DETERMINATION OF ARCHAEOLOGICAL STRUCTURE BY SPATIAL ANALYSIS AND ASSOCIATION

M.A. Kelly and J.G.B. Haigh†

School of Archaeological Sciences,
University of Bradford, Bradford, West Yorkshire BD7 1DP

1. Introduction

Spatial analysis may be described as the objective determination of relationships within and between spatial distributions. Such relationships are important in archaeological practice, during the interpretation both of site and of regional survey. It is not surprising that statistical techniques have been increasingly applied over the past decade to problems of spatial analysis in archaeology, especially by adherents of the "New Archaeology" (Hodder and Orton, 1976). Yet the majority of archaeologists still make minimal use of such techniques. Apart from a general unease about the use of numerical methodology, there has also been some criticism of the techniques currently available, in respect of their applicability to the problems of spatial archaeology (Orton, 1980). The aim of this paper is to present a broad discussion of such criticism, illustrated by some distributions within sites.

Psychological models of human perception and cognition indicate that the unaided interpretation of spatial information, even from a small site, is extremely subjective and non-reproducible. Thus quantitative spatial analysis is required to provide an objective basis for interpretation. In its simplest form it may provide a means of describing the patterning within a single distribution; in more sophisticated forms, it may be used to compare two or more different distributions, so that tests of hypotheses relating to different pattern types may be constructed. The archaeologist can then attempt to identify the processes which produced the spatial distributions in his records, but in practice he may find that the usefulness of the quantitative techniques is somewhat limited.

The null-hypothesis for many of the tests is that of a random distribution over the region of investigation. This concept of randomness need not be unacceptable to archaeologists for, although individual human behaviour is deterministic, the composition of many activities tends to exhibit strong random characteristics. On the other hand randomness may not relate to the original human occupation of a site, but rather to its post-depositional history, so that it becomes necessary to model all processes right up to the time of excavation. Entropy might be a useful concept to employ in this connection; a high value for the entropy would indicate that a distribution is strongly disorganised and contains only a minimal amount of useful archaeological information. In general, the entropy

† Also attached to the School of Mathematical Sciences

of a site is expected to increase with time, as physical, chemical, biological and human agents disorganise the original archaeological structure.

In the case of nearest-neighbour analysis, the alternative hypotheses also present difficulty in their archaeological interpretation. Although familiar to many archaeologists, this technique was first developed in plant ecology where it is applied to a random sample taken from a quasi-infinite distribution. In archaeology it is normally applied to a census taken over a region where bounds are predefined. Clustering and regularity are usually taken as alternative hypotheses to randomness; of these, clustering is particularly sensitive to the density of the distribution, which in turn relates to the area of the census region and to the geometry of its boundaries. Although a number of methods have been put forward to account for boundary effects within a finite region (Donnelly, 1978), none of them gives an entirely satisfactory solution to the problems. In fact, when applied to a census, nearest neighbour analysis may be regarded as an indicator of distribution dispersion, rather than of intrinsic patterning. The archaeologist is, however, more concerned with local patterns within the site than he is with a description of the overall distribution.

The problems described above in applying parametric tests to spatial distributions become yet more acute when applying them to spatial associations. It is therefore necessary to seek alternative means of approaching the problem.

2. Non-parametric methods for spatial association

The obvious way of avoiding the complex procedures required to parametrise spatial associations is to turn to non-parametric statistical methods. Among a wide variety of non-parametric tests, the Mann-Whitney U-test is a good example that may be applied to the measurement of spatial association (Freund, 1979). The U-test enables one to examine two samples of a continuous variate and to determine whether the samples are likely to have been drawn from the same continuous distribution.

Given two spatial distributions (A and B) of discrete point events, one measures the distance between each point of distribution A and its Kth nearest neighbour in the same distribution A. Likewise one measures the distance between each point of distribution B and its Kth nearest neighbour in distribution A. Under the null-hypothesis of no association between spatial distribution A and spatial distribution B, the distribution of BA Kth neighbour distances should resemble the distribution of AA Kth neighbour distances. In reality, a slight difference between the two distributions is expected because of dependence within the AA distances, but this should have no significant effect on the results of the test. If the two spatial distributions are positively assoicated, then the BA distances could be expected to be smaller overall than the AA distances whereas, if the spatial distributions are dissociated, then the BA distances will tend to be greater than the AA distances.

Since the Kth nearest neighbour distance is a continuous variate, the Mann-Whitney test may be used to compare the two samples (AA distances and BA distances) by calculating the statistic

$$U = n_A n_B + \frac{n_A(n_A + 1)}{2} - R_A$$

where n_A = number of AA distances

n_B = number of BA distances

R_A = sum of the ranks applied to the AA distances where the AA and BA distances are ranked jointly.

Under the null-hypothesis of no association between the two spatial distributions, the sampling distribution of U is expected to have mean and standard deviation

$$\mu_U = \frac{n_A n_B}{2} ,$$

$$\sigma_U = \sqrt{\frac{n_A n_B (n_A + n_B + 1)}{12}} ;$$

furthermore, if n_A and n_B are both greater than eight, the sampling distribution of U is approximated closely by a normal distribution with standard deviate

$$z \approx \frac{(U \pm \frac{1}{2}) - \mu_U}{\sigma_U}$$

for smaller values of n_A and n_B, exact tests may be used, based on special tables.

Figure 1 shows a plan of the Iron Age/Roman cemetery at Wederath in Germany, which may be used to illustrate the Mann-Whitney test when the graves dated by coins are used as the study population. The association between graves dated 1-50 A.D. and graves dated 51-100 A.D. was tested up to fifth order nearest neighbours, and the results are shown in table 1a. The results of similar tests of association between graves dated 51-100 A.D. and graves dated 101-150 A.D. are shown in table 1b. The results of the first set of tests (table 1a) show some tendency towards dissociation, but the trend is not statistically significant for any single result and disappears with higher order neighbours. The second set of tests (table 1b) exhibit a statistically significant degree of dissociation at each of the five orders of nearest neighbour distance. Archaeologically, this could imply some form of spatial development of the cemetery over time.

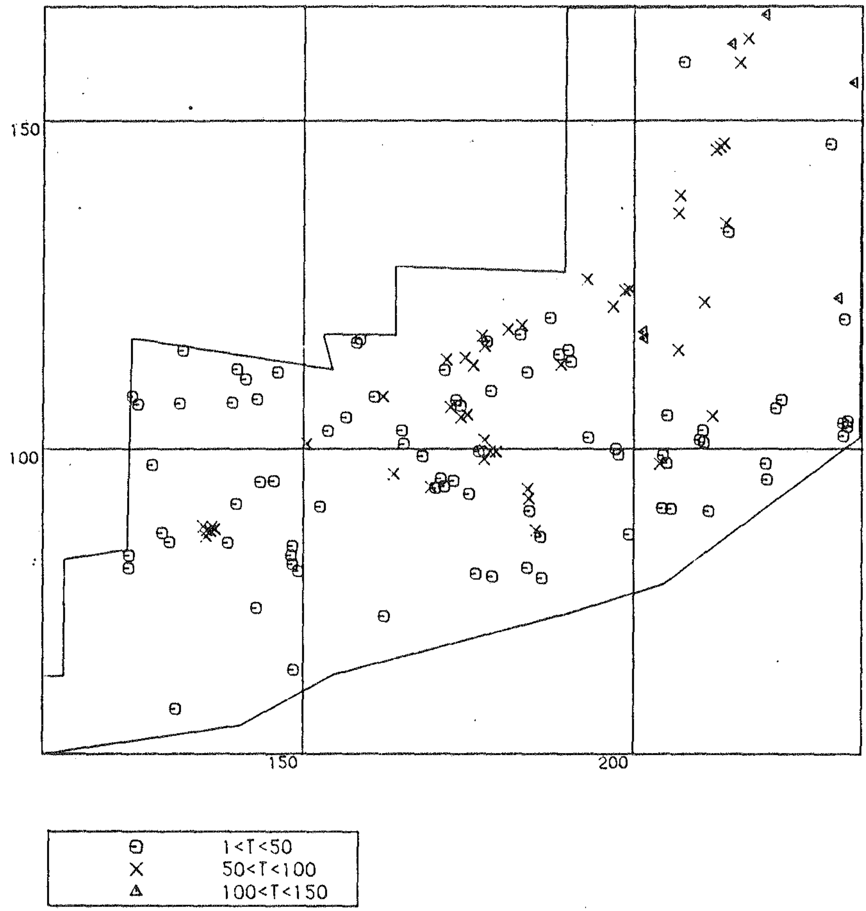

Figure 1: Iron age/Roman Cemetery at Wederath, Germany
Scale - 1:1000

nearest neighbour order	n_A	n_B	R_A	U	μ_U	σ_U	Z	P[U ≤ Z]
1	83	44	5093	2045	1826.00	197.37	1.1071	0.8659
2	83	44	5060	2078	1826.00	197.37	1.2743	0.8987
3	83	44	5069	2069	1826.00	197.37	1.2287	0.8904
4	83	44	5389	1749	1826.00	197.37	-0.3876	0.3492
5	83	44	5252	1886	1826.00	197.37	0.3015	0.6185

Table 1.a

Wederath association tests A = graves 1-50 A.D.
B = graves 50-100 A.D.

nearest neighbour order	n_A	n_B	R_A	U	μ_U	σ_U	Z	P[U ≤ Z]
1	44	6	1039	215	132.00	33.50	2.4630	0.9931
2	44	6	1038	196	132.00	33.50	1.8957	0.9710
3	44	6	1034	220	132.00	33.50	2.6122	0.9955
4	44	6	1049	205	132.00	33.50	2.1644	0.9845
5	44	6	1042	212	132.00	33.50	2.3734	0.9912

Table 1.b

Wederath association tests A = graves 50-100 A.D.
B = graves 100-150 A.D.

At first sight, the use of non-parametric statistics appears to provide an ideal solution to problems of spatial analysis in archaeology; the statistics are relatively easy to understand and it is possible to avoid the creation of complex models of probabilities. On the other hand, the results simply give a description of what associations are likely to be present, but give no idea of the actual spatial structure of such associations; this precludes any real understanding of the archaeological processes involved. Thus a proper understanding of the archaeological data still requires some form of parametric approach to the problem (Mood et al, 1979).

3. A simple method for mapping associations

Given two distributions (A and B) over a region of area α, a simple method for mapping the distribution of the association between them may be achieved as follows (Johnson, 1977). First the global density D_B of the distribution B is calculated:

$$D_B = \frac{n_B}{\alpha}$$

where n_B = number of points in distribution B.

Then a sampling radius of size r is chosen and a sampling circle is constructed around each point of distribution A. If it is assumed that there is no association between the two spatial distributions, then the number of B points within each sampling circle should follow a Poisson distribution, where the probability of obtaining K points within a given sampling circle is

$$p(K;\ \lambda) = \frac{\lambda^K}{K!} e^{-\lambda} \qquad \ldots \quad (1)$$

with $$\lambda = D_B \pi r^2 .$$

For each of the A points, the number of B points within its sampling circle is observed, and its cumulative probability is calculated on the basis of the probability distribution (1). Thus the cumulative probability is obtained at each of the A points, and may be mapped using an appropriate localised contouring program. The resulting map is used as a local measure of association.

As an example, the method is applied to the early mesolithic site of Plateau - Parrian at Mussidàn in France (figure 2). The association between the distribution of burins and scrapers (grattoirs) was mapped using a sampling radius of 2 metres and is shown in figure 3. To produce figure 3, the burins were supposed to form distribution A and the scrapers distribution B, but a similar diagram is produced when their roles are reversed. The contour map indicates two main loci of association, which may be interpreted as centres of areas used for the preparation of skins. When the association map of figure 3 is overlaid on to the excavation map, there is a clear relationship between the loci of association and the remains of habitation structures.

A problem arises in the choice of the sampling radius r; clearly different values of the parameter r will lead to varying interpretations of the intensity of association. In the case of the above example, the main loci of association remain effectively constant for values of r over the range 1 metre to 5 metres, showing that the significant features of the map are effectively independent of the choice of parameter.

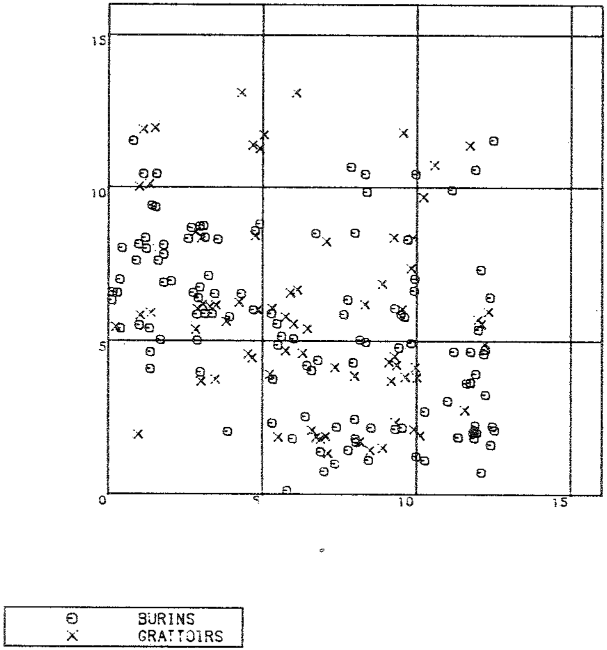

Figure 2: Early mesolithic site of Plateau-Parrian, France.

Scale - 1:200

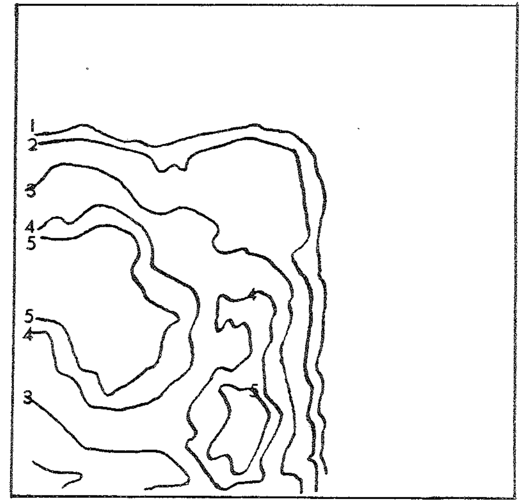

Figure 3: Mapped association for Plateau-Parrian. r = 2 metres. The contours represent the following cumulative probability values: 1 = 0.01, 2 = 0.05, 3 = 0.50, 4 = 0.95 and 5 = 0.99
Scale - 1:200

4. Summary

Although archaeologists may have been somewhat disillusioned by their first contacts with quantitative spatial analysis, there is a pressing need for such techniques to be applied to archaeological problems. Quantitative methods alone can give a proper assessment of relationships within and between distributions of artefacts. To achieve success, the available methods must be carefully adapted to the archaeological problem. Section 2 illustrates how existing methods can be effective in producing descriptive summary statistics. In most cases, however, the archaeologist is interested in examining the detailed structure of his distributions; in those circumstances he has to turn to more specific, and usually parametric, methods, such as that illustrated in section 3. There is a wide scope for the investigation of new techniques of this type.

References

Donnelly, K.P. 1978	Simulations to determine the variance and edge effect of total nearest neighbour distance. In Hodder, I.R. (ed.), Simulation studies in Archaeology, 91-95, Cambridge University Press
Freund, J.E. 1979	Modern elementary statistics, Fifth edition. London: Prentice Hall International
Hodder, I.R. and Orton, C.R. 1976	Spatial analysis in archaeology. Cambridge University Press
Johnson, I. 1977	Local density analysis COMPUTER APPLICATIONS IN ARCHAEOLOGY 1977, University of Birmingham
Orton, C.R. 1980	Mathematics in archaeology, Cambridge University Press
Mood, A.M., Graybill, F.A. and Boes, D.C. 1979	Introduction to the theory of statistics, 3rd edition. New York; McGraw-Hill.