

1988.132
Lit Bul
L5F

REGIONAL SURVEYS IN THE EASTERN UNITED STATES: THE STRENGTHS AND WEAKNESSES OF IMPLEMENTING SUBSURFACE TESTING PROGRAMS

Kent G. Lightfoot

The probability of detecting sites using subsurface testing programs is a serious concern for archaeologists working in the eastern United States. Some have suggested that current test-probe programs provide a poor method for estimating the frequency and distribution of sites. In this article I examine the usefulness of subsurface testing programs by comparing the results of an Eastern subsurface survey with a pedestrian surface survey conducted in the Southwest. The subsurface survey at Shelter Island, New York, was designed so that probability limits could be calculated for detecting sites of varying sizes. These probabilities were then employed to estimate the number and kinds of sites contained in sample units. When these results were compared with those from a pedestrian surface survey in northeastern Arizona, the results suggested that carefully designed subsurface surveys, although extremely labor-intensive, can provide settlement-pattern information as detailed as that collected in surface surveys.

An important consideration in contemporary archaeological survey work is the probability of detecting cultural remains. Recent overviews of archaeological survey methods (McManamon 1984a; Nance 1983; S. Plog et al. 1978; Schiffer et al. 1978) remind us that the probability of detecting archaeological remains varies substantially depending upon the discovery techniques employed and the nature of the cultural remains being investigated. The probability of detecting cultural remains of various sizes and compositions has important implications for estimating regional characteristics from archaeological survey data, for if inadequate discovery techniques are employed, then discovery biases will skew estimates on the number, sizes, and spatial distribution of remains found in sample units. The error will be compounded if the sample units were selected as part of a probabilistic sampling design in which the characteristics of the regional settlement pattern will be derived only from a sample of the units found in the study area.

The survey methods employed in the Eastern Woodlands are currently the subject of controversy. Although significant advances are being made in the discovery techniques used to find small buried sites (see, e.g., McManamon 1984a:244–262), Nance (1980), Wobst (1983), and others question whether current survey methods provide sufficient information to reconstruct regional settlement patterns. These critiques raise two important questions. First, are the discovery techniques employed by eastern archaeologists useful for estimating the frequency and kinds of sites existing within sample units? And second, can surveys conducted in the eastern United States reveal sites as effectively as can those projects undertaken in sparsely forested regions such as the American Southwest, where considerable success has been reported in the delineation of regional settlement patterns?

In this article I address these questions by comparing the results of two survey projects, one conducted in the American Southwest (northern Arizona) and the other in the American Northeast (eastern Long Island, New York). The comparison is instructive because I directed both projects and employed a similar survey design in each area, yet used different sets of discovery techniques—one designed to detect archaeological remains visible on the modern ground surface, and the other to reveal buried remains. Specifically, I examine whether *subsurface* testing programs can provide

Kent G. Lightfoot, Department of Anthropology, SUNY at Stony Brook, Stony Brook, NY 11794

American Antiquity, 51(3), 1986, pp. 484–504.
Copyright © 1986 by the Society for American Archaeology

estimates on the frequency and distribution of sites found within sample units that are comparable to those generated by conventional *surface* surveys.

THE SURFACE SURVEY

The selection of an appropriate discovery technique should be based on a number of factors (for detailed discussions see McManamon [1984a:242–250] and Schiffer et al. [1978:3–10]). One important consideration is the physical characteristics of the local environment, a variable that influences the visibility of archaeological remains. Visibility, directly related to the density of vegetation and the rate at which soil and humus accumulate over sites, is highest in arid areas of sparse vegetation, such as the American Southwest, where sites are often detectable on the modern ground surface with the naked eye. Where these optimal conditions prevail, the most common discovery procedure is the pedestrian surface survey, a technique that involves the systematic inspection of the ground surface at a given level of intensity.¹ Crew members spaced at regular intervals walk back and forth across the sample unit, noting artifacts, irregular topographic features (mounds, depressions), and unusual vegetation patterns, any of which may reflect the location of past human activities (Mueller 1974:10).

An important if obvious assumption underlying the pedestrian surface survey is that a representative sample of the target population of cultural remains will be detectable on the ground surface. Every category of the remains chosen for study should have an equal chance of producing surface manifestations readily detectable by archaeological crews; otherwise, one faces the problem of discovery bias, in which the most visible and accessible resources tend to be overrepresented at the expense of other, less conspicuous remains (Schiffer et al. 1978:6–7).

Surface Survey: An Example from Northeastern Arizona

The pedestrian surface survey is widely employed in the American Southwest where, given optimal environmental conditions, most archaeologists assume (either implicitly or explicitly) that most archaeological remains can be detected by careful inspection of the ground surface. I directed one such survey near the modern town of Pinedale, Arizona (Figure 1).² The survey was performed as part of a contract project designed to mitigate the impact of a proposed timber sale on the Apache Sitgreaves National Forests. The contract, which required that we locate cultural resources that might be adversely affected by logging activities, provided us with the opportunity to explore several research issues concerning the economic and sociopolitical organization of past Pueblo communities—issues that are considered in detail elsewhere (Lightfoot 1984a).

The fieldwork, which resulted in a 100% survey of a 41 km² area, began with the division of the study area into small, manageable sample units, each of which was surveyed by a crew of three or four people who walked contiguous transects back and forth between the boundaries of the unit. In an attempt to locate, inventory, and study archaeological sites of varying sizes and contents, we inspected each transect very carefully with crew members spaced 10 m from one another, so that a 60 to 80 m swath was covered on each transect. To ensure total coverage of the area, all transect boundaries were flagged with biodegradable paper. When artifacts—typically consisting of sherds and chipped or ground stone tools—were observed, crew members reported their locations and specifications.

Some artifact distributions appeared to represent non-site manifestations (S. Plog et al. 1978: 388–389; Thomas 1975), i.e., sparse concentrations of cultural material with broad spatial dispersion. At Pinedale, non-site manifestations were characterized by a dispersal of one or two artifacts per 50–100 m² area surveyed. Other cultural materials were found in definable clusters that appeared to represent discrete loci of more spatially restricted activity areas. Here a concentration of artifacts, often characterized by an artifact density of one or more per square meter, as well as occasional architectural features were delimited. These places were defined as sites. Differentiating site boundaries from more amorphous non-site manifestations necessitated difficult field decisions in some cases. We defined sites as all loci containing cultural materials of sufficient quantity and quality to

provide a good potential for interpreting the range of activities that once occurred there (see also S. Plog et al. [1978:389]).

The intensive survey of the project area yielded information on 152 site locations. Of these, 143 were prehistoric and 9 historic, the latter consisting mostly of the remains of logging camps and Mormon homesteads. Prehistoric sites ranged from special-purpose camps and one-to-two-room Pueblo units to large Pueblo settlements ranging in size from 40 to 200+ rooms. Large non-site manifestations, covering more than a 2 km² area, were also defined in the northern section of the project area, and appear to represent extensive resource zones utilized for hunting and gathering purposes. A full discussion of the spatial and temporal characteristics of the prehistoric Pueblo communities is presented elsewhere (Lightfoot 1984a).

THE SUBSURFACE SURVEY

In areas *not* characterized by sparse vegetation and minimal soil development, the use of the pedestrian surface survey is problematical. The Eastern Woodlands, where dense undergrowth obstructs a clear view of the modern ground surface and where leaf litter and topsoil can bury cultural remains rather quickly, is one such area. Here one cannot assume that all archaeological remains have an equal probability of producing detectable surface manifestations, a problem that may result in discovery biases in the kinds of sites observed in surface surveys (Alexander 1983; Birk and George 1976:3; Lovis 1976:364). Consequently, considerable energy has been devoted to developing discovery techniques that enable archaeologists to detect low visibility sites. Subsurface testing procedures designed specifically to reveal buried sites include test probes (shovel test units, soil cores, auger holes) of varying sizes and shapes (see McManamon [1984a:253–262] for a detailed discussion), the employment of chemical soil tests (Baker 1975:67–81; Casjens et al. 1978:91; Eidt 1973), and still-experimental geophysical survey techniques such as the use of resistivity devices (Carr 1977) and proton magnetometers (Aitken 1974; Tite 1972). Because the latter two techniques are most effective for intra-site exploration (see McManamon 1984a:247, 249), most regional subsurface surveys rely primarily on shovel probes and/or soil cores and augers.

In most respects, the designs of subsurface surveys are similar to those employed in surface surveys. Once the boundary of the study area is defined, the area is usually divided into smaller sample units. In most cases, these are arbitrarily defined spatial units whose size and shape (quadrats, transects, etc.) vary from project to project. Given the prohibitive labor costs of surveying an entire study area, typically only a sample of the total population of sample units is selected for field investigation (McManamon 1984b:26–27). This sampling process involves such considerations as choosing an appropriate sampling strategy (judgmental, random, stratified, systematic), defining the number of sample units to be surveyed (sample size), and defining the percentage of the study area to be surveyed (sample fraction) (see, e.g., Cable et al. 1977; Feder 1981; Lovis 1976; McBride and Dewar 1981; McManamon 1984b; Nance 1980).

The major difference between surface and subsurface surveys is the method of inspecting the sample units once they have been selected for study. Unlike surface work, in which the detection of archaeological remains is limited to a two-dimensional plane, subsurface surveys involve the inspection of three-dimensional space. In most cases the sample units are too large to excavate completely, and a testing program involving the placement of test probes across a fraction of the sample unit must be implemented. Viewed from this perspective, the space represented within a sample unit is being subsampled by a number of smaller test probes (Nance 1980:175).

The use of sample units raises the issue of “element” versus “cluster” sampling. There is currently some confusion about whether the subsampling of sample units represents element or cluster sampling, an important distinction because the statistical manipulations of the data produced by these two methods vary. In the clearest discussion of the problem to date, Nance (1983:297–298) points out that choosing either one or the other should depend upon the research objectives of the subsurface survey and, more specifically, upon the “elements” being studied—elements being the basic entities about which information is sought (see Mueller 1975:37). “Elements” may be sites, features, artifacts, sample units, or test probes, they may also be specific properties of sites, features, or artifacts located

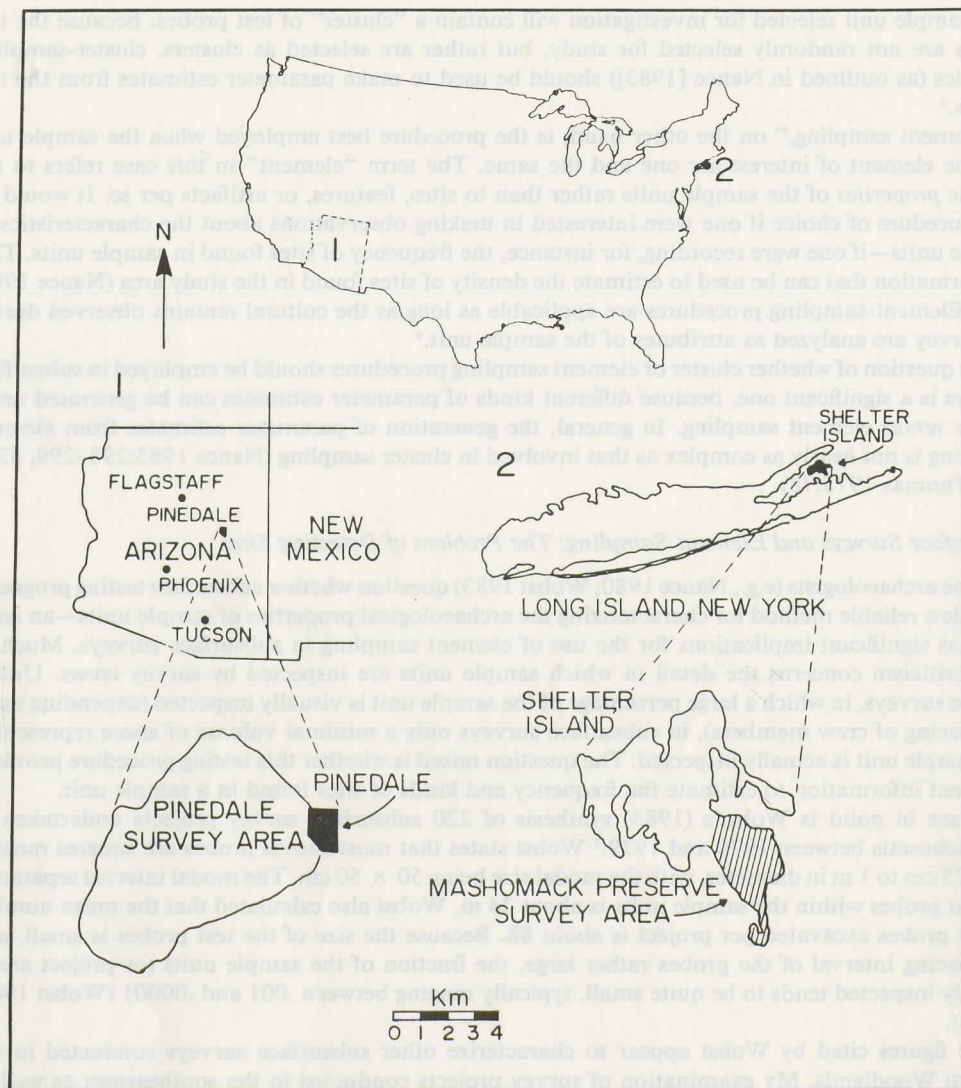


Figure 1. The location of the Pinedale and Mashomack Preserve Survey Areas.

within a larger survey unit. Depending upon which of these is under investigation, the subsurface surveyor should choose to pursue either cluster sampling or element sampling.

Cluster sampling should be pursued when the sample units contain collections, or "clusters," of elements. For example, one may be interested in comparing sites of varying sizes and contents found within a study area. Potentially, each sample unit can of course contain more than one element of the population of interest (Nance 1983:297). If the survey design involves the random selection of sample units rather than of elements per se, then elements are sampled from the study area as *clusters* rather than as independent entities (Thomas 1975:78).

Cluster sampling is also applicable when the *properties* of test probes are the element of interest. Here the cultural remains recovered during the survey are viewed as attributes of the test probes. One may be interested in estimating the artifact *frequency* for test probes, or in calculating the *density* of artifacts per unit of volume tested. In either case, cluster sampling is appropriate, because

each sample unit selected for investigation will contain a "cluster" of test probes. Because the test probes are not randomly selected for study, but rather are selected as clusters, cluster-sampling statistics (as outlined in Nance [1983]) should be used to make parameter estimates from the test probes.³

"Element sampling," on the other hand, is the procedure best employed when the sample unit and the element of interest are one and the same. The term "element" in this case refers to the specific *properties* of the sample units rather than to sites, features, or artifacts per se. It would be the procedure of choice if one were interested in making observations about the characteristics of sample units—if one were recording, for instance, the frequency of sites found in sample units. This is information that can be used to estimate the density of sites found in the study area (Nance 1983: 297). Element-sampling procedures are applicable as long as the cultural remains observed during the survey are analyzed as attributes of the sample unit.⁴

The question of whether cluster or element sampling procedures should be employed in subsurface surveys is a significant one, because different kinds of parameter estimates can be generated using cluster *versus* element sampling. In general, the generation of parameter estimates from element sampling is not nearly as complex as that involved in cluster sampling (Nance 1983:298–299, 320–326; Thomas 1975:78).

Subsurface Surveys and Element Sampling: The Problem of Detecting Sites

Some archaeologists (e.g., Nance 1980; Wobst 1983) question whether subsurface testing programs provide a reliable method for characterizing the archaeological properties of sample units—an issue that has significant implications for the use of element sampling in subsurface surveys. Much of their criticism concerns the detail in which sample units are inspected by survey crews. Unlike surface surveys, in which a large percentage of the sample unit is visually inspected (depending upon the spacing of crew members), in subsurface surveys only a minimal volume of space represented in a sample unit is actually inspected. The question raised is whether this testing procedure provides sufficient information to estimate the frequency and kinds of sites found in a sample unit.

A case in point is Wobst's (1983) synthesis of 220 subsurface survey projects undertaken in Massachusetts between 1970 and 1979.⁵ Wobst states that most shovel probes are squares ranging from 25 cm to 1 m in diameter, with the modal size being 50 × 50 cm. The modal interval separating the test probes within the sample units is about 24 m. Wobst also calculated that the mean number of test probes excavated per project is about 88. Because the size of the test probes is small, and the spacing interval of the probes rather large, the fraction of the sample units (or project areas) actually inspected tends to be quite small, typically ranging between .001 and .00001 (Wobst 1983: 60–61).

The figures cited by Wobst appear to characterize other subsurface surveys conducted in the Eastern Woodlands. My examination of survey projects conducted in the southeastern as well as the northeastern United States indicates that most shovel probes range in size from 30 × 30 cm to 1 × 1 m squares. Most shovel probes are placed either along a single transect or square grid system, with the interval separating probes ranging from 5 m to 100 m. In most cases the volume of space inspected for either the sample unit or the entire project area is minimal (Birk and Douglas 1976:3; Cable et al. 1977:5; Feder 1981:139; House and Ballenger 1976:44; McBride and Dewar 1981:40; McManamon 1978:81, 1981:199–203; Nance 1980:172; Warner 1978:69).

The minimal volume of space inspected within sample units has prompted Wobst, Nance, and others to point out the following three problems.

1. *Site Size Bias.* Nance (1979:173–174) argues that subsurface testing is inherently biased toward the discovery of sites characterized by large artifact scatters. That is, because one is sampling three-dimensional space, then the greater the volume and spatial extent of a site, the better the probability that at least one test probe will intersect the boundaries of the site. Nance suggests, consequently, that most subsurface survey designs tend to inflate estimates of the frequency of very large sites, at the same time underestimating the frequency of smaller sites found in a study area.

The problem is exacerbated by the nature of the archaeological remains found in many areas of

the Eastern Woodlands. Wobst (1983:68–69) points out that in comparison to the American Southwest, where long-standing sedentary societies produced relatively permanent structures and facilities, the Northeast is characterized by a rather ephemeral settlement pattern. Here populations practiced a variety of residential strategies that often produced minimal archaeological remains characteristically found in low-density scatters. Such remains can be difficult to detect, even when test probes are placed directly over the sites. This is especially true when soil augers (diameters 7–15 cm) and soil cores (diameters ca. 3 cm), which sample a relatively small volume of soil (see McManamon 1978:82, 1981:204), are used, but it also pertains to shovel probes (see Lynch 1980; Stone 1981; Wobst 1983:68–70). As the size of the shovel probe increases, so does the probability of locating artifactual material from low-density scatters (see Krakker et al. 1983:476; Stone 1981:182); but even large probes placed over known sites have occasionally produced no artifactual material. Recent evaluations of shovel probes indicate that, on the average, 23% to 40% of the test units placed over known sites may not produce any artifactual material (see McManamon 1984a:260; Nance and Ball 1986).

2. *Density Estimates.* Another problem concerns the ability of archaeologists to estimate the frequency and spatial distribution of sites found within sample units. This problem has implications for the use of element sampling designs employed to generate regional site densities and settlement patterns from sample unit data. Wobst (1983:71) argues that generalizations based on small-size test probes 20 or more meters apart cannot be “sufficiently accurate . . . to be useful.” Given the minimal volume of space typically inspected within sample units, he believes that subsurface surveys are inadequate for generating parameter estimates concerning regional site distributions and site densities. Both Nance (1979, 1980, 1983:319–320) and Wobst (1983) are skeptical about the use of element sampling designs intended to delineate the regional settlement pattern of sites based on information collected from the subsurface testing of large sample units.

3. *Labor Intensity.* One possible solution to this dilemma is to increase the size of the test probes or to decrease the intervals between them (thereby raising the intensity at which a sample unit is inspected). However, the benefits of increasing the survey intensity are not always clear-cut; while the probability of detecting cultural remains will be enhanced, the costs of undertaking the survey work will increase dramatically. Thus, while increasing the intensity of inspection may enhance our ability to make parameter estimates, it is not clear, in many cases, whether the information return per unit of labor is practical (Krakker et al. 1983:469; McManamon 1981:199). Subsurface testing programs, when compared to surface surveys, are extremely labor-intensive. Although few comparisons exist between surface and subsurface surveys (Claassen and Spears 1975:126–127; McManamon 1978:81), it is generally recognized that test probes represent a major expense of both time and labor (see McManamon 1984a:262–268). The interval between test probes may be shortened from 24 m to 10 m or 5 m or even 2 m, but the amount of work involved in surveying a unit of land increases astronomically—to the point where the costs become prohibitive for most archaeological projects.

Discovery Probabilities: Correcting for Archaeological Biases

A major issue underlying the use of subsurface testing programs is whether the archaeological biases outlined above can be minimized or even controlled. Clearly, the excavation of more test probes per sample unit is not a very practicable solution in most cases. An alternative solution is to calculate discovery probabilities that correct for the biases associated with subsurface testing. By using discovery probabilities outlined by Krakker et al. (1983), Nance and Ball (1986), and McManamon (1984a:274–276), one can make statistical adjustments that provide a more reasonable data set for estimating the frequency and distribution of sites found in a sample unit. Such a practice will facilitate the use of element sampling designs, since the investigator can define, with a known degree of confidence, the characteristics of the archaeological remains found in a sample unit. This information will, in turn, provide a more solid data base for estimating the regional parameters of the study area based on a sample population of survey units.

Although estimating regional population parameters from a sample is an important component

of contemporary survey work, this issue is not specifically addressed here because other archaeologists have described probabilistic sampling in detail (e.g., Judge et al. 1975; Mueller 1974; Plog 1976). Instead, I address the use of discovery probabilities to refine estimates of the frequency and distribution of sites present in sample units.

To make statistical adjustments based on discovery probabilities, the investigator must consider two related aspects of his or her subsurface testing program. First, the investigator must determine what kinds of cultural remains are to be detected in a sample unit. This decision involves a consideration of the physical properties of the archaeological remains (see McManamon 1984a:226–234). If the object of the survey is to define the density of sites found in sample units, then one must delineate the physical properties (site size, artifact density) of those sites that the investigator wishes to detect with some confidence. What, for example, is the smallest site with a given artifact density that the investigator wishes to discover with subsurface testing?

Second, given a target population of sites, the investigator must weigh factors of labor intensity against acceptable discovery probabilities and design a subsurface testing program accordingly. As outlined below (and in Krakker et al. [1983] and Nance and Ball [1986]), there is a direct correlation between the size and spacing of test probes and the ability to detect, with a given degree of confidence, variously sized sites found in a sample unit. The greater the survey intensity, the better the chances of detecting small buried sites. Thus, one can define the specific probability of discovering a site of a given size and artifact density by altering the spacing and size of test probes within the sample unit. There is, however, clearly a trade-off required between the chances of detecting sites of various sizes and the costs of undertaking the survey. Ultimately, given a specified target population of sites, the investigator must determine the most efficient discovery probabilities, based on a realistic level of survey intensity. Given budget considerations, what discovery probabilities are acceptable? Does the investigator intend to find every site of a specified size and content in the sample unit, or is he or she willing to find, on the average, 1 out of 10 such sites, or only 1 out of 100 such sites?

Subsurface Survey: An Example from Long Island

The factors outlined above were considered in designing a subsurface testing program for eastern Long Island, New York. The research goal addressed by the subsurface survey was a very basic one: to define the spatial and temporal parameters of prehistoric Long Island settlement patterns. A review of the published literature indicates that more than 80% of the sites described for coastal New York are shell middens (Lightfoot et al. 1985), a class of sites characterized by the presence of Virginia oyster (*Crassostrea virginica*), hard clam (*Mercenaria mercenaria*), soft clam (*Mya arenaria*), and bay scallop (*Pecten irradians*), occasional pit features, and assorted artifacts. Salwen (1970:3–4), Wyatt (1977:408), and Saxon (1975:1–3) have all questioned whether this high proportion of shell midden sites is a product of field investigation bias, rather than a true reflection of their abundance in the archaeological record. These authors point out that shell middens are among the largest and most visible prehistoric cultural resources in coastal New York, a fact that might have facilitated their discovery. Despite the destruction of the largest middens by farmers and later housing developments (see Ceci 1982:23), in some undeveloped areas of Long Island one can still observe midden deposits along productive estuaries. Some of these surficial scatters are quite extensive, with the white shell debris standing out in vivid contrast to the surrounding sand/gravel matrix.

Whether or not shell middens have received preferential treatment by archaeologists because of their high visibility is a real concern. In reviewing the published literature, we found that more than 90% of the published sites were discovered by construction workers building houses, roads, and sewers, or by individuals undertaking surface reconnaissance work (Lightfoot et al. 1985). Considering the problem of leaf litter and soil accumulation, as well as the dense vegetation in many areas of Long Island, one doubts that such methods have produced a representative picture of the regional aboriginal settlement pattern; small, buried sites have probably been overlooked at the expense of large shell middens. Given this problem, a number of questions are suggested. Are shell middens the primary sites found on Long Island, or has the almost exclusive consideration of midden sites

perpetuated a skewed perspective of the range of activities and settlement types that date to the prehistoric period? Are other kinds of settlements, which have yet to be widely recognized by area archaeologists, extensively represented on the island? If other settlement types do exist, then what kinds of activities are represented on these sites? Finally, how are archaeological resources, including shell middens, spatially distributed across the coastal and interior landscape, and how did the articulation of these settlement systems change over time?

The Department of Anthropology, SUNY/Stony Brook, recently initiated a regional survey project designed to gather the kinds of settlement information needed to address these questions. The project incorporates an intensive subsurface testing program specifically developed to detect small, buried sites. Our ultimate goal is to delineate the prehistoric settlement pattern of the interior woodlands and the diverse coastal environment of eastern Long Island. The ideal strategy, therefore, is to place a number of survey blocks along the north shore, south shore, and interior sections of Long Island.

The first study area chosen for our survey project was Shelter Island, a small island located between the northern and southern forks of eastern Long Island (Figure 1). The Nature Conservancy, a nonprofit environmental organization, granted permission to undertake survey work on its Mashomack Preserve, a 2,039-acre tract of land encompassing a considerable stretch of coastline, including numerous salt marshes, estuaries, and coastal embayments, as well as hundreds of acres of interior oak, hickory, and beech forest. The area has been protected from commercial development, so that its archaeological resources have not been disturbed by recent construction projects.

The Survey Design: Stage One. We employed a two-stage survey design to begin the study of the Mashomack Preserve. The first stage consisted of dividing the study area into manageable sample units. Initially we planned, rather optimistically, to survey the entire Preserve; accordingly (as was done in the Pinedale survey) the area was partitioned into 15 survey blocks whose size and boundaries conform to existing roads, ecological conditions, and topographic features. However, it became apparent that the time and labor involved in surveying the Preserve completely were prohibitive, and our research design was modified accordingly. Each block was subdivided into 25-acre (10-hectare) sample units, and a sample of these units was then selected for intensive study. The sample units encompass sufficient space to facilitate the study of inter-site relationships and associations, which may provide important insights into the organizational parameters of prehistoric Long Island communities.

If a complex and extensive settlement system were defined in one area of the Preserve, additional sample units could be surveyed to increase the total area covered. At the same time, the survey of a 25-acre block can be completed by a single crew during a summer field school. We estimate that a four-member crew can complete a block in three to four weeks.

A stratified judgmental sampling strategy based upon ecological considerations was employed to select sample units for intensive study. The study area was initially stratified into two zones representing the coastal strip and the interior woodlands. The coastal strip, in turn, was divided into two zones based on the density of vegetation, especially cat briar (*Smilax rotundifolia*), which produces a thick ground cover and seriously impedes the progress of survey crews. (The interior woodland is characterized by relatively open ground cover.) In consultation with the Nature Conservancy, we selected from each stratum sample units that characterized the ecological diversity of the Preserve. Three units were selected from the interior strata of the island (A, former agricultural fields; B and E, oak-hickory woodland) and three from the coastal strata (C, estuary, moderate underbrush; D, estuary, dense cat briar stands; F, coastal swamp, moderate underbrush) (see Figure 2). It was believed that by proceeding in this manner we could assess whether the density and settlement configuration of archaeological resources found along the coast of Shelter Island varied markedly from those of the interior. Furthermore, we would be able to determine the costs of undertaking survey work under diverse ecological conditions.

Subsurface Testing: Stage Two. Next, each of the six sample units selected was intensively surveyed, using a subsurface testing program. We designed our program to record spatial information on both isolated artifacts and sites. As in the Pinedale study, we wanted to identify both non-site manifestations, which may represent resource zones sporadically utilized by hunting and/or gathering parties, as well as more discretely occupied places, traditionally defined as sites. As already noted,

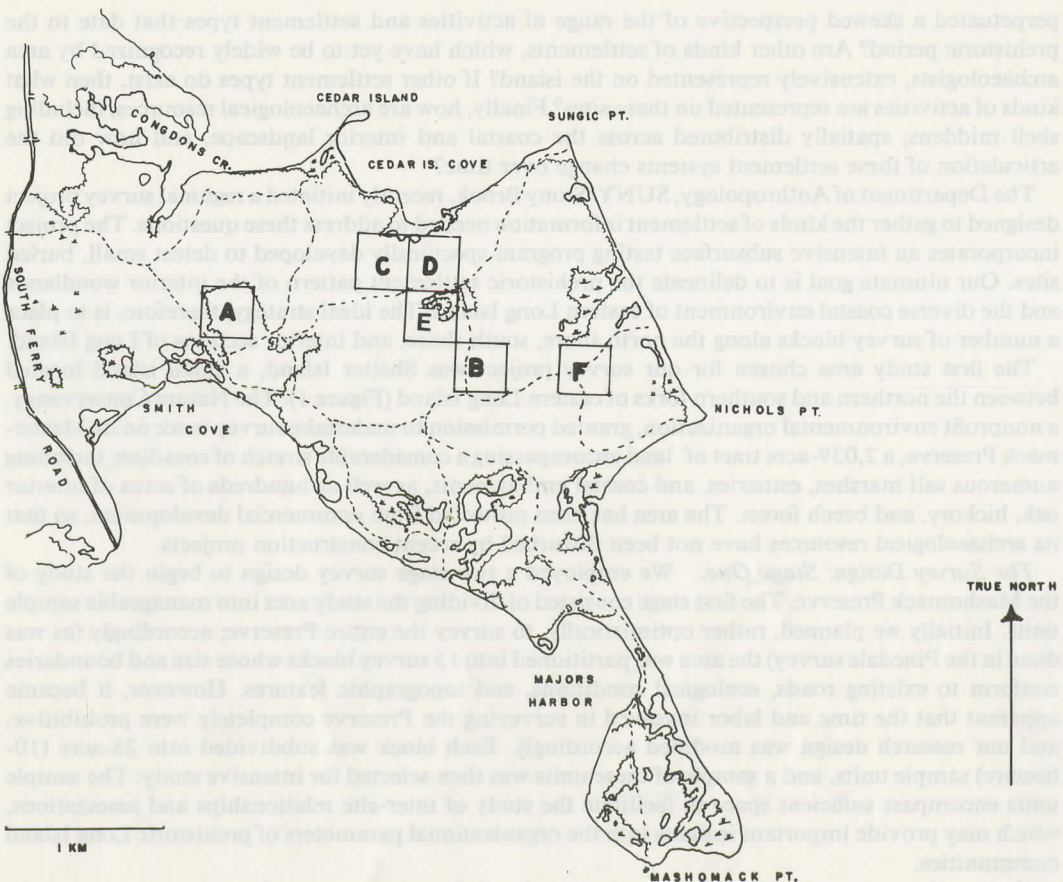


Figure 2. The sample units defined for the Mashomack Preserve.

the latter are expected to produce multiple artifacts in a specific place, which should provide enough information to make behavioral interpretations concerning site function and occupation duration.

We employed a systematic aligned sampling strategy characterized by shovel probes arranged along a square grid. The rationale for using this strategy as opposed to other available options was two-fold. First, such a design ensures an even distribution of test probes across the sample unit, thus facilitating the detection of large artifact scatters. Second, this sampling strategy allows the investigator to generate discovery probabilities for detecting sites of varying sizes and artifact densities.

Krakker et al. (1983) discuss in detail how discovery probability limits can be assigned when test probes are arranged in hexagonal or square grid patterns. Employing either pattern, the probability of encountering a site of x size and y artifact density is directly related to: 1) the interval separating the test probes, and 2) the size of the test probes. The probability limits outlined by Krakker et al. are based on a site characterized by a circular dispersal of artifacts, although sites of other shapes may also be considered. Theoretically, if the radius of a circular cluster of artifacts is greater than the midpoint of the square composed of any contiguous four test probes, then no matter how the site is placed in relation to the square grid, at least one test probe should pierce the site and lead to its discovery. The radius of the largest site that cannot escape detection is defined by:

$$r = \frac{i}{\sqrt{2}}$$

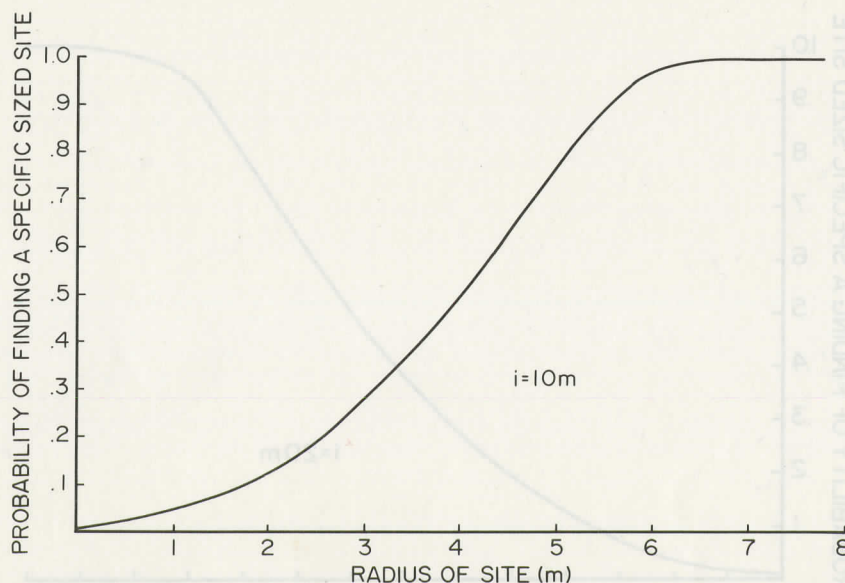


Figure 3. The probability of locating a specific sized site given a grid interval of 10 m.

where r = radius of the site, and i = interval spacing. In addition, one can estimate the probability of detecting sites smaller than the grid interval ($r < i/\sqrt{2}$). Assuming a random distribution of sites across the sample unit, then the probability of detecting a site of a specified size, following Krakker et al. (1983:471–472), is:

$$p = \pi r^2 / i^2.$$

Using this simple geometric formula, one can calculate the probability of encountering sites of various sizes, given a particular sampling interval.

The discovery of a site is dependent, of course, upon the recovery of at least one artifact from a test probe. However, even sites with relatively high artifact densities can be missed with subsurface probes. One means of addressing this problem and augmenting our recovery capability was to screen the contents of each probe through ¼-inch or finer wire mesh. Another method was to increase, within reason, the volume of dirt excavated from each test probe.

Still another means of addressing this problem would be to calculate probability limits for recovering one artifact per test probe given various artifact densities. The probability of recovering artifactual material is affected by: 1) the artifact density of the site, and 2) test probe size (see Krakker et al. 1983:476–477; McManamon 1984a:269–274; Stone 1981:516–517). We have begun to collect information on artifact densities employing different-sized test units (17 cm augers and 30 × 30 cm and 1 × 1 m shovel probes). From this information, we hope to estimate the range of artifact densities associated with Shelter Island sites. Eventually these density ranges will be used in calculating the probability of missing various classes of archaeological resources on the Preserve. We emphasize that the parameter estimates in this article should be viewed as conservative figures, because adjustments are not made for the probability of missing very low-density scatters.

After some experimentation in the field, we chose a sample unit interval of 10 m and a shovel probe size of 30 × 30 cm. Each sample unit was surveyed by a crew of four or five people, who walked contiguous transects back and forth between the unit boundaries. Crew members were spaced 10 m apart. Initially we excavated a shovel probe every 10 m, raking away leaf litter and vegetation (see Raab 1977) every 3 m to expose the ground surface. This practice was discontinued halfway through the field season, as it failed to reveal any cultural material. Test probes were excavated to sterile Pleistocene deposits, which varied in depth from 25 to 80 cm, although most ranged from

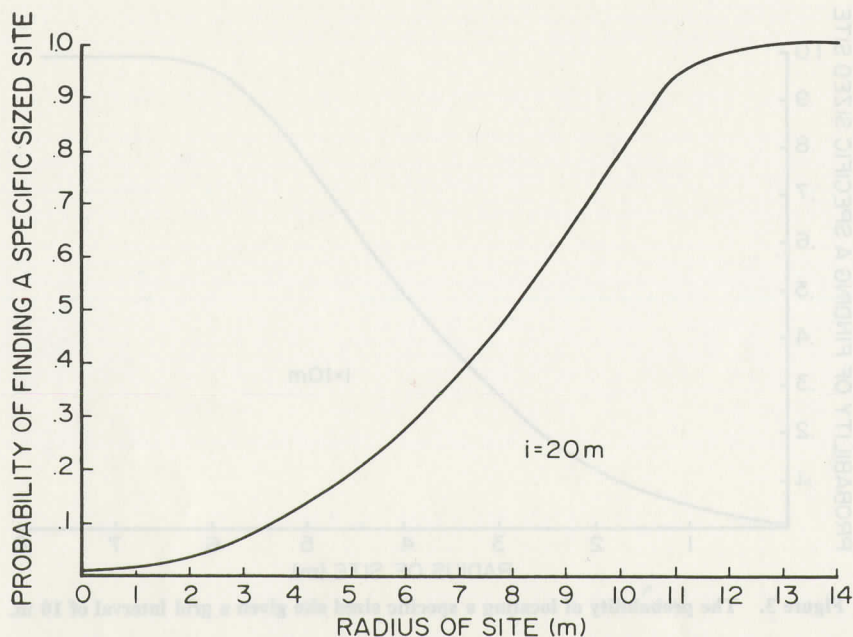


Figure 4. The probability of locating a specific sized site given a grid interval of 20 m.

30 to 50 cm. The excavation of shovel probes was greatly facilitated by the soil conditions of eastern Long Island; not only is the sandy soil relatively shallow and easy to dig, but the soil can be readily screened through $\frac{1}{4}$ -inch wire mesh. We found that crew members on an average day could excavate and screen the contents of 20 to 30 test probes, each measuring 30×30 cm.⁶

Using a grid interval of 10 m, we were relatively confident of our ability to detect circular sites with radii of 7 m or more, or square sites with sides measuring 10 m or more, which represent 154 and 100 m² areas, respectively. Although we would have preferred to locate smaller sites with a greater degree of confidence, we found the labor expenditure necessary to excavate more than one shovel probe per 10 m prohibitive. Employing the formula given above, we calculated the probability limits of discovering small sites, and these are presented in Figure 3. For example, given a site size of 50 m² (radius = 4 m), we expected to detect, on the average, one out of every two sites present in the sample unit. For a site size of 12 m² (radius = 2 m), we expected to detect, on the average, only one out of every eight sites.

Clearly, the probability of detecting small sites is closely related to the size of the grid interval. Had we employed a grid interval of 20 m (a fairly typical interval for eastern survey projects), we could have been confident of locating circular sites with radii of 14 m or more, or 615 m² in area. The probability of discovering sites smaller than this using the 20 m grid interval decreases rapidly; as depicted in Figure 4, there is only a one in eight chance of discovering a site of 50 m² (radius = 4 m). With a 30 m interval separating test units, we could be confident of locating only those sites with radii greater than 21 m and areas of 1,385 m² or larger. And the probability of locating small sites declines at a depressingly rapid rate, as illustrated in Figure 5, until there is only a 1 in 20 chance of discovering a site of 50 m² (radius = 4 m).

When artifactual material was found in a test probe, we then laid out an "iron cross"—a survey configuration that greatly enhanced our ability to differentiate isolated finds from sites. This procedure consisted of excavating test probes every 2 m in the four cardinal directions (cf. Chartkoff 1978; McManamon 1981a:205). If no further cultural material was located in the "iron cross," we defined the initial probe material as an isolated find. It was collected, its location noted on the topographic base map, and a brief description of the find recorded. By noting the spatial distribution

of such isolated finds, we were able to define several low-density scatters (non-site manifestations) characterized by one or two artifacts for every 100 to 500 m² covered by survey crews, i.e., one or two artifacts for approximately every 18 to 90 shovel probes excavated.

When cultural material was found in one or more axes of the "iron cross," we extended the testing procedure along those axes at 2 m intervals as far as the distribution of artifacts continued. In areas that we defined as sites, a concentration of artifacts and occasional features was delineated along the axes of the cross. As in the Pinedale survey, the sites appear to represent discrete loci of spatially restricted activity areas. By extending the "iron cross" in the direction(s) of the artifactual dispersal, we quickly determined the size and depth of the site, and collected a sample of artifacts from across the site area.

Employing these methods, we intensively surveyed a total of 235,618 m² (58.2 acres) during the 1983 field season. This work yielded a total of 20 sites (two historic and eighteen prehistoric), as well as a number of isolated finds. The survey was undertaken by a field crew of 20 people, who spent approximately eight days (160 person days) surveying the six sample units. Although three units (B, C, D) are close to completion, none of the sample units was actually finished. A major priority for future fieldwork is to complete these units, and—on the basis of our first season's results—to select several more sample units for intensive study.

SUBSURFACE AND SURFACE SURVEYS: A COMPARISON

My purpose here is not to discuss the specific results of the Shelter Island survey, which are presented in detail elsewhere (Lightfoot et al. 1985), but to evaluate the problems inherent in subsurface surveys. One expects, on the bases of Wobst's and Nance's critiques, that the results of a subsurface survey will produce skewed site densities and sizes that reflect the location of only a relatively few large sites. In comparing surface and subsurface surveys, one expects the range of site densities and sizes revealed by the two kinds of projects to vary considerably. To evaluate whether significant discovery biases had plagued the Shelter Island survey, I compared it with the results of the Pinedale project. If such biases do exist, then I expected (controlling for other factors to the extent possible) that the range of site sizes should be larger for Shelter Island, and that the site density should be considerably lower. At the same time, I expected there would be far greater labor invested per unit of area surveyed at Shelter Island. Below, I compare the two projects with respect to: 1) site-size bias, 2) density estimates, and 3) labor intensity.

Site-Size Bias

Does the Shelter Island survey over-emphasize the recovery of large sites? The results suggest that this is not the case. Figure 6 is a histogram of the prehistoric site sizes recorded for the survey; they range from 4 m² to 200 m² ($\bar{x}$ = 30 m², SD = 45.5 m²). Significantly, the majority (67%) of the sites are under 20 m² in size (radius = 2.5 m). In addition to these sites, we recovered numerous isolated artifacts, most of them quartz flakes and projectile points.

In comparison, the artifact scatters defined for the Pinedale survey and included in the settlement pattern analysis (see Lightfoot 1984a:130–134) range from 35 m² to 20,000 m² ($\bar{x}$ = 1,834 m², SD = 3,451 m²). When scatter size is examined as a histogram, three major classes of site size can be defined: 100 to 400 m², 600 to 800 m², and 1,000 m² or greater (Figure 7). Clearly, there is little overlap between the site sizes recorded at Shelter Island and those observed at Pinedale. Two factors may help to explain the discrepancy. First, the criteria employed to define sites at Pinedale may have thwarted the recognition of small scatters. This source of error, however, should be minimal, because the same criteria were employed to define sites in both projects; that is, a major goal of both was to locate sites of varying sizes, including small, limited-activity loci. The second and more important factor is that the archaeological remains found on Shelter Island and at Pinedale differ markedly in their variety and spatial extent. Scatters found at Pinedale are larger, possibly because they reflect different ways of utilizing space, different kinds of activities, and different occupation durations. Nevertheless, we did not expect that the Shelter Island survey would reveal a much smaller range of site sizes than the Pinedale survey.

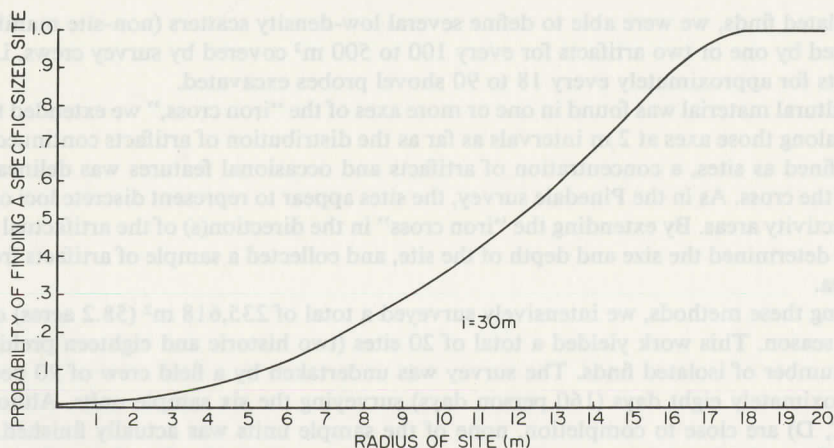


Figure 5. The probability of locating a specific sized site given a grid interval of 30 m.

Other intensive surface surveys conducted in the American Southwest with crew members spaced 10 m or less from one another have yielded results similar to those produced at Pinedale. The average size site recorded during the Chevelon project was 500 m² (see S. Plog et al. 1978:390)—a figure well above that recorded for Shelter Island. And even the Hay Hollow project, one of the most intensive surveys ever attempted in the Southwest (S. Plog et al. 1978:391–392), is characterized by a similar pattern. Although a number of small scatters was recorded, over half of the sites fall within the size classes of 15 to 75 m², 63 to 100 m², 100 to 1,000 m², and 900 m or greater (see Plog 1974:102–103).

Density Estimates

Are the results of our subsurface testing program useful for generating site-density estimates for Shelter Island? This is a difficult question to answer with any degree of certainty, because the total population of archaeological resources on the island remains unknown. Despite this difficulty, I strongly believe that the use of intensive subsurface sampling programs will eventually provide a more realistic picture of Long Island's past than that painted by traditional interpretations. I base my argument on two points.

First, subsurface testing will shift our focus away from the exclusive study of large shell middens. Significantly, of the 18 prehistoric sites located, only 5 can be defined as shell middens. In fact, a wide variety of activities appears to be represented by the sites, which include quarry areas, lithic workshops, camp sites, and amorphous lithic and lithic/ceramic scatters. The results of the subsurface sample suggest that shell middens do not represent the most common site type found on Shelter Island, but rather one component of a much larger settlement system. Interestingly, most sites located in our sample units had no clearly-defined surface manifestations. Only one site was visible on the ground surface; the remainder were buried 10 to 15 cm below the modern surface. These findings suggest that the majority of archaeological remains in our sample units are small, buried sites. If this finding holds true for other areas of Long Island, it raises serious doubts about traditional interpretations of the regional aboriginal settlement pattern based on chance discoveries and casual surface reconnaissances.

Second, subsurface testing provides the groundwork for making quantifiable parameter estimates of Long Island's archaeological resources. As more probability sampling is initiated on Long Island, parameter estimates will be re-examined and refined. For example, the density and distribution of sites recorded for Shelter Island suggest a relatively complex settlement system. Site densities were calculated in two ways. One, the unadjusted density, was computed by dividing the actual number of sites found by the total area of sample units surveyed. The other, the adjusted density, was

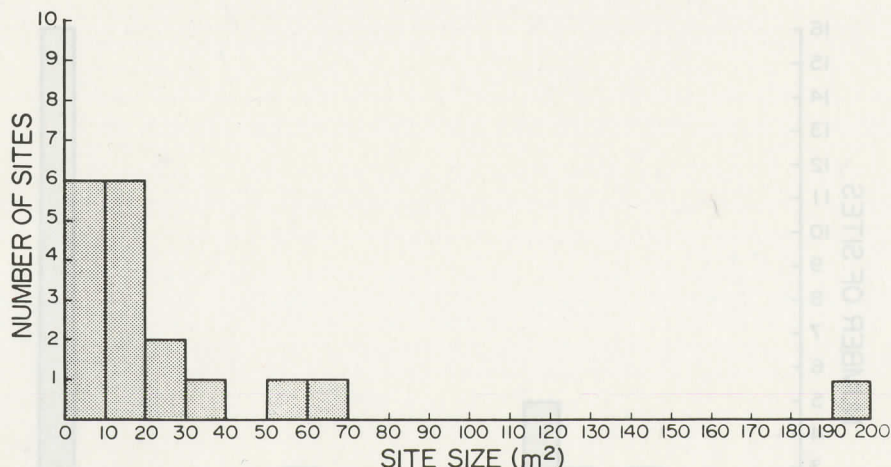


Figure 6. Histogram illustrating the range of site sizes recorded for the Mashomack Preserve Survey.

computed by dividing the number of sites *expected*, per unit of area surveyed, by the total area of sample units surveyed. The adjusted density takes into account the probability of locating sites smaller than the grid interval. The expected number of sites is obtained by summing the number of sites actually found and the number of small sites ($r < i/\sqrt{2}$) estimated to have been missed. Thus a site with a radius of 2.5 m has a .2 probability of being located (see Figure 3); that is, for every one site found, four others, on the average, will be missed by our test probes.

The site densities (both unadjusted and adjusted) vary tremendously between the sample units (see Table 1). There is a strong tendency for site densities to be lowest in the interior parcels (unadjusted and adjusted density range between 0 and .3 site/acre), and highest among the coastal sample units (unadjusted density: .41 to .54 site/acre; adjusted density: 2.75 to 5.1 sites/acre). In addition, the number of isolated finds is much higher for the coastal units. This pattern suggests that fewer activities resulting in recognizable remains were performed in the interior of the Preserve. These activities appear to be related to hunting, an interpretation based on the recovery of isolated points. On the other hand, the coastal strip is characterized by several non-site manifestations that cover areas of up to 7,500 m².

Dispersed around these large non-site manifestations are a number of smaller and more discrete activity areas that were defined as sites, including shell middens, lithic workshops, quarry sites, and lithic scatters, all located a short distance from one another. Most of the sites tend to represent small, special-purpose camps, although a few larger settlements are also present. The preliminary impression one gleans from the sample units is that the coastal strip was extensively utilized over a long period by small groups of lithic knappers, fishermen, and gatherers, who made occasional forays to the interior of the island. As more of the Preserve is surveyed, this preliminary assessment may be refined.

The most striking aspect of the survey results concerns the high site density. The average unadjusted site density for the six sample units is .31 sites per acre (76 sites/km²), and the adjusted density is 2.35 sites per acre (570 sites/km²) (see Table 1). These density figures are based on the excavation of more than 3,000 shovel probes representing a .0016 subsurface sample of 3.5% of the Preserve's surface area. However, these density figures should be viewed as preliminary. On the one hand, because of the boundary effect of the sample units, one must be careful about the over-inflation of density figures using either transects or quadrats (S. Plog et al. 1978:397-401). Only the percentage of site areas actually found within the boundaries of the sample unit should be included in the calculation of density figures. In addition, one must be very careful about interpreting adjusted site densities, because they are calculated on the assumption that sites with radii less than $i/\sqrt{2}$ are, in fact, present in the area sampled. If in reality the smaller sites are not present in the manner

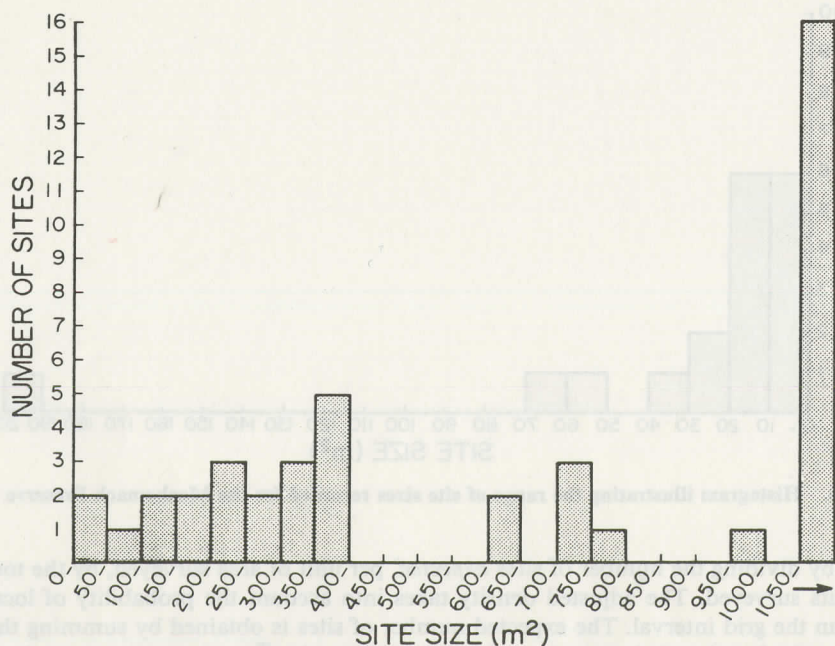


Figure 7. Histogram illustrating the range of site sizes recorded for the Pinedale Survey.

predicted, then the density estimates will also be inflated. On the other hand, the unadjusted and adjusted density figures may err on the conservative side, because we have yet to address the probability of detecting low-density scatters using shovel probes.

When the Shelter Island density figures are compared to those of Pinedale, the former represents a different level of magnitude. The overall prehistoric site density for Pinedale is approximately 3 sites per km² (.012 site per acre), which is 25 times lower than the unadjusted density recorded at Shelter Island (190 times lower than the adjusted density). At Pinedale there is a tendency to record fewer and larger sites, while at Shelter Island the trend is toward locating numerous small sites.

Pinedale does not represent an anomaly in the survey work published for the American Southwest. The intensive surveys of Chevelon Canyon and Hay Hollow Valley yielded site densities of 15/km² (.06 site per acre) and 25/km² (.078 site per acre), respectively (S. Plog et al. 1978:391). These figures are five and three times lower than the unadjusted density recorded at Shelter Island (38 and 23 times lower, respectively, than the adjusted density).

Labor Intensity

There is no question that subsurface survey is labor-intensive. We carefully monitored the labor input for each sample unit, and not surprisingly there were significant differences in the rate at which fieldwork proceeded across the Preserve. The surface area covered per hour of work varied from a high of 421 m² per person in sample unit B to a low of 91 m² per person in sample unit A. The overall average rate for the six sample units was 247 m² per person per hour of work. That is, on the average, one person can survey (and this includes the excavation of 30 by 30 cm shovel probes every 10 m) approximately 2,000 m² (.49 acre) per day.

There are three variables that we found significant in determining survey rates. One, of course, is the composition of crews; some people work faster than others. But other variables, such as the density of underbrush and the quantity of cultural material found on a transect, are equally important. For example, sample unit B, characterized by the most efficient survey rate, contains few sites and

Table 1. Unadjusted and Adjusted Site Densities for the Six Sample Units Surveyed in the Mashomack Preserve.

Sample Unit	Area	Pre-his- toric Sites (No.)	Estimated No. of Sites	Unadjusted Site Density	Adjusted Site Density
A. Interior	14,098 m ² (.014 km ²) 3.5 acres	0	0	0	0
B. Interior	66,915 m ² (.067 km ²) 16.5 acres	0	0	0	0
C. Estuary	59,586 m ² (.06 km ²) 14.7 acres	8	67/8 = 75	134/km ² .54/acre	1,250/km ² 5.1/acre
D. Estuary	58,750 m ² (.059 km ²) 14.5 acres	6	39/6 = 45	102/km ² .41/acre	750/km ² 3.1/acre
E. Interior	12,960 m ² (.013 km ²) 3.2 acres	1	1	77/km ² .30/acre	77/km ² .30/acre
F. Coastal Swamp	23,305 m ² (.023 km ²) 5.8 acres	3	13/3 = 16	129/km ² .52/acre	696/km ² 2.75/acre
Total of six Sample Units	235,618 m ² (.24 km ²) 58.2 acres	18	137	76/km ² .31/acre	570/km ² 2.35/acre

only minimal vegetation coverage. In contrast, less efficient rates recorded for sample units C and E are a product of high densities of cultural material and/or dense concentrations of cat briar.

When compared to the Pinedale project, the survey rate recorded for Shelter Island was discouragingly slow. An average crew member at Pinedale covered 26.6 acres per day or about 50 times more surface area than a similar crew member at Shelter Island. Other intensive surveys conducted in the Southwest bear out these figures. The survey rates for the Chevelon and Hay Hollow projects are approximately 16 and 8 acres per person per day, respectively (S. Plog et al. 1978:391). These rates show that crew members covered about 32 and 16 times more space per unit of time than crew members at Shelter Island.

EVALUATING THE COMPARISON

Contrary to prior expectations, the modal size of sites is substantially smaller for Shelter Island than Pinedale. At the same time, the site density recorded for Shelter Island is considerably higher than that recorded at Pinedale. Archaeologists may be inclined to dismiss these findings as comparing apples and oranges; certainly the long-term prehistoric adaptations to the Pinedale region and Shelter Island were very different, and one would expect marked differences in the archaeological resources found in each of these areas. Yet the survey results illuminate a very interesting point of correspondence between them. The preliminary findings at Shelter Island suggest that subsurface surveys can provide results that are as detailed and fine-grained as those of the most intensive surface surveys conducted in the American Southwest or, for that matter, anywhere in the world.

Although our findings are no doubt due in part to the nature of the prehistoric occupations of the two regions, I believe the variation between surveys can also be attributed to another factor—

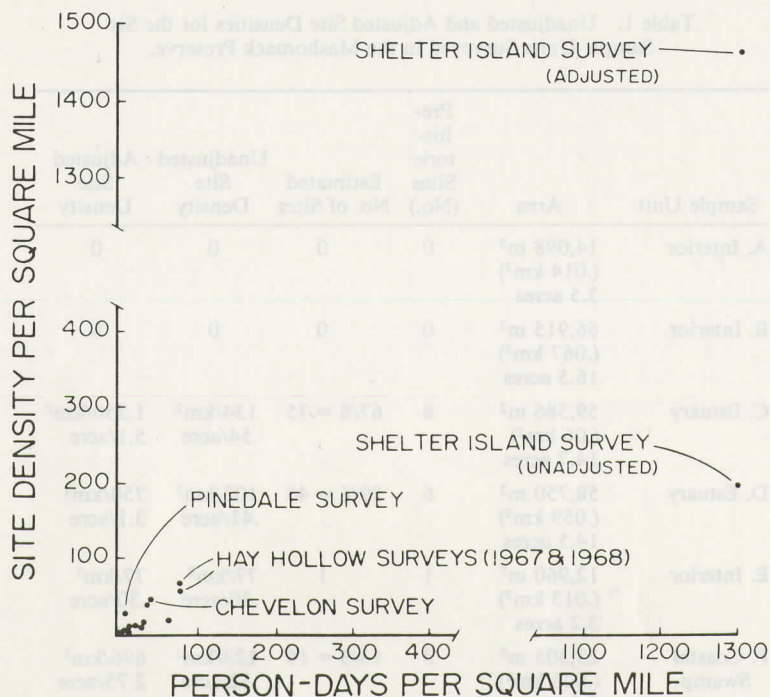


Figure 8. Scattergram showing the relationship between survey intensity and site density for the 12 Southwestern surveys described by S. Plog et al. (1978:391), the Pinedale survey, and the Shelter Island survey. Both the unadjusted and adjusted site densities are illustrated for the Shelter Island survey.

survey intensity. S. Plog et al. (1978:391-393) have neatly demonstrated that site densities in the American Southwest correlate directly with the labor invested on survey projects. Employing a linear regression plot ($r = .888$; $p < .01$ for prehistoric sites), they show that surveys of low intensity (crew members spaced 50 to 100 m apart) are significantly associated with low site densities, while high-intensity surveys (crew members spaced 10 m apart or less) are associated with greater site densities. They conclude that the greater the survey intensity, the greater the probability of detecting small sites of low visibility.

Intensive *subsurface* surveys such as the one conducted at Shelter Island appear to support the findings of S. Plog et al. (1978). If the adjusted site density and person days necessary to survey 1 mi² are extrapolated for Shelter Island (1,476 sites/mi² and 1,300 person days, respectively), then the Long Island survey represents a continuation of the pattern described for the American Southwest. Figure 8 illustrates the site density/labor intensity relation for the 12 Southwestern surveys reported by S. Plog et al. (1978:391), the Pinedale survey, and the Shelter Island survey.

Although the Shelter Island survey was 16 times more labor-intensive than any of the Southwestern projects examined by S. Plog et al. (1978), it produced an adjusted site density 23 times greater than any of the Southwestern surveys. These figures illustrate an important aspect of intensive subsurface testing programs: they involve a very critical examination of the landscape. Subsurface survey crews cover the ground surface at the same intensity as most surface survey crews, yet they also sample the subsurface soil strata periodically. At each test unit, a careful screening of the soil is undertaken to detect cultural material. If an isolated artifact is observed, the search procedure is concentrated in a limited area as test units are laid out around the spot at 2 m intervals. In contrast, I noticed when rechecking Southwestern survey areas that trained surface surveyors, myself included, did not always observe spatially-limited artifact concentrations (see also S. Plog et al. 1978:386-

387). This is an important observation because a major difference between the Pinedale and Shelter Island surveys lies in the delineation of small sites—a factor that greatly increased site density for Shelter Island. Unless the surface survey is undertaken on hands and knees (a level of intensity not widely employed), surveyors tend to miss sparse artifact scatters.

The main point I wish to make is that intensive subsurface testing programs, especially when designed around discovery probability limits, may provide just as solid a foundation for making parameter estimates as most surface surveys conducted in arid areas of sparse vegetation. In making this point I do not wish to minimize the important issues raised by Wobst, Nance, and others concerning the problems of subsurface surveys. I do suggest that as we continue to refine our subsurface testing techniques, we can begin to control these discovery problems in a more rigorous manner. This is significant, since an inferiority complex can sometimes be detected in Eastern survey reports, especially when compared to Southwestern projects. This complex is no doubt referable to 1) a recent historical trend, beginning in the late 1960s and early 1970s that saw Southwestern archaeologists publishing most of the seminal survey articles; and 2) the ideal conditions for survey work found in the Southwest.

In evaluating this issue, one must recognize that Southwestern survey work is characterized by its own problems—many of which parallel the problems raised for subsurface surveys. Some have been mentioned here; e.g., the bias in favor of detecting larger sites with greater frequency than smaller sites (S. Plog et al. 1978:400). Also, the assumption that most archaeological remains will produce detectable surface manifestations is questionable. As Schiffer et al. (1978:6–7) point out, surface visibility varies tremendously, depending upon such local factors as vegetation, soil accumulation, and alluvial deposition. Surface surveys may be missing many of the early Archaic and Paleoindian sites that are buried beneath alluvial depositions of prominent watercourses. Even in the case of later Pueblo settlements, there is a tendency to overlook smaller, special-purpose camps at the expense of large, multi-room Pueblos (S. Plog et al. 1978:386); minimal amounts of soil accumulation bury small artifact scatters while sites with masonry rubble are still detectable on the ground surface (Plog 1981:135).

By concentrating on large villages containing recognizable house features, plazas, and ceremonial architecture, some Southwestern surface surveys may have skewed our perspective on the sizes and settlement patterns of prehistoric populations. Recent studies, such as those of F. Plog et al. (1978) and Hantman and Neitzel (1984), indicate that the majority of people in the prehistoric period lived in small homesteads or camps. Other studies indicate that the degree to which Southwestern groups practiced a sedentary existence may be also over-represented in the literature. With increasing frequency archaeologists have begun identifying residually mobile populations who were dispersed across the regional landscape in seasonal villages and short-duration camps (see Lightfoot 1984b; Powell 1980; Reid 1982:193–194; Upham 1984).

In light of these findings, an important area of research is the comparison of subsurface and surface results in the American Southwest. If subsurface surveys are undertaken, will they produce different estimates on the number and kinds of sites than those generated by traditional surface surveys? For the Pinedale study area, I expect that an intensive surface survey would detect a much larger number of small artifact scatters than that recorded by the surface survey.

CONCLUSION

Archaeologists working in the American Southwest and the Eastern Woodlands are faced with their own sets of survey problems, but in each area these problems can be minimized through the implementation of carefully-designed regional surveys. When faced with the problems of detecting sites of low visibility, subsurface surveys offer a useful method for estimating the frequency and kinds of sites present in sample units. By defining the discovery probabilities of detecting a target population of sites within sample units, adjustments can be made that provide a solid data base for generating regional parameter estimates. In such a manner, element sampling procedures can be used to draw an accurate picture of the settlement pattern of a study area. In addition, subsurface surveys provide more detailed information on site boundaries and site sizes than is available for

surface surveys, because subsurface artifact distributions tend to be less affected by erosion and other natural formation processes.

Labor costs are the major shortcoming of subsurface surveys, and the investigator must weigh both these costs and the probability of discovering small sites of low visibility with the overall research objectives of the project. Clearly, subsurface testing programs should be an integral component of survey work undertaken in the Eastern Woodlands, with the size and spacing of test probes varying with the specific research objectives. In the American Southwest, pedestrian surface surveys represent a very efficient method for detecting sites, so long as the research objectives focus on sites of high visibility, such as those produced by sedentary groups. If, on the other hand, the research objectives involve the study of the mobile hunters and gatherers who probably occupied extensive areas of the Southwest during much of the prehistoric period (see Upham 1984), then surface surveys may not provide enough information to generate reliable parameter estimates on these populations. In this case, some form of subsurface testing will probably be required to detect, with a given level of probability, the small artifact scatters that represent seasonal residential bases and special-purpose camps.

Acknowledgments. The Pinedale and Shelter Island survey projects could not have been undertaken without the full cooperation of the United States Forest Service and The Nature Conservancy, respectively. I am especially indebted to Dee Green and Bruce Donaldson of the U.S.F.S. and Michael Laspia, Mashomack Preserve Director, for their guidance and support. Finally, I would like to thank Gretchen Gwynne, Jeff Hantman, Roberta Jewett, Frank McManamon, and Steve Plog, as well as the reviewers for *American Antiquity*, for reading and commenting upon various drafts of this article. Their insightful comments greatly strengthened and clarified the major points made in the paper. Any shortcomings, however, remain solely my responsibility.

REFERENCES CITED

- Aitken, M. J.
1974 *Physics and Archaeology*. Oxford University Press, Oxford.
- Alexander, Diana
1983 The Limitations of Traditional Surveying Techniques in a Forested Environment. *Journal of Field Archaeology* 10:177-186.
- Baker, Charles
1975 Arkansas Eastman Archaeological Project. *Arkansas Archaeological Survey, Research Report No. 6*. Fayetteville.
- Birk, Douglas A., and Douglas C. George
1976 A Woodland Survey Strategy and Its Application in the Salvage of a Late Archaic Locus of the Smith Mounds Site, (21-KC-3) Koochiching County, Minnesota. *Minnesota Archaeologist* 35:1-30.
- Cable, John, James L. Michie, and Stephen M. Perlman
1977 An Archaeological Reconnaissance of the Gaffney By-Pass, Cherokee County, South Carolina. Institute of Archaeology and Anthropology, *Research Manuscript Series No. 121*. University of South Carolina, Columbia.
- Carr, Christopher
1977 A New Role and Analytical Design for the Use of Resistivity Surveying in Archaeology. *Midcontinental Journal of Archaeology* 2:161-193.
- Casjens, Laurel, Garth Bawden, Michael Roberts, and Valerie Talmage
1978 Field Methods in New England Cultural Resource Management. In *Conservation Archaeology in the Northeast: Toward a Research Orientation*, edited by Arthur Spiess, pp. 87-94. Peabody Museum Bulletin 3, Harvard University, Cambridge.
- Ceci, Lynn
1982 Method and Theory in Coastal New York Archaeology: Paradigms of Settlement Pattern. *North American Archaeologist* 3:5-36.
- Chartkoff, John
1978 Transect Interval Sampling in Forests. *American Antiquity* 43:46-53.
- Classen, Cheryl, and Carol S. Spears
1978 An Assessment of Ambulatory Site Surveying, Shovel Testing and Augering. In *Arkansas Eastman Archaeological Project*, edited by Charles Baker, pp. 123-127. *Arkansas Archaeological Survey, Research Report No. 6*. Fayetteville.
- Dincauze, D. F., M. Wobst, R. Hasenstab, and D. Lacy
1981 Retrospective Assessment of Archaeological Survey Contracts in Massachusetts 1970-1979. Report prepared for the Massachusetts Historical Commission, Boston, Massachusetts.

- Eidt, R.
1973 A Rapid Chemical Field Test for Archaeological Surveying. *American Antiquity* 39:206-210.
- Feder, Kenneth L.
1981 The Farmington River Archaeological Project: Focus on a Small River Valley. *Man in the Northeast* 22:131-146.
- Hantman, Jeffrey L., and Jill E. Neitzel
1984 Demographic Reconstruction in the Northern Southwest. Ms. on file, Department of Anthropology, University of Virginia, Charlottesville.
- House, John H., and David L. Ballenger
1976 An Archaeological Survey of the Interstate 77 Route in the South Carolina Piedmont. Institute of Archaeology and Anthropology, *Research Manuscript Series No. 104*, University of South Carolina, Columbia.
- Judge, W. James, James I. Ebert, and Robert K. Hitchcock
1975 Sampling in Regional Archaeological Survey. In *Sampling in Archaeology*, edited by James W. Mueller, pp. 82-123. University of Arizona Press, Tucson.
- Kraker, James, Michael J. Shott, and Paul Welch
1983 Design and Evaluation of Shovel-Test Sampling in Regional Archaeological Survey. *Journal of Field Archaeology* 10:469-480.
- Lightfoot, Kent G.
1984a *Prehistoric Political Dynamics: A Case Example from the American Southwest*. Northern Illinois University Press, DeKalb.
1984b The Duncan Project: A Study of the Occupation Duration and Settlement Pattern of an Early Mogollon Pithouse Village. *Anthropological Field Studies No. 6*. Arizona State University, Tempe.
- Lightfoot, Kent, Robert Kalin, Owen Lindauer, and Linda Wicks
1985 Coastal New York Settlement Patterns: A Perspective from Shelter Island. *Man in the Northeast* 30: 59-82.
- Lovis, William, Jr.
1976 Quarter Sections and Forests: An Example of Probability Sampling in the Northeastern Woodlands. *American Antiquity* 41:364-372.
- Lynch, Mark
1980 Site Artifact Density and the Effectiveness of Shovel Probes. *Current Anthropology* 21:516-517.
- McBride, Kevin A., and Robert Dewar
1981 Prehistoric Settlement in the Lower Connecticut River Valley. *Man in the Northeast* 22:37-66.
- McManamon, Francis Patrick
1978 Site Identification in the Northeast. In *Conservation Archaeology in the Northeast: Toward a Research Orientation*, edited by Arthur E. Spiess, pp. 78-86. Peabody Museum Bulletin 3. Harvard University, Cambridge.
1981 Probability Sampling and Archaeological Survey in the Northeast: An Estimation Approach. In *Foundations of Northeast Archaeology*, edited by Dean Snow, pp. 195-227. Academic Press, New York.
1984a Discovering Sites Unseen. In *Advances in Archaeological Method and Theory*, vol. 7, edited by Michael Schiffer, pp. 223-292. Academic Press, New York.
1984b Method and Techniques for Survey and Site Examination. In *Chapters in the Archaeology of Cape Cod*, I, vol. 1, edited by Francis McManamon, pp. 25-44. U.S. Department of the Interior, Boston.
- Mueller, James W.
1974 The Use of Sampling in Archaeological Survey. *Memoirs of the Society for American Archaeology* 28.
1975 Archaeological Research as Cluster Sampling. In *Sampling in Archaeology*, edited by J. Mueller, pp. 33-41. The University of Arizona Press, Tucson.
- Nance, Jack D.
1979 Regional Subsampling and Statistical Inference in Forested Habitats. *American Antiquity* 44:172-176.
1980 Non-site Sampling in the Lower Cumberland River Valley, Kentucky. *Mid-Continental Journal of America* 5:169-191.
1983 Regional Sampling in Archaeological Survey: The Statistical Perspective. In *Advances in Archaeological Method and Theory*, vol. 6, edited by Michael Schiffer, pp. 289-356. Academic Press, New York.
- Nance, Jack D., and Bruce F. Ball
1986 No Surprises? The Reliability and Validity of Test Pit Sampling. *American Antiquity* 51:457-483.
- Plog, Fred
1974 *The Study of Prehistoric Change*. Academic Press, New York.
1981 *Cultural Resources Overview: Little Colorado Area, Arizona*. USDA Forest Service Southwestern Region, Albuquerque.
- Plog, Fred, Richard W. Effland, and Dee Green
1978 Inferences Using the SARG Data Bank. In *Investigations of the Southwestern Anthropological Research Group: The Proceedings of the 1976 Conference*, edited by R. C. Euler and G. J. Gumerman, pp. 139-148. Museum of Northern Arizona, Flagstaff.

- Plog, Stephen
1976 Relative Efficiencies of Sampling Techniques for Archaeological Surveys. In *The Early Mesoamerican Village*, edited by Kent Flannery, pp. 136–158. Academic Press, New York.
- Plog, Stephen, Fred Plog, and Walter Wait
1978 Decision Making in Modern Surveys. In *Advances in Archaeological Method and Theory*, vol. 1, edited by Michael B. Schiffer, pp. 383–421. Academic Press, New York.
- Powell, Shirley
1980 Material Culture and Behavior: A Prehistoric Example for the American Southwest. Ph.D. dissertation, Department of Anthropology, Arizona State University, Tempe.
- Raab, L. Mark
1977 An Archaeological Sample Survey of the Caddo Planning Unit, Ouachita National Forest, Arkansas. In *Archaeology and National Forest Land Management Planning*, pp. 23–34. USDA Forest Service, Southwestern Region, Archaeological Report 16.
- Reid, J. Jefferson
1982 *Cholla Project Archaeology*, vol. 1, *Introduction and Special Studies*. Arizona State Museum, *Cultural Resource Management Division Archaeological Series No. 161*. Tucson.
- Salwen, Bert
1970 Cultural Inferences from Faunal Remains: Examples from Three Northeast Coastal Sites. *Pennsylvania Archaeologist* 40:1–8.
- Saxon, Walter
1975 Archaeological Surveying and Testing in Connetquot River State Park. Ms. on file, Nassau County Museum, Long Island, New York.
- Schiffer, Michael, Alan P. Sullivan, and Timothy C. Klinger
1978 The Design of Archaeological Surveys. *World Archaeology* 10:1–28.
- Stone, Glen Davis
1981 On Artifact Density and Shovel Probes. *Current Anthropology* 22:182–183.
- Thomas, David Hurst
1975 Nonsite Sampling in Archaeology: Up the Creek Without a Site? In *Sampling in Archaeology*, edited by James W. Mueller, pp. 61–81. The University of Arizona, Tucson.
- Tite, M. S.
1972 *Methods of Physical Examination in Archaeology*. Seminar Press, New York.
- Upham, Steadman
1984 Adaptive Diversity and Southwestern Abandonment. *Journal of Anthropological Research* 40:235–256.
- Warner, Frederic W.
1978 Relative Effectiveness of Sampling Strategies at a Connecticut Site. In *Conservation Archaeology in the Northeast: Toward a Research Orientation*, edited by Arthur Spiess, pp. 67–78. Peabody Museum Bulletin 3, Harvard University, Cambridge.
- Wobst, Martin
1983 We Can't See the Forest for the Trees: Sampling and the Shapes of Archaeological Distributions. In *Archaeological Hammers and Theories*, edited by James Moore and Arthur Keene, pp. 37–85. Academic Press, New York.
- Wyatt, Ronald J.
1977 The Archaic on Long Island. *Annals of the New York Academy of Sciences* 288:400–410.

NOTES

¹ Intensity is defined as the spacing of crew members across the sample unit (S. Plog et al. 1978:389–390).

² Fieldwork was undertaken during the summer months of 1977 and 1978 in collaboration with Fred Plog and the Anthropology Department, Arizona State University.

³ A good example of cluster sampling procedures is provided by Nance's (1979, 1980) "non-site" study of the Lower Cumberland River region in Kentucky, in which he calculated the relative proportion of surface area expected to be the locus of artifactual material based on test probe information.

⁴ A good example of subsurface element sampling is McManamon's (1984b:33–35) survey of the Cape Cod National Seashore, a project in which the frequency of prehistoric sites found in different ecological zones was estimated, based on information obtained from a stratified random sample of sample units.

⁵ The information on the 220 survey projects was originally gathered by D. Dincauze, M. Wobst, R. Hasenstab, and D. Lacy. See Dincauze et al. 1981.

⁶ This survey rate is similar to that reported by McManamon (1984a:266–267) for the Cape Cod National Seashore Archaeological Survey.