9

An Application of Neural Networks to Use-Wear Analysis. Some Preliminary Results

J. A. Barceló,[1] A. Vila[2] and J. Gibaja[2]

[1] *Universitat Autonoma de Barcelona;* [2] *Consell Superior De Investigacions Cientifiques (Barcelona)*

9.1 Introduction

When studying lithic remains, archaeologists need to determine whether these remains have been used as tools for some specific task. The best way to arrive at this goal is through the analysis of macro- and microscopic traces generated by the tools' use. We should investigate, therefore, the formation process, that is, the relationship between the actual state of the material we are observing, and what produced that state.

Although the history of use-wear analysis is not very long, its has generated much debate, not only concerning technical questions, but also its application and results. Very often, this debate has been based on irrational hopes in *magical* methods. There are not universal answers for ascertaining 'use' from a set of descriptive features, but there has been a great deal of work trying to relate how specific descriptive features are related to specific uses of lithic tools.

Some authors have tried to quantify the description of use-wear evidence, starting from the processing of macro- or microscopic images (see, among others Grace *et al.* 1985, 1987; Rees *et al.* 1988; Yamada & Sawada 1993)). We have also used this approach (Vila & Gallart 1993), obtaining interesting results by quantifying and discriminating among images. The potential of image processing to use wear is, therefore, well understood and does not need more discussion here.

However, it seems obvious that any image of macro- or microscopic use-wear traces contains some level of redundancy. This redundancy is so great that most image discrimination methods cannot develop all their potential. Standard image processing methods are based on the correlation of all features in the image, those that are the result of the artifact's use, *and* those that have been produced by other kinds of processes (naturally induced wear, degree of preservation, post-depositional alterations, *etc.*). Furthermore, it is inappropriate that any discrimination method should be based on the assumption that *all* observed traces (whatever the scale, macro- or microscopic) have *equal* relevance or that they contribute in the same way to discriminating between different uses.

To change data (micro- and macroscopic images, use-wear trace descriptions) into knowledge we need ways to discern the useful information from the useless information. In this paper, we propose a mathematical methodology, inspired by Artificial Intelligence research, to discriminate between variables, and not only between images. The final goal is to obtain a set of probabilistic estimates of the differential contribution of each variable to each action performed with that tool.

9.2 Unsupervised *v.* supervised learning: the rôle of archaeological experimentation

Use-wear traces have been studied by means of various statistical classification methods. The purpose has been to produce a set of different possible uses for a lithic tool, and then to decide which one is appropriate for a tool with a specific set of descriptive features (use-wear traces).

These 'classifications' proceed by detecting statistical regularities within set of objects in a feature space, where every object is described in terms of these features. The purpose is to obtain a classification of objects according to the statistical pattern of regularity discovered. Usually archaeologists divide the population of objects into classes using a polithetic approach, that is: *tools that have been used for a specific purpose have a lot of features in common with all tools that have been used for the same activity, but not any single feature appears simultaneously in all objects.* As a direct consequence of such an axiom, all features should have exactly the same relevance, because there is not any single attribute, or attribute set that discriminates *better* between tool-sets.

How are the different classes, and therefore possible uses, calculated? It is achieved by generating a partition in the original data set in such a way that all objects used for the same activity are in a single class whose internal variance is lesser than the variance among different classes. The output of the

partition algorithm are usually *fuzzy* or *continuous* classes (De Gruijter & McBratney 1988), whose essential characteristic is that the membership is a continuous function of the descriptive features used to define individuals. The centre of gravity of such classes gives a summary of their meaning, given the fact that this centre is constituted by the most repetitive pattern of descriptive features. Statistical techniques like Correspondence Analysis, Principal Component Analysis, and others have been used to reduce the dimensionality in the original data set, allowing a more structured description of the centre of gravity in any class.

However, the resultant fuzzy classes are often difficult to interpret conceptually, and may not reflect the goals we want to develop. The centre of gravity in a continuous class of lithic tools (a specific activity performed with that tool) does not define, necessarily, the task undertaken with that tool, because it does not *only* contain those features related with use-wear. The statistically obtained relationships from the combination of attributes do not 'explain' the potential meaning of the objects they describe. In a classification of lithic tools the rules of syntactic well-formedness (statistical partition algorithms) have less importance than the search for rules of semantic manipulation, together with rules for converting between surface form and semantic structure. We need semantic properties which identify meanings, and not only syntactic properties. Consequently, the meaning of an ordination (classification) depends on the goal of classification, which is usually not defined in the algorithm which produces the ordination. If we want to define differences among tools based on their use, we need to extract only those variables that were related with that activity, considering also, that those features were *irregularly* related with that activity: some of them were produced by a single action, whereas other features could be the result of more than one action. This is not exactly the classical fuzzy logic notion of 'intensities' of objects within the same class (in terms of the Theory of Probabilities), but the notion of logical dependency between an action (use) and its result (trace produced directly or indirectly by that action — see Barceló 1996).

To carry out this task, we should select the descriptive features of the analysed entities according to their predictive power to make useful declarative statements. We should reject the classical definition of attribute in the Numerical Taxonomy paradigm ('any modality of a character used to describe a set of objects'), and prefer a more 'archaeological' definition, such as "any descriptive element with an 'archaeological' meaning" (Adams & Adams 1991; Clarke 1978; Klejn 1982). Therefore, use-wear traces are not a sufficient and necessary condition for the existence of the concept of 'use', although it is something which any kind of use empirically has. These sufficient and necessary conditions are not in the nature of the object, but in the archaeological problem we want to solve.

These questions affect directly the purpose of our analysis. Statistical similarity between the image of macro- or microscopic use-wear traces has not any meaning by itself. We need to know what we are looking for when we try to recognise some pattern discovered in the microscope. The really difficult problems of perception will not be solved by any simple classification of images; we need a generalisation of some empirical relationships or processes which runs in combination with *a priori* knowledge.

There is a family of classification algorithms that allow such a generalisation. They are part of the *supervised learning* category of machine learning methods. These systems process a preliminary data set — a learning set — in which the relationship between causes and effects is known. In our case, this is the action performed with a tool, and a description of use-wear in that tool. The purpose of a supervised classification is to generalise (learn) what we have discovered with the learning set to the entire data set — the archaeological data. Mathematically it can be done by translating the information recognised in the first sample, into a symbolic description that represents our *a priori* knowledge in an operational way. Therefore this approach has also been called the symbolic approach in clustering and classification (Kodratoff 1986; Michalski & Stepp 1983; Weiss & Kulikowski 1989).

In other words, we are looking for a mechanism to recognise patterns in data, using already stored patterns from a previous analysis of a sample of known cases. The method is very simple. Some learning elements are presented to the system as a positive or negative instance of some concept (for instance, a tool used to cut a piece of meat); the computer must find a concept description that effectively partitions all available lithic tools in two groups: positive and negative, those that were used to cut meat, and those that were used for another purpose. All instances in the positive region are believed by the learning system to be examples of the single concept. A description is consistent with the instances if it identifies all the positive instances and excludes all the negative ones.

The goal is not to explain a data set by means of a set of statistical rules, but given a badly defined concept (to cut meat, for example), to explain it through a combination of its positive and negative instances (use-wear traces that are exclusively associated with tools used to cut meat). Consequently, the method we are looking for would use:

- a set of lithic tools;
- a set of attributes (use-wear traces) to characterise these tools;
- prior knowledge: a set of possible tasks performed by these tools.

From these it should be able to find:

- a rule or discrimination function relating attributes with prior knowledge, that is to say, relating use-wear with categories of use.

What we are trying to do is turning a passive description of experimental data into an active classificatory rule that allows the correct recognition of new (non-experimental) data. Obviously, this will only be possible if there is a sub-set of features defining the experimental data.

Consequently, the classificatory process may be broken down into the following tasks:

1. definition of a description list (attributes present in the set);
2. selection of those attributes whose function is known;
3. determination of the correlation level of each attribute with the function that generated it;
4. calculation of the centre of gravity in the class of artifacts, based on the different contribution of each attribute to the discrimination function.

It is easy to see that in this approach not all variables, features or attributes should contribute in exactly the same way to the classificatory task (explanation).

There are several different methodologies for supervised classification (Bietti 1993; Bietti *et al.* 1985):

- Fisher linear discriminant;
- Linear maximum likelihood, minimising the Mahalanobis distance between the objects of the set;
- Quadratic maximum likelihood;
- Bayes, *i.e.,* classification according to the rules of Bayes theorem: one has to indicate *a priori* probabilities for each class.

Each of these methods has its own advantages and disadvantages. In this paper we have decided to use neural networks, as a tool to implement supervised learning, largely because we are working with binary data. In this case, classic statistical programs fail because mathematical models to fit the qualitatively described data are not available. Neural networks are therefore used here because they do not require an underlying prediction model.

9.3 Neural networks and experimental archaeology

Among the central problems of pattern recognition is the transition from a numerical to a symbolic representation. Archaeologists usually want to group several different numerical representations to the same symbolic representation. However, in this paper we want to explain how to transform a set of symbolic description terms (presence/absence of use-wear traces) into a series of numerical measures (the 'intensity' of each binary attribute helps to discriminate between different actions). A neural network can be trained through supervised learning to do this task (see, among others, Barceló 1995; Carpenter & Hoffman 1995; Caudill & Butler 1992; Kempka 1994a,b; Pao 1989; Zeidenberg 1990). The system we want to build is a diagnosis machine that predicts the probability that a lithic tool was used in some way, given a set of inputs (a binary description of some macro- and microscopic use-wear traces). This prediction does not follow a rigid algorithm in producing an answer based on a given input, but it is taught through training examples.

The network consists of many simple, but individual processing elements ('nodes') arranged in one or more layers, and a system of connections. These connections transmit the signals, which the nodes manipulate. A transfer function contained in each node governs this manipulation. The nodes sum weight-adjusted inputs, add a bias value, and finally pass the result through an activation function (also called a transfer or squashing function) to be used by other neurons or offered as an output.

Learning is performed in the network of connections. Although a network's transfer functions usually do not change, the connection strengths do change during learning. These changes result from the network making predictions based on the training examples, which contain known outputs based on real inputs. In our case, training examples are pairs of archaeological experimentation results, that is, the descriptive features observed in those lithic tools that were used for some specific activity in the laboratory.

Information is input to the first layer and then propagates through the structure of connections and nodes. When the information (pulses) finally reach the output nodes in the final layer, these units produce an answer (a number reflecting the intensity of the function in each unit), which is the network's prediction of the output based on the given input data. The predicted output at every node in the final layer is then compared to the correct (known) output at every node. Errors are generated as the difference between the correct output and the network's predicted output. These errors propagate backwards through the network, modifying the connection weights based on a mathematical equation that defines what is described as the learning rule. This process continues until the user is satisfied with the accuracy of the network's predictions.

Once the network is satisfactorily 'trained', it is put into actual use. The network is fed only input data, preferably data it has never seen before. Feeding the network the exact same data that was used during training only tests the network's ability to 'memorise' data. A useful network can accurately predict output for data it has never seen before.

The real challenge in developing a useful neural network system is the training process. Back-propagation (the learning algorithm used in this paper) represents a specific example of applied gradient descent to optimise a system. The training of the network is an iterative process based on large numbers of data samples representing the traffic flow within a certain region. Using standard connection weights, the network computes a set of outputs that is moved from the input layer to the hidden layer. The system compares this set of outputs with the input values by calculating a root mean square difference (or global error) and modifies the connection weights to displace the outputs toward the expected values. If the training is successful, the global error is reduced. In over-simplified terms, gradient descent works to optimise a system by minimising a given function. In the case of back-propagation, network error is minimised by optimising the weight values of the connections among nodes. Total network error is minimised by following the gradient (actually followed down towards a minima, hence descent).

The neural network we use in this paper has three layers: the input layer transforms qualitative data (presence/absence of binary descriptive features) into numerical data (the intensity of 222 nodes or processing elements, one for each descriptive feature in the database). The output layer contains the predictions of the system in 19 nodes. In the middle, there is a set of 67 'hidden' nodes, whose function is to store the specific mapping between inputs and outputs. Each feature detector in the middle layer looks for a key feature or features in the input pattern, and reacts strongly when one is found. The output layer can then construct an appropriate output pattern based on the particular combination of features the middle layer has detected. Such an appropriateness is defined in terms of the quantity of input received by an output node. The functioning of a neural network can be seen, then, as a process establishing an appropriate set of feature detectors in the middle layer and the proper response in the output layer.

9.4 Learning the relationship between qualitative description of tools and prehistoric activities: the training set

We have used the results of a previous project on experimental archaeology (*Generation, Description and Quantification of Use-Wear in Prehistoric Lithic Tools* DGICYT: PB91-0130) to obtain an adequate learning set and generalise the differential relevance of a set of 222 descriptive features, including microscopic and macroscopic observations.

Among these features, we can mention:

- general information about the shape and typology of the tool;
- microflaking: regularity, location, morphology, and size in both the dorsal and ventral surfaces;
- striations and linear features: location, ordering, shape, and length in both the dorsal and ventral surfaces;
- micropolish: location, reflectivity, regularity, thickness, invasiveness, stage of development *etc.* in both the dorsal and ventral surfaces.

We have randomly selected 55 experiments with tools, from a database with *c.* 1000 experiments, all of them undertaken with exactly the same kind of flint, and manufactured in our laboratory. All efforts have been made to ensure the right replication of these tools.

Actions on five raw materials were experimented with: fresh bone, wood (soft and hard, fresh and dry), meat and skin (from little and big mammals, fresh and dried). We also controlled the duration of each experiment: 5 minutes, 15 minutes and 30 minutes for all materials and tools used. The 19 nodes in the output layer represent, consequently:

- raw material: wood, skin, bone, meat
- condition of raw material: fresh, dry
- action: motion or use action undergone with each tool
- angle of working edge contact: dorsal, ventral
- time: short (1–10 minutes), medium (11–20 minutes), long (more than 20 minutes)

Our random sample contains a great number of experiments with wood processing, and fewer examples of the other raw materials. Consequently, this training set cannot be considered as a universal learning set, but a mechanism to refine some hypotheses. Specifically, that the variability among lithic tools used to process wood is so great that it can overlap the variability among other kinds of tools.

9.5 Redundancy and conflicting data

Our neural network cannot be used as an Expert System. It has produced around 60% correct predictions, and it cannot converge to an acceptable level of error. There is something in the database or in the feature list (input layer) that prevents learning. And this 'something' is, obviously, *redundancy* or redundant data.

At the beginning of the paper we stressed the fact that in use-wear analysis some activities have a larger

influence on some wear traces than on others. Redundancy occurs when many indicators may have the same bad influence on the output. Determining a good set of indicators is crucial in generating good predictions as well as reducing training times to reasonable levels.

Since many indicators at first glance appear to be relevant, we should perform sensitivity analysis with respect to the different inputs. This involves noting the percentage change in the output caused by a specific percentage change in one of the inputs, keeping all the other inputs the same. But we have also included the possibility of non-linear interactions, that is, changes to two or more inputs in tandem can have a different effect from that of changes to one input alone. Redundancy has not been deleted, because it was one of the goals in our analysis to evaluate if redundancy affects the quality of the classification. We have carried out only a preliminary sensitivity analysis in order to drop features that do not produce enough information. We have used the results of a binary classification of input and output units, deleting all inputs whose similarity with all output units was zero. This was achieved using Jaccard's Coefficient.

There are, however, more problems in our neural network. Conflicting data appear when two identical or 'similar enough' input patterns have entirely different outputs. The neural network cannot learn both patterns because they are contradictory. The cause of this problem may be found in the modelling process. Our task is to map input patterns into one of the 19 output categories. A modelling problem occurs if not enough fields are included in the model to distinguish adequately between the output categories, or if the output categories are correlated in some way, that is, a discrimination rule cannot be found to relate some inputs with some outputs. Programmers have tried to solve this problem by eliminating the conflicting patterns or by rebuilding the descriptive features and output units sets (Versaggi 1995). Once again, we have tried not to solve the source of data conflicts, but to study how those conflicts affect the relationship between binary description and functional attribution.

Consequently, we are not trying to automatise the work carried out by the expert archaeologists, but to study the potentials of erroneous attribution due to redundancy, conflicting data and modelling problems. Results from our research are very interesting in this respect.

- It is very difficult to discriminate between actions. Wood processing tools can have wide intra-group differences that mean some of them are like the tools used in other activities. For example, skin processing tools have only been differentiated by the high degree of wear, compared with low and middle degrees on wood, bone and meat processing tools. But a difference at this level contrasts with the lack of difference between wood and skin processing tools in characteristics like the degree of linkage. The same is true for bone processing and meat processing tools, although in this last case (meat), we have observed that the location of microflaking tends to be different, and that striations are significantly shorter than in other cases. To sum up, variability in each functional class is so great that we have not yet discovered any clear discrimination between binary descriptors.
- The condition of worked material, fresh or dry, is also difficult to discriminate between. Tools used to process fresh materials have some minor differences (*e.g.*, presence of negative edges, continuous microflaking, some specific distribution of micropolish), but most descriptors are common to both conditions.
- The other characteristics (angle of work, contact between tool and worked material, duration) are even less successfully discriminated.

We do not pretend to say that binary descriptors are not correlated with activity and conditions of work, but that the observed variability in an experimental context is so great, that we cannot build then directly into some discrimination functions or an Expert System based only on a few replications.

9.6 Statistical analysis *v.* Artificial Intelligence

Why are the neural network results so bad? Or we could ask, what is the interest in creating a computer program that cannot explain the use of lithic tools? Our intention in the first part of this project is not to produce a successful Expert System, but to study why archaeologists make mistakes, even though they mostly do not know that they are mistaken.

Our neural network is a simulation of expert archaeologists who cannot explain their data. Mistakes and mis-attributions made by the program are like errors made by the archaeologist. This is the interest of this part of the project: to simulate archeological reasoning in order to discover where the error is.

Consequently we have compared the neural network's actual state of activation (after *c.* 900 iterations and one modification of its topology) with a statistically obtained fuzzy cognitive map. For practical reasons we have only processed five output units: raw material worked with that tool, and condition of this raw material. The purpose was to investigate the relationship between the most relevant inputs and this short list of outputs, and how the relationship between outputs affected the discrimination rule. We have built a hybrid neural/statistical machine, where the link weights have not been trained, but they have a fixed value: Jaccard's coefficient between every input and every output. That is to say, we have calculated the similarity (the number of cases where both

features appear together) between the description and the experiment that would have produced that wear. Results are only relatively similar with those obtained through back-propagation.

First of all: the neural network could not learn to discriminate meat processing tools because there were not enough cases in the random sample of experiments. The neural network has tried to discriminate based on what it knew about other raw materials, but program's guesses have been demonstrated to be wrong.

Secondly, there is some degree of discrimination between tools, although the neural network was unable to discover it. The processing of wood, bone and skin produce different use-wear traces:

- Wood: one edge employed only, right edge, little and regular microflaking, with semicircular shape, and single ending, some micropolish in edges.
- Bone: more than one edge employed, big, irregular and continuous microflaking, with trapezoidal shape and abrupt ending, short striations all over the tool, evidence of micropolish in marginal location.
- Skin: irregular edge, distinctive microflaking, strong evidence of micropolish.

Why did the neural network not discover this discrimination? It did not because units in the output layer are correlated. That is, because the way the action was performed affects the traces produced by the task itself. In other words, in the sub-set of results analysed here, the variability of use-wear traces among tools used to process dry raw materials has no linear discrimination between wood and bone processing tools. We see that some input units (descriptors) are positively correlated with one output unit, but when we analyse the binary correlation (Jaccard's coefficient) between these input units and more than one input (for instance, tools used to process dry wood, or tools used to process fresh bone), the discrimination rule changes significantly. Any discrimination among tasks disappears when we introduces the discrimination in other experimental conditions. These different sources of discrimination are not linearly separable. It is not the same to learn to discriminate between wood and bone processing as to discriminate between cutting fresh-wood during 15 minutes, and cutting dried wood during 30 minutes.

9.7 Conclusions

If we want to understand the specific relationship between wear evidences (at the microscopic or macroscopic level) and function, we have to take into account all conditions and circumstances in which the tool was used, and not only the use alone. We still think that a neural network approach is the best way to discriminate the influences of all factors affecting wear descriptors, but only if we eliminate all potentials for redundancy and conflicting data.

It is important to realise, however, that sometimes the multidimensional nature of the wear/function relationship has been confused by an erroneous understanding of redundancy. We think that redundancy exists at the level of descriptors, and not of explanations. The final version of our neural network will include only those binary descriptors related significantly with tasks, and a totally different output layer, where units will be unrelated. For example:

Unit 1 experiments made on fresh meat during 30 minutes, and in conditions x, y, and z.

Unit 2 experiments made on fresh meat during 30 minutes, and in conditions a, b, and c.

Unit 3 experiments made on fresh bone during 30 minutes, and in conditions x, y, and z.

Unit 4 experiments made on fresh bone during 30 minutes, and in conditions a, b, and c.

Our research has demonstrated that we need more experimental control among functional classes if we want to build efficient discrimination rules. Use-wear analysis is more complex than we usually believe. There are statistical and mathematical tools to process this complexity, but we as archaeologists should take into account the potential of redundancy, conflicting data and modelling problems before building impressive but inefficient expert systems.

References

ADAMS, W. Y. & E. W. ADAMS 1991. *Archaeological Typology and Practical Reality.* Cambridge University Press, Cambridge.

ANDERSON, P. C., S. BEYRIES, M. OTTE & H. PLISSON (eds.) 1993. *Traces et Fonction: Les Gestes Retrovés.* Études et Recherches Archéologiques de l'Université de Liège. CNRS, Liège.

BARCELÓ, J. A. 1995. 'Back-propagation algorithms to compute similarity relationships among archaeological artifacts.' *In* J. W. Wilcock & K. Lockyear (eds.), *Computer Applications and Quantitative Methods in Archaeology 1993*, pp. 165–176. British Archaeological Reports International Series 598, Oxford.

BARCELÓ, J. A. 1996. 'Heuristic classification and fuzzy sets. new tools for archaeological typologies.' *In* H. Kamermans & K. Fennema (eds.), *Computer Applications and Quantitative Methods in Archaeology CAA95*, pp. 155–164. Institute of Prehistory, University of Leiden, Leiden. Analecta Praehistorica Leidensia 28.

BIETTI, A. 1993. 'Supervised pattern recognition and experimental archaeology.' *In* Anderson *et al.* (1993), pp. 529–538.

BIETTI, A., A. BURANI & L. ZANELLO 1985. 'Interactive pattern recognition in prehistoric archaeology: some applications.' *In* A. Voorrips & S. H. Loving (eds.), *To pattern the past : symposium organized by the Commission IV of the International Union of Pre-and Protohistoric Sciences (UISPP) at the University of Amsterdam, Amsterdam, May 1984*, PACT 11, pp. 205–228. Council of Europe, Strasbourg.

CARPENTER, W. C. & M. HOFFMAN 1995. 'Training backprop neural networks.' *AI-Expert* 10(3): 30–33.

CAUDILL, M. & C. BUTLER 1992. *Understanding Neural Networks.* MIT Press, Cambridge, Massachusetts.

CLARKE, D. 1978. *Analytical Archaeology.* Methuen, London, second edition.

DE GRUIJTER, J. J. & A. B. MCBRATNEY 1988. 'A modified fuzzy k-means method for predictive classification.' *In* H. H. Bock (ed.), *Classification and Related Methods of Data Analysis*, pp. 97–104. North-Holland, Amsterdam.

GRACE, R., I. D. GRAHAM & M. H. NEWCOMER 1985. 'The quantification of microwear polishes.' *World Archaeology* 17(1): 112–120.

GRACE, R., I. D. GRAHAM & M. H. NEWCOMER 1987. 'Preliminary investigations into the quantification of wear traces on flint tools.' *In* G. Sieveking & M. H. Newcomer (eds.), *The Human Uses of Flint and Chert*, pp. 63–69. Cambridge University Press, Cambridge.

KEMPKA, A. A. 1994a. 'Activating neural networks. Part I.' *AI-Expert* 9(6): 33–37.

KEMPKA, A. A. 1994b. 'Activating neural networks. Part II.' *AI-Expert* 9(8): 42–49.

KLEJN, L. S. 1982. *Archaeological Typology.* British Archaeological Reports International Series 153, Oxford.

KODRATOFF, Y. 1986. *Leçons d'Apprentissage Symbolique.* Cepadues Editions, Toulouse.

MICHALSKI, R. & R. E. STEPP 1983. 'Automatic construction of classifications: conceptual clustering versus numerical taxonomy.' *IEEE Transactions in Pattern Analysis and Machine Intelligence* 5(4): 396–410.

PAO, Y. H. 1989. *Adaptive Pattern Recognition and Neural Networks.* Addison-Wesley, Reading, Massachusetts.

REES, D., G. G. WILKINSON, C. R. ORTON & R. GRACE 1988. 'Fractal analysis of digital images of flint microwear.' *In* S. P. Q. Rahtz (ed.), *Computer and Quantitative Methods in Archaeology 1988*, pp. 177–183. British Archaeological Reports International Series 446, Oxford.

VERSAGGI, M. R. 1995. 'Understanding conflicting data.' *AI-Expert* 10(4): 21–25.

VILA, A. & F. GALLART 1993. 'Caracterización de los micropulidos de uso: ejemplo de aplicación del análisis de imágenes digitalizadas.' *In* Anderson *et al.* (1993), pp. 459–465.

WEISS, S. M. & C. A. KULIKOWSKI 1989. *Computer Systems That Learn.* Morgan Kaufmann, San Francisco.

YAMADA, S. & A. SAWADA 1993. 'The method of description for polished surfaces.' *In* Anderson *et al.* (1993), pp. 447–457.

ZEIDENBERG, M. 1990. *Neural Networks in Artificial Intelligence.* Ellis Horwood, Chichester.

J.A. Barceló
Universitat Autonoma de Barcelona
Divisió de Prehistòria
Facultat de Lletres
Edifici B
08193 Bellaterra
Spain

ilphd@blues.uab.es

A. Vila and J. Gibaja
Consell Superior de Investigacions Cientifiques
Institució Milà I Fontanals
Laboratori d'Arqueologia
c/ Egipcíaques 15
08001 Barcelona
Spain