *Neural computing, theory and practice*. New York: Van Nostrand Reinhold.

Paul M. Gibson  
Department of Archaeology  
University of York  
Micklegate House  
Micklegate  
York YO1 1JZ  
United Kingdom  
e-mail: GBYORK04.GIBSONPM@WCSMVS.INFONET.COM