A Simple Mathematical Procedure for Estimating the Adequacy of Site Survey Strategies

Linea Sundstrom


Stable URL:
http://links.jstor.org/sici?sici=0093-4690%281993%2920%3A1%3C91%3AAASMPFE%3E2.0.CO%3B2-6

*Journal of Field Archaeology* is currently published by Boston University.

Your use of the JSTOR archive indicates your acceptance of JSTOR’s Terms and Conditions of Use, available at http://uk.jstor.org/about/terms.html. JSTOR’s Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://uk.jstor.org/journals/boston.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.
A Simple Mathematical Procedure for Estimating the Adequacy of Site Survey Strategies

Linea Sundstrom
Milwaukee, Wisconsin

A set of easily calculated formulas estimate the probability that archaeological sites of a given size and shape were intersected by survey transects. Estimates derived from the formulas can be used by archaeologists in planning surveys and by cultural resource management personnel in evaluating survey proposals and results.

Introduction: The Myth of the “100%” Survey

A great deal of interest in site discovery methods is evidenced by the archaeological literature of the last two decades (for a sampling, see Dunnell and Dancey 1983; Goodyear, Raab, and Klinger 1978; Judge 1981; King 1978; Kintigh 1988; Kohler and Parker 1986; Krakker, Shott, and Welch 1983; Lightfoot 1986, 1989; Lynch 1980; McManamon 1984; Mueller 1975; Nance 1983; Nance and Ball 1986, 1989; Plog, Plog, and Wait 1978; Plog, Weide, and Stewart 1977; Redman 1987; Schiffer and Gumerman 1977; Schiffer, Sullivan, and Klinger 1978; Short 1985, 1989; Stone 1981; and Wandsnider and Ebert 1988). This interest seems to have been sparked by two developments: first, the interest in “scientific” methods introduced as part and parcel of the New Archaeology; and second, the need for cost-effective field methods by archaeologists involved in cultural resource management (CRM). Much of this literature concerns sampling theory; other important topics have been site versus non-site approaches and subsurface exploration, especially the effectiveness of shovel testing.

Without entering into the theoretical debates that have characterized the literature on sampling and site discovery, this paper presents a simple, straightforward means of estimating the adequacy of site survey plans. To those archaeologists well-versed in statistical and sampling theory, the points made here may seem overly basic. Nevertheless, even a glance at recent CRM reports will show that the myth of the “100% survey” is alive and well, and, tragically, still contributing to the loss of archaeological resources through ignorance (or fear) of the basic mathematical relationships between site area and transect spacing. The idea that “intuition” alone is an appropriate method for evaluating some or all survey plans also continues to crop up (Kintigh 1988: 686).

While the concepts and formulas presented here are basic (and require nothing more than a pencil and paper or calculator), they have several advantages. First they are more than adequate to bring meaning to frequently used, but infrequently defined, terms like “completely surveyed” or “100% coverage” (e.g., Judge 1981; Plog, Plog, and Wait 1978; Ebert 1988). They also give CRM personnel a non-subjective basis for gauging whether research designs are adequate to accomplish stated goals of site discovery. These concepts and formulas can be used in setting realistic minimum data recovery standards (as opposed to the often impractical 100% standard) and can address the cost-effectiveness of survey designs (Redman 1987). Finally, they can provide a viable justification for the resurvey of areas thought not to have been adequately examined by earlier projects.

Standard, directly comparable methods of surface survey are lacking in American archaeology (Wandsnider and Ebert 1988). In general, each project uses its own combination of sampling methods and survey techniques, be they systematic or intuitive. This often forces CRM personnel charged with evaluating projects to accept the principal investigator's or project director's own “guesstimate” of the effectiveness of a project strategy. Since the evaluator lacks independent measures of survey strategy effectiveness, the evaluation that results is often meaningless.

Archaeologists conducting field searches for archaeological sites have adopted systematic search patterns, generally in the form of equally-spaced parallel transects. While this
is an improvement over randomly inspecting the area to be surveyed, decisions about the spacing of transects have generally relied more upon the intuition of the field director than on an understanding of the mathematical relationships between the transect pattern and the probability of finding sites of a given size (Plog, Plog, and Wait 1978; Judge 1981; Plog, Weide, and Stewart 1977; Ebert 1988). Survey reports often contain a statement that 100% coverage was attained by the survey, when, in fact, the probability that the survey was adequate to find all sites present in the area was much lower. The phrase "100% coverage" too frequently is merely the opinion of the field director of the adequacy of the survey, rather than an actual quantitative measure (Plog, Plog, and Wait 1978; Ebert 1988).

Few systematic studies of survey effectiveness have been conducted. One exception is presented in Dunnell (1988). This study, based on resurvey of a controlled area (a plowed field) over three years, yielded one finding of particular relevance to the method proposed here for evaluating surveys. Consistent (i.e., "accurate") survey results were obtained when a uniform survey strategy was applied, in spite of annual variations in erosion and plant cover. The number of artifacts located was directly related to the width of the transects, as well as to the more elusive factor of time spent on each transect. In other words, "surface collection techniques can be controlled" (Dun nell 1988: 34).

Estimating Survey Adequacy

A simple mathematical method can be used to evaluate the adequacy of survey strategies. (These formulas are adapted from Davis 1986). The probability of discovering a target (in this case, an archaeological site) is a function of the size and shape of the target and the spacing of the search pattern (in this case, parallel transects). Assuming an elliptical target,

\[ P = \frac{s}{\pi d}, \]

where \( P \) is the probability that all sites of size \( s \) were intersected by survey transects spaced \( d \) distance apart. The perimeter \( s \) is calculated by:

\[ s = 2\pi \sqrt{\frac{a^2 + b^2}{2}}, \]

where \( a \) and \( b \) are the semiaxes of the ellipse. The formula for \( P \) can be reduced to

\[ P = \frac{2\sqrt{(a^2 + b^2)/2}}{d}, \]

or, if the target is circular,

\[ P = \frac{2r}{d}, \]

where \( r \) is the radius of the target.

For a linear target, the formula is

\[ P = \frac{2l}{\pi d}, \]

where \( l \) is the length of the linear target.

One or the other of these formulas should apply to any expected site shape. The elliptical site model is applicable to most archaeological sites, since even a linear site such as a road or trail produces an elliptical scatter of associated artifacts and features. There may, however, be instances in which the linear site model would yield more accurate results. As might be expected, the probability of intersecting a site decreases as sites become more linear and increases as sites become more nearly circular; thus, assuming an elliptical site shape may lead to overestimation of the probability of intersection, if a large number of the sites are actually linear, rather than elliptical or round.

Application of the Formulas

The use of these formulas can be illustrated by several hypothetical examples.

Suppose that prior surveys in an area have indicated that sites with buried components are typically round or elliptical in their surface manifestation and cover at least 50 sq. m. Following this, the minimum site size we are interested in locating is 10 \times 5 m. The survey proposal calls for parallel transects spaced 20 m apart. The probability that all sites of this size or larger will be discovered is calculated as:

\[ p = \frac{2\sqrt{(a^2 + b^2)/2}}{d} = \frac{2\sqrt{(10^2 + 5^2)/2}}{20} = .79. \]

The minimum site size chosen, of course, depends on the goals of the survey; for example if the survey is intended to find only village sites, a very large \( a \) and \( b \) (or \( r \)) may be used. On the other hand, if historical or archaeological data indicate that very small and/or linear sites can be expected in an area, the variables and formulas can be adjusted accordingly. For example, suppose one of the goals of the survey is to determine the types of split-rail fences used by early settlers in an area. Short, linear fence fragments are the expected surface remains of this site type. A length of 5 m is arbitrarily chosen as the minimum size of these linear features. Retaining the 20 m transect interval, the formula for linear targets yields:
\[ P = \frac{2l}{\pi d} = \frac{2(5)}{\pi(20)} = .16. \]

Whether this relatively low probability of discovery is acceptable depends on the preservation goals of the particular contracting agency involved. In this case, a more intense survey of part of the project area, or alternative site location methods, such as use of old township maps, might be called for.

The formulas make no assumptions regarding the expected distribution of sites (clustered, patterned, or random), making them ideal for CRM surveys where this is an unknown. The method thus avoids the tendency for archaeologists to use wider transects in areas where they “know” no sites will be found. Such a self-fulfilling prophecy obviously contributes nothing to a real knowledge of site distribution within an area (Ebert 1988; Plog, Weide, and Stewart 1977; Shott 1985).

The phrase “100% coverage” implies a 100% probability that all sites of a significant size were found by a survey. For this to be true, both common sense and the formulas presented above tell us that transects must be spaced 20 m (or less) apart to discover all sites 20 m in diameter or larger, assuming the sites are readily detectable:

\[ P = \frac{2r}{d} = \frac{10(2)}{20} = 1. \]

Any survey report that claims 100% coverage was attained using transects 150 m apart implies that the only sites of importance to the survey are those 150 m in diameter or larger! Cultural resource management personnel reviewing such reports can calculate the probability that smaller sites, say, 20 × 15 m, were found, as discussed above.

\[ P = \frac{2 \sqrt{(a^2 + b^2)/2}}{d} = \frac{2 \sqrt{(20^2 + 15^2)/2}}{150} = .24. \]

This probability can then be compared to a standard, whether 100% or lower, in order to judge the adequacy of the survey strategy. This method may be especially useful in determining whether an area should be considered for CRM purposes to have been surveyed. For example, an agency charged with monitoring compliance with cultural resource laws and regulations may decide to consider as surveyed only those areas in which surveys probably discovered (i.e., intersected) at least 90% of all sites 20 m in diameter or larger.

The method can also be applied in advance of the fieldwork by using the same formulas. For example, if the survey is to attain a 90% probability of finding all sites larger than 10 m in diameter, we substitute for

\[ P = \frac{2r}{d} = .90 = \frac{2(5)}{d} \]

\[ = d(.90) = 10 \]

\[ = d = \frac{10}{.90} = 11.11. \]

Thus, an adequate survey plan will call for a maximum transect interval of 11.11 m. Wider transects would not be adequate to attain the stated level of coverage.

Just as importantly, narrower transects would be a waste of time and money given the stated objectives of the survey design. These equations assume that the target is smaller than the spacing between lines. When this assumption is not met—when the minimum site size is larger than the transect interval—the formulas yield “probabilities” greater than 1.0. This indicates that the transects are so closely spaced that each site is likely to be intersected more than once. For example, in trying to relocate the remains of a historic house foundation, one project used a shovel-test transect interval of 3 m. Applying the formula for a 4 × 3 m elliptical target, the expected dimensions of the foundation, the probability of intersection is calculated as:

\[ P = \frac{2 \sqrt{(a^2 + b^2)/2}}{d} = \frac{2 \sqrt{(4^2 + 3^2)/2}}{3} = 2.36 \]

Assuming that the site would have been visible at any point of intersection, this indicates that about twice as many transects were used as were necessary to locate the site. In some situations, this level of survey intensity might be appropriate—for example, if site visibility was expected to be low or if locating the site was of extreme importance to preservation plans for the area. In other situations, a more efficient use of survey resources might be needed.

**Intersection versus Discovery**

This method for evaluating survey plans makes two important assumptions, which must be assessed on the basis of the individual project situation. The first assumption is that the geometric relationship between the axes of the targets (i.e., sites) and those of the parallel transects is random. If for some reason, this assumption is violated, the estimates of site intersection will be either too high or too low. For example, a survey of old rail lines or trails might tend to miss sites if the transects were placed parallel to the edge of stream valleys or terraces, following the natural contours of the land. On the other hand, such a survey might require fewer transects than the formulas would indicate to achieve adequate coverage, if the transects were placed so as to cross-cut natural stratigraphy. In general, however, this assumption will not present problems for most CRM surveys; special cases can be dealt
with by altering the orientation of survey transects to achieve a more random pattern with respect to topography, natural resources, or other relevant factors.

The second, and more troublesome, assumption of the method proposed here is that intersection is equated with discovery. This can be expressed two ways. The first way of expressing this assumption is that sites are visible at the point at which the transect intersects them. This means that artifacts or features are spaced densely enough that the visual inspection or shovel test will detect them anywhere in the site. (For a discussion of how artifact density affects site visibility see Lynch 1980; Shott 1985; Krakker, Shott, and Welch 1983; Lightfoot 1986, 1989; Dunnell 1988; Nance and Ball 1989; and Kintigh 1988). The formula, conversely, does not take into account the possibility that in unvegetated areas sites may be visible for several meters beyond the actual transect lines.

The second expression of the assumption that intersection equals discovery is that survey techniques employed are adequate to detect any site actually crossed by a transect. If some sites are not exposed on the surface, shovel testing and/or trenching must augment surface reconnaissance. Such tests must be spaced no farther apart than the transect interval and must be of sufficient depth and size to intersect the visible manifestations of the site (artifacts or features). If the distribution of artifacts and/or features is not continuous, the test interval must be shortened accordingly. Determining whether subsurface testing is necessary, and, if so, the necessary depth of the excavations, relies on an understanding of the geomorphology and sedimentology of the survey area, as well as the patterns of feature and artifact distribution typical of sites in the project area.

Obviously, these conditions—high site visibility and highly-productive site detection methods—are violated more often than not in the real world of site and artifact distribution. For this reason, the results of this method should be considered estimates only. The results of the formulas presented above are the probabilities that sites of a given dimension were found given an ideal (i.e., continuous and highly visible) distribution of artifacts or features. Standards set for CRM surveys must take this factor into account, as real-world site discovery probabilities will almost always be lower than the probabilistic estimates derived from these formulas. (The opposite may also hold true in the rarer case of eroded areas with sparse vegetation.) As Redman (1987: 249) so aptly stated, “successful application of currently accepted scientific approaches to field archaeology requires the constant intervention of thinking human beings.”

The method for evaluating surveys presented here may contribute to the over-emphasis on “sites”—that is, dense clusters of artifacts—at the expense of isolated finds and small clusters. The significance of the “intersite” archaeological record has been articulated by many recent researchers (Dunnell and Dancy 1983; Dunnell 1988; Ebert 1988; Talmage, Chesler, and Stoff 1977; Thomas 1975). If cultural resource managers use artifact density and site size as the sole criteria of significance, much significant archaeological material will be lost, perhaps to the permanent detriment of the discipline (Ebert 1988).

The problem of recognizing the intersite archaeological record is beyond the scope of this paper. Nevertheless, by implicitly recognizing that survey methods can contribute to an incomplete or distorted view of the archaeological record, the method proposed here may help CRM personnel sort out some of the complexities of both intersite and intrasite data. Certainly, if a survey is not adequate to recognize a large target (site), it will not yield an accurate picture of the distribution of small targets (isolates and small clusters), either. While the formulas presented here cannot solve the problem of interpretation of various kinds of surface data, they can at least allow reviewers to estimate the probability that various kinds of data were collected during a given field project.

Those CRM personnel who require more exact estimation of survey strategy effectiveness are referred to recent articles by Kintigh (1988); Krakker, Shott, and Welch 1983; Nance and Ball (1986, 1989); Shott (1985, 1989). The present study is presented as something of a compromise between the complex models presented in recent CRM literature (e.g., Kintigh 1988) and the personal intuition upon which CRM project evaluations too often rely. It is hoped that application of this survey evaluation method can remove some of the guesswork from management of our cultural resources and help ensure their protection for future generations.

In closing, it should be reiterated that the formulas presented here are intended only as a tool to help CRM personnel make decisions concerning survey proposals and evaluations. The estimates they yield are merely rough guidelines, which in every case must be adjusted to the particular situation at hand, taking into account the vegetation, geology, physiology, cultural history, preservation goals, and level of archaeological knowledge of the specific project area. (To avoid the circular reasoning of some recent CRM surveys, environmental factors should be viewed in terms of their effects on site visibility, and not as “predictive” factors [Ebert 1988].) Unfortunately, there is no “easy” formula to account for the complex combinations of these variables faced by real-world archaeologists. The method proposed here may, however, provide
them with a starting point, from which fair and informed evaluations can be developed.

Linea Sundstrom received her Ph.D. from the University of Kansas in 1989. She has been involved in numerous archaeological projects in the Great Plains region and has published three books and several articles on Plains archaeology and rock art. She is currently serving as a consulting archaeologist and is conducting research on Plains Indian ethnohistory.

Mailing address: 1320 East Lake Bluff Boulevard, Milwaukee, WI 53211.


Talmage, V., O. Chesler, and J.A. Stoff

Thomas, David Hurst

Wandsnider, LuAnn, and James I. Ebert, eds.