REPLICATED COLLECTION OF SITE SURFACES

A. J. Ammerman and M.W. Feldman

The replicated collection of the surface of a site provides one way of learning more about a site during survey work. Examples of the use of this technique at 2 sites in southern Italy are presented. The suggestion is made that the site surface in some cases can be viewed as operating as a biased multinomial sampling process. A statement of this hypothesis in formal terms is given.

The surface of a site is the one part of a site that the archaeologist has direct and also economical access to. Although improvements have been and are being made in the way survey work is conducted in the field, our ability to interpret material recovered from the surface of a site is still quite limited in most cases. Excavated material continues to provide the major source of evidence for archaeological research in most parts of the world. At the same time, parallel refinements in excavation methods imply, somewhat paradoxically, that only a few of the sites in a given region can be excavated and that very few are likely to be opened up on more than a limited scale. Moreover, if we are interested in developing efficient strategies of excavation, it turns out that a prerequisite is often prior knowledge derived from survey work. As more emphasis is placed on work of this kind, there is a need to ask basic questions about what site surfaces represent. What are the factors operating in the generation of scatters of material on the land surface? Which techniques and strategies of collection are most productive at different kinds of sites? What types of analyses are appropriate when it comes to looking for patterns in surface data? These are questions that tend to become increasingly important as more survey work is carried out each year and more of the archaeological record consists of sites known only in this form.

Various aspects of survey methodology have been the subject of active interest during the last 10 years (e.g., Redman and Watson 1970; Binford et al. 1970; Reid et al. 1975; Flannery 1976). One of the main approaches has been that of comparing patterns obtained from systematic surface collection at a site with those revealed during subsequent excavation. Here the surface is treated in some respects as a "predictor," which is to be tested against subsurface information. What has received little attention so far, however, is questions related to the expected behavior of the "predictor" and the nature of the correspondence between the 2 domains of information. One way of gaining some insight into this problem is through the replicated collection of site surfaces. By repeating at different times the collection of sets of grid squares at a site, it is possible to learn something about the kinds of variation that surface patterns are subject to. The use of this approach is best suited to situations where the land surface is stable in geomorphological terms and there is regular, light plowing of the land. By providing a means of checking on the consistency between sets of surface data, the comparison of collections can contribute to the interpretation of spatial patterns at sites. Replicated collection also focuses attention on a problem of wider interest: the operation of the surface as a sampling process (see Doran and Hodson 1975:42-43) with respect to material in the ground. How this sampling process operates—whether in selecting those pieces to appear on the surface from the set of pieces occurring in the ground, it behaves uniformly or differentially with regard to objects of different size, for instance—is a central question which has implications for many aspects of survey work. In this preliminary report, examples of replicated collection at 2 sites in the region of Calabria in southern Italy will be presented. A brief discussion of the kind of sampling process that is suggested for surface material will be given in a final section.

The first case concerns a Neolithic site located near the town of Amantea on the west coast of Calabria. The surface scatter, which consists for the most part of obsidian flakes and chips and

A. J. Ammerman, Department of Anthropology, SUNY—Binghamton, Binghamton, NY 13901
M. W. Feldman, Department of Biological Sciences, Stanford University, Stanford, CA 94305

also includes a small amount of worked chert and impressed-ware pottery, covers a roughly circular area about 120 m in diameter. The site is situated on a flat alluvial terrace along the coast and has a sandy clay soil. The grid system employed in covering the site was based on squares measuring 12 m on a side and tied in directly with the trees in a geometrically laid out olive grove, making it possible to relocate grid squares accurately and economically during replicated collections. Relatively large squares were used in both case studies in order to hold down the level of random variation occurring in surface material, which tends to increase as square size decreases. The first collection was carried out in October, 1974, on plowed fields that had been exposed to heavy autumn rains. The strategy used in selecting squares for collection, which will not be described in detail here, involved an initial, rapid inspection of the full set of grid squares over the site area with a qualitative attribution being made (without collection) of the amounts of obsidian visible on the surface of each square. A series of squares was then sampled to evaluate the reliability of these attributions, which turned out to be reasonably consistent in providing a rough idea of spatial patterns for the site area as a whole. Subsequent judgment sampling was oriented mainly toward the further definition of spatial patterns at the site. The collection procedure involved a team of 2 or 3 students working systematically back and forth over a grid square (first in one direction and then again at right angles) with the aim of collecting all of the lithic material on the surface. The second collection was made in October, 1975, under basically the same field conditions and using the same collection procedure. The site area was plowed twice during the time between the two collections. The second collection was undertaken without direct reference to previous results, and a new numbering system was used for the set of grid squares at the site in an effort to make the second collection as "blind" as possible with respect to earlier work. The relationship between the new square numbers and the original grid system was worked out only after the second collection had been completed. No counting of material was done until after the completion of work in the field. In Figure 1, the counts of obsidian for those squares collected on both occasions are given. While there is some variation that can be observed between the 2 collections, the patterns are in reasonably good overall agreement. This would clearly seem to be a promising result. A statistical comparison indicates that the 2 sets of counts, treated as paired observations following Wilcoxon's signed-ranks test, are not significantly different. (Wilcoxon's signed-ranks test is a distribution free or nonparametric test, which is described by Siegel [1956] and Sokal and Rohlf [1969] among others; here those tests with values

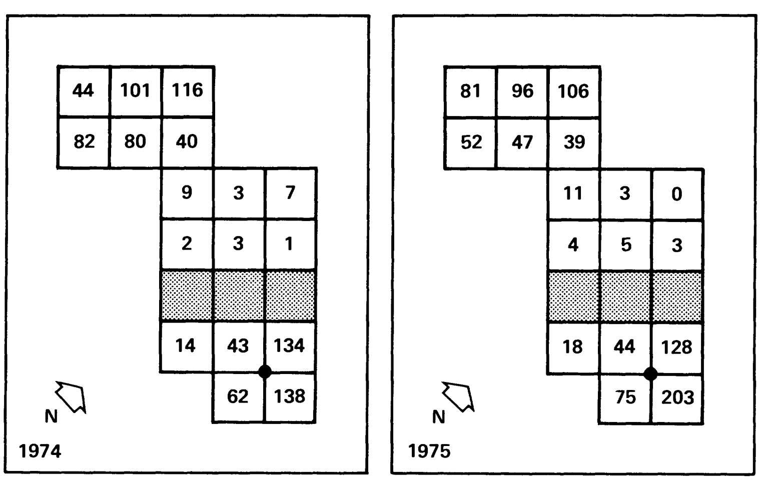

Figure 1. Replicated collection of grid squares at a Neolithic site located near Amantea in Calabria carried out, respectively, in October, 1974, and October, 1975. The numbers in the squares are counts of obsidian pieces. The three shaded squares have a farm path running through them and were plowed only in part.

at the 0.05 level or less are considered to be statistically significant.) Both collections support the interpretation of the site in spatial terms as consisting of 2 main concentrations of obsidian separated by an area of low density.

We planned to conduct a third collection at the site in October of 1976, but this plan had to be modified because only a small area on the south side of the site was found to have been plowed. The late onset of the rainy season in the fall of 1976 made it possible to carry out another "experiment" in surface collection, however. Experience in doing survey work over a 2-year period in Calabria had led to the impression that surface conditions appeared to play an important role in the amounts and even kinds of material that could be recovered during surface work. A series of 8 adjacent squares on the south side of the site were collected before the autumn rains arrived and then again about a week later after a moderate rain storm had occurred. It should also be added that the first collection in 1976 was made on a bright sunny day, while the second was carried out on an overcast day, which seems to offer better light conditions for seeing pieces of obsidian on the ground. The same collection procedure was again employed, and separate numbering systems for squares were used during the 2 collections. The counting of material was done only after the completion of work in the field. The results of the 2 collections are shown in Figure 2. Perhaps the most striking feature here is the amount of new material that had become "visible" and was recovered during the second collection. The obsidian counts are roughly equivalent to those obtained during the first collection. Another feature of interest is the fairly good correspondence between the 2 patterns: squares that contained many pieces the first time also produced a large number when the surface was covered a second time.

In statistical terms, the 2 sets of counts are again not significantly different from one another. Where they differ substantially, however, is in the sizes of the pieces recovered. This can be seen in Table 1, where the weights of obsidian pieces are used as a general measure of size. It is worth noting 2 main trends in the frequency distributions here: (1) among the larger pieces, most were recovered during the first collection, and few new pieces were obtained when the surface was systematically covered again; and (2) many more small pieces were recovered the second time. This result should come as no real surprise to any archaeologist who has walked over fields after it has rained. Essentially, it translates into quantitative terms an expectation about the visibility of material on a site surface that we have had all along.

What is less obvious is the implication that under varying surface conditions the chance of a given piece being recovered from the surface is *not* independent of its size. Larger pieces will have a greater chance of being found even when conditions are less than ideal. The recovery of smaller pieces is much more likely to vary with the quality of collection conditions. This raises certain problems when it comes to the analysis of surface material, especially if we plan to look at surface data in composition form. We would get somewhat different pictures of obsidian density and weight distributions (see Fig. 2 C and the last columns in Table 1) from the same site surface if we were to take the collection made before the rain (A) and compare it with that obtained after

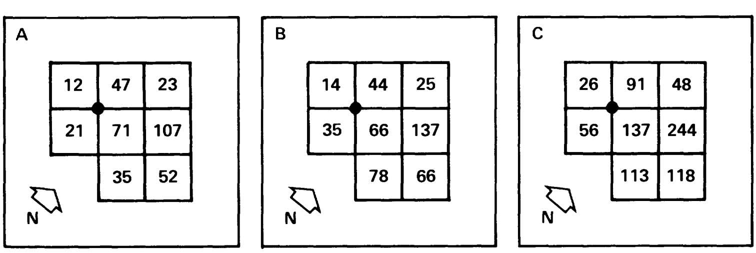

Figure 2. Replicated collection of grid squares carried out in October, 1976, at the same site shown in Figure 1. The numbers in the squares represent counts of obsidian. The squares were collected before it had rained (A) and then after a moderate rain (B). On the right (C), the combined counts for the two collections are given. The black dot corresponds in its position to the dot shown in Figure 1.

Table 1. Frequency Distributions of the Weights of Pieces of Obsidian Collected from 4 of the Grid Squares Shown in Figure 2.

Weight (g)	Grid square 01		02		03		04		Total		
	A	B	A	B	A	B	A	B	A	B	A + B
0.0–0.1	10	18	5	4	6	7	3	4	24	33	57
0.2–0.3	10	16	6	11	6	15	3	4	25	46	71
0.4–0.5	8	7	5	4	1	8	0	3	14	22	36
0.6–0.7	10	6	0	4	7	3	2	2	19	15	34
0.8–1.1	7	6	1	3	3	3	1	1	12	13	25
1.2–1.5	11	6	1	6	4	2	1	0	17	14	31
1.6–1.9	5	2	1	2	1	1	0	0	7	5	12
2.0–2.9	3	3	1	0	3	2	2	0	9	5	14
3.0–3.9	4	2	0	0	3	2	0	0	7	4	11
4.0–4.9	2	0	0	0	3	0	0	0	5	0	5
5.0–6.9	0	0	1	1	5	0	0	0	6	1	7
≥7.0	1	0	0	0	5	1	0	0	6	1	7
Total	71	66	21	35	47	44	12	14	151	159	310

Note: A represents the first collection and B is the second collection, which was carried out after it had rained. It is worth noting the J-shaped character of the distributions here.

the rain (A and B combined, making the assumption that the pieces found before the rain, if left on the surface and not collected, would also be found after the rain). The main differences in the weight frequency distributions would be a much larger number of small pieces and a much lower relative proportion of large pieces in the latter case. If comparisons of any refinement are to be made between sites on the basis of surface material, it is important that collection conditions be comparable or that allowances be made for such differences. It is worth adding that comparability can be a potential problem even at a given site if conditions change appreciably during the course of collection or if different parts of the site are collected at different times.

A case where much more variability between collections can be observed and where the question of comparable conditions may be involved is provided by a second Neolithic site, which is located near the village of Acconia on the west coast of Calabria. The surface scatter here consists mainly of worked obsidian and impressed-ware sherds and covers an area measuring about 350 m in length (E–W) and 60 m in width. The site is situated near the edge of a reasonably flat area of stabilized sand dunes having a well-developed soil formation. Three collections—in April, 1975, October, 1975, and April, 1976—each separated by an interval of about 6 months (during which time the soil was plowed at least once) were made at the site. The obsidian counts for these collections are given in Figure 3. Variation between the collections is much more marked: there is little correspondence, for example, between those squares that are most abundant in 1 collection and those in other collections. If a pairwise comparison is made between the sets of obsidian counts, there is a statistically significant difference between the first and second collections (A and B) and between the second and third collections (B and C). The 1 case where the 2 sets of counts are not found to be significantly different concerns the first and third collections (A and C), which incidentally were made at the same time of year. But even in this case, the level of variation is still quite high. If we look at any one of the first 3 rows in Figure 3 by itself, it is tempting to recognize patterns in spatial terms. But a comparison with collections made at other times cautions us against this. Of the spatial patterns available to us here, probably the most reliable is the one offered by the fourth row (D) or the 3 collections taken together. This represents another way in which replicated collections can be of use.

There has been a tendency to regard surface material as representing a rather simple or straightforward transformation of the material occurring in the ground or, more specifically in the examples presented above, in the plow zone. As we have seen, a certain amount of variation

A	16	6	9	8	15	13	7	5	10	14	16	11	13	15	6
B	2	5	12	8	4	5	9	5	10	6	7	4	2	4	4
C	2	9	22	5	13	11	10	13	20	18	18	24	6	8	9
D	20	20	43	21	32	29	26	23	40	38	41	39	21	27	19

Figure 3. Replicated collection of a series of grid squares at a Neolithic site located near Acconia in Calabria. The numbers in the squares are obsidian counts. The three collections were made respectively in: (A) April, 1975, (B) October, 1975, and (C) April, 1976. The combined counts for each square from the three collections are given in the last row (D). The rows run in orientation from west to east.

can be expected, since appearance on the surface represents a sampling process. A further complication needs to be considered in addition. There are reasons for suspecting that at least in certain cases, the sampling process that we are dealing with is not strictly a random one: the chance of a piece in the ground being recovered from the surface may not be independent of its size. A statistical comparison (based on a test of independence using 2-way contingency tables) of the frequency distributions shown in Table 1, for example, does not support the assumption of size independence. When the different pairs of frequency distributions are examined in this way, the null hypothesis of independence is rejected for grid square 03 and the combined counts for the 4 squares at the 0.025 level (with *chi*-square values of 22.06 and 24.49 with 11 degrees of freedom). If the weight classes are divided into 2 main groups along the lines of a threshold model (e.g., with the first 3, 0.0–0.5 g, in 1 and the remainder in the other), the 2-by-2 contingency table for the 4 squares taken together from the respective collections yields a *chi*-square value of 14.77, which is significant at the 0.001 level.

What is suggested here is a biased multinomial sampling process, where pieces belonging to different size classes have different chances of making their appearance in surface collections (for a general discussion of multinomial distributions, see Johnson and Kotz 1969). It is of some importance to determine which of 2 hypotheses—that the surface is operating as a random sampling process or as a biased multinomial one—is correct over a range of different site contexts, keeping in mind the possibility that the answer may not be the same in all situations.

As a basis for developing tests, it is worth briefly describing the 2 hypotheses in formal terms. Let us start by considering that in the plow zone and the surface there are k size classes (say of obsidian pieces on the basis of weight or some other measure), with $N_1 \ldots N_k$ being, respectively, the number of pieces present in each class. The total number of pieces (i.e., in the plow zone and surface) is then given by: $N = \sum_{1}^{k} N_i$. The probability of finding a piece in the surface belonging to class i (any 1 of the classes) would be $p_i = N_i/N$, under the assumption that the surface is a random sample of the material in the plow zone. The probability distribution of obtaining a surface sample which contains $n_1, n_2, \ldots n_l$ with $\Sigma n_l = n$ (i.e., with n being the total number of pieces observed in the surface) is given by:

$$\left(\frac{n!}{n_1!\ n_2!\ \ldots .\ n_l!}\right) p_1^{\ n_1} \ldots p_l^{\ n_l} \qquad l<k\,, \tag{1}$$

which represents a formulation of the hypothesis that all pieces have an equal chance of occurring on the surface regardless of their size.

If the probability of finding a piece on the surface depends on size, then the expected distribution might change as follows. We will consider a simple threshold model where if the size of a piece is greater than or equal to size class T, the threshold, the chance of a piece being found goes up. Conversely, if the size is less than T, then the chance goes down. If we start by listing the largest size class first, then the respective probabilities $p_1, p_2, \ldots p_T, p_{T+1} \ldots p_k$ under the assumption of random sampling would become modified to

$$\frac{p_1(1+s)}{P}, \frac{p_2(1+s)}{P}, \ldots \frac{p_T(1+s)}{P}, \frac{p_{T+1}}{P}, \ldots \frac{p_k}{P}, \qquad (2)$$

where $P = (1+s)[p_1 + \ldots + p_T] + [p_{T+1} + \ldots + p_k]$ and s is a coefficient representing the increased visibility or recoverability of pieces in the larger size classes. (The coefficient used here is analogous to a selection coefficient in population genetics; the model could be extended by attaching a different selection coefficient to each size category if it were found that more than a simple threshold phenomenon is involved.) This set of modified probabilities would replace those used in (1) to obtain the expected distribution among surface samples under the hypothesis of multinomial sampling. A model of this kind could be used to explain the results observed in Figure 2 and Table 1. The specific alternative presented here belongs to a family of multinomial models; an evaluation of which model or type of model is most appropriate can be attempted when more evidence is available.

While it would seem best to break out of the circular practice sometimes encountered in archaeology of basing the test of a model on the same data set which suggested the model in the first place, it may be useful to make some comments with regard to the testing of the hypothesis. Perhaps the most efficient approach would be one that takes an "experimental" form, as explained below, where the number of pieces belonging to different size classes and circulating in the plow zone would be fully known. Excavation offers one means of establishing the actual size distribution of the pieces occurring in the ground. One limitation possibly arising here is that it may take some work to define the plow zone or the depth of deposit over which pieces have a reasonable chance of making their appearance on the surface. There is the additional drawback that excavation effectively terminates the possibility of further study of surface-to-subsurface relationships for the areas involved. An alternative would be to conduct what amounts to a "marked release" experiment where a given number of marked pieces (such as resistant plastic gaming pieces) of different sizes are placed (in a random fashion) in each of a series of grid squares. After several rounds of plowing have been carried out, the surface can be collected and the ratio of "captured" pieces to those initially "released" for a given size class can be calculated for each square. The mean values of the ratios for the different size classes can also be computed for the set of squares and these values compared to see if the ratios are significantly different. If a further test is required, this procedure can be repeated, when the field is subsequently plowed again.

The discussion in this report is confined to the level of the archaeological context (Schiffer 1976) or the mechanics of the relationship between material in the plow zone and that visible on the site surface. In practical terms, replicated collection provides a means of checking on the consistency of the patterns obtained from survey work. It should be emphasized that there is potentially a wide range of replication designs that can be employed at a site. For example, different artifact classes such as pottery and obsidian can be collected either singly or in combination during a series of visits to a site. There is also the possibility of adopting a finer spatial "grain" during 1 of the collections at a site: the squares can be subdivided into 4 quadrants, for example, during a second collection, with the comparison of replicates being done at the original grid size (i.e., the 4 quadrants taken together in the latter case). In the choice of a design, much will obviously depend on the specific questions of interest at a site. Replicated collection can also make a contribution on a broader methodological level. The variability observed in replicated surface collections and the suggestion that the surface may operate as a biased multinomial sampling process have wide implications for the kinds of analyses and inferences that can be made with surface material. Before definite conclusions can be drawn, however, there is a need for the technique to be applied to a wider range of sites and contexts and for comprehensive testing along both "experimental" and excavation lines. Among the factors (other than those involved in the initial deposition of material at a site) that may influence the behavior of the surface as a sampling process and that require further examination are collection conditions, soils, square size, and the density of material in the ground.

Acknowledgments. The research reported in this paper was supported in part by grants from the N.S.F. (BNS 76-15095 and DEB 77-05742) and the N.I.H. (10452-13). We would like to thank the Superintendent of Antiquities in Calabria and the University of Calabria for their collaboration with regard to the field work.

REFERENCES CITED

Binford, L. R., S. R. Binford, R. Whallon, and M. A. Hardin
1970 Archaeology at Hatchery West. *Memoirs of the Society for American Archaeology* 24.
Doran, J. E., and F. R. Hodson
1975 *Mathematics and computers in archaeology.* Harvard University Press, Cambridge.
Flannery, K. V.
1976 *The early Mesoamerican village.* Academic Press, New York.
Johnson, N. L., and S. Kotz
1969 *Discrete distributions.* Houghton Mifflin, New York.
Redman, C. L., and P. J. Watson
1970 Systematic, intensive surface collection. *American Antiquity* 35:279–291.
Reid, J. J., M. B. Schiffer, and J. M. Neff
1975 Archaeological considerations of intrasite sampling. In *Sampling in archaeology,* edited by J.W. Mueller. University of Arizona Press, Tucson.
Schiffer, M. B.
1976 *Behavioral archaeology.* Academic Press, New York.
Siegel, S.
1956 *Nonparametric statistics for the behavioral sciences.* McGraw-Hill, New York.
Sokal, R.R., and F.J. Rohlf
1969 *Biometry.* Freeman, San Francisco.

ERRATUM

In the article, "Archaeology Beyond Anthropology," by George J. Gumerman and David A. Phillips, Jr. (*American Antiquity,* Vol. 43, No. 2, April 1978, pp. 184–191), the following 3 entries should be added to the list of references cited:

Gumerman, George J. (editor)
1971 *The distribution of prehistoric population aggregates: proceedings of the Southwestern Anthropological Research Group.* Prescott College, Press, Prescott, Arizona.
Gumerman, George J.
1973 The reconciliation of theory and method in archeology. In *Research and theory in current archeology,* edited by C. L. Redman, pp. 287–289. Wiley, New York.
Gumerman, George J., and R. Roy Johnson
1971 Prehistoric human population distribution in a biological transition zone. In *The distribution of prehistoric population aggregates: proceedings of the Southwestern Anthropological Research Group,* edited by George J. Gumerman. Prescott College Press, Prescott, Arizona.