PROCEDURES FOR THE SURFACE SAMPLING OF ARCHAEOLOGICAL SITES

J.G.B. Haigh

Schools of Mathematical Sciences and Archaeological Sciences,
University of Bradford, West Yorkshire BD7 1DP

The aims of surface sampling

During recent years archaeologists have tended to move away from the study of the temporal development of single sites and towards the investigation of the spatial relationships between contemporary sites spread over a wider region. For the newer type of investigation to be successful it is necessary to have some means of locating all the sites within the study region and then to be able to extract some useful information about them. When surveys in western Europe consistently reveal densities of around 5 to 10 sites per square kilometer, it is obviously impractical to investigate every site by the traditional means of extensive excavation. Archaeologists have therefore turned to other methods for the preliminary investigation of site-geophysical survey, aerial photography and the collection of surface sherds. The present paper is concerned with the last technique, principally in application to Mediterranean sites of the classical periods, and with the use of sampling to maximise the useful information obtained for given effort.

The literature now contains many papers on archaeological sampling techniques, both over a region and within a site (Mueller 1975, Plog 1976, Plog et al. 1978, Cherry et al. 1978, Bellhouse 1980). These articles tend to fall into two varieties: the first group apply some predetermined sampling technique to a group of sites and then publish the results, often with some rather arbitrary estimate of the likely errors; the second group consider some site from which all the relevant information has previously been obtained, apply a variety of different sampling schemes, and then come to some conclusion about the "best" scheme.

There is some lack of discussion about how to choose in advance sampling schemes likely to be suitable for a given site, and of which procedures should be sufficiently robust and flexible to give good results at each of a range of sites. A major source of difficulty may be that sampling theory tends to assume a reasonable homogeneity among the population to be studied, whereas archaeological sites are often inhomogeneous in nature. Indeed one could argue that it is heterogeneity which is of real interest to archaeologists.

Some previous results

When considering the sampling of classical Mediterranean sites, many workers expect to gather a reasonable proportion of diagnostic sherds. In fact most of what they gather is likely to be tile - a fairly intractable material from the archaeological point of view, since manufacturing methods have varied little over the ages and thus it provides little evidence of date. It is almost unheard of for diagnostic materials, usually potsherds, to form more than 10% of the sherds recovered; it is not uncommon for the proportion of diagnostic sherds to be less than 1% of the total. Under these circumstances it is very difficult to make satisfactory estimates of the ratios of numbers of different types; statistical techniques tend to be confined to estimating the total number of sherds on site.

The statistical analysis of the results of the 1979 Boeotian survey (Haigh 1980) was largely concerned with estimating total numbers of sherds. Examination of the results shows something of the problems involved when a single sampling technique is to be applied to all the sites in a region. The sites ranged in size from small scatters containing only a few hundred sherds at a density of about one per square meter, to the ancient city of Thisbi which contained around 4,500,000 sherds in 36 hectares. There were two sites which each covered about 2 hectares, one of which contained an estimated 150,000 sherds whereas the other had only about 10,000 sherds. Clearly it is necessary to turn to some form of sampling technique if information about the larger sites is to be obtained with reasonably small effort.

As well as varying in size and intensity, the sites also show enormous variation in visibility and surface condition. Some are on recently ploughed land, which is ideal from the point of view of sample collection; others may be under crops, on rough pasture, in dense undergrowth or in completely impenetrable scrub. Any predetermined sampling scheme should be able to cope with all these different conditions and still provide statistically reliable results. If any adjustments are necessary to suit particular conditions, then it must be within the competence of the field teams to make them.

The model site

	0·37 5	0·10 1	0·02 0						0·74 4	1·36 9	
	0·75 5	0·78 0	0·01 1					0·00 0	9·11 12	5·28 9	1·26 9
4·08 19	2·39 23	0·23 1	0·09 1	0·00 0	1·23 0	0·00 0	0·00 0	0·00 0	6·17 12	2·75 10	1·10 2
3·98 14	1·36 14	0·29 4	0·41 2	0·59 4	0·00 1	0·01 0	0·55 3	3·66 6	7·37 6	0·84 2	
	0·46 8	2·72 7	2·59 10	0·54 1	0·09 4	2·28 0	0·77 3	23·06 26	16·00 33	0·91 1	
	0·33 2	1·90 13	5·63 63	1·78 15	0·72 6	0·39 1	5·29 20	18·43 69	2·27 7	0·20 0	
	0·23 3	1·88 9	4·21 36	2·94 35	1·72 12	0·60 12					

Upper value is weight of tile (Kg) in each 10m square.
Lower value is number of potsherds in each 10m square.

Figure 1: Sherd distribution on the Cors site.

The results from the Boeotian survey were quoted with 95% confidence limits, which were calculated from standard formulae based on the internal structure of the samples taken. Since no attempt had been made to record a site completely, there was no direct means of checking the reliability of the confidence limits. The opportunity to test the sampling techniques, and the associated confidence limits, came from a classical site at Cors, in

Catalonia, northern Spain. The Ampurdan Survey undertook a total collection of the surface sherds, recording the weight of tile and the number of potsherds found in each 10 meter square within the site. A summary of the results is shown in Fig. 1. It is apparent that the surface scatter has considerable internal structure, and is by no means homogeneous.

In their original form the results are not suitable for testing the sampling methods, since much more detailed spatial information is required. A simulation of the site was produced in which the information was taken to be expressed in terms of 2 meter squares. Since it is impossible to estimate how weights of tile might be distributed, without additional statistical information, it was assumed that all the tile was divided into sherds, each of 25 grams in weight. The tile sherds within each 10 meter square could then be distributed across the twenty-five 2 meter squares, according to a probability distribution chosen to match the apparent structure of the scatter. The weight of 25 grams per sherd seemed reasonable on empirical grounds, and meant that the Cors site was estimated to contain about the same number of sherds as a middle-sized Boeotian site.

The number of potsherds (575 in all) was less than 10% of the number of tile sherds; since one might expect at most half of the potsherds to be diagnostic (they were not in fact subjected to detailed examination), the proportion of diagnostic sherds on site may be regarded as typical. When such small numbers of diagnostic sherds are found, it is apparent that any sample covering only a small proportion of the site will produce only a handful of sherds attributable to any given period. Thus there is very little chance of obtaining indications of the dates when the site was occupied and of its relative usage at different periods, since ratio estimates based upon such small numbers would be unreliable statistically. For the time-being the simulation will be confined to the sampling of tile sherds, but some suggestions on how to make use of the information from diagnostic sherds will be put forward in the last section.

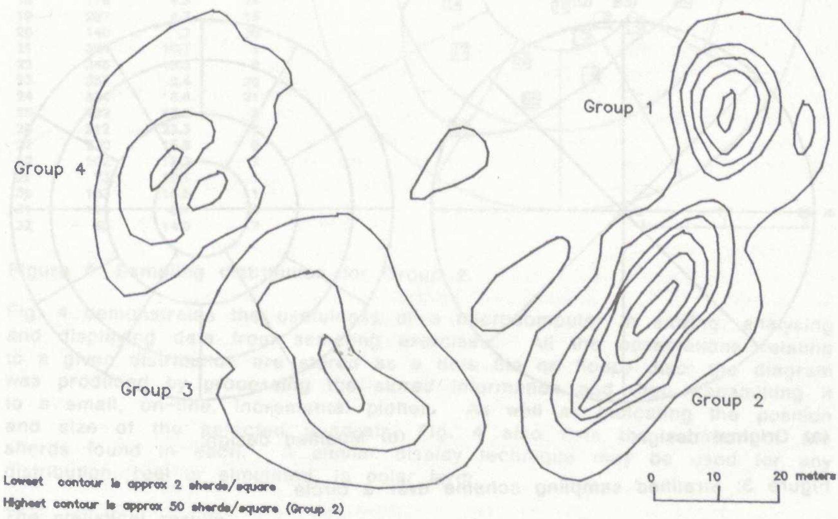


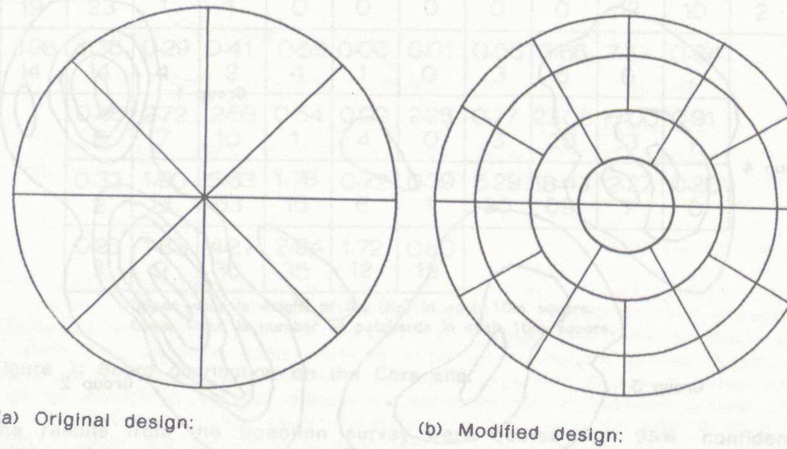
Figure 2: Distribution of sherd densities in the Cors simulation.

Being far from homogeneous, the simulated site effectively separates into at least four separate scatters, as is indicated in the contour diagram in Fig. 2. It is apparent that any sampling scheme which used quadrats scattered over the entire site would include a large number of empty quadrats (null-quadrats). Close examination of the Boeotian results suggests that the sites there contain similar inhomogeneities, although the Boeotian samples seem to include rather fewer null-quadrats. The reasons for this last difference have not yet been ascertained; it is possible that they may lie in the collection technique, since gathering surface sherds is by no means a simple operation and needs to be carefully monitored.

In the simulation it was decided to sample each of the four major groups as though it was a separate scatter. This was useful for two reasons:
(i) because the polar sampling procedures, as devised for the Boeotian survey, were more suitable for small scatters than for an extended site
(ii) because it gave four separate opportunities to test the statistical methods involved.

The polar sampling technique

The sampling method employed on each group was devised in response to a request from the directors of the Boeotian survey for procedures which do not require the establishment of a rectangular grid. The positions of the quadrats are specified as polar coordinates distributed over a circular area, in contrast to the more traditional methods where they are specified as cartesian coordinates distributed over a rectangular area. The field teams are given computer print-outs listing the coordinates of thirty-two randomly located points; the use of polar coordinates enables the quadrats to be located by the use of compass and tape-measure. The lists are arranged so that it is possible to take fewer than thirty-two quadrats and still obtain a satisfactory random distribution over the site, but thirty-two quadrats provide a good working minimum for statistical purposes.



(a) Original design; (b) Modified design;

Figure 3: Stratified sampling scheme over a circle.

Fig. 3a shows the original form of the scheme, in which the site was divided into eight octants and the stratified random sample contained four points from each octant. When it was realised that this scheme did not provide a sufficiently uniform distribution of quadrats, a more complicated scheme was introduced where the circle was divided into thirty-two separate segments, as in Fig. 3b, and a single point was chosen at random from each segment.

In order to apply the distribution over the circle to an actual scatter, it is necessary first to define the centre and the perimeter of the scatter. The radial coordinate is interpreted, not as a true radius, but as a fraction of the distance from the centre to the perimeter. This is illustrated in Fig. 4, which shows a typical sampling distribution for group 2; the site perimeter corresponds roughly with the lowest contour surrounding group 2 in Fig. 2. The locations of the quadrats have been obtained by taking a distribution over the outer circle and then compressing it along the radii into the region within the perimeter of the scatter. This compression has resulted in an uneven distribution of quadrats, the sampling intensity being highest where the radius to the perimeter is small and lowest where the radius is large.

SAMPLE	ANGLE	RADIUS	NUMBER
1	273	5.6	19
2	17	9.5	52
3	32	7.6	48
4	337	2.5	59
5	126	11.5	1
6	164	13.6	0
7	162	8.8	4
8	140	6.0	4
9	267	14.8	1
10	297	10.8	1
11	272	10.0	4
12	253	13.6	6
13	8	15.2	16
14	33	20.0	13
15	49	15.0	22
16	24	13.7	32
17	141	4.4	17
18	178	4.3	24
19	257	6.7	15
20	140	.7	30
21	324	10.7	1
22	345	13.3	9
23	357	9.4	26
24	324	6.8	21
25	192	16.6	2
26	212	23.3	1
27	230	15.8	5
28	202	12.2	7
29	73	19.4	1
30	103	15.5	1
31	114	8.4	4
32	65	14.0	7

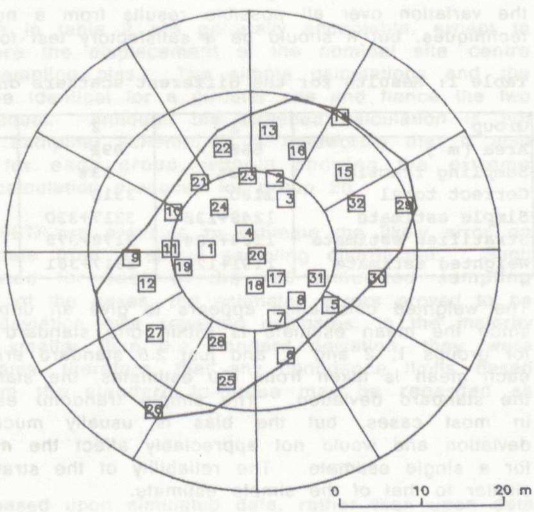


Figure 4: Sampling distribution for Group 2.

Fig. 4 demonstrates the usefulness of a microcomputer in storing, analysing and displaying data from sampling exercises. All the observations relating to a given distribution are stored as a data file on floppy disc; the diagram was produced by processing the stored information and then transmitting it to a small, on-line, incremental plotter. As well as indicating the position and size of the selected quadrats, Fig. 4 also lists the number of tile sherds found in each. A similar display technique may be used for any distribution, real or simulated, in polar form.

The statistical results

The results obtained from the selected quadrats may be used to estimate the total number of sherds within the scatter. Three different methods have been used to make this estimate. The first method is on the basis of simple random sampling, and is clearly subject to error because of the sampling bias arising from the varying radius to the perimeter; this bias was illustrated in Fig. 4. The second is on the basis of stratified random sampling with the octant design in Fig. 3a; this method was employed in the analysis of the Boeotian results and is included here for comparison, although it is not strictly applicable to the modified design of Fig. 3b. The third method is a weighted calculation which is designed to account for the bias arising from the variation in radius.

In order to test the reliability of the three estimates, each of the four sherd groups on the simulated Cors site was sampled one hundred times, the three estimates of the number of sherds being calculated each time. The results are summarised in Table 1, where the mean and standard deviation over each set of 100 estimates are shown. The means are compared with the correct, known number of sherds, which is itself subject to some small uncertainty, since it is difficult to determine precisely how the computer treats quadrats which are intersected by the site perimeter. The program followed here is less rigorous than that of Bellhouse (1980), who considered the variation over all possible results from a number of standard sampling techniques, but it should be a satisfactory test for the present methods.

Table 1: Results for the different scatters on the simulated site.

Group	1	2	3	4
Area (m)	668	991	1173	861
Sampling fraction	19%	13%	11%	15%
Correct total	1180	3310	1100	650
Simple estimate	1248+132	3217+330	1085+117	635+72
Stratified estimate	1184+124	3178+375	1066+121	620+80
Weighted estimate	1191+124	3317+381	1101+137	630+81

The weighted calculation appears to give an unbiased estimate of the total, since the mean estimate is within one standard error of the correct value for groups 1, 2 and 3, and just 2.5 standard errors out for group 4. Since each mean is taken from 100 estimates, the standard error is one tenth of the standard deviation. The simple (random) estimate shows a slight bias in most cases, but the bias is usually much less than the standard deviation and would not appreciably affect the mean square error expected for a single estimate. The reliability of the stratified estimate seems to be similar to that of the simple estimate.

The correct definition of the centre and perimeter of the scatter is of crucial importance when obtaining these results. Table 2 shows the results for group 2 in comparison with results for other definitions of the same group. Group 2a has its perimeter enlarged to include the small group to the upper left of group 2 in Fig. 2; group 2b is identical to group 2 except that the nominal centre has been moved 4 meters to the left; group 2c has a circular perimeter based on the largest radius for group 2.

Table 2 shows again that the weighted calculation appears to give an unbiassed estimate of the total number of sherds, since in each case it is within two standard errors of the correct value. On the other hand the standard deviation of the estimate increases markedly with the defined area of the scatter; indeed, the result for group 2b shows that there is a

Table 2: Results for the different definitions of Group 2.

Group	2	2a	2b	2c
Area (m)	991	1352	991	1810
Sampling fraction	13%	9%	13%	7%
Correct total	3310	3470	3310	3500
Simple estimate	3217+330	3327+417	2744+288	3573+542
Stratified estimate	3178+375	3359+464	3035+431	3373+474
Weighted estimate	3317+381	3503+488	3266+434	3573+542

significant increase in standard deviation simply because of a poor definition of the group centre. The dependence of the likely error of the estimate on the proper delimitation of the area to be sampled is one of the major problems of this type of sampling procedure. It arises in part from the fact that too generous a definition of the delimited area will incorporate a number of null-quadrats and hence lead to an increase in the variance over the sample. The problem is by no means confined to the polar sampling schemes discussed here; in the more usual rectangular schemes it is necessary to fit a box round the delimited area, thereby increasing the number of null-quadrats still further.

The simple estimates shown in table 2 are generally acceptable, except in the case of group 2b where the displacement of the nominal site centre has produced a massive sampling bias. The simple calculations and the weighted calculations become identical for a circular site and hence the two results for group 2c are equal. Although the stratified calculation is not optimised for the modified sampling scheme, it is noteworthy that it also gives satisfactory results for each group, without showing the extreme variation which the simple calculation produces for group 2b.

Standard formulae (Haigh 1981) are available to estimate the likely error on a simple or stratified estimate from a single sampling distribution. Such likely errors were calculated for each of the 700 simulated sampling distributions. In about 95% of the cases, the estimated errors proved to be greater than the standard deviation over repeated estimates; in the majority of cases when they were smaller than the standard deviation, they were only marginally so. It seems, therefore, that any confidence limits based upon errors estimated from the standard formulae may be regarded as generous ones.

Conclusions and prospect

All the above results are based upon simulated data, rather than upon data taken directly from a real site; it is necessary to make two reservations about their interpretation. The first reservation is that the site was divided into discrete 2m squares, and a selected quadrat is taken to be the nearest square rather than a quadrat centred upon the random sampling point. Since each of the designated groups covers a sufficiently large number of squares for the results to encompass a full range of variability, this reservation should not have a significant effect. The second, more serious reservation, concerns the transformation of the data from 10 meter squares to 2 meter squares; this can only be made on the basis of certain assumptions about the structure of the site. The fact that the structure has not been entirely accounted for is indicated by one or two discontinuities along the edges of the original 10 meter squares. Nevertheless the simulation should give a reasonable picture of the sort of features that might be expected in this type of site. The following conclusions should be applicable to other similar sites:

(i) The calculation based on simple random sampling gives a biased estimate of the number of sherds in a scatter. This was foreseen from the known sampling bias of the polar distribution but in many cases the simple random calculation did give a reasonable estimate in terms of the mean square error.

(ii) The weighted calculation appears to give an unbiased estimate of sherd numbers, but with a variance which is often greater than that for the simple estimate and sometimes greater than the mean square error of the simple estimate. The weighted estimate is more reliable, however, in those cases where the simple estimate may go far astray. The stratified estimate, based on the octant design, is partially successful in accounting for the bias and should be reliable, particularly for sampling distributions which incorporate the same design.

(iii) Confidence limits based upon the standard estimates of likely errors may be regarded as generous. Since the stratified random sampling, as applied to the Boeotian results (Haigh 1981), should account for most of the sampling bias, the estimates and confidence limits derived from it should be reliable. Hence one may place some confidence in the hierarchy of scatters which seemed to be emerging from the Boeotian observations.

(iv) The accuracy of the results depends upon the proper definition of the boundary of the scatter. The underlying problem had already been noted by fieldworkers in both Boeotia and Spain; they found considerable difficulty in determining at what point the distribution of sherds in the scatter finally merged with the background distribution.

It is possible to contemplate a number of further improvements to the polar sampling scheme. When the quadrats are selected on a stratified random basis using the modified design in Fig. 3b, an uneven distribution of quadrats may occur; a better spread would result from a more systematic selection of quadrats within the segments. If a microcomputer were available to select the positions of the quadrats after the site perimeter had been defined, then it would be possible to devise a sampling distribution which both accounted for the radial bias and minimised the likely error. If any further reductions in error were required after these improvements, then it would be necessary to increase the sampling fraction, either by selecting more quadrats or else by taking larger quadrats.

On the other hand, knowing that the accuracy of the results depends on the proper definition of the site perimeter, it seems more expedient to devote any extra effort to defining the extent of the scatter, rather than to the details of the internal sampling scheme. One possible means to that end might be to use a rectangular systematic scheme in which 2 or 2.5m quadrats are selected at 5m intervals. The advantage of such a scheme over total surface collection is that one needs small quadrats to obtain accurate spatial information, while the extraction of the whole of a large site in terms of such small quadrats would be prohibitively time-consuming.

The results of such systematic sampling would be in a similar format to data from a geophysical survey. The Archaeological Sciences group in Bradford is already involved with methods for the analysis and display of geophysical data, possibly by means of a microcomputer on site; the same methods could be used to determine the positions of any sherd concentrations. A variety of methods should be available for displaying the spatial distribution of sherd densities, including varying levels of grey on a video monitor by means of a standard graphics package, and contour diagrams, in the style of Fig. 2, which might be produced on the incremental plotter.

A project currently being considered is to survey sites both through surface collection and by geophysical techniques. It would then be interesting to evaluate the degree of spatial correlation between the two sets of results. Once the basic structure of the site has been determined, it would be interesting to examine the spatial distribution of the diagnostic materials and to determine whether they provide any evidence of the date and usage of different sections of the site. If useful results are to be obtained in this way, it may be necessary to undertake a more intensive collection of diagnostic materials than that of the tile sherds.

To suggest the use of an alternative systematic scheme is not to deny the efficacy of the polar sampling scheme. Many circumstances may be envisaged where there would be definite advantages in using the polar scheme, especially in cases when the use of a rectangular grid was undesirable. However, given the problem of defining the edge of the scatter and with a team who are willing to use rectangular grids, there is a strong case for systematic sampling. The successful completion of the first simulation study indicates how to use the data from other sites to create more simulated distributions. It is hoped to test both the polar and the systematic sampling schemes on the basis of new simulations, before making new observations in the field.

It is necessary to take a pragmatic approach to fieldwalking and surface collection, in order to ascertain which procedures are most efficient for detecting sites and providing information about them. The observational evidence cannot yet be related directly to the original state of the site, since the condition of the surface is determined by many factors, including geology, climate, site history and current land use. To use theoretical terms such as 'target population' does not seem very helpful at this stage, since it presupposes that one understands the relationship between the evidence provided by the new techniques and that from traditional archaeological investigation.

A comparison was made earlier between a systematic sampling scheme and geophysical survey. What sampling schemes have in common with geophysical survey, and with aerial photography, is that they provide different facets for archaeology; the evidence which they provide is of a different quality from that which comes from excavation. It is only with care and experience that archaeologists will learn how to reconcile the evidence from all these related facets.

Acknowledgements

The author is grateful to the directors and staff of the Ampurdan survey for their cooperation in providing the data upon which the results of this paper are based; in particular, to Mr Martin Millett, who was responsible for the collection of sherds at the Cors site.

BELLHOUSE, D.R. 1980 Sampling studies in archaeology. Archaeometry 22, 123-132.

CHERRY, J.F., GAMBLE, C. & Shennan, S. (eds) 1978 Sampling in contemporary British archaeology. Brit. Arch. Reps. 50.

HAIGH, J.G.B. 1981 A scheme for regional survey and site sampling. Revue d'Archeometrie 5, 1-9.

MUELLER, J.W.(ed.) 1975 Sampling in Archaeology. Univ. Arizona Press.

PLOG, S. 1976 Relative efficiencies of sampling techniques for archaeological surveys. in Flannery, K.V. (ed) The early Meso-american village, 136-158. Academic Press, New York.