Fuzzy Logic Application to Artifact Surface Survey Data

Emeri Farinetti,[1] Sorin Hermon[2] and Franco Niccolucci[2]

[1]University of Leiden (NL)
[2]P.I.N. Prato (I)
sorin.hermon@pin.unifi.it, niccolucci@unifi.it

Abstract. Survey archaeologists, studying human activities over space during time, need to assign a chronological framework to their field data and consequently produce maps showing sites and their chronology, determined mainly by the cultural material. Specialists who assign chronological attributions, have difficulties with rigid chronological categories, and tend to create additional ones, in order to match their data. Moreover, they face materials with uncertain, or multiple chronology, and end up with large chronological ranges. Survey archaeologists need to extract meanings out of those data and, using traditional classification methods, are forced to reduce those classes into fewer, and often not representative of the surface data. We propose to use a fuzzy logic approach in order to give data more transparency and to present a more realistic map, according to the real nature of the data.

1. Fuzziness in Artifact Surface Surveys Datasets

In the methodological framework of intensive systematic surveys, the distribution of material over the space allows us to interpret, at several degrees, the distribution of human activities in the landscape, during time. Once recognized an activity focus – or site – in the landscape, by using criteria related to density of surface material (corrected by several variables) and geomorphological characteristics of the landscape, we then go much deeper into the 'site', recording and collecting material at a higher level of intensity. Result of the process are assemblages of material, spatially linked to a survey unit (geometric-regular or irregular according to cases), that have to be processed and studied in order to produce meaningful results on the occupation of the landscape on that site-area. Of different nature and entity are the problems that a landscape archaeologist has to deal with while managing data from a systematic intensive artifact surface survey.

Data representativity, as well as data reliability and period visibility are, for instance, crucial issues to be considered while interpreting data and extracting meanings from different landscape datasets.

In this paper, we would like to examine one of the many critical issues linked to the methodology of data retrieving from survey data. We will concentrate especially on the fuzziness of chronological attribution, crucial factor in order to recognize the presence of an activity focus in a certain period, as well as to assess the continuity of occupation of a certain site.

Time, as well as space, are distinctive factors in the archaeological research and in the interpretation of the human landscape. Time, seen as the chronological attribution to material culture, is always characterized by a fuzziness, that can be named 'temporal fuzziness'. This has to do with our imprecise knowledge of the material culture in the different periods of the past, but also with the need, for research purposes, to look at chronological ranges, since we cannot go back to the individual moment in which, for instance, a pot was made. We need, therefore, to classify the chronological attributes. This process ends up with the creation of a huge variety of classifications, that become even more if one takes into account the terminology in use for the diverse chronological classifications. Often the chronological systems in use in different areas differ quite sensibly, and many examples of either slight or substantial diversion between chronological classification systems can be found. This has to do with the fuzzy nature of period terminology: in Greece, for instance, discrepancies can be found in the ways of defining the Hellenistic period. Some scholars use the chronological range HL as corresponding to the Hellenistic period according to historical sources (which means from the death of Alexander the Great, dated 323 BC ...). On the other hand, in terms of material culture/archaeological material, and in terms of settlement pattern, EHL is differentiated from LHL: EHL goes often together with LC and is differentiated from LHL and then from the ER period. The LC is a period of peak in the settlement pattern and high densities on the landscape, while LHL and ER periods are characterized by a decrease in site density and appearance of larger estates in the landscape.

We must therefore devise – as stated recently by Van Leusen (2002: 238) – a system of fuzzy dates that would allow us not to loose the fuzzy nature of period terminology. Using fuzzy logic we would try to assess a better categorization of real data that better matches the actual reality of occupation, and the degree of occupation in the landscape for each period.

In the following section, we will focus on the problem of chronological attribution to sherd assemblages.

2. Fuzziness in Chronological Attribution to Sherds

During surveys, surveyors collect material, more or less diagnostic, from the surface. In the case of site samples, different and various assemblage of sherds constitute the data-sets to be processed. During this process, each sherd is given a unique number and is examined. All characters, corres-

ponding to different attributes/fields assigned to each single sherd, are inserted into a database. One of these attributes is chronology, that could also be assumed from a set of given attributes, already inserted into the database by students and by the pottery specialists themselves.

Since it depends on many different attributes (fabric, inclusions, shape, type, decoration, etc), and especially as it depends on the subjectivity of each single specialist and each single specialist's knowledge, chronological attribution is marked by a strong degree of subjectivity.

Therefore, the chronological attribution assigned to the diagnostic materials has to be considered as a fuzzy property. Several are the possible truths as far as the chronological attribution of an individual sherd (and the object once it had belonged to). In other words, we deal with large chronological ranges, in which a type of pot might have been produced and used. Pottery specialists, asked to give a chronology to each individual sherd, can date some sherds (decorated, for instance) with much more certainty than others. Certain categories (specific wares, as black glazed pots or medieval decorated wares, etc.) can be assigned with precision to a period. On the other hand, other sherds (coarse wares, typical finds of a surface survey, not easily recognizable, small pieces, often in very bad conditions) might be assigned to larger chronological categories, comprehensive of more chronological periods, either because of the uncertainty of the attribution or because of the continuous production and use of a type of pot during a large span of time.

Therefore, instead of dealing with a usual Boolean logic: TRUE or FALSE, we face attributes which are, in the majority of cases, neither TRUE nor FALSE as far as an individual period, but TRUE for a period and TRUE also for another, and therefore we deal in the majority of cases with a 'multiple truth' at the end (a sherd can be attributed to 2 or more than 2 periods). The result is a very fragmented picture if we take into account each chronological attribute assigned (Fig.1).

Sherds that can be assigned a sure and unique chronological attribution (lc, ehl, lhl, etc) are very few, in percentage. Therefore, we cannot work out any meaningful and representative picture of the chronology of sherd-assemblages on the basis of the clear (no-fuzzy) chronological attributions only.

A sherd dated by the pottery specialist as A-H can be assigned either to the chronological range A or C or LC or EHL. On the other hand, a sherd dated as (c) h-r can belong either to the chronological range LC or EHL or LHL or ER (less probably though to the LC or EHL period). In the same way, a sherd dated as (c)-h can be assigned either to the chronological range LC or EHL, but more probably to the HL period, etc.

As we just saw in the examples, those wider ranges can be split up, conceptually, into all the periods that constitute them (all the periods archaeologically meaningful). Therefore, in many cases (the majority sometimes) a sherd can be classified into more than one chronological range, according to the 'degree of belonging' of the item/sherd to a particular chronology.

Following this logic, the alternative proposed in this paper to usual chronological attribution, is to apply fuzzy logic concepts since the data set is marked by a high degree of fuzziness.

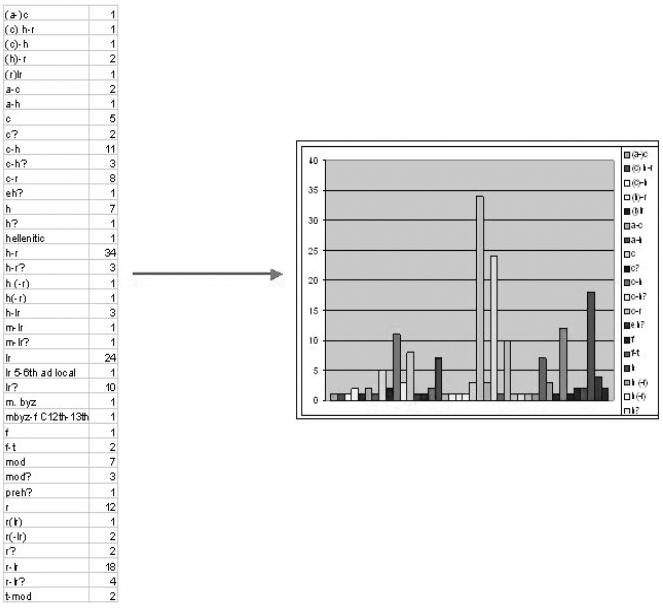

Fig. 1. Fuzzy fragmented picture from the potsherds dataset.

3. The Fuzzy Logic

Fuzzy logic (Zadeh 1965, Novak 1989, Kosko 1993, McNeill and Freiberger 1993) aims to extend ordinary deductive methods by assigning a numerical degree of truth or falsity to statements not completely true nor completely false, otherwise indiscriminately cast into the "neither" category, or into the 'multiple' category, probably wasting valuable information (Copeland 1997) or constituting the multiple field not searchable through queries. The range of truth values is the closed interval [0;1] of real numbers. The degree of fuzzy belonging of a member to its set is also given by a real number between 0 and 1 (Kosko 1993, Hermon and Niccolucci 2002).

The application of fuzzy logic to chronological attribution to archaeological material in a artifact surface survey context implies using a matrix A with items (individual sherds) as rows and chronologies as columns, as inventory catalogue of the assemblage under study. The matrix elements are the fuzzy coefficients, expressing our certainty that each item belongs to each chronological range. It is also defined the reliability index R for each chronological attribution, for each item and for the entire assemblage, according to this formula:

$$R(x_1, x_2, \ldots x_m) = \frac{\left(\max_k x_k\right)^2}{\sum_k x_k} \qquad (1)$$

The fuzzy method may help to identify sherd assemblages 'difficult' as far as chronology. We can define a reliability index of each item which takes into account jointly the spread of possibilities assigned to that item (i.e. the number of different chronological ranges considered as possible for it) and the value of the most reliable assignment.

In order to process and analyze our data-sets, the first thing to do could be therefore to 'deconstruct' the chronological ranges in their minimum components (i.e. into the individual relative chronological ranges).
In fuzzy (logic) terms, in the matrix A with sherds as rows and chronologies as columns, for the sherd dated lc-ehl the 'r' index is 1 because the sherd can be dated LC or EHL with the same index of reliability , on the other hand for the sherd dated as R-(LR) the 'r' index is 1 (ER) – 1 (MR), but is 0.7 (LR). This is because some attributes of the sherd do not seem to correspond to the LR characteristics, while others do. The same happens, for instance, in the case of sherds dated (r)-lr. On the other hand, the Reliability index can be 0.3 when a sherd very probably cannot be attributed to a period, but it has few characteristics that cannot totally exclude its attribution to that chronological range (Fig.2).

	Chron 1	Chron 2	Chron 3	Chron 4	Chron 5	Chron 6
Sherd 1	0.5	1	0.1	0	0	0
Sherd 2	0	0	1	0	0	0
Sherd 3	0	0.2	0.7	0.7	0.3	0
Sherd 4	0.7	0	0	0	0	0
Sherd 5	0	1	0	0	0	1

Fig. 2. Potsherds attribution to chronological ranges using fuzzy logic.

The interpretation process for the assignation of the fuzzy R index to each chronological attribution is based on and takes into account the subjective process which lead the specialist to that chronological attribution, which in other words means also: What does the chronological expression R-LR, for instance, correspond to in the mind of our pottery expert?

Dealing with fuzziness in chronological attribution to sites

We come now to our second point: the chronological attribution to sites.
The survey archaeologist's urge of giving a reliable chronological attribution to sherd assemblages is strictly linked to the necessity of giving a reliable chronological attribution to the sites (or probable sites, or small activity focuses) that fill in the landscape under study.
As we just saw before, the chronological attribution to each single sherd is characterized by strong subjectivity, resulting in a certain degree of fuzziness. Consequently, subjectivity works also at the level of dating the site itself.
What survey archaeologists have been doing traditionally is to group, in a subjective way, several sherd assemblages, dated in diverse chronological sub-periods, into larger chronological ranges meaningful for the study of the settlement pattern, often interpreting an impression they had from the landscape.
In such a grouping process, for instance, in the category a-eh will fall all sherds dated as c, hl, a-c, c-hl.
That way, often a-systematic as well as totally subjective, we would not monitor the process and the resulting picture could be biased. We could for instance enhance the biases inner in the knowledge of pottery, in the attempt to get rid of difficulties, instead of dealing with the uncertainty and subjectivity linked with the interpretation of data.
Taking out the categories with a higher degree of uncertainty we would take out of the picture less known periods, that would be that way yet more forgotten.
Moreover, the result is, as we see in Fig.3, that each site (VM1, VM29, VM61, etc. in the picture) has got a different periodisation, in the practice, and we face a non consistent system of periodisation. When preparing final maps and graphs for publication, then, survey archaeologists reduce the periods of site occupation to few chronological ranges, equal for all sites, comprehensive of a larger time-span, therefore attributing the site to certain periods, even if there is a degree of uncertainty concerning the individual periods of the site's life.
That way, a landscape history is reconstructed during the different periods. Examples come, in the case study presented here, from multi-period sites found in the Valley of the Muses (Boeotia – Central Greece). In the picture, you can see the area in the Preh periods, Neolithic, EBA-MBA, and LBA, and in the Historic periods A-G, C-HL, and LHL-ER, LR.
We use data coming from a survey carried out in the '80s, a quite old set of data, not yet published. The survey was jointly directed by J.Bintliff and A.Snodgrass[1]. Within the datasets, chronological attribution really varies from sherd to sherd, without following any structured categorization, and the ceramic material from each site is collected in small quantity, often to small to operate statistical analysis.
The presence of a site in a certain period, even if clearly recognized in the landscape by surveyors then, turns out to be less clear in the final picture, once processed all the material coming from the site and plotted the result on a map, especially because the material is fragmented among several periods, once taken out the undiagnostic material.
Fuzzy logic approach, considering the fuzziness inner to each chronological attribution, and looking at each individual case, could help us to extract meaningful traces of occupation in each period. Furthermore, not being a statistical approach, it could help us to deal with very small sets of data.
In figs. 4–5, we compare site data (from different sites found in the valley), analyzing them according to the 'until now used' grouping of chronological ranges, just examined before,

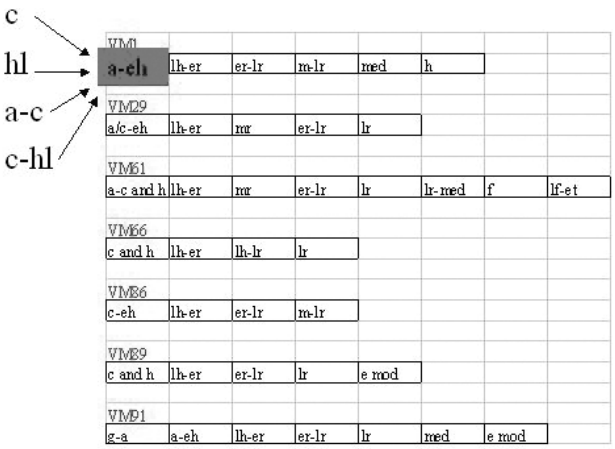

Fig. 3. Chronological attribution to sites using a non consistent periodisation system.

and then sub-dividing the ranges into their individual components and finally applying fuzzy logic.

In the example in Fig.4, we can see how fuzzy logic can help to differentiate a site's life into shorter time-ranges within a larger chronological time-span.

Fuzzy logic can also help to appreciate and better evaluate the transition from one phase to the other.

In Fig.4 (site VM1), for instance, we can clearly recognize in the site's life two main peaks (applying both methods) but through fuzzy logic we appreciate better the transition phases.

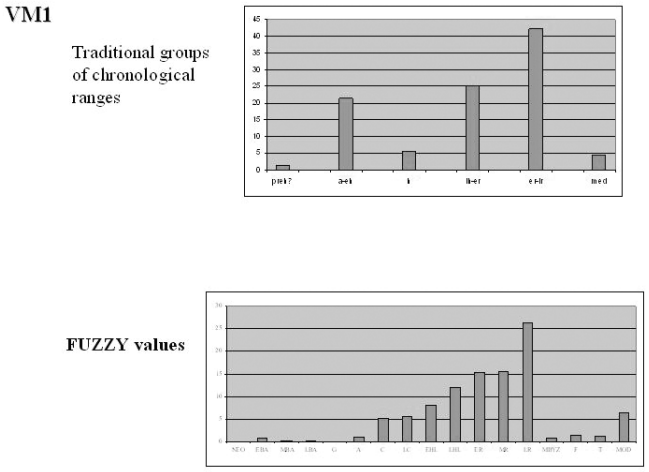

Fig. 4. VM1 site example.

This site shows a clear peak of occupation in the earlier historical periods, as well as the site VM89. The site VM89 is a clear example of farm-site of C period, with a recognizable peak of occupation towards the end of the C period and the beginning of the HL, which is the usual picture in the Greek landscape and it is clearly recognizable in this case only with the application of fuzzy logic.

On the other hand, in the case of the site VM61 the C occupation is less clear. Traditional grouping gives us the idea of a clear presence of a site in earlier historical periods, while, applying fuzzy logic, we see that we would probably need a further look into the material and into the landscape to verify reliability of data. The same happens in the case of site VM86. Fuzzy logic helps therefore also to verify the representativity of data and to focus the problems still to be solved, and to answer questions like: can a sherd concentration be recognized as a site also in earlier (or later) period of occupation?

In a GIS environment, re-reversing the database into the landscape, plotted data deriving from the calculation of sherds for each period would constitute layers period by period. The picture would change a lot, according to the classification methods used to define chronology (of sites or of site areas). Therefore, within the GIS system, the subsequent analysis, carried out using the base chronological maps, would easily result distorted if data are biased since the beginning.

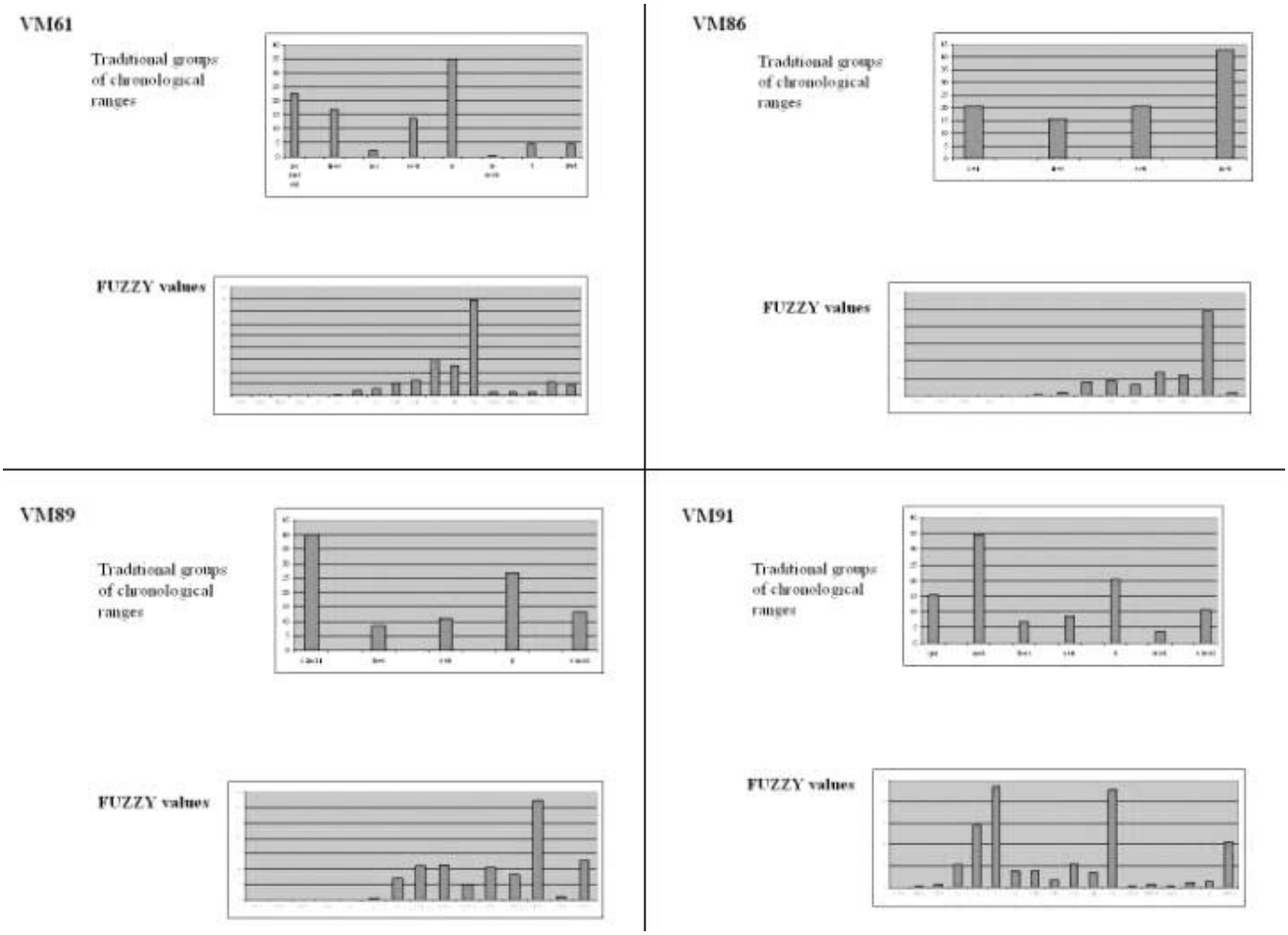

Fig. 5. Four other site examples.

4. Conclusions

To sum up, we therefore should deal with the degree of uncertainty since the beginning of the processing of our collected datasets, and therefore at the sherd database level, as we saw, in order to get less biased results at the end of the process.

We think that fuzzy logic could be therefore considered as a useful tool to evaluate settlement history, to make decisions and to throw light on the uncertainty, dealing in a positive way with the fuzziness inner to the available datasets.

References

Bintliff, J., 1996. The archaeological survey of the Valley of the Muses and its significance for Boeotian History. In Hurst, A. and Schachter, A. (eds), *La montagne des Muses*, Genève. 193–210.

Copeland, B. J., 1997. Vague identity and fuzzy logic. *The Journal of Philosophy* 94 (10), 514–534.

Hermon, S. and Niccolucci, F., 2002. Estimating subjectivity of typologists and typological classification with fuzzy logic. *Archeologia e Calcolatori* 13, 217–232.

Kosko, B., 1993. *Fuzzy Thinking: The New Science of Fuzzy Logic*, Hyperion, New York.

McNeill, D. and Freiberger, P., 1993. *Fuzzy logic*, New York, Simon and Shuster.

Niccolucci, F., D'Andrea, A. and Crescioli, M., 2001. Archaeological applications of fuzzy databases. In Stančič, Z. and Veljanovski, T. (eds), *Computer Archaeology for Understanding the Past CAA2000, Proceedings of the 28th Conference, Ljubljana, April 2000*, Oxford, ArchaeoPress BAR International Series 931, 107–116.

Novak, V., 1989. *Fuzzy Sets and their Applications*, Adam Hilger, Bristol.

Van Leusen, M., 2002. Pattern to Process. PhD dissertation, University of Groningen (NL).

Zadeh, L. A., 1965. Fuzzy Sets. *Information and Control* 8, 338–353.