Design and Evaluation of Shovel-Test Sampling in Regional Archaeological Survey

James J. Krakker
Michael J. Shott
Paul D. Welch

Museum of Anthropology
University of Michigan
Ann Arbor, Michigan

Shovel-test sampling, the excavation of small test units at regular intervals along survey transects, is a widely used technique for archaeological survey in heavily vegetated areas. In order to achieve efficiently the archaeological goals of a survey employing the technique, the survey should be designed with consideration of the statistical properties of shovel-test sampling. In this paper we examine the effects that test–unit size, spacing, and patterning have on the discovery of archaeological sites of varying size and artifact density. This examination presents some simple procedures both for the efficient design of surveys and the evaluation of existing survey results.

Introduction

Shovel-test sampling is an increasingly common technique used in the archaeological survey of forested regions. The technique consists of excavating small units, usually about 1 sq. ft. in area, at intervals along transects, and is used when vegetation cover limits exposure of the ground surface and therefore prohibits survey by simple surface inspection.

Although the technique is widely used, little is known about the properties of shovel-test sampling for discovering sites and estimating the density of sites within an area. Frequently, critical aspects of a survey based on shovel-test sampling are established without regard to their suitability for the task. For example, the spacing, size, and pattern of test units in a survey area are often determined by factors not related to the archaeological properties of the area or to the aims of the survey. Under these circumstances, there is no guarantee that the survey will yield results suitable to the investigator or representative of the kind and number of sites in the area.

Only in recent years have archaeologists started to employ shovel-testing as a probabilistic sampling technique in forested areas.[1] Even more recently, some thought has been given to the problems associated with the use of shovel-test sampling[2] and critical evaluations of the method as a regional sampling technique have begun to appear.[3] For the most part, however, these studies emphasize the shortcomings of the method but do not offer solutions to the problems. Intuitive suggestions for improvements, such as digging more and more units at shorter and shorter intervals, rarely prove to be practical. Given the familiar material constraints on archaeological research, it is incumbent upon archaeologists to explore more closely the characteristics of shovel-test sampling and to attempt to devise optimal sampling strategies based on the efficient use of scarce archaeological resources. In what follows, we attempt an initial examination of the problem.

Our discussion is restricted to statistical aspects of

1. See W. A. Lovis, "Quarter Sections and Forests: an Example of Probability Sampling in the Northeastern Woodlands," *AmAnt* 41 (1976) 364–372, for one of the first uses of shovel-test sampling as a probabilistic method in regional survey. See also Thomas R. Scott, Michael McCarthy, and Mark A. Grady, *Archaeological Survey Methods in Cherokee, Smith and Rusk Counties, Texas: a Lesson in Survey Methods. Southern Methodist University Archaeological Research Program Research Report 116* (Dallas 1978) 16–20; John H. House and David L. Ballenger, *An Archaeological Survey of the Interstate 77 Route in the South Carolina Piedmont. University of South Carolina Institute of Archaeology and Anthropology Research Manuscript Series No. 104* (Columbia 1976) 43–49.

2. B. Mark Lynch, "Site Artifact Density and the Effectiveness of Shovel Probes," *CA* 21 (1980) 516–517; Glenn D. Stone, "On Artifact Density and Shovel Probes," *CA* 22 (1981) 182–183.

3. For example, Jack D. Nance, "Regional Subsampling and Statistical Inference in Forested Habitats," *AmAnt* 44 (1979) 172–176.

shovel-test sampling that can be derived from geometry and probability theory. We do not report substantive results of actual or simulated shovel-test surveys. In part, this focus is forced upon us because the data necessary for a field test of our results do not, to our knowledge, exist. Specifically, we do not know of any large, heavily vegetated area for which the true site locations, site sizes, and densities of artifacts are fully known. Such a research universe could, of course, be constructed on a computer, but since we do not have empirical knowledge of all the parameters of such a population there is no guarantee that such a computer-based study would be relevant to real populations of archaeological sites.

For our purposes, shovel-test sampling may be viewed as a specific form of point sampling on a grid. This treatment enables us to utilize in the following sections the extensive body of literature on sampling theory and the relationship of sampling to statistical inference that has been developed in a variety of fields.[4]

Probabilities of Site Discovery

Definitions and Assumptions

In order to analyze the statistical properties and limitations of shovel-test sampling, we must be explicit about precisely what we are sampling. Shovel-test sampling usually is employed to discover the location and/or extent of archaeological sites. For present purposes, it is possible to leave aside the issues of the cultural or behavioral import of the term "archaeological site", and instead consider an archaeological site simply as a target or place with definable physical properties. Our interest in this paper is the relationship between the properties of the target and the method—shovel-test sampling—used to provide information about these properties.

Initially we assume that when a shovel-test unit is placed within a site, the site will be discovered. This assumption carries with it a set of implications concerning artifact densities, intrasite distributions, and recovery rates, which we discuss later. Our initial assumption is, of course, unrealistic, but we employ it because it permits a clear exposition of certain statistical properties of shovel-test sampling.

The second assumption we employ is that sites are circular. This is merely a convenience, since the mathematical relationships we discuss can be reformulated for any geometrically definable site shape. For example, Igor Savinskii[5] and Donald Singer[6] show how some of our results can be extended to ellipses of varying size and eccentricity.

Finally, we assume that sites are randomly or independently distributed within the survey area. The validity of this assumption is difficult to verify. Intuitively, one would expect a nonrandom distribution of sites, since archaeological experience has shown that topographic features, soil types, and other factors can often strongly condition the location of sites. On the other hand, while the distribution of sites in a functioning settlement system is almost certainly nonrandom, the cumulative effect of repeated occupation of an area over several millennia may produce essentially random distributions of site aggregates.[7] While most theoretical models make the assumption of independence, statistical sampling theory has investigated the effects on sample results of various nonrandom spatial patterns.[8] It should be possible, therefore, to model the relationship of nonrandom patterns of site distribution to sampling strategies. In the meantime, we assume independence of site locations in the following discussion.

We are not concerned in this paper with the discovery of archaeological sites by the detection of soil-chemical anomalies. Although such by-products of past human occupation are widespread[9] and efficient field methods for their detection have been developed,[10] our discussion treats sites as clusters of artifacts. The relationships we present between the properties of archaeological sites and the number, spacing, and patterning of shovel-test

4. William G. Cochran, *Sampling Techniques*, 3rd edn. (Wiley and Sons: New York 1977); George Koch and Richard Link, *Statistical Analysis of Geological Data, Vol. II* (Dover: New York 1971); Bernard O. Koopman, *Search and Screening: General Principles and Historical Applications* (Pergamon Press: Elmsford, New York 1981); R. B. McCammon, "Target Intersection Probabilities for Parallel-Line and Continuous-Grid Types of Search," *Journal of the International Association for Mathematical Geology* 9 (1977) 369–382.

5. Igor D. Savinskii, *Probability Tables for Locating Elliptical Undergound Masses with a Rectangular Grid* (Consultants Bureau: New York 1965).

6. Donald A. Singer, "Relative Efficiencies of Square and Triangular Grids in the Search for Elliptically Shaped Resource Targets," *United States Geological Survey Journal of Research* 3 (1975) 163–167.

7. This may be viewed as a temporal correlate of the concept of random spatial economy; cf. Leslie Curry, "The Random Spatial Economy: an Exploration in Settlement Theory," *AAAG* 54 (1964) 138–146.

8. Cochran, op. cit. (in note 4) 212–221.

9. Christopher G. Carr, *Handbook on Soil Resistivity Surveying: Interpretation of Data from Earthen Archaeological Sites* (Center for American Archaeology Press: Evanston, Illinois 1982); Fekri Hassan, "Sediments in Archaeology: Methods and Implications for Palaeoenvironmental and Cultural Analysis," *JFA* 5 (1978) 197–213.

10. Robert C. Eidt, "A Rapid Chemical Field Test for Archaeological Site Surveying," *AmAnt* 38 (1973) 206–210; Fekri Hassan, "Rapid Quantitative Determination of Phosphate in Archaeological Sediments," *JFA* 8 (1981) 384–387; I. W. Woods, "The Quantitative Analysis of Soil Phosphate," *AmAnt* 42 (1977) 248–252.

units should apply whether sites are considered as artifact clusters or soil-chemical anomalies. Where we consider artifact density within sites, however, the two views are no longer commensurate. For this reason, our discussion is confined to sites as artifact clusters.

Sampling on a Grid: Grid Size

Since shovel-test sampling is used primarily as a method for discovering sites, our first problem is to determine the probability that a survey employing shovel-test sampling did, in fact, find the sites existing in the survey area. This probability depends on the site size and the sampling interval (i.e., the distance between shovel-test units). Intuitively, one would expect big sites to be found while small sites might be missed. Similarly, the smaller the sampling interval the smaller the sites that will be discovered with certainty. To present the specific algebraic form of these relationships we first show how to calculate the size of the largest circular site that could not escape detection by shovel-test sampling.

Within an area covered by a square grid of shovel-test units, the largest site that can be missed lies at the center of a grid square (FIG. 1). A simple application of the Pythagorean Theorem shows that the radius of the circle illustrated equals the length of the square's side divided by $\sqrt{2}$. To be certain of finding a site of radius r, therefore, the sampling interval i is calculated as follows:

$$i \leqslant r\sqrt{2}.$$

Thus, with an interval of 100 ft. the largest site that can possibly be missed has a radius of 71 ft. Conversely, if we wish to find every site with a 50 ft. radius or larger, we must use a sampling interval of no more than 71 ft. Notice that it is possible in fact to overlook a site that is wider than the sampling interval.

In addition to encountering every site with a radius at least equal to the sampling interval divided by $\sqrt{2}$, a survey employing shovel-test sampling will encounter some of the smaller sites present in the survey area. It is unlikely that shovel-test sampling will encounter all of the small sites unless the sampling interval is extremely (and impractically) small. It is important, therefore, to consider the probability of finding these small sites. This issue has been addressed by researchers in other fields.[11] The probability of encountering a circular site is a function of the ratio between the radius of the site and the sampling interval. Obviously, as the radius of the site approaches the sampling interval divided by $\sqrt{2}$, the probability of encountering the site approaches 1.0. The relationship is not linear, however, as is shown in Curve A of Figure 2. (Yellen apparently has observed a non-linear relationship in his southern African archaeological data.[12]) This curve was prepared from probability tables calculated by Savinskii,[13] who also presents probability tables for encountering ellipses of varying size and eccentricity.

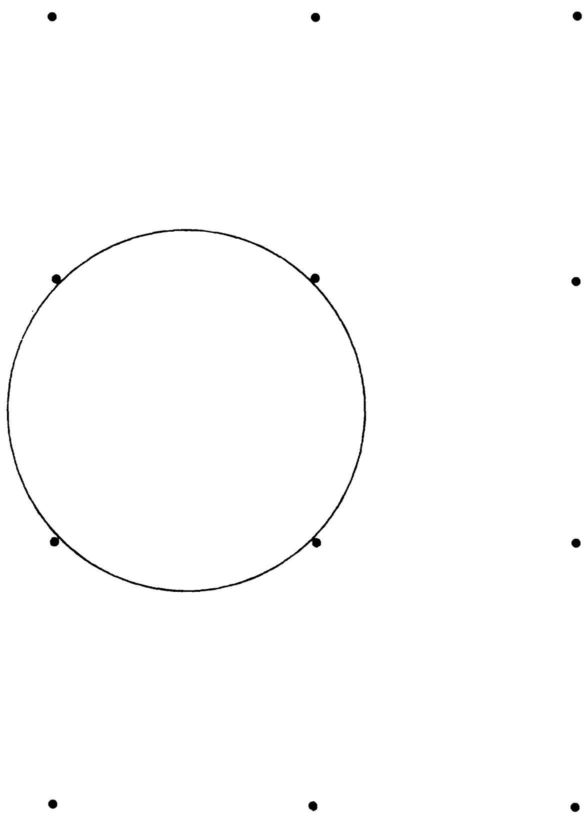


Figure 1. Circular site in center of square formed by four shovel-test units, showing largest site that can escape detection in a shovel-test survey. Circle radius equals $i\sqrt{2}$.

The figure may be read in two ways. First, it indicates the probability of encountering any particular site of a given size using a given sampling interval. Second, the figure indicates what proportion of all sites of a given size would, on average, be encountered with a given sampling interval. It should be noted, however, that as overall site density in a survey area declines the sampling distribution of the estimate of proportions departs from normality and approaches a Poisson shape. This means

11. Savinskii, loc. cit. (in note 5); Lawrence J. Drew, "Grid Drilling and its Application to the Search for Petroleum," *Economic Geology* 62 (1967) 698–710; Singer, loc. cit. (in note 6).

12. John E. Yellen, *Archaeological Approaches to the Present: Models for Reconstructing the Past* (Academic Press: New York 1977) 81.

13. Savinskii, op. cit. (in note 5) Tables 1–18.

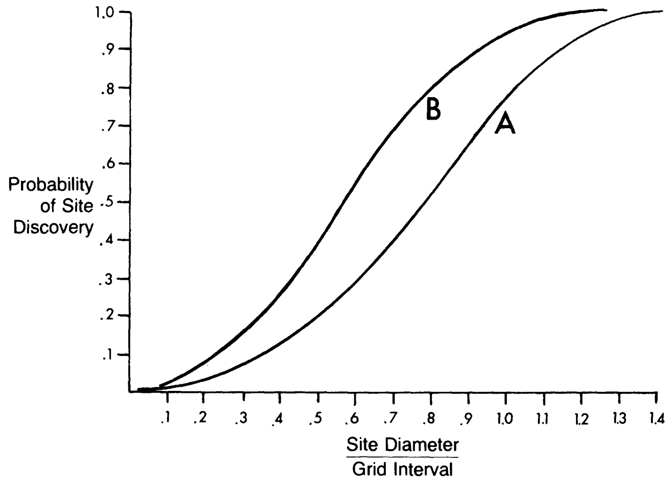


Figure 2. Relationship of the ratio of site diameter:grid interval to the probability of discovering a site, using a square grid (Curve A), and a staggered grid (Curve B).

that for low site densities the number of sites actually encountered is likely to be even less than what can be calculated from Figure 2.

There is a .78 probability of finding a site with a diameter equal to the grid interval. Perhaps more significant to note is that there is a .78 probability of discovering a site that has only one half the area of the target size that is certain to be discovered. As site diameter declines below the grid interval, the probability of discovery declines rapidly. For sites with a radius less than one half the grid interval the probability of discovery can be calculated as follows:

$$p = (\pi r^2)/i^2$$

where r is the radius of the site and i is the interval between tests.

Sampling on a Grid: Grid Pattern

For finding circular targets, a square grid is not the most efficient way to distribute shovel-test units. For example, rather than using a square grid the test units could be dug in staggered rows (i.e., with the distance between rows equal to the interval between tests in a row, but with tests in adjacent rows offset one half this interval). With such an arrangement one could be certain of finding sites of the same minimum size, but with only 77.6% as many test units as would be necessary using the square grid. Theoretically optimal, though only slightly better than the preceding, is a true hexagonal grid, which would require only 76.4% as many shovel-test units as would a square grid to discover circular sites of a given minimum diameter.[14] Though the hexagonal grid is much more efficient than the square grid for discovering circular sites, the relative efficiency of the hexagonal grid decreases as the sites become increasingly elliptical.[15]

Constructing a hexagonal grid in the field is simple. The offset, or staggered, grid described above is close to being a hexagonal grid, but differs from the true hexagonal spacing in having the interval between rows equal to the distance between the tests on a row. The proper spacing between rows for a true hexagonal grid is slightly less than the interval between tests on a row. For a hexagonal grid, a triangle drawn between two tests on a row and the intervening test on the adjacent row is an equilateral triangle. The spacing between rows of tests is thus the height, h, of an equilateral triangle having sides of length i, the interval between adjacent tests:

$$h = 1/2\, i\sqrt{3}.$$

For example, where $i = 50$, $h = 43.3$.

Not only does the use of staggered shovel-test units increase efficiency in regional survey, but it also increases the probability of site discovery for a given ratio of site diameter:grid interval. Curve B in Figure 2 shows the improved relationship. The probability values along this curve are calculated as the ratio between that portion of the area of a circular site lying within the triangular grid (which is created by the staggered units) and the area of the grid triangle, assuming the circle is centered equidistant from the vertices of the triangle. This is a modification of the calculations presented by L. B. Slichter.[16] Curve B shows the probability values associated with the use of a staggered grid of units. The values for a perfect hexagonal grid of units would lie slightly above Curve B, but have been omitted for purposes of visual clarity.

Using a grid of staggered units, where the ratio of site diameter:grid interval is 1.0, one can attain only a .94 probability of discovery. As we saw previously, the comparable probability for an aligned grid is .78. At

14. The discussion of this point by Koch and Link (op. cit. [in note 4] 217–218) is in error both in the equations used and in the value used for the hexagonal grid interval. The relative efficiency values we have presented are calculated by comparing the respective densities of shovel-test units per unit area necessary to discover with certainty a circle of a given radius.

15. Singer, op. cit. (in note 6) 164.

16. L. B. Slichter, "Geophysics Applied to Prospecting for Ores," in A. M. Bateman, ed., *Economic Geology*, 50th Anniversary Volume, *1905–1955*, Part 2 (Economic Geology Publishing Co.: Urbana, Illinois 1955) 885–969.

lower values of the ratio, the differences between discovery probabilities for staggered and aligned grids are even greater. Thus, where the ratio is .8, the site will be discovered with a probability of .81 using a staggered grid, but only a .50 probability using an aligned grid.

This comparison clearly illustrates the superior ability of staggered grids to discover sites. Especially noteworthy is the fact that the greatest improvement is found for the range of sites with diameters between ca. 0.4 and 0.9 of the grid interval. This is a range of relatively small sites (for a given grid interval) that may often be of great importance to the success of a regional survey.

Surveying Transects: Site Discovery Probabilities and Sampling Interval

Frequently the area to be surveyed by shovel-test sampling is a long, narrow strip, such as a highway right-of-way or a pipeline easement. In these cases the shovel-test units are usually spaced at regular intervals in 1 to 3 parallel lines.[17] This sort of transect sampling differs from grid sampling in several ways. The difference arises from the effect of the edges of the survey area. When shovel-test sampling large blocks of land one can ignore this boundary effect, or, if desired, compensate by additional testing along the boundary. The boundary effect assumes major importance when one is surveying transects.

The term "boundary effect" here refers to the fact that, regardless of the sampling interval, even very large sites impinging on the transect can be missed if they lie only along the edge of the transect. This situation is depicted in Figure 3, which shows how a single line of shovel-test units at 50 ft. intervals along a 100-ft.-wide right-of-way would miss the part of Teotihuacan crossed by the right-of-way. Intuitively, one might think that the way to avoid a problem like this is to use two lines of shovel-test units, one along each edge. This will work, of course, but at the cost of doubling the number of shovel-test units that must be excavated. If we use two lines of shovel-test units but lengthen the sampling interval to 100 ft., so that the number of shovel-test units excavated is the same as in the original case, the survey could completely miss a 140-ft.-wide site centered on the right-of-way. The question confronting the archaeologist is this: how to minimize the boundary effect with the fewest shovel-test units without exceeding a stated size of site that could be missed in the center of the transect. A couple of very simple equations provide the answer.

Consider first the case of a single line of shovel-test units. If some part of a site lies within the survey area, but no part of the site is crossed by the transect of shovel–test units, the site will not be detected. If the line of shovel-test units does cross some part of the site, it may be detected, but we are certain to detect the site only if that part of it crossed by the transect is wider than the sampling interval. Using our assumption that sites are circular, we can begin to calculate the largest site size that we could possibly miss with a single transect. To make the calculations easier and, more importantly, archaeologically reasonable, we will also assume that we are not concerned with sites lying more than half outside of the survey area. (After all, continuing the Teotihuacan example, we would not be terribly concerned if we were to nick a 5-ft.-wide slice off the edge of a site that covers several square miles.) With these assumptions, the solution to our problem is shown in Figure 4. The largest site that can be missed is centered on the edge of the survey area equidistant from consecutive shovel-test units. Applying more high-school geometry shows that, if the shovel-test units are spaced at intervals of i units apart in a line down the middle of a band of width w, the radius r of the largest site that can be missed is:

$$r = 1/2 \sqrt{(w^2 + i^2)}.$$

17. For example, see James J. Krakker, Michael J. Shott, and Paul D. Welch, *Report of Phase I Archaeological Investigation of the Huron Valley Wastewater Control System, Wayne County, Michigan* (Division of the Great Lakes, University of Michigan Museum of Anthropology: Ann Arbor 1981) 45–48.

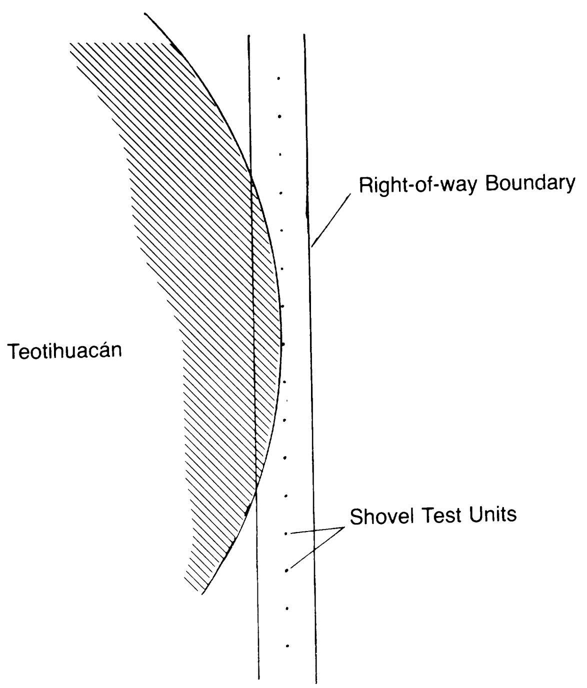
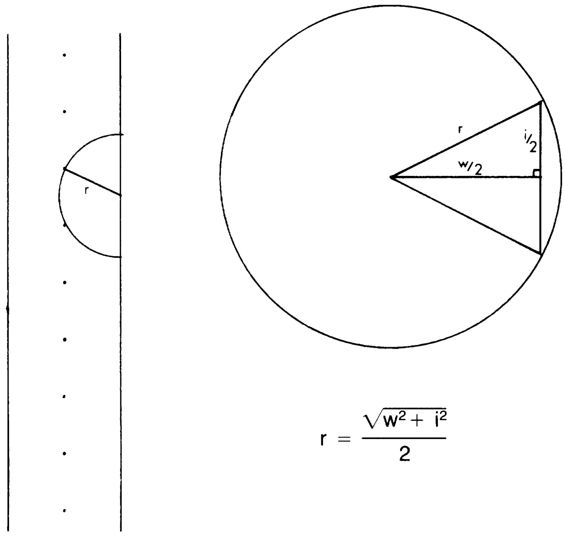


Figure 3. The boundary effect in transect surveying.



$$r = \frac{\sqrt{w^2 + i^2}}{2}$$

Figure 4. Derivation of maximum site size that can be overlooked using a single line of shovel-test units.

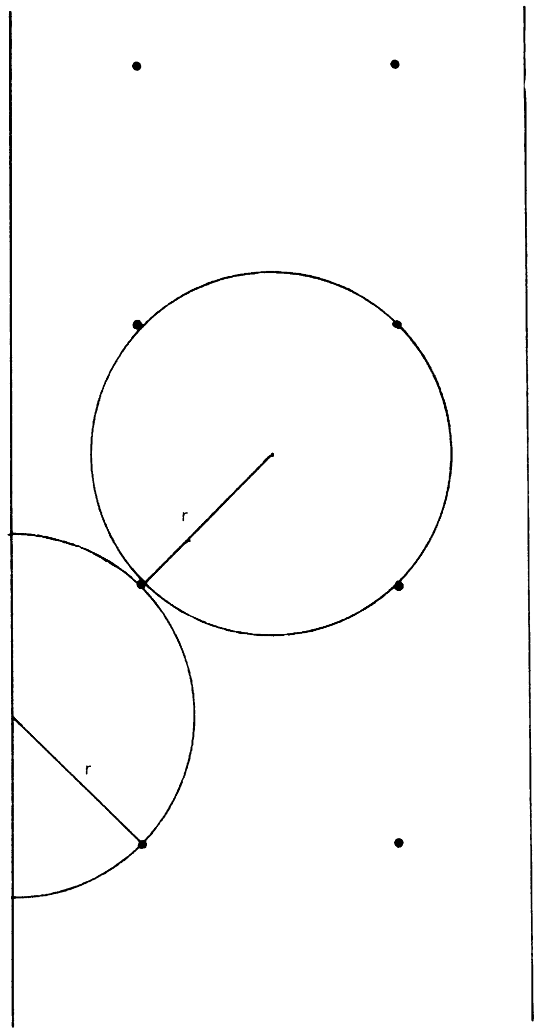


Figure 5. Optimal spacing of aligned units, two-transect case.

Thus, with a line of shovel-test units 50 ft. apart down the center of a 100-ft. right-of-way, the smallest site that can be discovered with complete certainty measures 56 ft. in radius. It is important to note that simply decreasing the sampling interval will not entirely alter the situation.

If we are concerned with the prospect of overlooking a 110-ft.-wide site even though the shovel-test units are placed at 50-ft. intervals along a 100-ft.-wide strip, an obvious solution is to excavate two transects of shovel-test units. It is not so obvious, however, how far apart the transects should be, nor what interval to use between shovel-test units. One solution is depicted in Figure 5. In this case the optimal interval is attained when the two circles have equal radii. The reason for this is that we want to be equally certain of finding a site in the center of the survey area as we are of finding a site of the same size centered along the edge of the area. Our now well-honed high-school geometry indicates that this arrangement is attained when the interval i between shovel-test units on each transect equals the distance between the transects and when this distance i is one–half the width w of the survey area. With this arrangement the largest site that can be overlooked has a radius r of the sampling interval divided by $\sqrt{2}$. This, of course, is the same relationship as in the case of testing a wide area with a square grid.

Surveying Transects: Optimal Sampling Pattern

Just as a square grid is not the most efficient way to test a wide area, the arrangement shown in Figure 5 is not the most efficient way to position shovel-test units along the transects. It is more efficient to stagger the intervals of two transects than to have the intervals aligned. That is, rather than placing shovel-test units abreast of each other, it is better to offset the units one half the distance between consecutive units on a single transect, as is shown in Figure 6. As before, the optimal patterning is achieved when the two circles shown have equal radii. The equations relating the sampling interval i to the width w of the survey area, however, and the radius of the largest site that can be overlooked, are quite different from the preceding case.

To show how the equations are derived, in Figure 7 we have expanded part of the previous figure and labelled some of the geometry. The large triangle connects two consecutive shovel-test units on one transect with the intervening shovel-test units on the second transect.

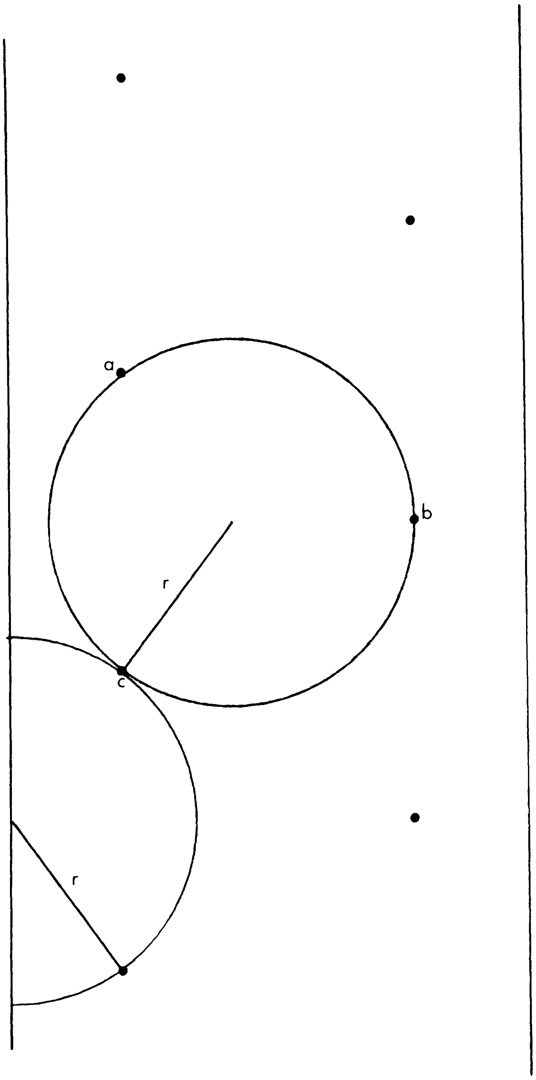
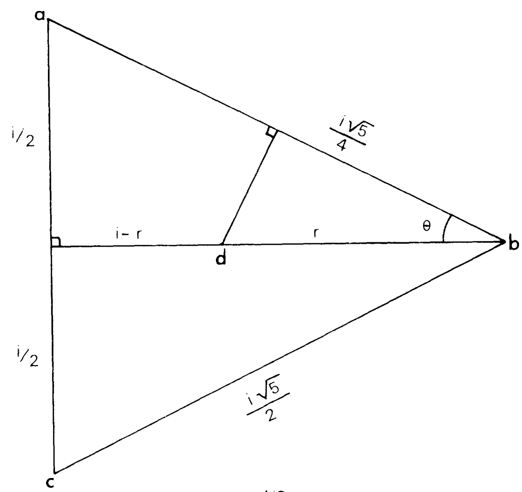


Figure 6. Optimal spacing of offset units, two-transect case.



$$\tan \theta = \frac{i/2}{i} = .5 \qquad \text{hence } \theta = 26^\circ\, 34'$$

$$\cos \theta = \frac{(i\sqrt{5}/4)}{r} \qquad \text{hence } r = \frac{i\sqrt{5}}{4 \cos \theta} = .625i$$

Figure 7. Derivation of optimal spacing of offset units, two-transect case.

The largest site that can be missed by these shovel-test units has its center at the point d, which is located at the intersection of orthogonals drawn through the midpoints of the sides of the large triangle. Some simple trigonometry allows us to calculate the value of r, which is .625i. The largest site, therefore, that can be missed between staggered shovel-test units along two transects has a width of 1.25 multiplied by the sampling interval, as compared to 1.4 multiplied by the interval when the shovel-test units are aligned.

A further difference between the aligned and staggered shovel-test units is the relation of the sampling interval to the width of the survey area. For aligned shovel-test units the sampling interval i should be half the width w of the survey area, while by comparing Figures 6 and 7 it can be seen that for staggered shovel-test units the sampling interval i should be .571 the width w of the survey area. In simple terms, this means that for a given width of the survey area, when compared to aligned spacing the staggered spacing will be equally certain of detecting sites of a given size and will do so with 12% fewer shovel-test units.

The equations used to calculate the optimal interval between staggered shovel-test units can be generalized for cases where more than two transects are needed. The derivation of the equations is quite simple, as can be seen by comparing Figures 6 and 8. Where two transects are used (FIG. 6), the width w of the survey band equals the distance i between transects plus the distance .375i on the outer side of each transect ($w = i + .375i + .375i$). Where three transects are used (FIG. 8), the width w equals two times the interval i plus twice .375i ($w = 2i + 2\,[.375i]$). In general then, the width w can be expressed as:

$$w = (t - 1)\,i + 2\,(.375i)$$

where t is the number of transects. This allows us to calculate the optimal grid interval for a band of given width to be surveyed with a given number of transects. The width of the survey band, of course, is usually specified *a priori*. The number of transects to be used can be determined from the size of the smallest site one wishes to be absolutely certain of discovering.

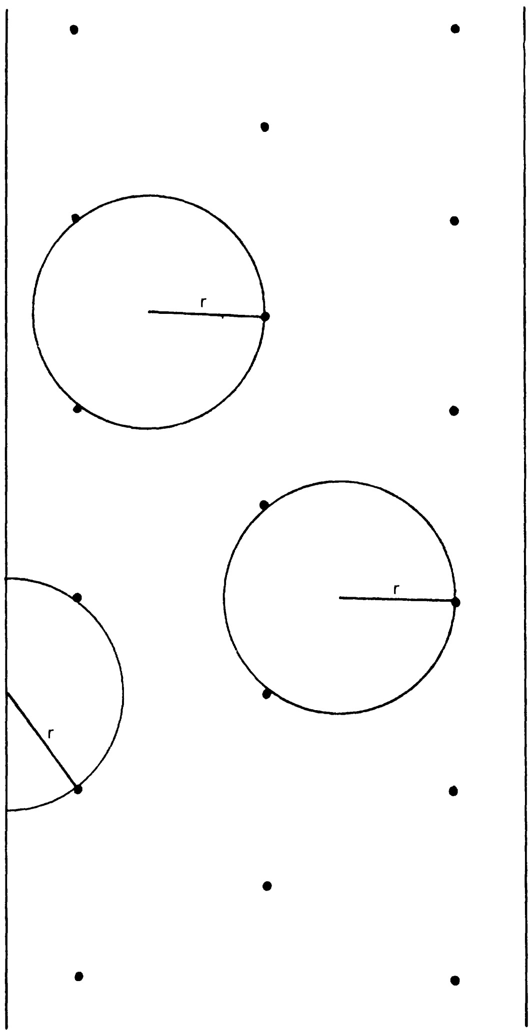


Figure 8. Optimal spacing of offset units, three-transect case.

In Figure 9 we have graphed the relation between site size, the width of the survey band, and the minimum number of transects that must be used to be certain of detecting sites of a given size. To use the graph, find on the vertical axis the width of the band to be surveyed. Follow that horizontal line rightward to its intersection with that slanting line which represents the minimum site size the survey must be certain of detecting. From this intersection read down to the horizontal axis. The minimum number of transects that must be used is the next whole number greater than or equal to the point reached on the horizontal axis.

Density and Detection of Artifacts

Thus far we have assumed that a site is discovered when a shovel-test unit is placed within it. This assumption itself involves others concerning artifact densities within sites and the probability of actually detecting artifacts present in a test unit. The empirical validity of such assumptions, of course, is open to question. Lynch, for instance, has shown that sites that by many standards would be considered fairly large and dense can escape detection in a shovel-test sample.[18] Apparently, real archaeological artifact densities and probabilities of detecting items present in test units are highly variable. In this section, we examine the effects of these factors on the probability of discovering sites.

Artifact Density and Probability of Site Discovery

By making the optimistic assumption that all artifacts present in a shovel-test unit will be discovered, we note that the minimum artifact density that can be found with certainty is one item per test unit area. Sites with densities below this threshold will not be found with certainty. In general, then, the shovel-test unit size should be no smaller than the area in which there is a high probability of finding one item.

A Poisson distribution may be used to estimate the probability of finding at least one item in a test unit for a specific *average* debris density.[19] The frequency distribution of occurrences per observation unit is expected to follow a Poisson distribution when the mean occurrences per observation unit are low.[20]

Using a Poisson distribution, we find that the probability of locating at least one item in a test unit is estimated by:

$$p = 1 - e^{-ad}$$

where a is the area of the test unit and d is the artifact density.[21] If the average density is one item per 1 sq. ft. and test units of 1 sq. ft. are used then the probability of finding at least one item in a test unit is .632. For a specified probability, p, of finding at least one item at a specified average density, the area that would require testing is estimated by:

$$a = (-\ln(1 - p))/d.$$

18. Lynch, op. cit. (in note 2) 517.

19. Stone, op. cit. (in note 2) 182.

20. Robert R. Sokal and James F. Rohlf, *Biometry: The Principles and Practice of Statistics in Biological Research* (Freeman and Company: San Francisco 1969) 81.

21. Ibid. 84–86. Note that Stone (loc. cit. [in note 19]) presents the formula as $p = 1 - e^{-a/d}$ and defines artifact density, d, as the area containing one artifact. This is the inverse of our definition of artifact density—i.e., number of artifacts per unit area—which we believe is the more common usage of the term.

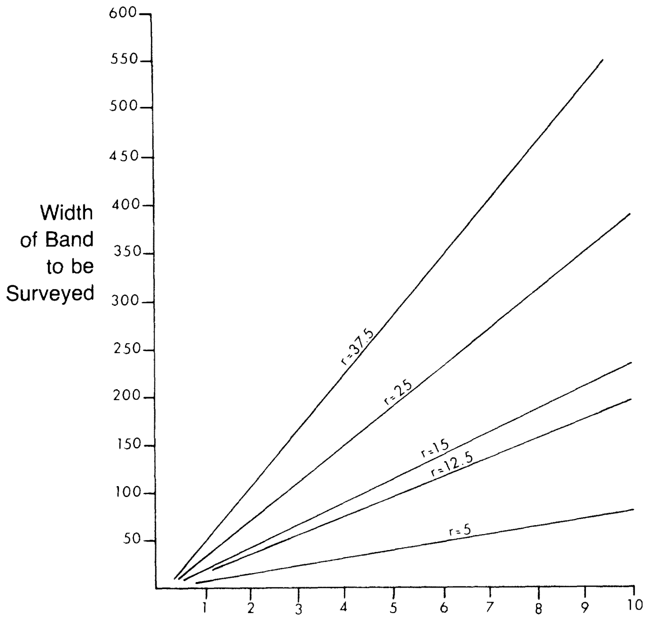


Figure 9. Relationship between number of transects, width of survey band, and radius (r, in same units as band width) of minimum site size to be discovered with certainty.

Therefore, to achieve a .95 probability of finding at least one item requires testing about 3 sq. ft. when the artifact density is one item per 1 sq. ft.

The Poisson sampling distribution assumes a random spatial distribution of artifacts where the variance of the distribution equals the mean.[22] We consider it unlikely, however, that artifacts are randomly distributed within a site. Departure from a random spatial distribution would be expected to be in the direction of a more clustered or heterogeneous distribution. The other alternative, a uniform spatial distribution, does not seem likely. The effect of a departure from random toward heterogeneous distribution on the sampling distribution of artifacts among test units would be to increase the number of units with no items and number of units with more items than expected.[23] Departure from the assumption of a random distribution, therefore, makes the Poisson-based estimate of unit size too small and underestimates the probability of not finding any items in a unit.

In reality, the areas of artifact clusters are likely to be much larger than the size of the test unit. Even in cases where there is obvious clustering of artifacts, the distribution of their occurrences in small units, at least experimentally, fits a Poisson distribution.[24] At any rate, the Poisson is probably a suitable model for present purposes. Further research and simulation, however, may indicate that a negative binomial or other density function provides a better fit to actual artifact density distribution.

As we have seen, in practical terms we define sites by assuming a minimal mean density of artifacts, or threshold density. A site may have an average density above a given threshold, yet, if there are large areas within it below the average density the site may not be found. The patterning of variability of intrasite density has some obvious influences on the effective site size that can be discovered with a given probability.

To begin with, artifact density should normally decline towards the margins of a site (FIG. 10). Therefore, the effective size of the site that can be found with certainty is reduced to the area of the site with an artifact density exceeding the threshold that can be found given the shovel-test unit size.

22. Sokal and Rohlf, op. cit. (in note 20) 85.

23. Peter Greig-Smith, "The Use of Random and Contiguous Quadrats in the Study of the Structure of Plant Communities," *Annals of Botany* 16 (1952) 293–316; Kenneth A. Kershaw, *Quantitative and Dynamic Plant Ecology*, 2nd edn. (American Elsevier: New York 1975) 133–134.

24. Greig-Smith, op. cit. (in note 23) 307–308; Kershaw, op. cit. (in note 23) 136.

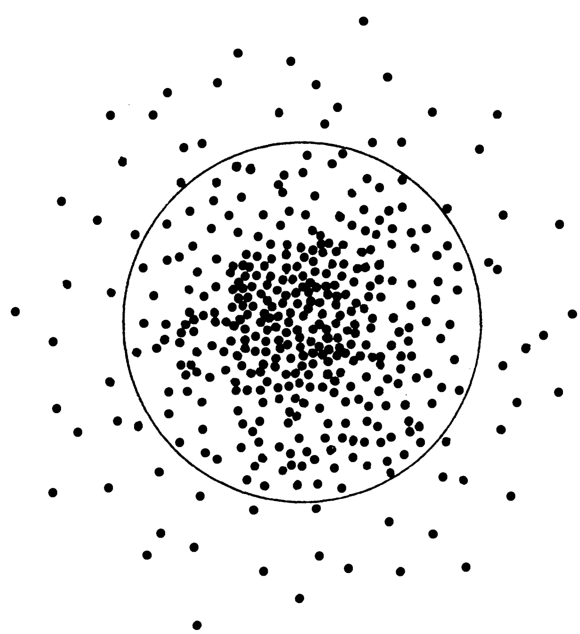


Figure 10. Simple pattern of variability of artifact density.

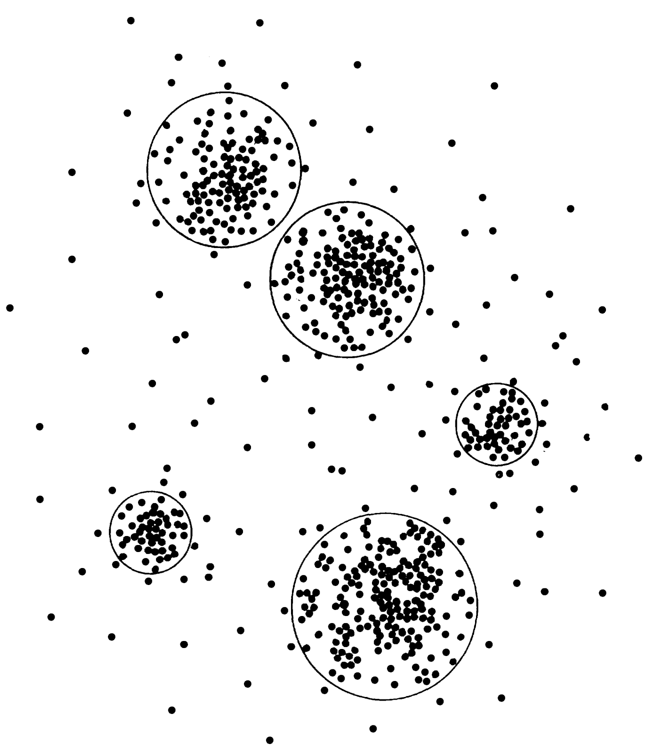


Figure 11. Complex pattern of variability of artifact density.

Of course, the internal structure of sites is not expected to be limited to the simple bull's-eye pattern of maximum density in the center and declining density towards the margins. Within sites there are likely to be several areas of high density of artifacts (FIG. 11). Provided that the artifact density between these concentrations is higher than the minimum density that can be detected with certainty given the test–unit size, the site will still be found. If, however, there are areas within a site below the density threshold, then the probability of discovering the site is reduced. It is easy to visualize sites consisting of clusters of artifact concentrations or overlapping concentrations, with artifact density below the threshold in between. In such a case, the site in effect consists of a cluster of small sites, each of which may be smaller than the minimum size that can be discovered with certainty given the sampling interval used in the survey.

Detection of Artifacts and Test-Unit Size

Up to this point we have assumed that artifacts present in a shovel-test unit will be detected. Intuitively, it seems likely that this assumption may not always be warranted, since many artifacts, such as flakes and sherds, are small and can be difficult to detect in a test unit. We now examine the effects of differential artifact detection on the probability of discovering a site, and make an initial attempt to deal with this problem.

Since the discovery of an archaeological site requires finding only one artifact in a shovel test, imperfect detection of those artifacts actually present will not be a critical problem when artifact densities are very high. Imperfect detection of artifacts, however, will cause an archaeological survey to miss many sites of low density. Since imperfect artifact detection results in decreased *effective* artifact density, the shovel-test unit size calculated by the methods discussed above must be adjusted.

To adjust the shovel-test unit size, we may use a binomial model of the probability of detecting those artifacts actually present. We assume that the probability of detecting items is independent; i.e., the probability of detecting one item is not related to the probability of detecting any other. If the probability of finding an item is p, then the probability of not finding it, q, is $1 - p$. In n trials the probability of finding one or more items is found from the binomial expansion $1 - q^n$.[25] If the area of a shovel-test unit has two items and their independent probabilities of detection are .9, then the probability of discovering one item is .99. If the probability of item detection is .8, the probability of detecting at least one item is .96. Hence, at detection rates near 100%, there is near certainty of discovering a site of the minimum artifact density if the shovel-test units are about twice the size that would be used assuming 100% detection. Similarly, for test units of a given size and detection rates of .8 or .9, the only sites virtually certain of discovery are those having two or more items per test unit.

If the detection probability is as low as .5 (we hope that it is not lower than this) then the probability is .75 of discovering a site that has two items per test unit. Obviously, if the detection probability is well below 1.0, test units much larger than 1 sq. ft. must be used to be nearly certain of detecting sites that have only one item per 1 sq. ft. Similarly, for low-detection probabilities there is a rather high probability of missing even those sites with two items per test unit.

Another problem to consider is that the probability of detection may vary with artifact type. For example, it seems likely that large artifacts like fire-cracked rocks should be found with a greater probability than smaller artifacts like flakes. The implication, if detection probabilities vary among artifact types, is that some sites are more likely to be discovered than others of the same density, only because of their different artifact composition.

Screening of shovel-test units, of course, greatly increases the probability of detection. Perhaps an equally

25. Sokal and Rohlf, op. cit. (in note 20) 71.

significant feature of screening is the reduction of the bias produced by different observers' abilities and differences among the visibilities of various types of artifact. Whether or not screening is more efficient than excavating larger shovel-test units can be empirically determined and depends on how much lower is the probability of finding items without screening.

In sum, the implication is that shovel-test units should be of some minimum size, and that they must be adjusted in size depending on the average minimum density of artifacts one would like to be certain, or nearly certain, of discovering. In addition, the size of shovel-test units must be adjusted to account for imperfect detection of artifacts.

Discussion

Having established some theoretical relationships between site sizes, shapes and densities on the one hand, and strategies for grid sampling on the other, we consider how these relationships work out in practice and explore the implications they have for the use of shovel-test sampling.

Let us assume for our purposes that it is important to discover a circular archaeological site 50 ft. in diameter having an average density of one artifact per 1 sq. ft. Determining whether this is a relatively large or dense site would require systematic information on regional site sizes and artifact densities, but for many regions such a site intuitively seems reasonable. The area of the site is 1,963 sq. ft., and it contains 1,963 artifacts, which we assume are randomly distributed within it. Referring to Figure 2, we see that a shovel-test survey using a square grid of 1-sq.-ft. units at 50 ft. intervals has roughly a .80 probability of actually encountering the site. If a test unit does happen to encounter the site, the Poisson sampling distribution indicates there is a .63 probability that the unit will actually contain at least one artifact. If we assume a .9 probability of detecting those artifacts present, then the actual probability that the shovel-test survey will discover the site is roughly .45 (i.e., .9 × .63 × .8). Conversely, we expect that such a shovel-test survey would discover only 45% of such sites actually present.

Using the procedures we have presented, we can design a shovel-test survey that would have a far higher probability of discovering our hypothetical site. First, the test units can be distributed more efficiently. Using a square grid with staggered or offset test units, we can have a 100% probability of encountering the site if we shorten the sampling interval to 40 ft. (using a square grid with aligned test units we would have to shorten the interval to 35 ft.; for a hexagonal grid the interval would be very slightly above 40 ft.). Second, we can adjust the test–unit size. Using the Poisson sampling distribution, we can calculate the size of unit necessary to achieve a specified probability that at least one artifact will be in the unit(s) that encountered the site. For a 90% probability that the unit contains at least one artifact, the unit must be 2.3 sq. ft.; for an 80% probability the figure is 1.6 sq. ft. Third, we can compensate for the assumed .9 probability of detecting those items actually present. The compensation can be made by increasing the size of the test unit or by increasing the detection rate through use of screens, or both. If we do not make such compensation, the probability of discovering the site is .72 (.8 × .9) for test units of 1.6 sq. ft. (Actually, this is a fairly conservative estimate of the probability of discovering the site. While it is certain that at least one test unit will encounter the site, it is probable that two or more units will encounter the site and in each case there is a .72 probability that the site will be detected.)

With such a design, searching for the site in a survey area of 1/4 sq. mi.[26] will require the excavation of 4,356 shovel-test units. Even without exact measurement of the intervals and with only limited notetaking, it is optimistic to assume a rate of 10 units per person per hour. At this rate, over 54 person-days are required for the survey of the 1/4 sq. mi.

Because there is a tremendous labor investment required for the survey of extensive areas, shovel-test sampling will probably have prohibitively high labor requirements if small intervals are used and large areas are surveyed. It is probably a practical method to use only in the investigation of fairly small areas measuring perhaps up to several acres in extent.

To increase accuracy and to reduce labor requirements, two-stage or stratified sampling designs would be useful procedures to use. A first-stage survey sample may be used to estimate the distribution of site size, site density, artifact density, and spatial distribution of sites. This information would be used in designing a second-stage survey that would achieve a specified accuracy level. Prior knowledge, survey priorities, or results from an initial survey may provide a basis for stratification of a survey area and allocation of survey effort more efficiently.

At some point shovel-test survey becomes excessively labor intensive in comparison to the research goals, and in each survey the decision whether or not to use shovel-test sampling should probably rely at least in part on archaeological judgement. We emphasize, however, that rigorous sampling procedures permit the exact calculation of probabilities of site discovery and thus make eval-

26. Survey units are frequently one-quarter mile in area (e.g., Lovis, loc. cit. [in note 1]).

uating the results of surveys possible. Furthermore, rigorous sampling is essential for comparison between the results of different surveys. In contrast, reliance on non-probabilistic methods, while possibly justified in some cases, renders comparison and rigorous evaluation of results impossible.

Conclusion

Shovel-test sampling is a useful discovery and estimation technique in regional surveying, but there are some serious problems connected with its use. We caution against the uncritical use of results produced by shovel-test sampling; like any research technique the results are constrained by the design. Survey procedures should be modified to be consistent with the goals and circumstances of a research project.

Under most circumstances, archaeologists using shovel-test sampling can expect to discover only a portion, perhaps a very small percentage, of the sites that are present in their survey areas. Fortunately, in many cases they should be able to estimate reliably the actual number present.

The purpose here has been to provide a quantitative basis for adjusting shovel-test survey procedures to research goals and labor constraints. In using these methods, archaeologists are concerned with establishing grid intervals, with patterning of test units, and with adjusting test-unit size to achieve some specified level of certainty for finding sites of a specified size and artifact density. The same methods may be applied to evaluate the results of a completed survey or to compare the results of different surveys.

We hope that this discussion will stimulate further research on the methods and problems of shovel-test sampling. Such research should introduce greater efficiency, accuracy, and precision to the collection of archaeological survey data.

Acknowledgments

Dr. Richard I. Ford and Dr. Henry T. Wright of the University of Michigan, Dr. William McDonald of Columbia University, Dr. William A. Lovis of Michigan State University, and Larry Gorenflo of the University of California at Santa Barbara read and commented on earlier versions of this paper. We are grateful for their advice and assistance. Any errors or omissions in the text, however, are solely our responsibility.

James J. Krakker, Michael J. Shott, and Paul D. Welch are doctoral students in Anthropology at the University of Michigan. They have jointly conducted fieldwork in Michigan that served as the inspiration for this paper. Krakker is completing his dissertation on social organization in the late prehistory of SE Michigan, Shott is studying the adaptations of Preceramic foraging societies in the Central Andes, and Welch is conducting research on the Late Woodland-Mississippian transition in the SE United States.