# COMPUTER ANALYSIS OF POST HOLE DISTRIBUTIONS

C.D. Litton and J. Restorick

Department of Mathematics, University of Nottingham

Nottingham NG7 2RD

## 1. Introduction

Our interest in the analysis of post hole distributions was aroused by members of the Trent Valley Archaeological Committee who wished to interpret the site plan of an Anglo-Saxon settlement at Catholme, Barton-under-Needwood, Burton-on-Trent, Staffordshire. Part of the site plan is illustrated in Fig.1. As can be seen, several rectangular "buildings" have been identified and are indicated by the joined post holes. However there are many post holes not joined for this purpose. In fact the whole site has well over three thousand post holes and there is a real problem in determining all the possible positions of rectangular buildings. Two problems posed by such data are:=

(i) With such a distribution of post holes, what is a sensible criterion for deciding whether several post holes that do lie in a particular rectangular pattern of interest do in fact represent a building?

(ii) Furthermore, since there is evidence that there were several phases of occupation of the site, the orientation of buildings is of interest. Buildings constructed in different phases of occupation would be built in different directions. The question is whether these directions can be determined.

In recent years several papers have been published concerned with the detection of patterns within post hole distributions. For example, Fletcher and Lock (1980, 1981) describe a method suitable for searching for circular arrangements and discuss its application to an Iron age site. Coghill (1980) discussed the use of pattern recognition techniques to identify circular huts at a Middle Bronze age site. In all of these papers some attempt is made to discuss the statistical significance of the results.

The purpose of this paper is twofold. Firstly we wish to describe two possible methods for detecting rectangular patterns corresponding to rectangular buildings. Secondly we wish to suggest a simple statistical method of analysing the significance of our results. In one sense, the latter is probably far more important than the former as our statistical method is applicable to most search methodologies. In particular it would be readily applied to the methodologies of Fletcher and Lock and Coghill.

In the rest of the paper we shall be concerned with analysing a two dimensional map of post holes, where each post hole is represented by a point. To simplify the problem, we are deliberately ignoring much useful information such as the depth of each post hole, its diameter, whether it is in a shallow trench etc. If necessary, the methodology could easily incorporate this additional information. Also we are assuming that a building does not require special types of post holes in certain positions, e.g. door posts, or corner posts.

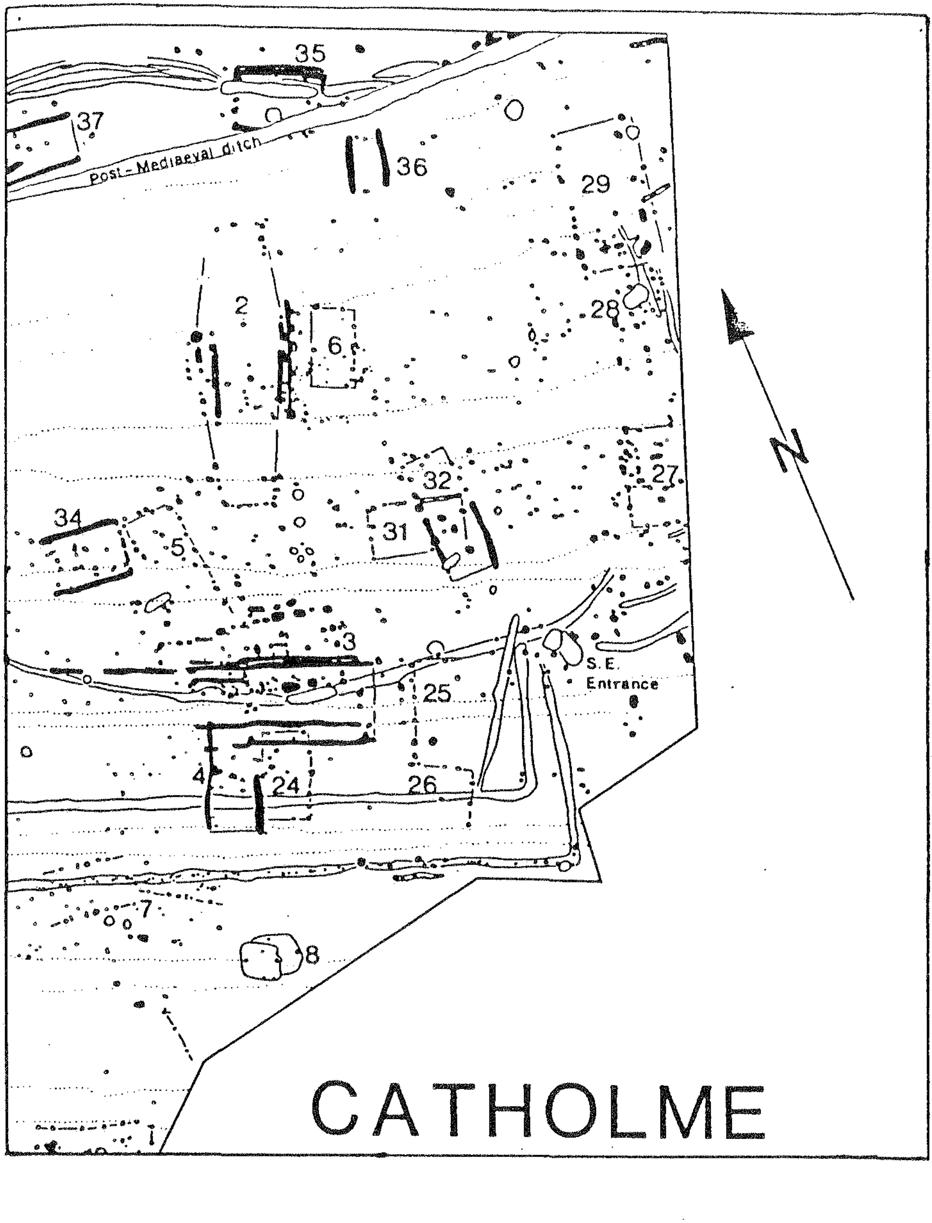


Figure 1   Site plan of part of the site at Catholme showing the distribution of post holes and possible buildings.

## 2. Method One

We suppose that we are interested in detecting a rectangular pattern of fixed dimensions. This would correspond to searching for buildings whose dimensions were known in advance, perhaps from other sites. To do this we propose to use a frame of external length L, external breadth B and width W (see Fig.2). Then for a particular site plan, we move the frame around in some systematic way to discover the positions in which the frame contains a large number of post holes. Such positions would suggest possible sites of buildings. However we need to decide what is a sensible number of post holes that are required to be within the frame and how they are to be distributed around the frame. Two possible criteria are:-

(i) a minimum of n post holes should lie within the frame;

(ii) a minimum of $n_1$ post holes should lie within each of the larger sides of the frame and a minimum of $n_2$ in each of the shorter sides.

Of course there are many other possibilities. The choice of a suitable criterion must be discussed with the archaeologist. It must incorporate if possible his ideas of what is an acceptable number and/or what is an acceptable distribution around the frame for the type of site under analysis. It must also take account of the density of post holes in that if the density is high one would require n (or $n_1$ and $n_2$ ) to be relatively high also.

To illustrate the use of this method we have chosen for simplicity to use criterion (i). Since one of the questions posed earlier was related to the direction of structures we will analyse the data using a directional search. Here we fix a direction and align the larger sides of the frame parallel to that direction. The frame of fixed size is then moved in a systematic manner over the whole site plan with the direction of its sides fixed. At each position the number of post holes within a frame is recorded. If this number is greater than or equal to n a position of a possible building is noted. This search procedure is carried out by a FORTRAN program run on the ICL 2977 at the Cripps Computing Centre, University of Nottingham. Possible positions of buildings are then indicated on the graphical output produced by the program. This search is then repeated for different directions.

This technique has been applied to a part of the Catholme site consisting of about six hundred post holes. The results are shown in Fig.3. for $n = 13, 14, 15, 16$. Notice that in directions from 0.00 to 0.40 radians and from 1.60 to 2.00 radians from the base direction there appear to be more potential buildings than in the directions 0.40 to 1.60. Within the latter range the average number of potential buildings is about 50 to 60 when a minimum of 13 post holes is used. The two peaks of the graphs roughly occur at a separation of $\pi/2$ radians, thus corresponding to the frame being rotated by about $90°$. This indicates possible buildings of the same size but built at right angles to each other.

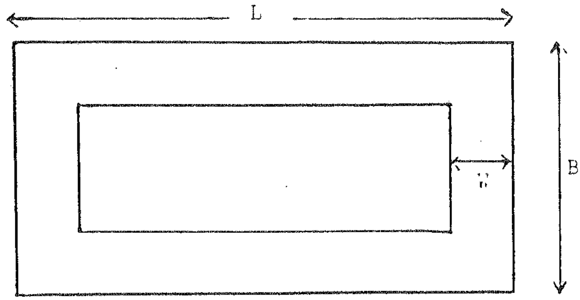


Figure 2   Frame of dimensions L, B and W.

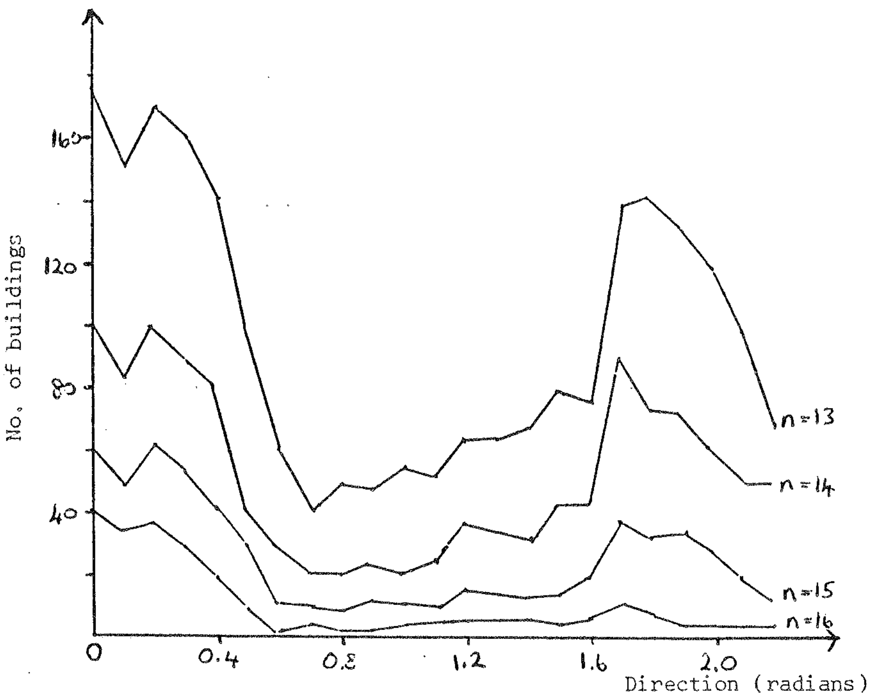


Figure 3   Number of potential rectangular buildings found in different directions using the criterion of a minimum of n post holes in the frame (n = 13,14,15 and 16).

## 3. Statistical Analysis of the Results

The problem now is to assess the significance of these results. The comparable method to that used by both Fletcher and Lock and Coghill would be to simulate the same number of points randomly distributed over the same region. Then the directional search is repeated and the results from the real data are compared with those from the simulated data. The rationale for doing so is that the simulated data is purely random and by determining the number of structures that can be "found" within random data, the number of "by chance" structures in the real data can be assessed. In other words, if the real data gives rise to far more buildings than the simulated data then it may be concluded that some of the buildings on the former are "real" and not due to chance alignment of post holes.

However in the Catholme case, and perhaps in others, this method would be incorrect and could lead to erroneous conclusions. In any settlement there is always some organisation of the buildings. For example, paths or trackways exist between buildings and very few post holes will be found there. Also for general topography of the land implies that building will be constructed in some positions and not in others. Simulating the same number of post holes over the same area destroys this overall organisation of the settlement.

For example, consider the arrangement of post holes from part of the Catholme site given in Fig.4. We observe that there are groups or clusters of post holes corresponding to possible buildings, with the areas between the groups containing very few post holes. Thus the density of post holes is by no means constant over the whole site. Now randomly simulated data will be more uniformly distributed over the whole area and thus removes not only any rectangular patterns but also removes the relative position of the groups or clusters. In other words the simulated data does not reflect the overall structure of the settlement and does not allow the density of post holes to vary from one part of the site to another.

Instead of this approach of simulating post holes we suggest that the position of each post hole should be perturbated slightly in some stochastic way. This slight perturbation should be large enough to break up any rectangular patterns in the data but small enough to preserve the overall distribution of groups of post holes. Therefore both the overall and local density of post holes will remain more or less the same.

We perturbate the data as follows. Let $(X_i, Y_i)$ be the coordinates of the ith post hole and let $(X_i', Y_i')$ be its coordinates after its perturbation. Then

$$X_i' = X_i + \varepsilon_{i1} \qquad i = 1,2,\ldots,m$$
$$Y_i' = Y_i + \varepsilon_{i2}$$

where m is the total number of post holes and $\varepsilon_{1i}$, $\varepsilon_{2i}$ are independent and identically distributed random variables. For our practical example we have chosen the $\varepsilon_{1i}$ and $\varepsilon_{2i}$ to have uniform distributions on the interval (-3W, 3W) where W is the width of the frame. The post hole

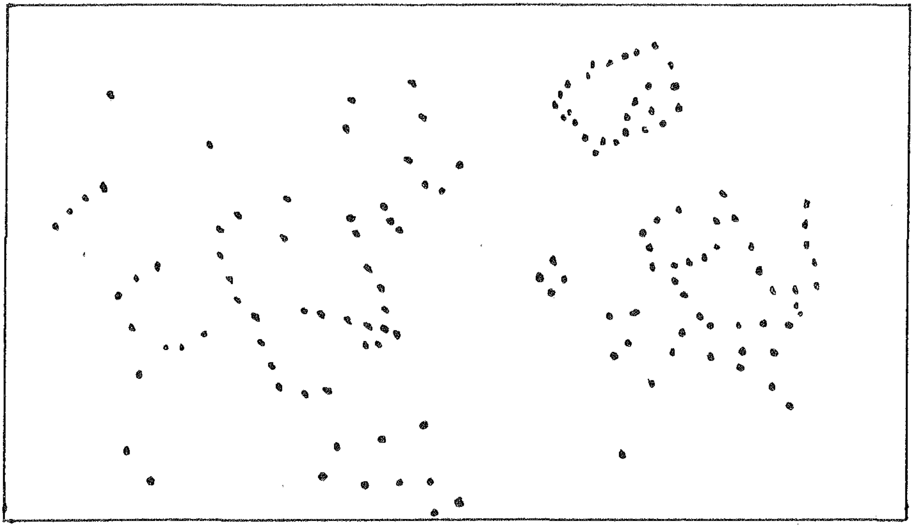

Figure 4 Part of the site plan illustrating the clustering of post holes.

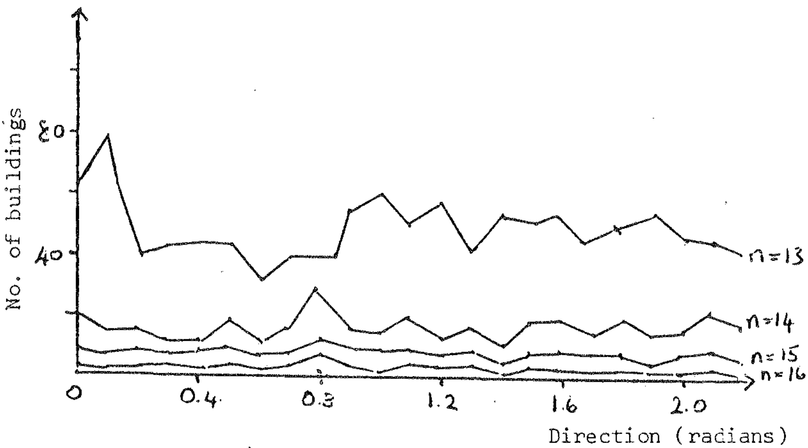


Figure 5 Number of potential rectangular buildings found in different directions after the post hole positions have been perturbated.

at $(X_i, Y_i)$ moves to $(X_i^{'} Y_i^{'})$ where $(X_i^{'}, Y_i^{'})$ is a randomly chosen point within the square of length 6W and centre at $(X_i, Y_i)$. Thus each point is perturbed slightly but overall the relative positions of clusters of points are preserved.

Having perturbed the data, the directional search with the frame is applied. The results for our previous considered set of post holes is presented in Fig.5. Comparing with Fig.3 it is immediately evident that the peaks in directions 0.00 to 0.40 and 1.60 to 2.00 have almost completely disappeared. What the perturbated data is telling us is that with this sort of data, arranged in these types of groups and with $n = 13$ as our criterion we would expect to get by chance about 50 to 60 potential buildings in any direction. This number is very similar to that found in the directions 0.40 to 1.60 in the original data, thus suggesting that these buildings are occurring "by chance". In the directions corresponding to the peaks, there is obviously some difference between the two sets of results, thus suggesting that there are some real buildings in these directions. Of course the simulation could be repeated several times to get some form of confidence interval for the numbers in each direction.

4. Method Two

The major drawback of method one is that the dimensions of the frame or equivalently the building must be specified in advance. This could be the case if one site has produced buildings of a particular size and one wished to determine whether a second site contained any buildings of the same size. However, usually we are interested in determining any rectangular structures. One possible way of doing so is to search using method one but varying the dimensions of the frame. However this would be extremely expensive in computer time.

A simpler and quicker method is to search for alignments of post holes at right angles to each other. This will then identify the corners of possible buildings of different sizes. To do this we use a cross with its arms (numbered 1, 2, 3 and 4) set at right angles (Fig.6.). Using this cross we can detect (i) four right angles, using arms 1&2, 2&3, 3&4 and 4&1, respectively, and (ii) two straight lines using arms 1&3, and 2 4 respectively. Again for a given size of cross we have to determine a crition in order to say that we have detected a corner or a straight line alignment. One that has proved quite successful in initial trials is as follows.

A right angle (or straightline alignment) has been detected in arms i and j of the cross ($i = 1,2,3,4$; $j = 1,2,3,4$; $i \neq j$) if the number of post holes is (i) in arm $i > n_0$, (ii) in arm $j > n_0$ and (iii) in arms i and $j > n_1$.

This method has been applied to part of the Catholme site and appears to detect buildings of different sizes quite successfully. However it is somewhat limited in that it frequently identifies say three corners and two sides of a structure but fails to identify the whole building. This is usually because the number of post holes at the missing corner is too small. To overcome this, the method needs to be more flexible and one possible way of doing so it to reduce the values of $n_0$ and $n_1$ when searching

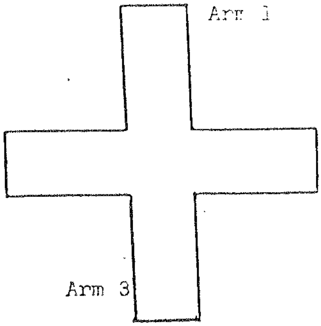


Figure 6 The cross.

for the fourth corner. Alternatively, since three corners more or less fix the dimensions of the building, the frame idea from method one could be used. At present work is in progress on this more dynamic method of searching and hopefully the results will be reported in some later publication.

## Acknowledgments

We thank the members of the Trent Valley Archaeological Committee for bringing this problem to our attention. We are extremely grateful to Dr R.R. Laxton of the Mathematics Department, University of Nottingham for his many useful suggestions.

## References

Cogbill, S. 1980 — Computer post-hole analysis with reference to the Bronze Age, COMPUTER APPLICATIONS IN ARCHAEOLOGY, 1980, pp 35-38, University of Birmingham ISBN 0-7044-0412-5.

Fletcher, M. Lock, G. 1980 — Computer assisted pattern perception within post hole distributions, COMPUTER APPLICATIONS IN ARCHAEOLOGY, 1980 pp. 39-48, University of Birmingham ISBN 0-7044-0412-5.

Fletcher, M. Lock, G. 1981. — Computerised pattern perception with post hole distributions, SCIENCE AND ARCHAEOLOGY, 23, pp.15-20.