Department of Defense
Legacy Resource Management Program

PROJECT 07-353

An Assessment of Archaeological Data Quality:

A Report Submitted in Partial Fulfillment of Legacy Resource Management Program Project "To Develop Analytical Tools for Characterizing, Visualizing, and Evaluating Archaeological Data Quality Systematically for Communities of Practice within the Department of Defense"

Michael P. Heilen, Christopher L. Nagle, and Jeffrey H. Altschul

December 2008

An Assessment of Archaeological Data Quality: A Report Submitted in Partial Fulfillment of Legacy Resource Management Program Project "To Develop Analytical Tools for Characterizing, Visualizing, and Evaluating Archaeological Data Quality Systematically for Communities of Practice within the Department of Defense"

Michael P. Heilen, Christopher L. Nagle, and Jeffrey H. Altschul

Submitted to
Michael L. Denight
Contracting Officer's Technical Representative
U.S. Army Corps of Engineers
Engineer Research and Development Center
Construction Engineering Research Laboratory

and

Michelle Wurtz
Geo-Marine, Inc.
2201 K Ave., Ste. A2
Plano, TX

Subcontract Agreement No. 2255sa07-A
Prime Contract No. DACA42-02-D-0012
Legacy Resource Management Program Project No. 07-353

Technical Report 08-65
Statistical Research, Inc.
Tucson, Arizona

An Assessment of Archaeological Data Quality: A Report Submitted in Partial Fulfillment of Legacy Resource Management Program Project "To Develop Analytical Tools for Characterizing, Visualizing, and Evaluating Archaeological Data Quality Systematically for Communities of Practice within the Department of Defense"

Michael P. Heilen, Christopher L. Nagle, and Jeffrey H. Altschul

Submitted to
Michael L. Denight
Contracting Officer's Technical Representative
U.S. Army Corps of Engineers
Engineer Research and Development Center
Construction Engineering Research Laboratory

and

Michelle Wurtz
Geo-Marine, Inc.
2201 K Ave., Ste. A2
Plano, TX

Subcontract Agreement No. 2255sa07-A
Prime Contract No. DACA42-02-D-0012
Legacy Resource Management Program Project No. 07-353

Technical Report 08-65
Statistical Research, Inc.
Tucson, Arizona

December 2008

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

The data quality project was conceived by Tad Britt of the Construction Engineering Research Laboratory (CERL) and Clay Mathers of Statistical Research, Inc. (SRI). The project was designed and executed under the direction of Jeff Altschul and Michael Heilen of SRI. We would like to acknowledge the support of Michael Denight, who served as CERL's technical representative after Mr. Britt left the agency. We would also like to thank Michelle Wurtz and Duane Peters of Geo-Marine, Inc., and Jay Newman of the U.S. Army Corps of Engineers, Fort Worth District, for ably administering the contract.

This study was allied with the Department of Defense's predictive modeling initiative. That work is being sponsored by Air Combat Command (ACC), supported by the Office of the Secretary of Defense, and led by Paul Green, Headquarters ACC Cultural Resources Manager. In particular, we would like to thank Maureen Sullivan, Federal Preservation Officer (FPO), and Brian Lione, Deputy FPO. We also want to thank Hillori Schenker, Cultural Resources Management Specialist, Legacy Resource Management Program, for her encouragement and patience.

We would like to thank the following individuals for providing data, information, or guidance, or comments on the draft report: Michael Baskerville (China Lake Naval Air Weapons Station); Jerome King and Amy Gilreath (Far Western Anthropological Research Group, Inc.); Brian D. Knight (Fort Bliss); Laurie W. Rush, Jim Rapant, and Jaime Ratner (Fort Drum); Mark Stanley and George Cole (Eglin Air Force Base [AFB]); Jaynie Hirschi and Kate Stratford (Hill AFB); David Doyel (Luke AFB); Sheri Mattoon-Bowden and Carl Rudeen (Mountain Home AFB); Helen O'Brien, Archaeology Center, Pima Community College; Prentice Thomas, Jan Campbell, and Jon Wesson (Prentice Thomas & Associates, Inc.); Keith Kintigh (School of Human Evolution and Social Change, Arizona State University); and James A. Zeidler (Center for Environmental Management of Military Lands, Colorado State University).

EXECUTIVE SUMMARY

The U.S. Department of Defense (DoD) recognizes the need to ensure the quality of archaeological data used in planning efforts. This project, funded by the Legacy Resource Management Program, explores the potential value and benefit to the DoD of establishing key quality indicators for archaeological data. Statistical Research, Inc. (SRI)—in partnership with the U.S. Army Corps of Engineers Engineer Research and Development Center Construction Engineering Research Laboratory and under subcontract agreement with Geo-Marine, Inc.—conducted this project using data from six installations. The installations used as case studies were chosen to represent a variety of Services, regions, environments, and cultures. These were the Barry M. Goldwater Range (Arizona), China Lake Naval Air Weapons Station (California), Eglin Air Force Base (Florida), Fort Bliss (McGregor Range, New Mexico), Fort Drum (New York), and the Utah Test and Training Range (Utah).

The goals of this pilot project were to address the following questions: How good are cultural resource management (CRM) data on military installations? Does the quality of the data affect decisions critical to military activities? To address these goals, this project illustrated the use of analytical tools for assessing the effect of standardized inventory and recording practices on the quality of archaeological data. The installations chosen for this project use a range of methodological approaches. Installations in the eastern United States, for instance, typically use shovel-test pits to discover sites, whereas installations in the western United States rely mostly on pedestrian survey. In addition, each installation implements standardized inventory and recording practices that are appropriate to regional characteristics of the archaeological record and the kinds of cultural resources requiring management.

SRI explored three data quality issues of major concern to DoD military installations: survey reliability, site location, and site boundaries. For each of the three issues, SRI illustrated the fallacy of obtaining "true" results and demonstrated the utility of using multiple observations to calculate precise estimates of numbers of sites, site locations, and site boundaries. Analytical tools suitable for use by military personnel for mission planning, training, and CRM purposes were discussed and explained. The study concluded with the following recommendations: (1) archaeological data and metadata on survey methods, recording techniques, and instrumentation should be regularly maintained and archived according to standardized formats; (2) metrics illustrated in the study should be used to evaluate data quality on other military installations; and (3) problems with the site concept should be addressed at DoD installations by focusing on regional patterns rather than specific site locations.

CHAPTER 1

Introduction

As part of a Legacy Resource Management Program (Legacy) project (No. 07-353), this pilot study explores the complex issues involved in assessing the quality of archaeological data used by the Department of Defense (DoD) to comply with federal statutes and regulations affecting cultural resources. Our goal is to determine the potential value and benefit of establishing key quality indicators for archaeological data. To do this, we examine the statistical implications of current approaches to finding and defining archaeological sites and the effects of these approaches on data quality. This project was conducted for the U.S. Army Corps of Engineers, Engineer Research and Development Center, Construction Engineering Research Laboratory under a subcontract agreement issued by Geo-Marine, Inc., (Subcontract No. 2255sa07-A) to Statistical Research, Inc. (SRI).

We begin this chapter with a brief review of methodological studies in American archaeology. We show that early concerns focused mainly on the conduct of field and analytical research. With the advent of federal and state statutes requiring cultural resource management (CRM), methodological studies moved from theoretical discussions to the development of standardized practices aimed at gaining the confidence of nonprofessional stakeholders in the results of field inventories. Today, these practices have become accepted but are subject to very little critical examination. Although accepted standards provide structure and consistency to the CRM process, they also hide serious logical and empirical faults in the underlying assumptions. Problems resulting from inaccurate assessment of archaeological data quality have caused the unexpected expenditure of time and money and have recently emerged as a major concern in DoD CRM. In this report, we identify and explore three issues in archaeological data quality of major concern to the DoD, in each case developing and presenting analytical tools that can be used to assess archaeological data quality: survey reliability, site location, and site-boundary definition. We then use data from military installations distributed throughout the United States to highlight the three topics of major concern to DoD and illustrate the use of tools for characterizing, visualizing, and evaluating archaeological data quality. Chapter 2 is devoted to survey reliability, Chapter 3 to site location, and Chapter 4 to site boundaries. We close in Chapter 5 with the practical implications of this study to DoD (in the form of a series of recommendations) and thoughts on future research directions.

Data Quality in American Archaeology

American archaeology was established as an academic discipline in reaction to antiquarian approaches to the past. Antiquarian approaches focused more on spectacular finds and speculation than on systematic data development, logical inference, and the application of scientific principles. Before World War II, methodological studies in American archaeology focused largely on issues of time and space, or culture history. Methods were developed to order artifacts in temporal sequence, and relative dating techniques, such as stratigraphy and seriation, were accepted as general practice in fieldwork and analysis. After the war, methodology was driven by changes in technology. For the first time, absolute dating methods, including radiocarbon dating and dendrochronology, allowed archaeologists to present the age of materials with greater confidence and on an absolute time scale. Studies of technology and subsistence were revolutionized by advances in high-powered microscopes, which allowed the detection and identification of pollen, microfossils, and the chemical composition of materials (Dean 1978; Pollard and Heron 1996; Trigger 1984, 1989).

The 1960s ushered in a period of great theoretical and methodological change during which the assumptions underlying archaeological interpretation were questioned (e.g., Binford 1962, 1965; White 1959; Willey and Phillips 1958). Culture histories, for example, were constructed from data acquired from a small set of type sites considered representative of a region's archaeological record. But how did archaeologists know that these type sites were truly representative?

An increasing number of archaeologists turned to statistical inference and logic to address epistemological issues. Archaeologists were particularly drawn to statistical sampling to answer such questions as these: How many sites exist in the region? How many site classes exist? What is the relationship between the distribution of sites and environmental features? How many artifacts or features exist on a site? How many artifacts or features do we have to study to define behavioral units?

For most of the 1960s and 1970s, methodological debates focused less on data quality than on sampling design. For the most part, archaeologists were willing to accept that data were essentially equivalent—a sherd was a sherd, a pit house was a pit house. The debate pivoted on how to relate data in ways that supported logical arguments. Systematic designs were preferred over judgmental observation and collection.

CRM brought new concerns to the study of the past. Archaeological data were no longer the sole province of academic researchers. Now, archaeological data were central to decisions about land use, development, and preservation. Nonprofessionals, such as federal land managers, Native Americans, and industry, wanted archaeologists to find, evaluate, and, if significant, treat "all" resources affected by an undertaking.

Archaeologists were slow to respond to the needs of these stakeholders for two reasons. First, archaeologists were not used to being held accountable by others. In the early days of CRM, archaeologists tended to be overly defensive; they were not used to people challenging their interpretations or offering their own ideas about the past.

A second reason was that archaeologists did not communicate well with nonarchaeologists. Archaeologists knew that a 100 percent survey did not result in the discovery of 100 percent of the sites. Survey techniques involve field archaeologists spaced at set intervals inspecting the ground either visually or with the aid of small subsurface probes, usually shovel tests. Sites that fall completely between observation units will not be found. Moreover, an archaeologist will find a site only if he or she detects something—for example, an artifact, a feature, or a hand-cut tree limb—that is the result of human behavior. Even the best field archaeologist does not detect everything.

Many factors affect site discovery and recording—for example, ground cover, deposition, preservation, disturbance, artifact density, and artifact size (Banning 2002; Ebert 1992; Fish and Kowalewski 1990; Judge 1981; Kintigh 1988; Krakker et al. 1983; Mueller 1975; Nance and Ball 1986; Plog et al. 1978; Rossignol and Wandsnider 1992; Schiffer 1976, 1987; Schiffer et al. 1978; Schiffer and Wells 1982; Shennan 1985; Shipman 1981; Shott 1989, 1992; Tainter 1983; Wandsnider and Dore 1995). Many of the variables affecting site discovery or recording have been explored in a general fashion and often in a cautionary way. Most studies have been directed to an archaeological audience and not to other parties with an interest in archaeological sites. With respect to DoD, few systematic studies have been undertaken that examine the relationship between archaeological data and specific factors affecting data quality on military installations (although see, e.g., Altschul 2000; Altschul et al. 2004; Homburg et al. 1994; Zeidler 1995, 2001).

Thus far, studies show that much potential for error exists in site discovery and recording. Although archaeologists strive to minimize errors, it is important to recognize that we cannot eliminate them. We can determine the severity of error and express it statistically, however. Probability theory provides us with the basis from which we can calculate the amount of confidence we have in sample estimates. Generally, these are expressed as confidence intervals (e.g., the number of sites per 100 acres is 10 ± 3). The great advantage of sampling theory is not that it provides us with the true value of the variable of interest (e.g., number of sites in a survey area) but that it provides us with the means to calculate how much confidence we can place in our estimate.

The major purpose of this study is not to point out problems with the quality of installation data but to provide ways of assessing the inherent errors in data that directly affect military objectives. Instead of assuming that cultural resource data are error free, installations would be far better off if they were to try to answer correctly the following questions:

- What is the proportion of archaeological sites found during inventory? How many are missed and of what type?
- Are archaeological sites located where we think they are? In other words, are our maps accurate?
- Can we accurately define the boundaries of archaeological sites?

These three issues can be subsumed under the categories of survey reliability, site location, and site boundaries; they form the subjects of our study.

Case Studies

As for this project, we modeled the effects of survey and recording methods on survey reliability, site location, and site boundaries using archaeological data from six installations: the Barry M. Goldwater Range (BMGR), China Lake, Eglin Air Force Base (Eglin AFB), Fort Bliss, Fort Drum, and the Utah Test and Training Range (UTTR) (Figure 1.1). The installations chosen provide a range of environments, culture histories, archaeological resources, and military services and missions.

In the East, we selected two installations: Fort Drum and Eglin AFB. Eglin AFB covers 464,000 acres in northwest Florida. With a diversity of environments ranging from coastline, swamps, marshes, and forests, Eglin suffers from poor site visibility. Shovel tests are used extensively as a site-discovery method. Archaeological sites include small artifact scatters, shell middens, and residential sites. Fort Drum in upstate New York covers 107,000 acres, which are mostly forested with deep ravines. Surface visibility is generally poor, and site-discovery methods rely heavily on shovel tests. Aboriginal sites range from small lithic scatters to longhouse villages.

In the West, we selected four installations: the BMGR, China Lake, Fort Bliss, and the UTTR. The McGregor Range of Fort Bliss covers about 700,000 acres of the Tularosa Basin in south-central New Mexico. Located in the Chihuahuan Desert, this installation has good surface visibility; pedestrian survey is used to find sites. Archaeological sites are dominated by small artifact scatters used primarily by foragers, along with a few residential sites associated with the farmers of the Jornada Mogollon. Farther west, we selected the BMGR of southwest Arizona. Of the 1,650,000 acres withdrawn for military use, we focus on portions of the 1,050,000 acres assigned to the BMGR East, which is administered by the 56th Range Management Office, Luke AFB (the remaining 650,000 acres are administered by the Marine Corps Air Station, Yuma). The BMGR is part of the Western Papaguería of the Sonoran Desert. With limited vegetative ground cover or deposition, surface visibility is excellent, and pedestrian surveys are adequate to inventory archaeological sites. Sites are mostly small and consist of rock features of various types and scatters of artifacts. Residential camps are present but relatively rare. In eastern California, we selected Naval Air Weapons Station China Lake. The 1.1-million-acre installation lies in the Mojave Desert. Also sparsely vegetated, China Lake has excellent surface visibility, and archaeologists rely exclusively on pedestrian surveys as a discovery method. The archaeological record is dominated by small artifact scatters. To represent the Great Basin, we chose the UTTR, administered by Hill AFB. Approximately 1.2 million acres, the installation lies in the Bonneville Salt Flats, where surface visibility is excellent. Pedestrian survey is used to discover sites, which are mostly small artifact scatters.

Each installation was chosen to illuminate aspects of one or more of the topics under study. The failure to use an installation to illustrate a point does not mean that it does not suffer from problems similar to those discussed. Instead, it means either that the data available are not amenable to examining a particular aspect

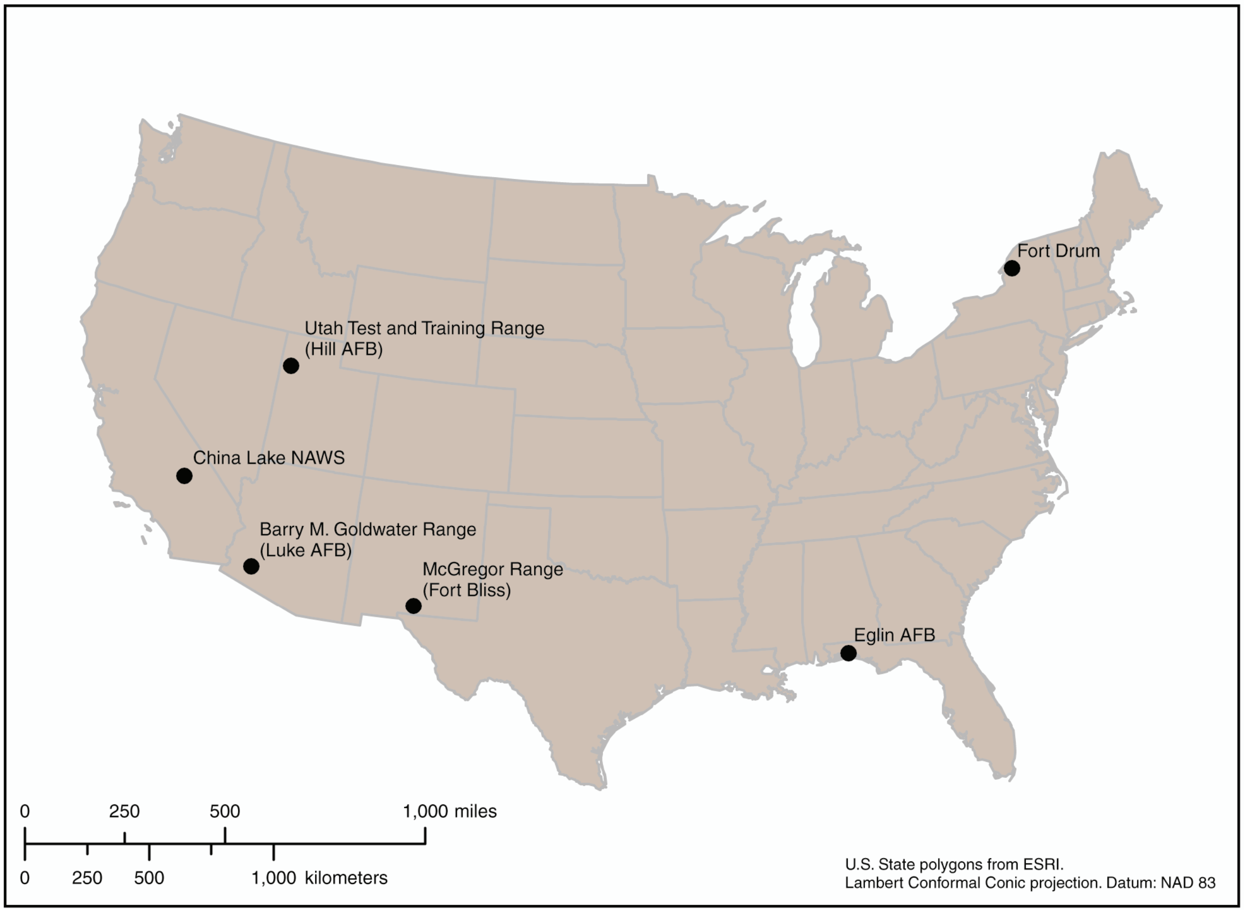

Figure 1.1. Participating military installations and ranges.

of data quality or that such a study would be redundant. The installations on which we focused our three major subjects of study were as follows:

- Survey reliability: Fort Drum, Eglin AFB, the BMGR, and China Lake
- Site location: the UTTR
- Site boundaries: Fort Drum, Eglin AFB, the UTTR, and Fort Bliss

A Note on the Case Studies

An examination of data quality has the danger of being misinterpreted as critical of the methods and practices of the installations used as case studies. Nothing could be further from the truth. We selected installations where archaeological data had been systematically collected to professional standards. Otherwise, we could not have performed our study. The errors we examine are not errors of competence but are those inherent in archaeological methods. Not all DoD installations have data of sufficient quality to have been used in our study.

CHAPTER 2

Survey Reliability

A Class III archaeological survey is defined as an intensive, full-coverage examination of a project area. Archaeologists, managers, and other interested parties (e.g., Native American tribes and state historic preservation officers [SHPOs]) often assume that 100 percent coverage means 100 percent detection of cultural resources contained within a project area. In truth, we never find all archaeological sites in a project area (Plog et al. 1978). More important, the fraction of archaeological sites found can vary tremendously. Understanding the reliability of archaeological surveys is essential for DoD CRM and land managers.

Although it is rarely discussed, perhaps the most important methodological decision a DoD archaeologist will make is how archaeological sites will be found at an installation. Pedestrian surveys—in which surveyors are spaced at regular intervals and visually examine the surface for indications of archaeological sites, mostly in the form of artifacts—are used throughout the western United States, where surface visibility is generally good. In many parts of the country, however, the surface is obscured by dense vegetation, or archaeological sites are buried, with no surface expression. In areas where sites are not visible at the surface, archaeologists use subsurface detection methods to discover sites. The most common method is to excavate shovel-test pits (STPs, or test units) placed along transects at regular intervals.

Archaeologists have long known that survey methods have a significant effect on the discovery rates of archaeological sites and on parameter estimates (for variables such as total number of sites, site density, and proportion of site types) that are derived from sample results (Altschul n.d.; Altschul and Nagle 1988; Homburg et al. 1994; Judge et al. 1975; Plog et al. 1978; Schiffer et al. 1978; Schiffer and Wells 1982). Nance and Ball (1986) have shown that the likelihood of discovering sites with subsurface tests varies directly with the artifact density and the size of the sites (see also Kintigh 1988). Another key variable in determining site-discovery potential is the intensity with which the fill of the test is inspected. The probability of site discovery increases dramatically with a shift from visual inspection of the fill to screening of the fill, and the probability increases still further as the size of the screen mesh decreases. It is worth pointing out, however, that even with a small interval between subsurface tests and screening through fine mesh, the likelihood of missing small low-density sites is usually very high, much higher than the probability of missing comparable sites on pedestrian surveys (Altschul and Nagle 1988:276).

Eglin AFB has provided a good example of the effect that small changes in methods can have on detection rates. Early surveys at Eglin AFB used STPs of approximately 25 by 25 cm. When survey methods switched to STPs of 50 by 50 cm, the number of archaeological sites discovered increased dramatically (Thomas and Campbell 1993). The size and shape of sample units have also been the subject of much discussion within archaeology. Two types of sample units are common in archaeology: square quadrats and linear transects. Much of the discussion surrounding whether to use quadrats or transects concerns the so-called edge effect. The edge effect refers to the effect of survey-unit size and shape on site discovery and parameter estimation. In surveys, all sites discovered in the sample units are usually recorded. Recorded sites include sites that are contained wholly within the unit as well as those that are only partially located in the unit. Thus, when sites are detected near the edges of a survey unit, the crew actually surveys an area that is somewhat larger than the survey unit itself. How much larger depends on the amount of recorded site area that falls outside survey areas. The edge effect is greater as site size increases, as the size of the sample unit decreases, and as the ratio of the length to width of the sample units increases (Figure 2.1).

Altschul and Towner (1994:287) explored this issue with the results from a survey of Area A in the northeastern section of the BMGR in southwest Arizona (now the Sonoran Desert National Monument).

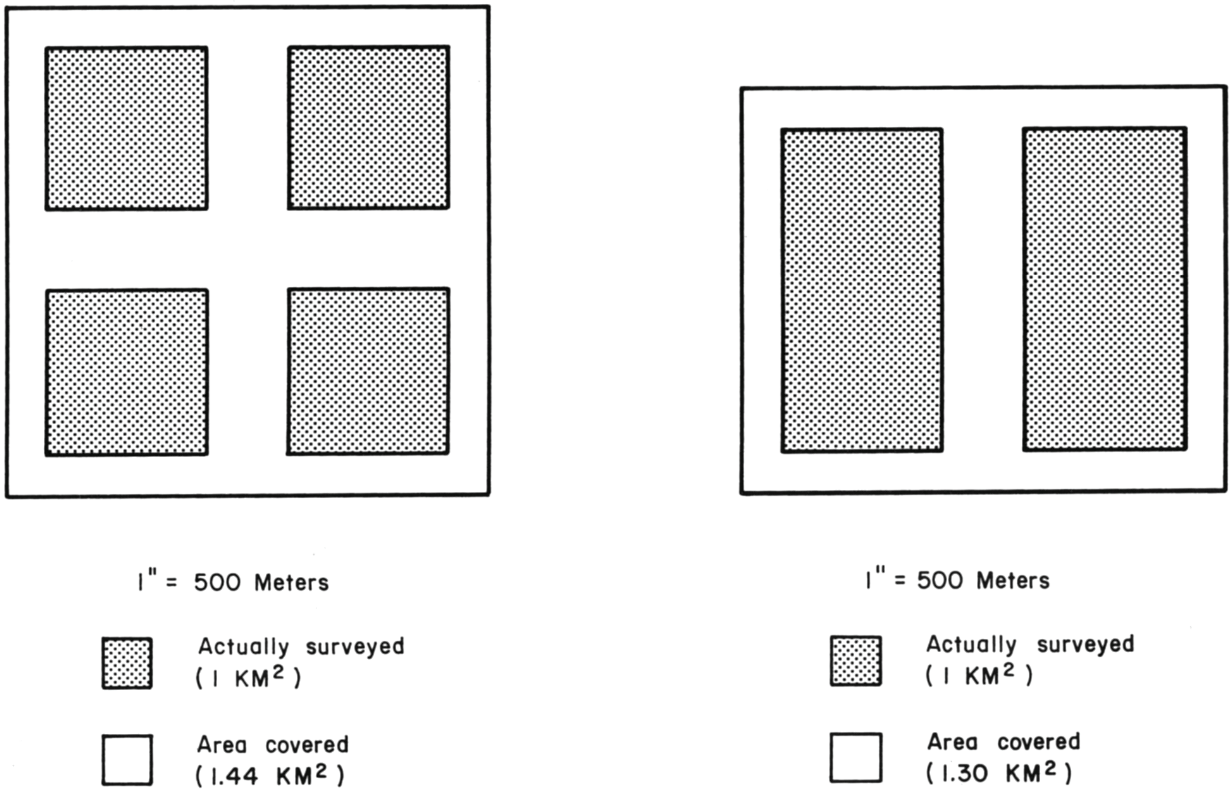

Figure 2.1. Edge effect for survey units with different sizes and shapes (Altschul and Nagle 1988:263). See text for explanation.

A systematic Class III survey of the area had found 130 archaeological sites. Altschul and Towner aimed to resolve the following two questions: (1) How many sites should have been found? (2) What types of sites were missed? By modeling survey reliability according to the edge effect and characteristics of the archaeological record, Altschul and Towner were able to show that nearly all large sites were discovered during the survey but around 10 percent of small sites were missed. They were also able to show that, because lithic scatters and plant-procurement or plant-processing sites were small or had low artifact densities, these site types were underrepresented by survey. By contrast, sites that were larger and had higher artifact densities, such as habitation sites and lithic quarries, were possibly overrepresented as a result of survey.

The survey-reliability problem intensifies as sites become smaller and artifact densities decrease. At Camp Pendleton, where sites are small (site size averages about 112.8 m in diameter, or 10,000 m^2) and have very low artifact densities (fewer than 30 artifacts on average), Altschul (2000) estimated that, in a 3,000-acre survey in which 109 sites were found, as many as 300 were missed. Knowing that site density in this part of Camp Pendleton is actually 1 ± 3 sites per 30 acres—as opposed to thinking that all sites in the project area have been found—can have a tremendous effect on the decisions made by military planners.

Another factor affecting survey reliability concerns decisions affecting the definition of "site." Archaeologists have argued continually over how to define sites (e.g., Dunnell 1971; Ebert 1992; Thomas 1975). Theoretically, we would like to match the definition with attributes of past human behavior that we wish to study. Unfortunately, the archaeological record is messy, and it is often not easy to determine in the field the relationship between artifacts and features, on the one hand, and the past behaviors to which they correspond, on the other (Schiffer 1987). In an attempt to systematize observations made by many different archaeologists over time, most states have taken it upon themselves to define what constitutes a site. Statewide definitions vary from the discovery of a single artifact to discovery of a minimum number of artifacts and artifact classes within a certain area (e.g., 15 artifacts from two artifact classes in 100 m^2).

Beyond official site definitions, which often focus on minimum elements, there is considerable variability in how sites are defined. In CRM, most contractors prefer to define sites as large entities. Site recording is expensive; therefore, the fewer sites that must be recorded, the fewer costs a contractor has to meet. Contractors, then, tend to lump artifact scatters together into one site, often designating each cluster of artifacts a locus. Managers, in contrast, prefer to define sites as small. Small sites result in a maximum of nonsite areas, which can be opened to military activities. Instead of two or three loci at a site, managers tend to prefer to have two or three separate sites, with nonsite areas between them.

At the BMGR, Tagg and Heilen (2008) showed that median site sizes for surveys conducted in the same tactical ranges decreased over time, probably as a result of pressure from land managers to split complex sites into several smaller sites. This methodological shift in site definition resulted in decreased site size, but survey parameters for other variables were affected as well. Because larger sites were split into smaller sites, discovery rates for both sites and isolates increased independently of what was surveyed. The effects of changes in site definition are not benign in terms of cultural resource management. Larger sites place more restrictions on military activities and require greater effort to meet compliance requirements. In contrast, the danger of decreasing the site footprint just to areas that contain surface evidence of cultural material is that significant deposits may be discovered during construction or impacted by other military activities. These unanticipated discovery situations are very costly because they require immediate attention and because military activity has to be stopped or severely restricted.

As discussed above, survey reliability is affected by discovery methods—including survey size, survey shape, sampling intervals, and screen size—and site-definition methods. Accordingly, to investigate data quality in survey reliability we did the following:

- Modeled the effects of pedestrian-survey and shovel-pit-survey methods on discovery rates of sites and on other population characteristics;
- Modeled the effects of survey shape and survey size on discovery rates of sites and on other population characteristics;
- Modeled the effects of any changes over time in site definition on site size, site density, other site characteristics, and the number of discovery situations; and
- Developed metrics to assess variation in site-discovery data stemming from the use of varying survey methods.

The Effects of Survey Methods on Discovery Rates

For decades, most land managers have labored under the false assumption that "full-coverage survey" covers 100 percent of the land within survey units. The reality is that all survey is sample survey; no plot of land is completely surveyed such that every piece of ground is directly observed. Instead, archaeologists sample relatively small areas within a survey unit and use those results to characterize the entire survey unit as well as areas outside survey units. Installation archaeologists rarely consider in any formalized manner the specific characteristics of those samples or the effect that sample characteristics have on issues such as parameter estimation (for variables such as site density and average site size) or site evaluations.

In this section we discuss several approaches to assessing survey reliability for shovel-test and pedestrian sample survey. Several metrics are introduced and their general implications for survey reliability discussed. These include the following:

- Sample fraction;
- Probabilities of intersection, detection, and discovery;
- Largest site missed;
- Test-unit density;

- The edge effect; and
- Percentage of sites missed.

Each of the above metrics is discussed below. Discussion of metrics is followed by discussion of how these metrics can be applied in the assessment of survey reliability at selected installations.

Sample Fraction

Sample fraction is simply the amount of area of a survey unit that is covered by transects or test units. For test-unit sampling, archaeologists excavate a hole of a standard size in locations specified by their sampling strategy. The sampling fraction for a particular survey unit is thus the product of the number of test units and the test-unit area, divided by the survey-unit area. Because of the small size of most test units used in sample survey, sample fraction is often quite low. A survey area that is 1,000 by 1,000 m in size might contain 400 units spaced 50 m apart. If these units were 50 by 50 cm, or 0.25 m^2, then the sample fraction would be 0.25(400) ÷ 1,000,000, or 0.0001.

Typically, sample fraction is considerably higher for pedestrian survey, although pedestrian survey also has its limitations, particularly given that landforms are not sampled at depth and sediments are typically not screened or closely inspected. For most pedestrian survey, the sample is essentially a series of rectangular swaths centering on the axis of each survey transect. Sample fraction is roughly equivalent to transect width (e.g., 3 m) divided by transect spacing (e.g., 20 m). When survey units are rectangular and a transect width of 3 m is assumed, the sample fraction for a pedestrian survey that employs 15-m intervals is approximately 0.2. A pedestrian survey that instead employs 30-m intervals results in a sample fraction of less than 0.1, or half that of the survey that uses 15-m spacing. Even less intensive, a pedestrian survey that employs 60-m spacing results in a sample fraction of approximately 0.05. If managers wanted to ensure a 50 percent sample and assumed, as we did, that transects are roughly 3 m wide, then field crew would have to survey transects spaced 6 m apart (Figure 2.2). Ultimately, as transect or test-unit spacing increases, the sample size decreases and the size and number of sites potentially missed by survey increases. Because the size of a site not intersected by at least one transect increases with spacing, small sites are sampled at a lower intensity than bigger sites. The implication is that small sites will be underrepresented by survey, particularly as transect or test-unit spacing increases.

Probabilities of Intersection, Detection, and Discovery

Much of the discussion that follows is predicated on several interrelated probabilities: those for intersection, detection, and discovery. The probability of intersection equals 1 when a site of a given size and shape will always fall within at least one test unit or transect. The probability is less than 1 when a site is small enough that it will not always fall within at least one transect or test unit. The probability of detection equals 1 when one or more artifacts will always be discovered within a test unit or transect of a given size. The probability of discovery is the probability that both of the following will occur: a site will be intersected and the elements of the site will be detected.

The probability of discovery is thus dependent on the probability of intersection and the probability of detection (Shott 1985:462). Shott, in fact, modeled the probability of discovery as the product of the probabilities of intersection and detection.

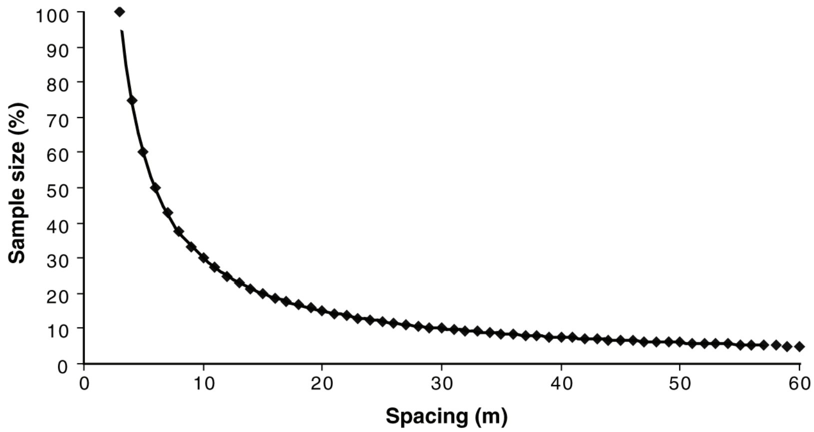

Figure 2.2. The relationship between spacing and sample size for pedestrian transects. Transect width is assumed to be 3 m.

The Probability of Intersection: Pedestrian Survey

Pedestrian survey is typically conducted in areas, such as the western United States, where surface visibility is relatively high. Most observations are made from a standing position on the ground surface immediately in front of each crew member (for exceptions, see Burger et al. 2002–2004; Ebert 1992; Wandsnider and Camilli 1992). Because most sites are discovered as the result of detecting one or more artifacts or subtle indications of cultural features, archaeologists literally discover diagnostic indicators of a site at their feet as they walk over them.

Most pedestrian surveys conducted within the past two decades in the western United States reported transect spacing of 15, 20, 25, or 30 m. Earlier surveys sometimes used transect spacing as wide as 50 or 60 m or varied spacing based on assumed differences in site density. That is, if site density was thought to be low, then a wide spacing interval was used to minimize the time wasted on areas of low site density and to maximize the time spent on areas considered more important, or on areas of presumed higher site density. As some investigators have pointed out, however, such an approach could just as easily have focused higher-intensity survey on areas hypothesized to be of low site density, in order to test assumptions about site density.

For regularly spaced pedestrian transects and circular sites, the probability of intersection (p_i) is simply equal to site diameter (d) divided by spacing (s) (Banning 2002), such that

$$p_i = d/s.$$

If we model transects as rectangles of standard width (w), rather than lines, then

$$p_i = d/(s - w).$$

For site diameters smaller than $s - w$, p_i increases linearly with site diameter (Figure 2.3).

The Probability of Intersection: Shovel-Test Survey

Determining the probability of intersection for shovel testing is somewhat difficult, because the relationship is nonlinear and an equation has not been developed to describe the entire relationship (Krakker et al. 1983). Furthermore, the probability of intersection varies depending on whether STPs are arranged according to square, rectangular, staggered, or hexagonal grid patterns (Kintigh 1988). Krakker et al. (1983:Figure 2)

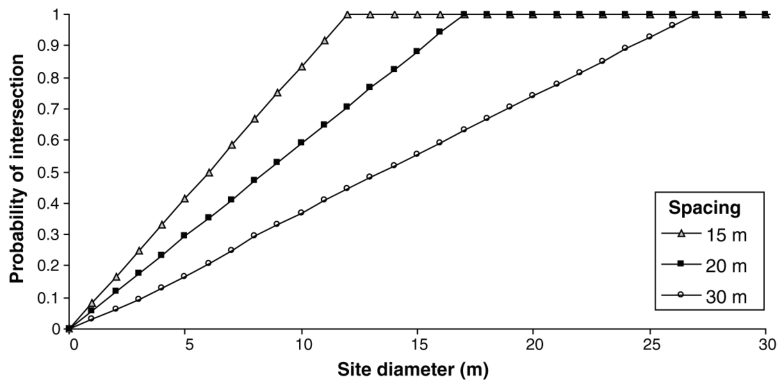

Figure 2.3. The probability of intersection for 15-m, 20-m, and 30-m spacing, when transect width is 3 m.

graphed the relationship between STP spacing and the probability of intersection for square grids and staggered grids using probability tables from Savinskii (1965). On the basis of their graph, Krakker et al. (1983) estimated that there is only a "0.78 probability of finding a site with a diameter equal to the grid interval" for square grids (Krakker et al. 1983). Sites half that size have a probability of intersection around 0.2. In order to have a 50 percent chance of intersecting a site, a site must have a diameter that is $^4/_5$ of the grid interval.

Although perhaps not as commonly used, staggered grids are more efficient than square grids in site discovery. Krakker et al.'s (1983:Figure 2) graph indicates that for staggered grids, the probability of discovering a site with a diameter equal to grid spacing is around 0.92; the probability of intersection is around 0.42 for sites half that size. In order to have a 50 percent chance of intersecting a site with a staggered grid, a site must have a diameter slightly larger than half the grid interval.

Krakker et al. (1983) were able to develop an equation for the probability of intersection for square grids for sites whose diameters are less than half the STP spacing. According to their equation, for sites with diameters less than half the survey spacing, p_i is equal to site area divided by a square with sides equal to s, such that

$$p_i = \pi r^2/s^2.$$

The probability of intersection for shovel-test survey is probably most accurate when simulated. Fortunately, Kintigh (1988) has produced a computer program that can be used to estimate the probability of intersection for circular sites of a given size (see below).

The Probability of Detection

Detection is highly dependent on factors such as artifact density, artifact or feature characteristics (e.g., size, shape, color, and preservation), vegetation, or burial. As a result, detection is affected by the complex interaction of multiple factors and is context-specific. One way to model detection is through use of probability distributions. The probability of detecting (p_d) a single artifact in a test unit can be calculated using the Poisson distribution, such that

$$p_d = 1 - e^{-kad},$$

where a is unit area, d is artifact density, and k is a factor that accounts for the effects of visibility, preservation, disturbance, and so on. Although the Poisson distribution can be considered an optimistic method for estimating the probability of detection, the probability of detection calculated using this method is quite low for characteristic test-unit sizes and artifact densities commonly found at military installations (Figure 2.4). For 30-by-30-cm units and $k = 1$, the probability of detecting a single artifact in a test unit is only 0.50 for artifact densities of 7.8 artifacts/m^2 and is 0.95 once artifact densities reach 33.3 artifacts/m^2. For a more typical artifact density of 1 artifact/m^2, the probability of detecting a single artifact in a 30-by-30-cm test unit is less than 0.09. The probability of detection increases substantially with test-unit size but not enough to elevate the probability of detection to a high level of confidence for most archaeological contexts at military installations. For 50-by-50-cm units and $k = 1$, artifact density must be greater than 12 artifacts/m^2 for a probability of 0.95 or more. For artifact densities of 1 artifact/m^2 and $k = 1$, the probability of detecting at least 1 artifact in a 50-by-50-cm unit is only 0.22. When factors such as visibility, preservation, or disturbance are taken into account (i.e., $k < 1$), the probability of detecting an artifact in a single unit is smaller still for typical archaeological contexts.

The low probability values for detecting artifacts in test units have serious consequences for site discovery. We used the probability of detection and the binomial distribution to calculate how many units (U) must be dug in order to guarantee detection of an artifact from a site, given a particular density, unit size, and confidence (c) level, such that

$$U = \ln(1 - c) \div \ln(1 - p_d).$$

For a 95 percent confidence level ($c = 0.95$), U is roughly equivalent to 3 divided by the product of unit area (a) and artifact density (d), such that

$$U = 2.9975 \div ad.$$

A practical implication of these calculations is that for large sites with low artifact densities, often only one or a few STPs will be positive within a large site. Most STPs falling within the site will be negative, because of the low probability of detection. There is a real risk in such situations of interpreting a positive test as either (1) an isolated find or (2) part of a much smaller site than the one actually tested. Installations have responded differently to this conundrum. Fort Drum, for instance, generally represents sites discovered through survey as points, so as not to assume that small numbers of positive tests correspond to a small site. Eglin AFB, by contrast, uses the location of positive test pits and other information to infer site boundaries where tests are positive.

For low-density sites and small unit sizes, more than 100 STPs could be required to discover a single artifact (Figure 2.5). In cases where a test is positive in a large low-density site, the likelihood that the footprint of the entire site could be inferred is extremely low. We can extrapolate from this scenario that many sites discovered through shovel-test survey could be substantially larger than the distribution of positive STPs indicates and that it may be very difficult statistically to differentiate between isolated finds and sites. In addition, areas that have no or few positive tests could still contain small to medium-sized low-density sites. Conversely, areas that have many positive tests could contain, rather than a handful of small sites and isolated finds, one or more large sites. One of the management implications of this situation is that the size and extent of many sites discovered are largely unknown and must be determined through additional work. In such cases, site boundaries may best be considered placeholders requiring further evaluation. Estimates of the number of units (U) based on the probability of detection theoretically could be used to determine the area that must be considered in order to determine site boundaries. We suspect that because of low detection probabilities in many installation contexts, the area under consideration should be substantially larger than the immediate area around one or few positive STPs.

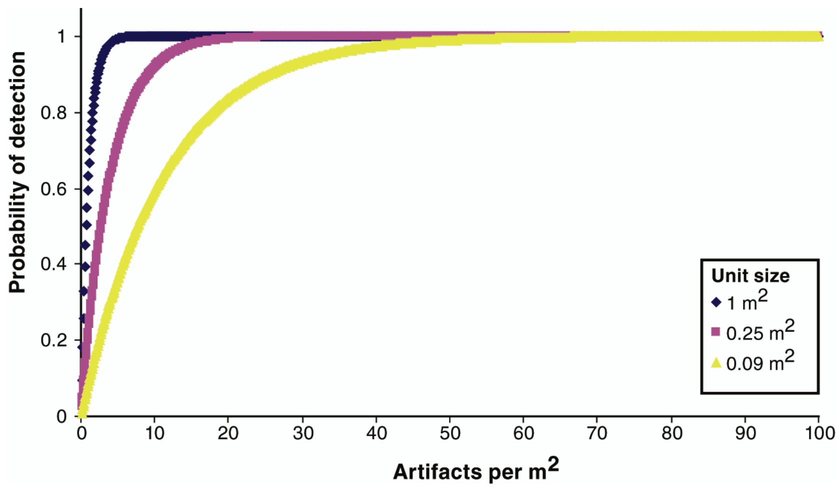

Figure 2.4. The probability of detecting at least one artifact for unit sizes of 0.09 m^2, 0.25 m^2, and 1.0 m^2.

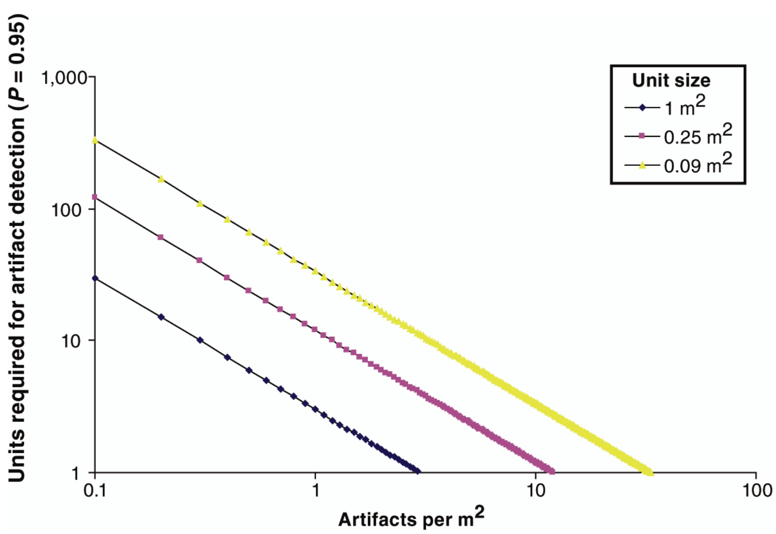

Figure 2.5. The number of units required to guarantee detection of at least one artifact for unit sizes of 0.09 m^2, 0.25 m^2, and 1.0 m^2 (*P* = 0.95). Both scales are logarithmic.

The Probability of Discovery: Shovel-Test Survey

As the foregoing discussion shows, there are serious inadequacies with shovel-test sampling in discovering sites, particularly due to the low probabilities of detecting artifacts in widely spaced test units. In the absence of other discovery methods, shovel-test survey cannot be considered a highly reliable method for site discovery. For artifact densities characteristic of many archaeological contexts, the probability of discovery falls well below the probability of intersection for sites of a particular size. As Shott (1985:466) noted:

> The chief problem with shovel-test sampling in regional survey is not the method's inadequacy in any objective sense, but the lack of congruence between what it is used for and what it can do. Shovel-test sampling is a perfectly valid method for the estimation of survey parameters, such as the density of cultural materials across a region. That is, results of shovel-test sampling are best viewed as an estimate of the proportion of the survey region occupied by cultural material. As the preceding discussion shows, however, it is poorly suited for the task of discovering specific sites. Unfortunately, most applications are concerned with the latter goal, usually because they are cultural resource management investigations. The chief goal of such investigations is to discover any sites that may exist in the survey area.

Given the number of variables involved, it is difficult to estimate the probability of site discovery using mathematical equations. Fortunately, Kintigh (1988) developed a computer simulation program that allows users to estimate the probability of discovery based on grid pattern, spacing, unit size, and artifact density. We use Kintigh's program below to estimate discovery probabilities for a variety of survey methods used at military installations. Unfortunately, to our knowledge a similar program is not available for pedestrian survey, although such a program would be very useful for assessing the reliability of pedestrian surveys.

The Probability of Discovery: Transects

DoD CRM managers commonly assume that pedestrian surveys are very reliable. Usually the reliability is overstated, however. For example, much of Camp Pendleton in southern California lies in coastal grasslands, which obscure the surface. Archaeological surveys were conducted for many years as though the visibility factor was not significant. After wildfires burned significant portions of the installation, archaeologists resurveyed areas, finding in some cases a tenfold increase in the number of sites. Installation managers believed that under these conditions they were finding a very high percentage of sites. Altschul (n.d.), however, was not convinced. He reasoned that, because sites averaged 10,000 m^2 in size and most sites had 30 or fewer artifacts, seven transects spaced at 15-m intervals would traverse a circular site whose center fell within one of the transects. With a transect width of 2 m, Altschul calculated that the transects, when combined, would sample 14.6 percent of the site and that on average 4.3 artifacts would be discovered within the sample. It is unclear how many artifacts must be encountered before archaeologists will recognize a site. Many of the artifacts at Camp Pendleton were small pieces of lithic debitage, many of which were dark in color and lying on a burned surface. Altschul suggested conservatively that archaeologists would have to pass over 8 artifacts within the discovery potential of their transects to recognize a site. According to this scenario, roughly 50 percent of sites would be discovered through survey.

We have brought a greater degree of rigor to Altschul's "thought problem" of survey reliability by using Kintigh's program to simulate pedestrian survey. We have adapted the program to pedestrian survey by stringing a series of 3-m-diameter units spaced 3 m apart along transects spaced according to standard spacing intervals.

Largest Site Missed

The largest site missed indicates the largest site size that can fail to be discovered through a given survey method. These are sites whose sizes and artifact densities are small enough that, given a particular discovery method, the probability of discovery is less than 1. As discussed above, simply intersecting a site with a test unit or a pedestrian transect does not guarantee site discovery. Components of the site, such as artifacts, features, or deposits, also have to be detected in order for a site to be discovered.

Values for specific survey variables, such as spacing and unit size, can be used to calculate the largest site that can potentially be missed by survey. This information can then be used to assess survey reliability for specific survey methods. The specific sizes of sites potentially missed by survey depend on site shape, the spacing of transects or test units, and, in the case of test units, the geometry of test-unit grids. For the purpose of this report, we model sites as circular. Some mathematical models have been developed for elliptically shaped sites, but these are somewhat unwieldy and less widely applicable (Banning 2002).

Estimates of the largest site missed that are based on the probabilistic discovery of at least one artifact are fairly optimistic because they assume that if one artifact is discovered, the corresponding site will also be discovered. The problem with this assumption is that unless artifacts are clustered, the discovery of one or more artifacts within a large low-density site may easily be interpreted as an isolated find. There is no clear information on how many artifacts must be discovered in order for investigators to recognize them as part of a site. Such a threshold will also vary according to bureaucratic definitions of sites. In some jurisdictions, a single artifact may considered a site; in other jurisdictions, a specific quantity of artifacts within a certain area, of associated features, or of associated features and artifacts is required to define sites.

It may also be possible to use U to estimate the size of a site that could potentially be missed, on the basis of the unit size and spacing. For instance, if $U = 23$, then in order to guarantee the detection of at least one artifact from the site, the site would have to be large enough to include 23 test units, based on a particular sampling strategy. It follows, then, that as spacing increases, the largest site missed will also increase, as larger sites will be required for at least 23 units to fall within a site.

Kintigh provided equations that can be used estimate the largest site *not intersected* for a variety of grid patterns. The diameter (d) of the largest site not intersected is

$$d = \lambda s,$$

where s is spacing and λ is $\sqrt{2}$ for square grids, 4/3 for hexagonal grids, and 1.25 for staggered grids. For arbitrary rectangular or staggered grids, both row spacing (s) and column spacing (i) must be taken into account. For arbitrary rectangular grids,

$$d = \sqrt{(s^2 + i^2)}.$$

For arbitrary staggered grids,

$$d = s + i^2/4s.$$

We use Kintigh's simulation program to calculate the largest site missed for surveys at Eglin AFB and Fort Drum.

Largest Site Missed: Pedestrian Transects

For regularly spaced parallel transects, the diameter of the largest site that can fail to be *intersected* is simply equal to the survey interval, or if transect width is taken into account, the survey spacing minus transect width (Figure 2.6). When both intersection and detection are taken into account, the largest site missed is more difficult to calculate. We use Kintigh's simulation program to estimate the largest site

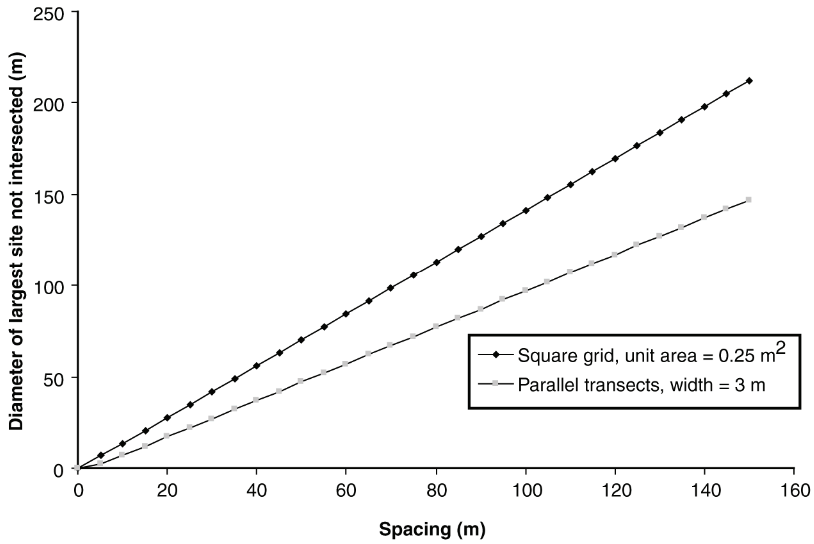

Figure 2.6. Comparison of the largest site not intersected for square grids and pedestrian transects, according to spacing.

missed for pedestrian transects, although a simulation designed specifically for simulating pedestrian survey would probably produce more-accurate results.

Test-Unit Density

STPs often are placed systematically, according to a regular pattern, such as a square, rectangular, staggered, or hexagonal grid. Each kind of grid pattern has an effect on the probability that a site of a given size will be intersected. In general, hexagonal and staggered grids are more likely than square or rectangular grids to find circular sites of a given size (Kintigh 1988; Krakker et al. 1983; Nance and Ball 1986; Shott 1985). Often, however, actual placement of STPs in a survey area deviates from the ideal pattern because of local variation in environmental or archaeological context, as well as inaccurate assessment of STP location. Sometimes, STP locations are moved to nearby locations or planned STPs are skipped, because of factors that make an STP impractical (i.e., dense vegetation, pavement, standing water), dangerous (i.e., animal nests, unexploded ordnance) or unnecessary (i.e., obvious disturbance) to dig. As a result, STP arrays sometimes follow a more irregular pattern, and because of the movement or avoidance of STP locations, some areas within survey units are not actually tested with STPs. Therefore, a useful proxy measure for the intensity of survey effort in a particular survey area is the number of STPs per acre.

The calculation of STP density, however, depends on the size of the area used to make the calculation. We suggest including in the calculation only areas where STPs were actually excavated. One way to calculate tested areas in geographic information systems (GIS) is to create a raster grid in which each cell records the distance to the nearest STP location. The grid should then be reclassified so that only cells that fall within a specified distance are made equal to 1 (i.e., where distance to STP is less than or equal to spacing). The reclassified grid can then be converted into a polygon coverage and the resulting polygons buffered by the negative of one-half of the test-unit spacing.

In addition to a more accurate estimate of STP density, calculation in GIS of the areas actually tested, based on the location of excavated test units, offers two other benefits. It allows estimation of the proportion of survey units subject to testing, and it provides sample areas that can be used in parameter estimation, without including areas within survey units that were not actually tested with STPs.

The Edge Effect

Estimates of the largest site missed as a result of survey methods strongly suggest that small sites and low-density sites are typically underrepresented by survey. The major differences between survey methods in the largest site missed result from differences in spacing, unit size, and artifact density. By contrast, the edge effect can result in the overrepresentation of sites discovered around the edges of survey units. Theory suggests that sites discovered through the edge effect tend to be larger on average, because larger sites centered outside survey units have a greater chance of extending into survey units. The edge effect is most pronounced for small survey units (e.g., small quadrats) and units whose perimeters are large with respect to survey unit area (e.g., long, linear units) (see Figure 2.1).

According to Plog et al. (1978), the edge-effect area can be calculated by estimating the average site diameter and using that site size to buffer survey areas. The amount of edge-effect area can then be expressed as the percent increase in effective survey area based on the edge effect.

Another way to measure the edge effect is to calculate in GIS the number of sites and the amount of site area discovered inside and outside survey units during a particular survey. The number of sites falling outside survey units can be expressed as a percentage of sites discovered during the survey. Similarly, the amount of survey area falling outside survey units can be expressed as a fraction of the total site area discovered during survey. These numbers can then be used to examine which survey areas had a discernible edge effect and which areas did not.

The edge effect can be significant with respect to parameter estimation, because discoveries outside survey areas that were recorded because of the edge effect may overrepresent sites or site characteristics not actually discovered within survey units. One way to account for the edge effect in parameter estimation is to use only portions of sites falling within the survey unit or only sites whose centers fall within the survey unit (Banning 2002). A problem with this approach, however, is that sites are often recorded as units, such that the different attributes or components of a site are treated as a whole. Site attributes or components may be difficult to separate on the basis of whether they fall inside or outside survey units, depending on how sites were recorded. If sites are recorded in great detail, and individual artifacts and features are accurately mapped, it may be possible to use only features and artifacts discovered within the survey unit in parameter estimation. Below, we estimate the edge effect for surveys at the BMGR and China Lake, showing that, as expected, the edge effect is most pronounced for small survey units and long, linear units. Units with a pronounced edge effect would have to be paid the closest attention in correcting for the edge effect on parameter estimation.

Percentage of Sites Missed

Correcting for discovery bias is not an easy task because so many variables are involved. One potentially useful way to correct for discovery bias, however, is to calculate, for each site, the probability of discovery, based on artifact density, size, and survey method. For eastern U.S. installations where shovel-test sampling is the norm, such an approach could be problematic because of a lack of accurate information on site size. In western U.S. installations where pedestrian survey is used, site size may tend to be more accurate but will nonetheless still vary according to recording methods and other factors. With these caveats in mind, the probability of discovery for each site in a survey unit can be used to estimate the number of sites that may have been missed during survey. Such an approach is most easily performed by organizing sites into discrete classes, such as small low-density sites (e.g., less than 20 m in diameter and with 0.2 artifacts/m^2), because discovery probabilities must be determined through simulation using specific site sizes and artifact densities or must be assumed on the basis of other information. Using the probability of discovery, installation archaeologists can calculate how many sites are expected and can use that information to calculate the percentage of sites missed. For instance, if five small low-density sites were discovered in a survey area and each had a 0.4 probability of discovery, we might expect that 12.5 similar sites were present in the survey area (5 ÷ 0.4). The sum, across sites, of the inverse of the probability of discovery for each site can be used as an estimate of the total number of sites within a survey area. The total number of sites missed can be calculated by subtracting the number of discovered sites from the estimated total number of sites. Those last two numbers can then be used to calculate the percentage of sites missed, as well as to estimate the total site area that could have been missed. Below, we estimate percentage of sites missed for surveys at the BMGR, Eglin AFB, and Fort Drum.

BMGR Survey Reliability

More than 180,000 acres have been surveyed at the BMGR East since the 1960s. Contractors have completed inventory of lands identified as being most affected by military activity, resulting in the identification of 1,269 sites and a large number of isolated occurrences. Most survey at the BMGR East was conducted by SWCA, Inc., SRI, or Dames & Moore in the 1990s or 2000s. More than 15,000 acres were surveyed by multiple organizations before 1992, mostly during the 1970s. For the purpose of this discussion, we focus on survey performed since 1992 by SWCA, SRI, and Dames & Moore. Most survey since 1992 has used 15- or 20-m transect spacing. SWCA's and Dames & Moore's survey in the 1990s employed 20-m transect spacing; SRI's survey between 1999 and 2004 employed 15-m spacing. Surveys conducted in the 1960s through 1980s employed highly variable spacing, ranging anywhere from 10 to 60 m. For some surveys conducted before 1992, spacing varied according to assumed site density. If resource density was expected to be low, then spacing intervals were increased, presumably to decrease the level of effort necessary to "clear" survey areas.

The BMGR data indicate that sites with maximum dimensions as small as 2 m have been routinely recorded. Since 1992, the smallest site recorded in any given year had an area of less than 12 m^2, suggesting that small sites are discovered and recorded. Many of these consist of a single feature, often a thermal feature, associated with one or more artifacts.

Edge Effect at the BMGR

There should be a detectable edge effect for BMGR surveys, given that survey areas vary considerably in size and shape. We expected that small survey quadrats and relatively linear survey areas would have a pronounced edge effect, whereas large survey blocks would have only a small edge effect. Following Plog et al. (1978), we estimated the average radius of sites at the BMGR East to be approximately 30 m and buffered survey-unit polygons using that fixed linear distance. Edge-effect area (the amount of additional area resulting from the edge effect) increases at a slower rate than survey-unit area. We calculated effective inventory area as a percentage of the survey-unit area, in order to gauge the increase in sample size represented by the edge effect; the result was a value that ranged from 3 to 5,822 percent. We found that for BMGR surveys, as survey size increased, the effective inventory area decreased by roughly the square root of the survey-unit area. As expected, large survey blocks had only a small increase in sample size due to the edge effect, whereas some small quadrats and linear survey areas had effective inventory areas many times the size of the survey unit.

Calculation of the amount of site area falling outside survey units bears this out well. Large survey blocks had little to no increase in the number of sites or amount of site area resulting from the edge effect. The edge effect was mostly discernible for small quadrats and linear survey units. We suggest that parameter estimation that involves these survey areas should take the edge effect into account by considering only site areas or centroids found within survey units (Figure 2.7; Table 2.1). For example, if three survey areas had substantial edge effects as determined by this method, then edge sites discovered during those surveys would have to be treated specially for parameter estimation. If site density is estimated, then only sites with centroids within survey areas should be included in the calculation. If, instead, site areas are used in calculations, then only site areas that fall within survey units should be included. If one wanted to calculate the proportion of different site types in a survey area or set of survey areas, for instance, one approach to

Table 2.1. Edge Effect for Surveys at the BMGR East

Edge Effect	Total Survey Acres	Percent Surveyed Area	Survey Units	Percent Survey Units
Little or none	85,536.7	46.5	70	26.6
Small	48,564.2	26.4	10	3.8
Medium	8,684.2	4.7	5	1.9
Large	20,518.1	11.2	20	7.6
No sites	20,494.5	11.2	158	60.1
Total	183,797.7	100.0	263	100.0

address the edge effect would be to calculate edge sites as fractions. For instance, if a survey had three flaked stone scatters, all of which were edge sites, and seven ceramic scatters entirely within the survey unit, then to calculate proportions of site types, each ceramic scatter would count as one site, but the flaked stone scatters would count as fractions. For instance, if exactly one-half of each flaked stone scatter fell within the survey area, then we would calculate that the survey discovered 1.5 flaked stone scatters (0.5 + 0.5 + 0.5) and seven ceramic scatters, rather than three flaked stone scatters and seven ceramic scatters. The proportion of flaked stone scatters would then be 0.18 (1.5 ÷ 8.5). The proportion of ceramic scatters would be 0.82 (7 ÷ 8.5). If the edge effect were not accounted for, then the estimated proportions would instead be 0.30 for flaked stone scatters (3 ÷ 10) and 0.70 (7 ÷ 10) for ceramic scatters. In other words, without accounting for the edge effect, the relative proportion of flaked stone scatters would be overestimated and the relative proportion of ceramic scatters would be underestimated.

Largest Site Missed and Percentage of Sites Missed

We used Kintigh's simulation program to estimate the discovery probabilities for transects spaced 15 m apart by stringing a series of 3-m-diameter units 3 m apart along transects (Figure 2.8). This approach assumes that field crew will stop and look at closely spaced units as they walk along transects. A simulation designed specifically for pedestrian survey would probably provide more-refined results, but the adaptation of Kintigh's simulation program to pedestrian survey provides at least a rough estimate of discovery probabilities. Because of limitations on the number of shovel tests that can be simulated with the program (n = 800), survey areas had to be smaller than those simulated for test-unit survey. Using this approach, the largest site missed for an artifact density of 0.1 artifacts/m^2 is around 27 m in diameter. Discovery probabilities are fairly close to intersection probabilities for artifact densities that are greater than approximately 0.5 artifacts/m^2. Detection is mainly a problem for very low artifact densities or low visibilities. If we assume, for instance, that artifact density is approximately 0.1 artifacts/m^2, but visibility is around 50 percent, than the effective density is 0.05 artifacts/m^2 and the largest site missed would be 36 m in diameter. If we instead assume visibility of 10 percent, then the largest site missed would be 96 m in diameter. Since visibility is a problem that diminishes probabilities of detection, particularly for pedestrian survey, we might expect that the largest site missed for a typical pedestrian survey using 15-m spacing would be on the order of 30–50 m in diameter.

Evaluation of the largest site missed for BMGR surveys suggests that large numbers of small sites and low-density sites could have been missed by survey (Figures 2.9 and 2.10; Tables 2.2 and 2.3). For particular organizations and installation areas, estimates of the number of sites not intersected average between approximately 13 and 18 percent; estimates of the number of sites not discovered average between approximately 31 and 36 percent. Many small sites at the BMGR are small flaked stone scatters, lithic scatters, ceramic scatters, flaked stone scatters with features, lithic scatters with features, or ceramic scatters

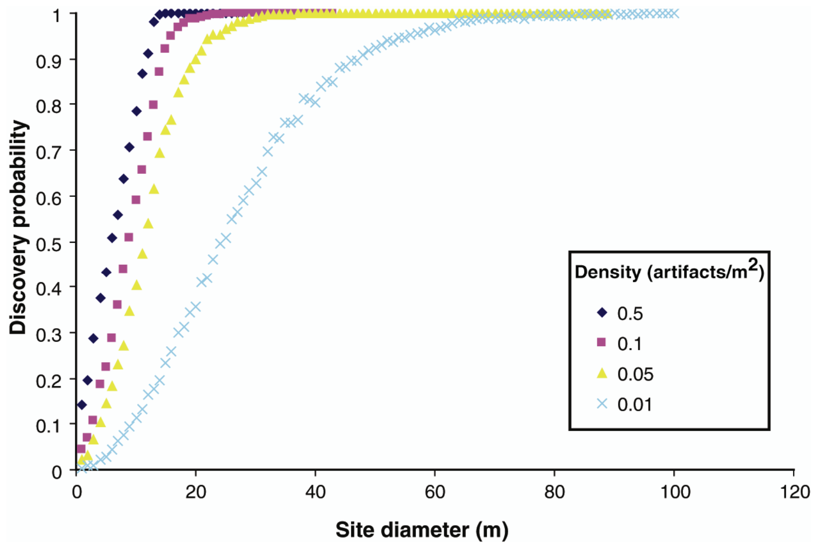

Figure 2.8. The probability of discovery for 15-m spacing of 3-m-wide pedestrian transects, according to artifact density and site size.

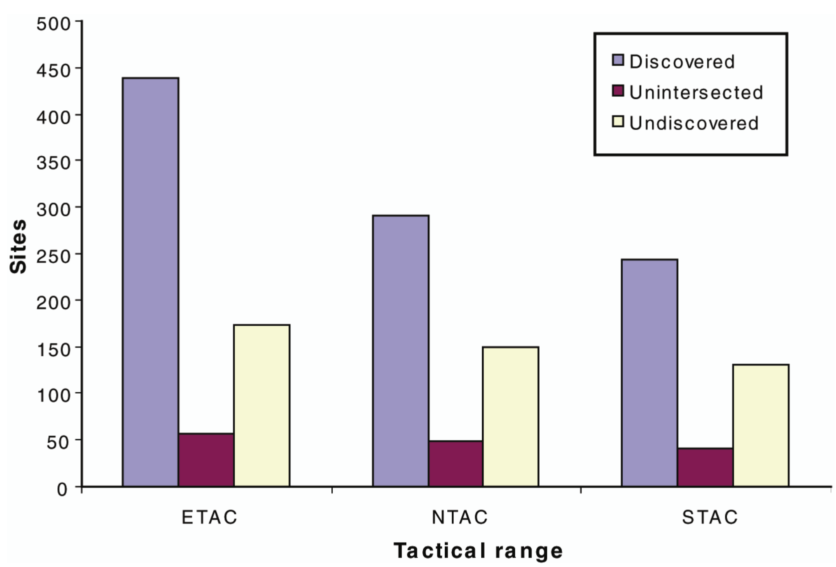

Figure 2.9. Discovered, unintersected, and undiscovered sites for surveys on ETAC, NTAC, and STAC at the BMGR East.

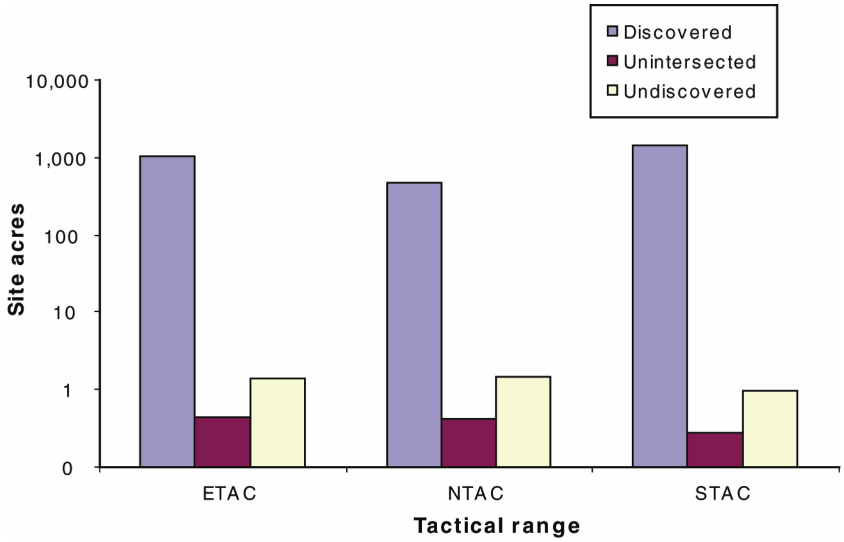

Figure 2.10. Discovered, unintersected, and undiscovered site acres for surveys on ETAC, NTAC, and STAC at the BMGR East. The *y*-axis is logarithmic.

Table 2.2. Estimates of the Number of Sites Missed by Surveys on ETAC, NTAC, and STAC at the BMGR

Tactical Range	Dames & Moore	SRI	SWCA	Total
	Discovered Sites			
ETAC	22	180	237	439
NTAC	32	138	121	291
STAC	13	103	128	244
Total	67	421	486	974
	Intersection Estimate			
ETAC	28.9	199.2	267.8	495.9
NTAC	36.6	160.8	142.4	339.8
STAC	17.6	113.2	154.8	285.6
Total	83.1	473.2	565.0	1,121.3
	Discovery Estimate			
ETAC	38.7	239.7	334.3	612.7
NTAC	42.6	215.3	182.1	440.0
STAC	22.3	139.5	213.2	375.0
Total	103.6	594.5	729.6	1,427.7
	Unintersected Sites			
ETAC	6.9	19.2	30.8	56.9

continued on next page

Tactical Range	Dames & Moore	SRI	SWCA	Total
NTAC	4.6	22.8	21.4	48.8
STAC	4.6	10.2	26.8	41.6
Total	16.1	52.2	79.0	147.3
	Percent Not Intersected			
ETAC	23.9	9.6	11.5	15.0
NTAC	12.6	14.2	15.0	13.9
STAC	26.0	9.0	17.3	17.4
Total	20.8	10.9	14.6	15.4
	Undiscovered Sites			
ETAC	16.7	59.7	97.3	173.7
NTAC	10.6	77.3	61.1	149.0
STAC	9.3	36.5	85.2	131.0
Total	36.6	173.5	243.6	453.7
	Percent Sites Not Discovered			
ETAC	43.1	24.9	29.1	32.4
NTAC	24.9	35.9	33.6	31.5
STAC	41.6	26.2	40.0	35.9
Total	36.5	29.0	34.2	33.2

Table 2.3. Estimates of the Number of Site Acres Missed by Surveys on ETAC, NTAC, and STAC at the BMGR

Tactical Range	Dames & Moore	SRI	SWCA	Total
	Discovered Site Acres			
ETAC	232.6	358.5	463.6	1,054.7
NTAC	85.7	177.3	218.2	481.2
STAC	766.3	432.7	259.3	1,458.3
Subtotal	1,084.6	968.5	941.1	2,994.2
	Estimated Total Site Acres (Intersection)			
ETAC	232.6	358.7	463.9	1,055.1
NTAC	85.7	177.5	218.4	481.7
STAC	766.3	432.8	259.5	1,458.6
Subtotal	1,084.7	969.0	941.8	2,995.4
	Estimated Total Site Acres (Discovery)			
ETAC	232.7	359.0	464.4	1,056.1

Tactical Range	Dames & Moore	SRI	SWCA	Total
NTAC	85.8	178.2	218.7	482.7
STAC	766.4	433.1	259.8	1,459.3
Subtotal	1,084.9	970.3	942.9	2,998.1
	Unintersected Site Acres			
ETAC	0.1	0.1	0.3	0.5
NTAC	0.1	0.2	0.2	0.5
STAC	0.0	0.1	0.2	0.3
Subtotal	0.2	0.4	0.7	1.3
	Percent Unintersected Site Acres			
ETAC	0.0	0.0	0.1	0.0
NTAC	0.1	0.1	0.1	0.1
STAC	0.0	0.0	0.1	0.0
Subtotal	0.0	0.0	0.1	0.0
	Undiscovered Site Acres			
ETAC	0.1	0.4	0.8	1.3
NTAC	0.2	0.8	0.5	1.5
STAC	0.1	0.4	0.5	1.0
Subtotal	0.4	1.6	1.8	3.8
	Percent Undiscovered Site Acres			
ETAC	0.0	0.1	0.2	0.1
NTAC	0.2	0.5	0.2	0.3
STAC	0.0	0.1	0.2	0.1
Subtotal	0.0	0.2	0.2	0.1

with features. Typically, these are sites with 30 or fewer artifacts and no features or one or more thermal features with few associated artifacts. Such sites form an important part of research at the BMGR, particularly because of how landscapes of the BMGR were used aboriginally. People often used the BMGR for resource procurement and processing, travel and transportation, and short-term habitation. Large-scale residential sites and evidence of more-intensive use of particular locales for agriculture and other activities do exist, but such sites are relatively rare with respect to other smaller, less obtrusive kinds. Understanding the distribution of small sites is crucial to understanding how landscapes were used at different points in time and in different installation areas and should also prove important in evaluating site significance (Altschul and Rankin 2008; Vanderpot et al. 2008).

Given how the landscape was used, large numbers of small sites probably have gone undiscovered through survey. Using the method proposed in the previous section, we estimated that on the order of 1,400 small sites could have been undiscovered as a result of surveys conducted by Dames & Moore, SRI, and SWCA on the East, North, and South Tactical Ranges (ETAC, NTAC, and STAC, respectively) of the BMGR. Altogether, these undiscovered sites probably constitute a small percentage (approximately 0.1 percent) of the overall site area discovered. In other words, of the almost 3,000 site acres discovered, only around 4 site acres are likely to have been missed using this method for estimating percentage of sites missed. Nonetheless, undiscovered sites do represent an important component of current research and have been in recent years the focus of more-intensive investigations, such as the limited testing of imperiled thermal features.

The numbers of large low-density sites that have gone undiscovered, however, are more difficult to gauge. The situation may be similar to what Altschul (n.d) determined for Camp Pendleton, where a typical site covered 10,000 m^2 and had fewer than 30 artifacts (d = 0.003 artifacts/m^2). Perhaps 50 percent or more of large very-low-density sites have gone undiscovered through survey. Given that artifacts or features in large low-density sites may be thinly spread over the landscape, discoveries made within such sites may often be recorded as isolated finds, rather than as sites. Again, use of discovery probabilities for individual site classes is probably a better means of assessing how many large very-low-density sites could have been missed by survey than the method we have applied here.

China Lake

The BMGR and China Lake are fairly similar in terms of basic survey methods and typical artifact densities. Therefore, many aspects of survey reliability will be roughly similar between the two installations. One substantial difference between them, however, is in the size and shape of survey units. Both installations employ a variety of survey unit sizes and shapes, but China Lake survey units tend to be smaller and often more linear in shape than those of the BMGR. We therefore felt it was appropriate to evaluate the edge effect for surveys conducted at China Lake in order to compare the results with those from the BMGR.

When this study was conducted, China Lake had GIS data for 469 survey-unit polygons and 1,905 site polygons. Because of how data were set up in the files provided to us, however, it was difficult to associate particular sites with particular surveys in many instances. To circumvent this problem, we selected a sample of survey units for which we could be reasonably certain of which sites were discovered as a result of a particular survey. This resulted in a sample of 226 survey units and 766 sites likely to have been discovered as a result of those surveys. As for the BMGR, we calculated the number of sites for each survey that fell completely within survey boundaries or on the edge of or outside survey boundaries, and we counted as edge sites those that did not fall completely within survey boundaries. We then calculated a percentage, per survey unit, of edge sites and categorized survey units according to edge-site percentages using the same cutoff points that we used for the BMGR. The results indicate that although an edge effect is detectable for surveys at both installations, the edge effect is more frequent and often larger for surveys at China Lake (Table 2.4; Figure 2.11). As for the BMGR, the implication is that parameter estimation based on China Lake surveys will have to account for the edge effect for at least a third of survey units. As one might expect, China Lake edge sites are substantially larger than sites discovered completely within survey units. The trimmed mean (10 percent, two-sided) of edge sites is 2.17 (n = 166), whereas the trimmed mean of sites discovered fully within survey units is 0.45 (n = 446). Clearly, the attributes of edge sites could be substantially different from those of sites discovered within survey units. Conceivably, large sites are over-represented by survey at China Lake to a substantial degree.

Eglin AFB

STP-location data and associated survey area data are available in GIS for surveys performed at Eglin AFB between 2001 and 2008. We used those STP locations along with survey-unit boundaries to explore survey reliability at Eglin AFB.

Table 2.4. Edge Effect for Select Surveys at China Lake

Edge Effect	Total Survey Acres	Percent Surveyed Area	Survey Units	Percent Survey Units
Little or none	10,056.7	24.4	43	19.0
Small	8,449.5	20.5	14	6.2
Medium	8,253.1	20.1	29	12.8
Large	3,069.0	7.5	39	17.3
No sites	11,307.0	27.5	101	44.7
Total	41,135.3	100.0	226	100.0

In order to evaluate survey reliability, we selected a sample of survey units in which GIS data for STPs were available. In order to select the sample, we converted multipart survey polygons to single-part survey polygons, so that all polygons for each survey unit could be treated individually (i.e., some survey units had multiple associated survey polygons). We then selected from that set all survey polygons within which STPs fell. Some additional survey polygons and STPs had to be removed from the sample if there appeared to be a discrepancy between the survey unit associated with an STP and the survey unit associated with a survey polygon. Some of the removed cases probably resulted from STPs dug outside survey boundaries; others may have resulted from misattribution of either the survey polygon or STPs in GIS. The sample consists of 313 survey polygons from 242 survey units containing a total of 32,986 STPs.

For many survey areas with GIS data on STP locations, arrays of STPs do not appear to follow a systematic grid pattern, such as a square grid or a staggered grid. In many cases, the locations of STPs align with topographic contours and avoid sloped areas, wetlands, or areas of heavy disturbance or dense vegetation (Figure 2.12). Since 1993, STPs have also avoided areas considered to have low probability for site discovery and have varied on the basis of expectations of site density. Conversations with installation archaeologists indicate that different STP densities (e.g., 1, 1.5, or 2 STPs per acre) were specified contractually for different projects, depending on the expected density of sites in an area to be surveyed; the expectation was based on prior experience with similar settings. In recent years, resurvey of low-probability areas has resulted in the discovery of additional sites, particularly ones with historical-period components. Although the location and density of STPs probably follow practical and realistic guidelines, the variability in STP location and density makes the objective assessment of survey reliability difficult. For installations like Eglin, the prior probability for discovery may be quite low in avoided areas and substantially higher in areas that are tested with STPs. Nonetheless, given that areas of apparently low probability are avoided or tested only to a limited degree, the reliability of survey in those areas differs from that of other areas that are more intensively tested.

Most methods for assessing survey reliability are based on the assumption that shovel tests follow a regular, systematic pattern, such as square, rectangular, isosceles triangular, or equilateral triangular STP grids, although on-the-ground placement of STPs are expected to diverge from the ideal pattern even in relatively flat and environmentally homogeneous landscape zones. Because STP locations at Eglin AFB follow a fairly irregular pattern that varies within and among survey areas, we performed a nearest neighbor analysis on STP locations. For each STP, we calculated the Euclidean linear distance to the nearest STP and analyzed the results according to survey polygon. Four survey polygons contained only a single STP and thus could not be evaluated with nearest neighbor statistics. For survey polygons with between 2 and 30 STPs (n = 108), the distribution of nearest neighbor distances was right-skewed and highly variable. These were mostly surveys in which a few widely spaced STPs were placed; nearest neighbor distances ranged between 16.4 and 644.0 m and averaged 112.0 ± 15.5 m. The two-sided 10 percent trimmed mean was 99.6 m. For survey polygons with more than 30 STPs (n = 201), the distribution was normal and highly modal. Nearest neighbor distances ranged between 22.7 and 83.6 and averaged 53.0 ± 1.1 m. Many surveys appear to have had 50-m spacing as the goal, but surveys that used 60-m spacing, or occasionally 30- or 40-m spacing, also appear to have been implemented. The distribution of nearest neighbor distances for individual surveys is often modal but broad for individual surveys, converging on a mean, but with a high standard deviation. Variability in nearest neighbor distances suggests that survey reliability varies within and among individual survey units, depending on the local density of STPs. Average nearest neighbor distance varied between 2001 and 2008, but not dramatically. Average nearest neighbor distance declined from 53.7 in 2001 to 48.5 in 2004 and then increased steadily each year until 2007, when it reached 58.0 m. Average nearest neighbor distance appears to have decreased again in 2008. We are not sure what the cause of this variation is, but we suspect that it may have to do with variation between survey areas.

For most of the surveys with more than 30 STPs, the minimum distance between STPs was between 10 and 30 m, with a trimmed mean of 23.4 m (n = 159). We suspect that some STPs were placed near positive STPs in order to seek additional information, but the distribution of STPs does not bear this out. Occasionally, an STP is closer to a positive find than the average distance, but often STPs appear to be closer

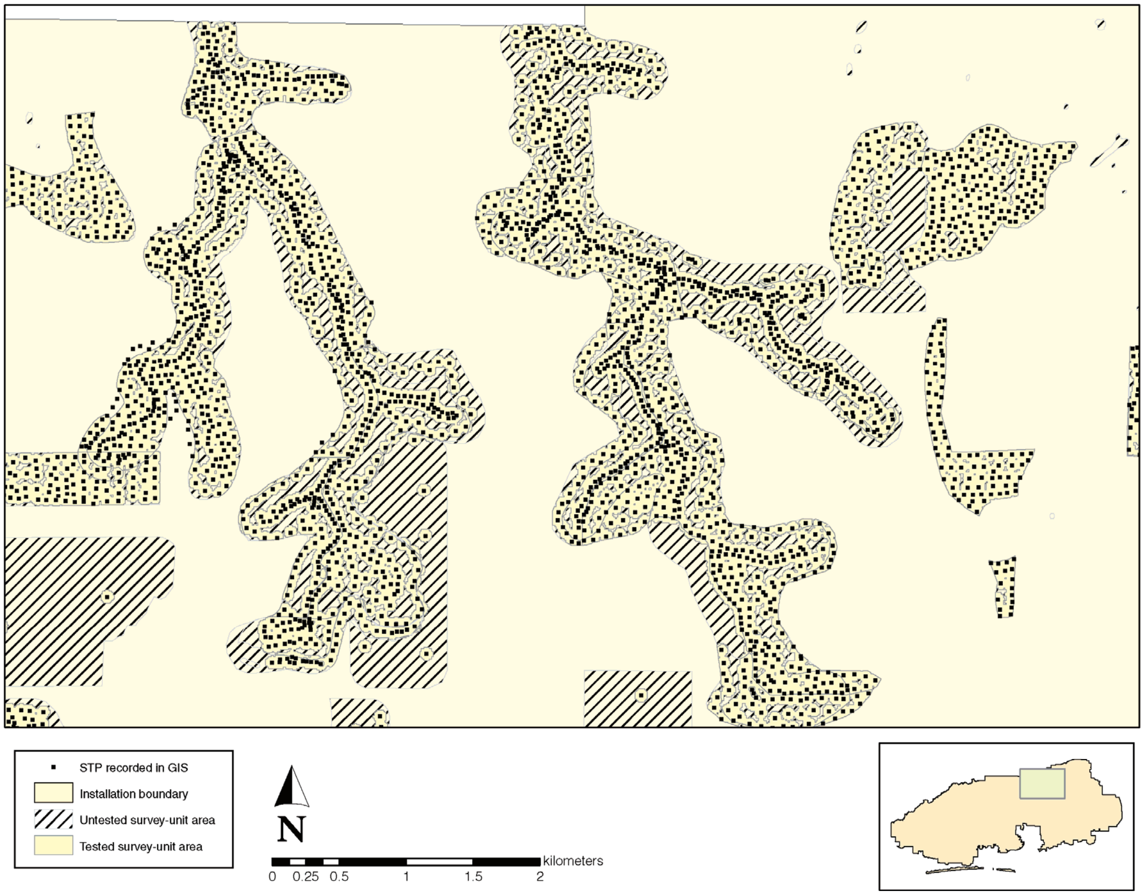

Figure 2.12. Example of STP distributions at Eglin AFB.

together because investigators must have thought a particular area or zone had a higher likelihood of site discovery. That is, STPs appear to be located at a higher density in some areas regardless of the distribution of positive tests. Our conversations with installation archaeologists suggest that this inference may be true. In many areas, 50-m spacing appears to have been fairly standard, but 30-m spacing was also implemented in areas considered to have a higher probability for site discovery. We understand from conversations with installation archaeologists that site discovery typically involves a discovery crew, who excavate the initial STP, and a recording crew, who later excavate additional STPs where sites were discovered, in order to determine site boundaries and other characteristics. From what we can determine, the GIS STP locations made available to us appear to correspond to the discovery crews' efforts, but not the recording crews' efforts. The former is perhaps the more appropriate sample for assessing survey reliability, because the initial discovery of sites is made before the efforts of the recording crew.

For most surveys of 30 or more STPs, the maximum distance between nearest neighbor STPs was between 50 and 200 m, with a trimmed mean of 98.3 (n = 159). The range of nearest neighbor distances was also quite large and variable among surveys, with a trimmed mean of 74.1. Coefficients of variation for all three variables (minimum nearest neighbor distance, maximum nearest neighbor distance, and range of nearest neighbor distances) were 30 or more percent, suggesting a large degree of variability in nearest neighbor distances. These data indicate that variability in the location of STPs was potentially large, depending on how STP locations were recorded. Some variability could have been introduced as a result of instrument (i.e., Global Positioning System [GPS]) error rather than actual variability in STP spacing. In general, however, test pits appear to have been placed judgmentally in order to cover particular landforms according to general spacing criteria that varied according to context.

For many surveys, the average spacing between STPs was approximately 50 m. We used Kintigh's (1988) Monte Carlo simulation program to estimate the probability of discovery for sites of varying sizes and artifact densities using 50-m spacing and 50-by-50-cm test units placed on a square grid. For each site size and artifact density, we simulated the discovery potential of 2,000 sites placed within a 1,000-by-1,000-m area and calculated discovery probabilities based on those simulations for artifact densities of 0.1, 0.5, 1.0, 5.0, and 10.0 artifacts/m^2. The results, shown in Figure 2.13, demonstrate a low likelihood of finding small sites and low-density sites based on survey method.

For 50-m spacing and 50-by-50-cm STPs, investigators can fail to intersect (let alone discover) circular sites almost an acre in size. For artifact densities below 5 artifacts/m^2, much larger sites will fail to be discovered even if they are intersected. For sites with low artifact densities, it is possible to miss sites from one to several hundred acres in size, depending on artifact density. For instance, a 1.9-acre site with an artifact density of 1 artifact/m^2 has a 54 percent probability of being discovered. If the same site had instead a density of 0.1 artifact/m^2, it would have only a 7.5 percent probability of discovery. Simulations suggest that with 50-m spacing on a square grid, discovery probabilities for artifact densities greater than approximately 5 artifacts/m^2 more closely approximate intersection probabilities for individual site sizes. Intersection and discovery probabilities become very similar for artifact densities at approximately 10 artifacts/m^2. Thus, we can expect to find high-density sites much more often than low-density sites, and sites below 0.82 acres in size can potentially be missed, even with very high artifact densities. Analysis of the largest site missed for the simulation results suggests that the largest site missed decreases allometrically as artifact density increases, or by approximately the square root of artifact density (largest site missed = $365.57d^{-0.5536}$, $r^2 = 0.9926$). This suggests that the largest site not discovered will equal the largest site not intersected when artifact density is around 23 artifacts/m^2. Conversely, the relationship suggests that as artifact density decreases, the largest site missed increases astronomically.

Obviously, more sites would have failed to be discovered during earlier surveys that used smaller units for site discovery than during later surveys. This will need to be taken into account for predictive modeling and other management efforts at the installation. By contrast, areas subjected to testing at smaller spacing intervals (i.e., 30 m) would have considerably higher probabilities of site discovery but would still fail to discover many small or medium-sized sites with low artifact densities.

As portions of some survey areas had no STPs placed within them, we calculated STP density per survey polygon, per areas of survey polygons actually tested with STPs, and per year. In order to estimate areas actually tested with STPs, we created polygons representing land parcels that have at least one STP within 50 m of the parcel's centroid. We used a distance of 50 m because nearest neighbor analysis suggested that 50 m was the modal distance for many surveys. Our conversations with installation archaeologists confirmed that 50 m was a standard distance between STPs for most surveys. Then, using the union command, we combined those polygons with survey polygons for which we had STP-location data. Using those polygon layers, we were able to calculate the number of STPs that fell within tested areas as well as the areas those STPs theoretically covered with, at most, 50-m spacing. The results demonstrate wide variability in the density of STPs per survey unit, the percentage of survey units actually covered by STPs, and the density of STPs within areas actually covered. These data suggest that there was wide variability in survey intensity for surveyed areas. Different survey units, and different parts of individual survey units, were surveyed at different intensities.

Of the 313 survey polygons in the sample, 23 percent had less than 50 percent survey coverage ([tested area ÷ survey unit area] × 100). Almost half of the survey units in the sample had less than 80 percent survey coverage. Only 22 percent of the sample had 90 percent or greater survey coverage. Overall, less than two-thirds of all survey-unit areas in the sample were covered by STPs spaced 50 m or less apart. Essentially, this means that a third of survey-unit areas were not tested with STPs. Examination shows that the distribution of percent survey coverage is bimodal (Figure 2.14). Some survey units contained only a few widely spaced STPs. Areas within those survey units could be inaccessible or considered areas with extremely low probability for site discovery, thus resulting in perfunctory testing. An alternative explanation is that the available GIS data do not reflect the number of STPs actually placed in survey areas with only a few STPs. When we remove survey units with apparently low STP densities, the result is a highly modal distribution

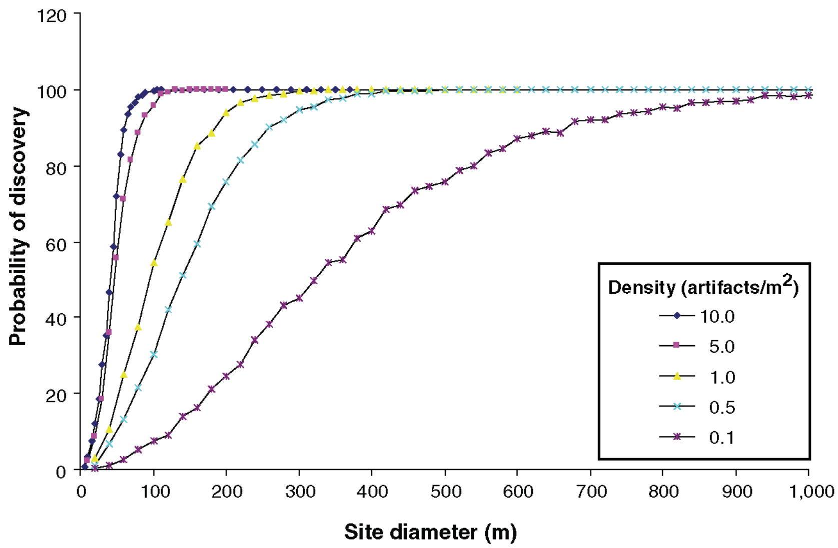

Figure 2.13. The probability of discovery for 50-m spacing on a square grid and 50-by-50-cm test units, according to artifact density and site size.

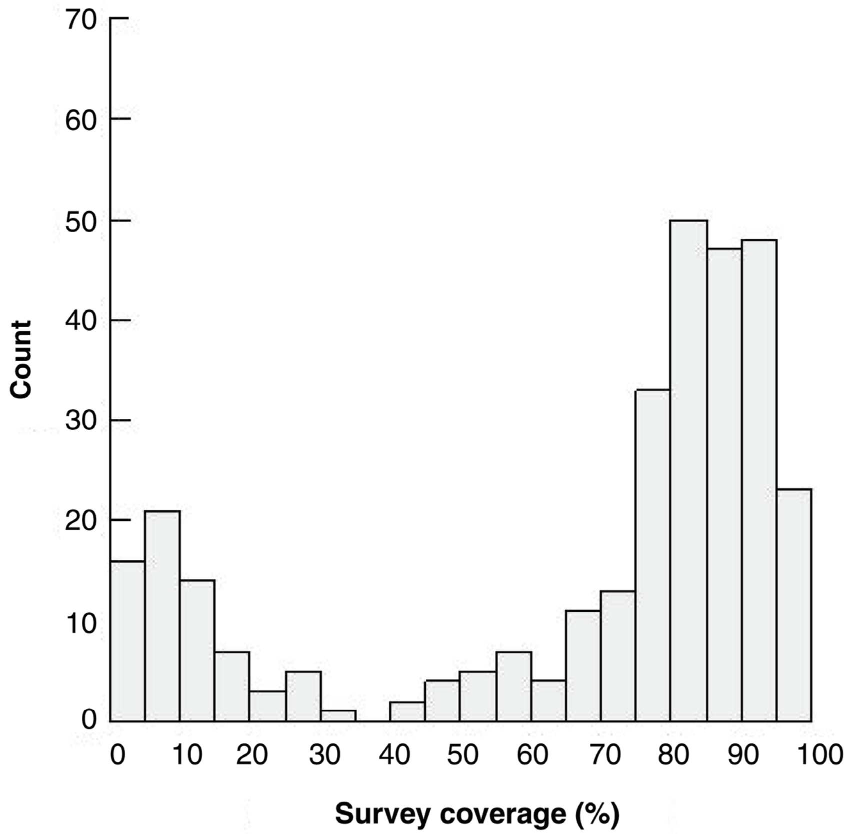

Figure 2.14. Distribution of survey coverage for Eglin AFB survey units with STP locations in GIS.

with a mean survey coverage of 82.4 ± 1.5 percent (n = 247). This indicates that for three-quarters of survey areas, around 18 percent of survey units are not covered by STPs spaced 50 m or less apart.

STP density is also variable. The arithmetic mean was 0.649 ± 0.047 STPs per acre with a range between 0.002 and 3.48 STPs per acre. The distribution is bimodal, however, with a large number of survey polygons with STP densities below 0.2 STPs per acre (n = 59) and several large outliers. When cases below 0.2 STPs per acre and above 2 STPs per acre were removed, STP density averaged 0.756 ± 0.027 STPs per acre. STP density in tested areas, or areas actually covered with STPs spaced 50 m apart, ranged from 0.5 to 3.6 STPs per acre. Almost 99 percent of those measurements fall below 2 STPs per acre, however. When the outliers are removed, the result is a highly modal normal distribution with a mean of 0.89 ± 0.04 STPs per acre and a range of 0.52–1.76 STPs per acre (Figure 2.15). STP density in tested areas varied from year to year, varying between 0.84 STPs per acre and 0.99 STPs per acre. STP density peaked in 2004 and declined between 2004 and 2007, essentially mirroring temporal change in nearest neighbor statistics (discussed above). It is unclear whether annual variation is the result of a programmatic change or the result of differences between survey areas that required different levels of effort.

Overall, estimates of STP density indicate that, even in areas actually covered by STPs spaced 50 m apart or less, STP density is substantially below 1 STP per acre. To put this into perspective, if we were to have 15-m spacing between STPs on a square grid, we might expect a density more than 20 times higher, or around 17.6 STPs per acre. In short, for the sample we analyzed, around 80 percent of survey areas were actually covered by STPs spaced 50 m or less apart, and the actual density of STPs was substantially less than expected. With 50-m spacing on a square grid, we might expect around 1.5 STPs per acre, instead of 0.86 STPs per acre.

One of the problems with this analysis is that many areas of Eglin have a low probability of site discovery. Because of a lack of GIS data describing these areas, we were unable to factor those low-probability areas into the analysis. If we had been able to base our calculations only on areas that were considered suitable for testing, we would probably have arrived at a substantially higher density of effort. That is, eliminating areas that cannot be tested through conventional methods (e.g., heavily disturbed or developed areas, wetlands) and areas that have a low prior probability of site discovery may considerably constrain the testable areas and thus substantially increase density of effort. Conversations with installation archaeologists suggest that this is indeed true. Unfortunately, many areas that could not be tested because of excessive slope, disturbance, or other factors are not mapped in GIS and thus cannot be used at this time to constrain the area available for testing.

Given the low probabilities of discovering artifacts in STPs and the sizes of sites that would fail to be intersected by survey, the results indicate that a very large number of sites could theoretically have been missed by survey. At Eglin AFB, if we assume that most sites have an artifact density of 1 artifact/m^2, then 98 percent of sites discovered thus far have discovery probabilities less than 1. If we instead assume that most sites have an artifact density of 10 artifacts/m^2, 90 percent of sites discovered thus far have discovery probabilities less than 1. In either scenario, most sites will fail to be discovered 100 percent of the time and there must be large numbers of undiscovered sites. If we assume that site size, as recorded, accurately reflects the "true" size of a site and that artifact densities range between 1 and 10 artifacts/m^2, we can estimate that perhaps 75–92 percent of sites (7,000–27,000 sites) have been missed. Unlike the situation at the BMGR, where most *missed* sites are probably small, missed sites at Eglin could be very large. As a result, perhaps anywhere from 700 to 5,000 site acres (14–56 percent) could have gone undiscovered in survey areas. The high end of these ranges probably overestimates numbers of sites and site acres missed. Because of low probabilities of detection, sites may often be larger than recorded. Thus, they will have higher probabilities of discovery than calculated using the recorded site size, a situation that ultimately translates to fewer undiscovered sites. Similarly, these estimates are based on discovery probabilities simulated for 50-m spacing, when we know that some installation areas were tested with 30-m spacing. This number would also decline significantly if we could use in the calculation only sites that were discovered through test-unit sampling methods, as some sites were evidently discovered through detection of historical-period artifacts or features observed at the surface and other information, such as historical records. Still, even if we were to reduce these figures by 50 percent, we would have on the order of 3,500–13,500 sites and 350–2,500 site acres potentially missed in surveyed areas.

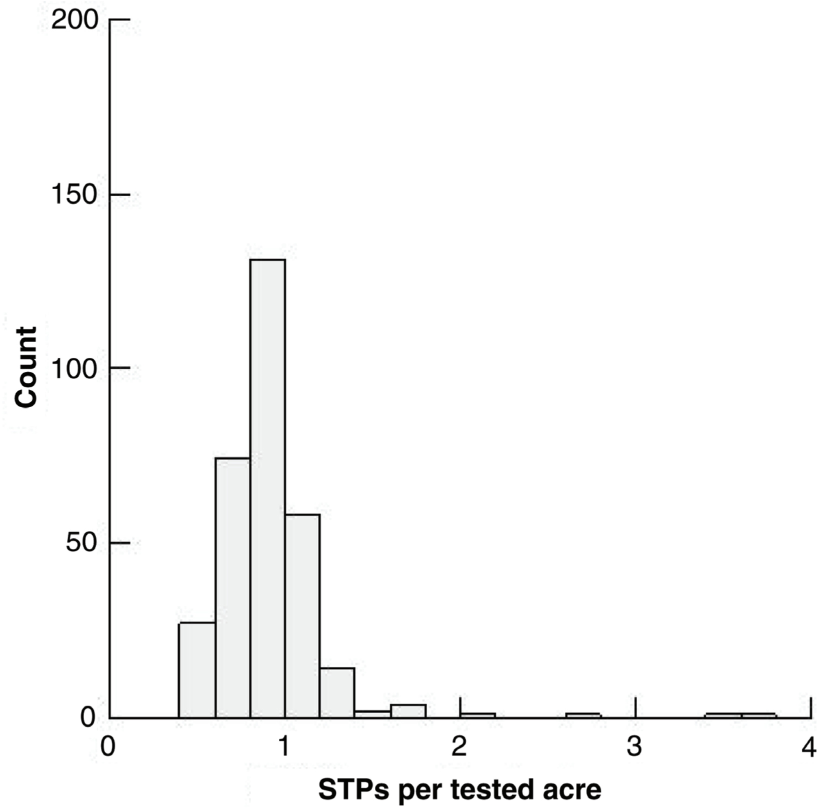

Figure 2.15. Distribution of STP density for Eglin AFB survey units with STP locations in GIS.

Fort Drum

The situation at Fort Drum with regard to STP data is markedly different from that at Eglin. At Fort Drum, GIS data have been developed for more than 138,000 STPs for surveys dating from 1993 through 2006. Each STP has a unique ID and many associated attribute data that allow for STPs to be evaluated according to transect, project, year, installation area, or other factors. Detailed attribute data for STPs are stored in a Microsoft Access database that can be linked to the GIS data using unique identifiers. Although we have not yet been able to harness the full power of this database, it is an impressive and highly useful resource for archiving data and for evaluating such factors as survey reliability. It also has strong potential as a research tool, as the specific attributes of individual test units—including artifact content, stratigraphy, depth, and location—can be manipulated in a highly flexible manner, allowing researchers to evaluate not only the distribution of positive and negative STPs but also the distribution of different artifact types, deposits, or other factors.

Another fundamental difference between Fort Drum and Eglin AFB STP data is in how they are represented in GIS. Eglin AFB's STP data appear to indicate the "true" location of STPs insofar as can be determined from mapping information. Fort Drum's STP GIS data, by contrast, are highly standardized for individual survey units, indicating that STP locations stored in GIS are more-idealized locations based on the systematic logic of individual survey designs (Figure 2.16). In some cases, the systematic locating of STPs in GIS is absolutely necessary because of a lack of accurate GPS measurements taken in the field. We expect, however, that the "true" STP locations probably depart somewhat from those depicted in GIS. This is not a problem, per se, only a mapping approach to be pointed out.

In addition, STPs stored in GIS include many that were planned but were not dug because of the presence of impenetrable vegetation, bedrock, buried cable, standing water, animal dens or hives, building foundations,

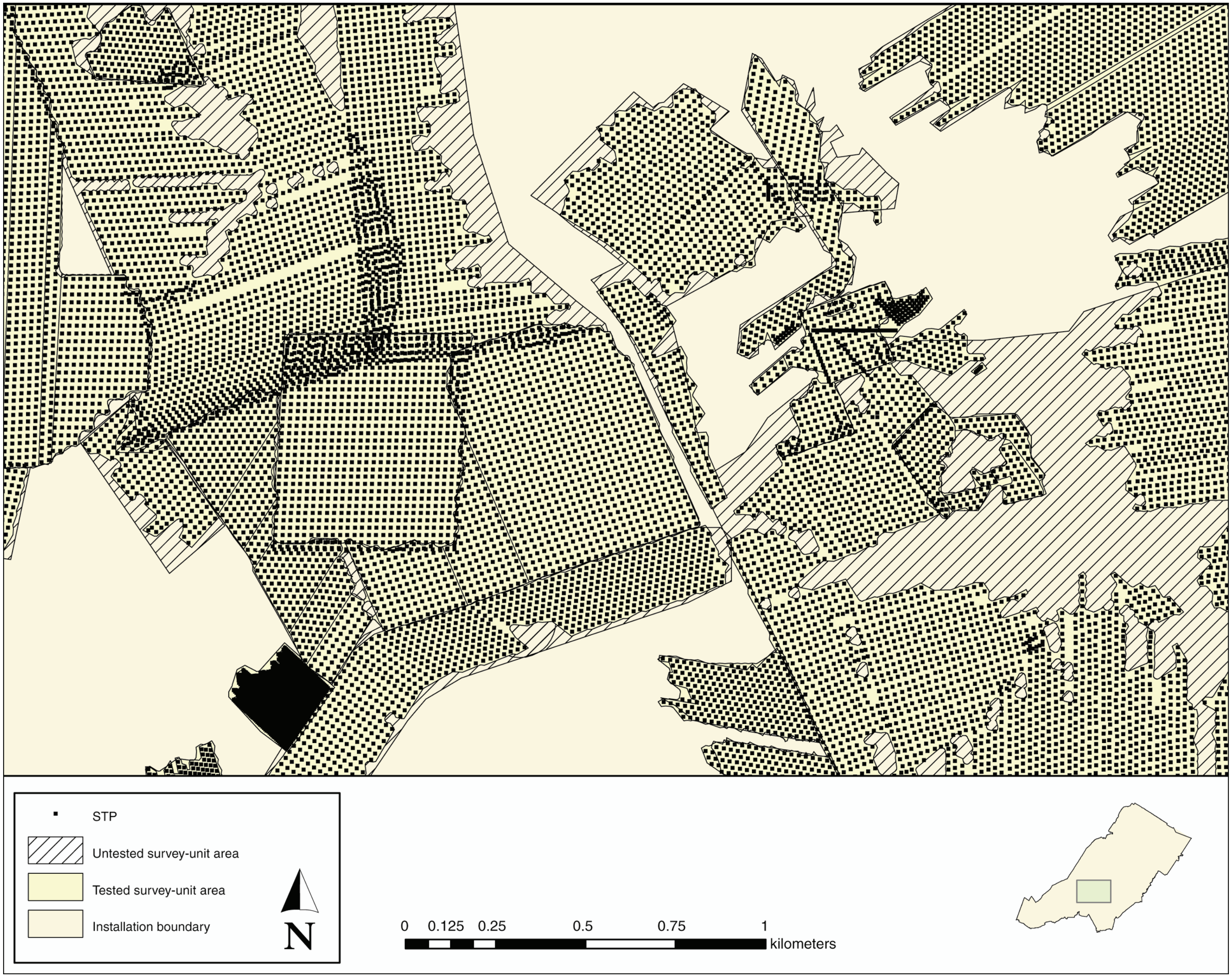

Figure 2.16. Example of STP distributions at Fort Drum.

pavement, or other obstacles. For projects conducted in the year 2000, for instance, more than 35,000 STPs were planned but fewer than 23,000 were actually dug. For individual projects during that year, anywhere from 11 to 100 percent of STPs were actually dug. The average was 64 percent. Percentage of planned STPs dug also varied from year to year and has generally increased since 1995 (Figure 2.17). We expect, then, that the reliability of survey is lower than is indicated by the standard spacing alone and is likely to vary among projects and years.

Using the number of STPs actually excavated each year and the area actually covered by STPs, we were able to calculate survey intensity per year for Fort Drum surveys. Following an approach similar to that used for Eglin, we calculated the areas actually covered by STPs spaced 20 m or less apart (the typical spacing for Fort Drum surveys). We then used these areas to calculate how much area of each survey unit was actually tested with STPs. The next step was to use the number of STPs actually excavated within survey units to calculate the density of STPs (per acre) for the years that data were available (1995, 1998–2005). The results show that the density of STPs varied from year to year and has gradually increased since 1995 (Figure 2.18). For each year, the density of effort was substantially lower than planned because of the removal of STPs that could not be excavated from the systematic grid of planned STPs. From these data, we can conclude that survey reliability would probably have varied from year to year and from project to project. As at Eglin, more-refined estimates of survey intensity could be achieved if areas that could not be tested, because of disturbance or environmental factors, were removed from the pool of survey areas.

Surveys at Fort Drum, as at Eglin AFB, have employed a variety of spacing intervals and testing strategies but have generally used smaller units placed closer together than those at Eglin AFB. STPs have generally been 30 by 30 cm in size, which we assume corresponds to a unit area between 0.07 and 0.09 m^2, depending on whether STPs were square or circular in plan view. Surveys at Fort Drum used a variety of different grid patterns, including square, staggered, and arbitrary rectangular grids. Spacing was generally between 5 and 20 m, depending in part on the particular grid pattern used. The variability in grid patterns and spacing suggests that the Fort Drum data could be very useful in evaluating the effectiveness of different survey methods in different installation areas, but such an effort is beyond the scope of the current project.

As for Eglin AFB, we calculated discovery probabilities for typical surveys conducted at Fort Drum using Kintigh's (1988) simulation program (Figure 2.19). For Fort Drum, we simulated square and staggered grids, as both grid patterns have been commonly employed at the installation. We used 15-m spacing and 30-by-30-cm units in the simulations, although we recognize that a variety of different spacings were used for different surveys and grid patterns. The results indicate that, as for Eglin AFB, low-density sites have a low probability of discovery. At Fort Drum, sites with artifact densities of 0.1 artifact/m^2 must be more than 160 m in diameter (5 acres) in order have a 50 percent chance of being discovered, and 270 m in diameter (14 acres) in order have an 80 percent chance of being discovered. Obviously, the probability of discovering a low-density site is high only for very large low-density sites. As discussed earlier, the likelihood that the "true" size or extent of such sites could be determined through STP survey is exceedingly low; therefore, these sites could easily be interpreted as either small low-density sites or isolated finds.

In contrast to the results for Eglin AFB, the probability of discovery approaches the probability of intersection at lower artifact densities. Also, the largest site not discovered through survey is lower for Fort Drum than for Eglin AFB (Figure 2.20). For instance, at Fort Drum a site that is perhaps 70 m in diameter with an artifact density of 5 artifacts/m^2 could potentially be missed. At Eglin AFB, a site with the same artifact density could be more than twice the size and still be missed. As stated above, survey reliability should vary among surveys at Eglin because of variation in spacing, grid pattern, and the percentage of STPs actually excavated. We expect that estimates of site discovery simulated for Fort Drum are relatively conservative, because many surveys had spacing smaller than 15 m and often employed staggered grid patterns. Given that large numbers of STPs were not excavated during particular surveys, however, the reliability of a survey will vary across survey areas, depending on whether STPs were actually excavated. Survey reliability as simulated thus applies only to areas where STPs were actually dug. Intervening areas obviously were not tested and could not be considered to have been surveyed with the same reliability as other areas.

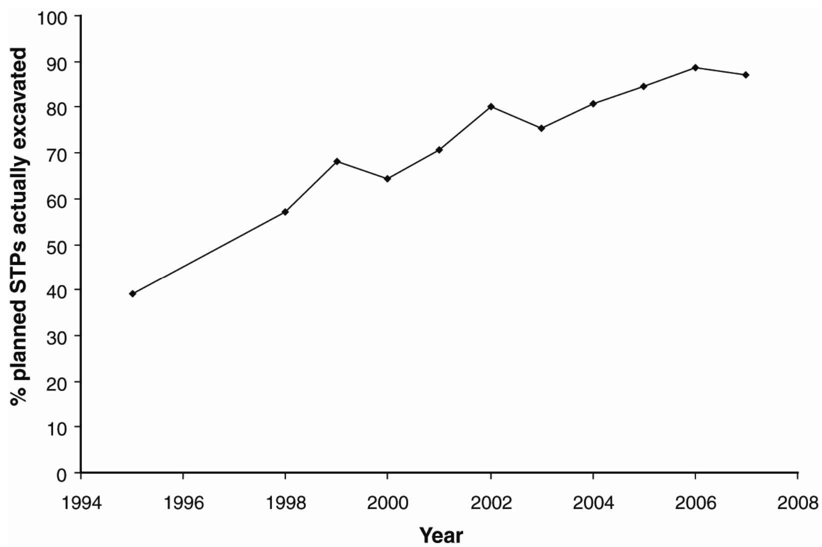

Figure 2.17. Percentage of planned STPs actually excavated at Fort Drum, according to year.

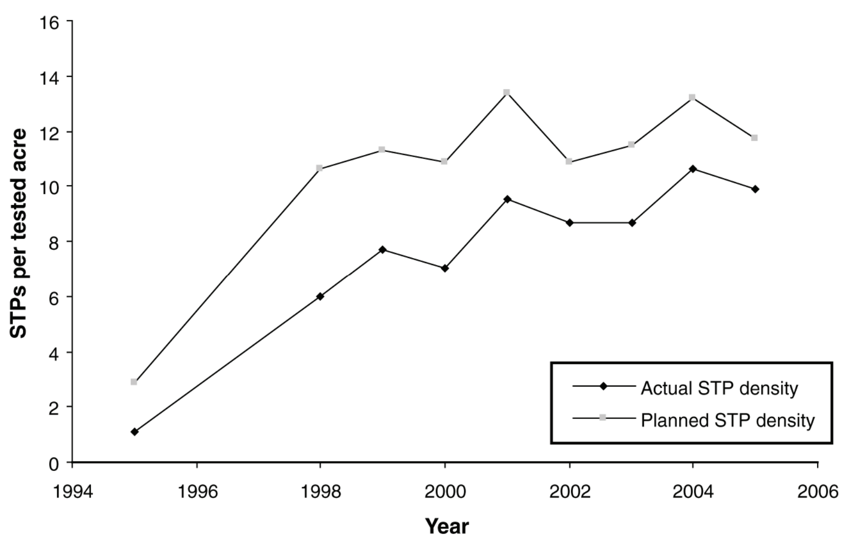

Figure 2.18. Planned and actual STP density in tested areas of Fort Drum, according to year.

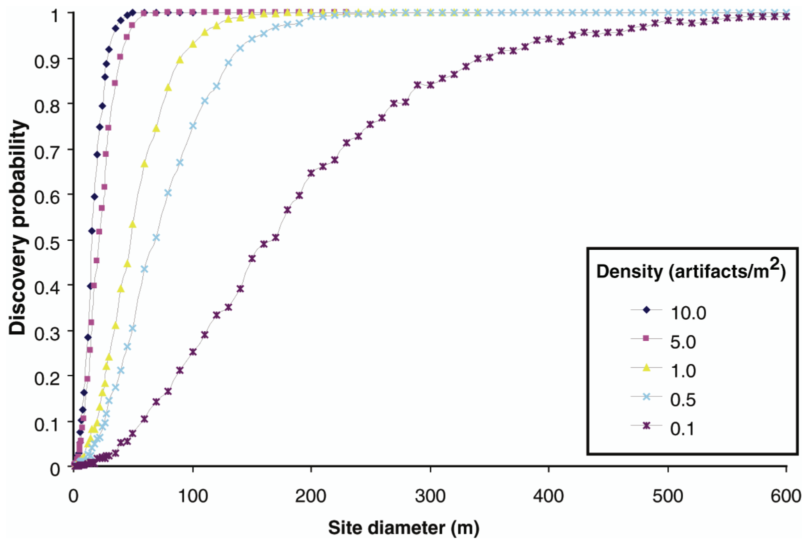

Figure 2.19. The probability of discovery for 15-m spacing on a square grid and 30-by-30-cm test units, according to artifact density and site size.

As for Eglin AFB, we calculated numbers of sites and site acres potentially missed by survey. Unlike those at Eglin AFB, data on Fort Drum sites are stored as points. We do not have GIS data on site size. Of the 717 site points, 502 fell within survey areas and could be inferred to have been potentially discovered through shovel-test survey, although some may have been discovered through other methods. If we assume that sites discovered at Fort Drum have the same size distribution as sites discovered at Eglin, we can provide a comparable estimate of sites missed for surveys at Fort Drum. We calculated the median site size for each decile of the Eglin size distribution and used those numbers to simulate the distribution of site sizes at Fort Drum. We then calculated the discovery probability for those site sizes for artifact densities of 1 and 10 artifacts/m^2, using the simulated discovery probabilities for a square grid with 15-m spacing. The discovery probabilities used are an underestimate for some sites, as spacing less than 15 m was used for some surveys. Thus, the percentage of sites missed is probably an overestimate. The results are very different from those for Eglin AFB, however, even though site-size distribution was held constant between the two installations. For Fort Drum, 200–1,600 sites (30–76 percent) may have been missed by survey. This corresponds to 20–200 site acres undiscovered in survey areas (1–12 percent site acres missed). The two installations differ in size and number of acres surveyed; therefore, the number of sites and site acres missed cannot be compared without qualification. If we ignore differences in size and number of acres surveyed between installations, the gross difference between the two installations in sites and site acres missed is roughly an order of magnitude. Obviously, many sites are missed by shovel-test survey at Fort Drum, but the problem of undiscovered sites is significantly smaller than it is for Eglin AFB.

Summary

We began this chapter by introducing the problem of survey reliability and identifying the issues we feel are particularly relevant to the assessment of survey reliability at military installations. Our discussion emphasized that the reliability of individual surveys is dependent on survey method. We then presented a series of metrics and methods that can be used to assess survey reliability. In our presentation of methods, we demonstrated the effects of changes in key variables—such as spacing, unit size, and site size—on survey reliability.

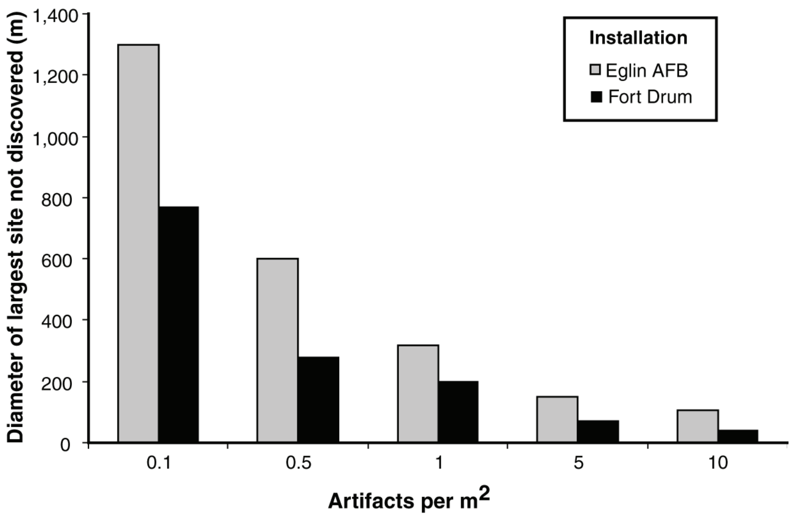

Figure 2.20. The largest site missed for surveys conducted with 50-m spacing and 50-by-50-cm test units vs. with 15-m spacing and 30-by-30-cm test units, corresponding to methods for some surveys at Eglin AFB and Fort Drum, respectively, according to artifact density.

What should be clear from our discussion is that there is no simple fix to the problem of survey reliability. The reader will note that although relationships between key variables can be simulated or computed, relationships are often nonlinear. Depending on the particular point in such a relationship, small change in an independent variable could cause comparatively large change in a dependent variable, or vice versa. As a result, one set of survey methods may be reliable for the discovery of certain kinds of sites (e.g., sites more than 50 m in diameter and average artifact densities of 5 artifacts/m^2) but less reliable for the discovery of other kinds of sites. Installation managers must decide how reliable surveys have to be for specific kinds of sites in order to apply effective methods for assessing survey reliability.

Another issue we have stressed is that, regardless of survey method, not all sites will be discovered in a survey area, and that survey results are likely to vary among surveys, independently of what is actually surveyed. If an area is surveyed multiple times, different numbers of sites of a given size and artifact density could be discovered each time an area is surveyed. Using the binomial distribution and discovery probabilities, Kintigh (1988) showed, for instance, that if a hexagonal grid with 31.5-m spacing is used to survey an area 200 times, different results will be obtained for individual surveys. For an area that contains five 20-m-diameter sites with an average artifact density of 10 artifacts/m^2, around 20 percent of surveys will result in no site discovered. Fewer than 1 percent of surveys will result in all five sites discovered. Thus, the likelihood of discovering all sites in the survey area is extremely low, even when artifact densities are relatively high. This model of site discovery also implies that site density can vary from survey to survey, independently of what was actually surveyed. In designing and evaluating surveys, installation managers must decide what level of confidence is acceptable for discovering sites with particular characteristics and must accept the fact that in no survey will all sites be discovered. This fact is particularly true for small or low-density sites, which are most likely to be missed.

Following our discussion of methods for assessing survey reliability, we evaluated survey reliability for surveys performed at four installations: the BMGR, China Lake, Eglin AFB, and Fort Drum. The two western installations, the BMGR and China Lake, use pedestrian methods for site discovery. Eglin AFB and Fort Drum use shovel-test sampling methods. Both methods have advantages and disadvantages. The advantage of pedestrian methods is that the sample fraction is typically much larger than that of shovel-test sampling. Therefore, pedestrian survey generally is better in discovering smaller and lower-density sites insofar as such sites have an adequate surface expression. A disadvantage of pedestrian survey is that survey reliability is substantially affected by change in surface visibility or variability in the potential for buried sites.

An advantage of shovel-test survey is that samples are taken at depth. Thus, buried or low-visibility sites can be discovered even when entirely obscured by vegetation or surface deposits. In addition, because excavated sediments are screened, more of the artifacts present in a particular location (i.e., in a test unit) are likely to be discovered than if the location was simply walked over. Shovel-test sampling thus overcomes some of the problems with visibility inherent in pedestrian survey. A major disadvantage of shovel-test survey, however, is the extremely small sample fraction. Coupled with relatively wide spacing, the small unit sizes implemented in shovel-test survey contribute to very low discovery probabilities for low-density sites and small-to-medium-sized sites.

In our evaluation of survey reliability at military installations, we generally found pedestrian survey to be more reliable in site discovery, but it must be noted that the scale of the problem identified above has yet to be determined for specific installations. In other words, how likely are we to miss or misinterpret low-density sites? If we follow Altschul's (n.d.) example for Camp Pendleton, the percentage of large low-density sites missed could be as high as 50 percent. This is not necessarily a problem, as long as a 50 percent sample is considered adequate for evaluating large low-density sites. This sample size is probably sufficient to address issues such as the types of environmental settings where we find these sites, the types of activities represented at these sites, and their place in an overall settlement system. The sample size becomes a problem, however, if installation managers represent the nature of the archaeological record differently from stakeholders: that is, if survey results are presented as a complete record of past activity in a project area. Evaluation of survey reliability is critically important to decision making and should be incorporated into inventory reports used for compliance purposes.

For both shovel-test and pedestrian-survey methods, we suggest that artifact discovery in low-density sites could easily be interpreted as isolated finds or portions of much smaller sites, rather than as portions of the sites of which they are actually a part. Because sites usually are managerial units that imperfectly describe concentrations of cultural resources, the fact that components of large low-density sites may be recorded as isolates may not be a problem for landscape-oriented research, as long as discoveries from both isolated finds and sites are considered for analysis. The classification of low-density site areas as isolated finds, however, is a management concern, as areas containing high densities of isolates may in fact contain several undiscovered low-density sites. In those cases, cultural resources recorded as isolates may not be managed effectively or may not be factored into the evaluation of sites and are more at risk of disturbance than similar resources classified as sites. Installation managers who face this problem must evaluate the level of confidence that we can place in the interpretation of apparently isolated discoveries as isolated finds, as small, discrete sites, or as larger, more inclusive sites. Statistical analyses such as the ones described above are a good way to develop this type of confidence.

Another issue that deserves further evaluation is the disparity between actual and planned survey intensity. As noted above, there are many ground conditions—such as vegetation, disturbance, or health hazards—that place practical limitations on where survey can actually be carried out. In the case of shovel-test survey, the planned locations of test units are either skipped or moved to more-promising locations to address this problem. Similarly, pedestrian transects may deviate from their planned alignments or be skipped because of comparable problems. The implication of this situation is that the actual intensity of survey efforts, and thus survey reliability, will vary within and among survey areas. In addition, some areas within survey units will effectively not be surveyed.

For the two installations for which we have GIS data on STP locations, we found that STP density varied considerably among survey areas and, in the case of Fort Drum, from year to year. For Fort Drum, the percentage of planned STPs actually excavated increased after 1995, as did STP density. The implication is that survey reliability generally increased through time, as fewer areas have been avoided and more-recently excavated STPs have been spaced more closely together. At Eglin AFB, STP density in tested areas varied from year to year, but it is unclear whether annual variation is the result of a programmatic change or the result of differences between survey areas. At Eglin AFB, STPs have often been placed with respect to variation in topography, vegetation, and other factors and have tended to be spaced more closely in areas considered to have a higher probability for site discovery. The result is that different survey areas will be tested at different intensities, depending on a variety factors.

For both Fort Drum and Eglin AFB, we simulated discovery probabilities and estimated the largest site missed using square grid patterns with spacings of 15 and 50 m, respectively. Grid patterns and spacing vary within a certain range at both installations. At Fort Drum, spacing tended to be between 5 and 20 m, according to square or staggered grid patterns, and often around 15 m. Units were approximately 30 by 30 cm in size, or 0.09 m^2. At Eglin AFB, spacing typically varied between 30 and 100 m and was often around 50 m. Units excavated since 1983 were on the order of 50 by 50 cm in size, or 0.25 m^2. We also simulated staggered grids with 15-m spacing for Fort Drum, but the results were similar overall to those for the square grids. The presentation of those results was considered redundant and was not included in this report. In comparing the survey methods at the two installations, we found that survey reliability was generally better for surveys conducted at Fort Drum, despite the use of smaller test units. The difference between the installations lies in the substantial difference in spacing, which at Fort Drum has a much higher potential for intersecting small sites as well as for placing more units within larger sites. At neither installation, however, do survey methods perform well in discovering low-density sites.

As discussed above, we applied similar methods in assessing survey reliability for surveys conducted at the BMGR. A simulation program similar to that for the shovel-test survey, however, does not exist for calculating discovery and intersection probabilities for pedestrian survey. We applied Kintigh's simulation program by stringing a series of closely spaced units along transects, under the assumption that field crew would stop and look repeatedly at discrete units as transects were walked. Although the approach has limitations, it suggests that low-density sites with diameters larger than spacing will indeed be missed, but not nearly to the degree that similar sites would be missed by shovel-test survey. For artifact densities of 0.1 artifact/m^2 and 15-m spacing, sites on the order of perhaps 30 m in diameter could be missed. By contrast, sites with similar densities and diameters 25–45 times larger could be missed by test-unit survey. In addition, as artifact density increases, discovery probabilities for pedestrian survey approach intersection probabilities much more rapidly. Large numbers of small low-density sites could easily be missed by pedestrian survey, but not nearly to the degree that they are missed by shovel-test survey. The problems of visibility and interpretation remain for low-density sites in areas where pedestrian survey is conducted. Low-density sites could be missed or misinterpreted as isolated finds or small sites, particularly as spacing increases. Installation archaeologists may feel that in these situations they have an adequate sample of small low-density sites. In such cases, survey reliability is still useful in assessing which kinds of sites are likely to have been under- or overrepresented by survey, so that these factors can be corrected for in parameter estimation. In addition, discovery probabilities can be used to estimate the number of sites potentially missed by survey.

A final aspect of survey reliability that we evaluated was the edge effect. We evaluated the edge effect for China Lake and the BMGR by calculating, for each survey, the percentage of sites that fell completely inside a survey unit or overlapped the edge of a survey unit. By examining the distribution of edge-effect percentages, we defined arbitrary cutoff points and applied them to both installations. We interpreted ranges of edge-effect percentage as corresponding to the categories of (1) no sites discovered, (2) little or no edge effect (0–10 percent), (3) small effect (10–25 percent), (4) medium effect (25–50 percent), or (5) large effect (50–100 percent). For the BMGR, we found that, as expected, the greatest edge effect was associated with small survey units and linear units. Many surveys at the BMGR, however, were large block surveys where the edge effect was minimal or nonexistent. At China Lake, because many survey units are either relatively small or linear, it was expected that the edge effect would be more pronounced. At China Lake, more than one-quarter of survey units sampled had an edge effect of 25 percent or greater. By contrast, around 10 percent of survey units at the BMGR had a similar effect.

Recommendations

One of the problems we encountered when conducting this study was in understanding and organizing the data in a manner that was conducive to analysis. Different installations have different ways of organizing data, and the data for different variables are in various states of completion. A standard set of data fields would be very useful in the assessment of survey reliability. For instance, STP locations could be attributed according to survey unit, dates of survey, survey name, unit dimensions, location methods, planned/excavated, results (e.g., positive or negative, content), and phase of investigation. Sites likewise could be attributed according to survey unit, recording dates, survey name, recording methods, location methods, artifact density, size, type, and the presence/absence of artifact or feature classes. Survey units could be attributed according to survey-unit number, survey name, dates of survey, survey methods (i.e., spacing, unit size), reference, and organization. With the above-listed kinds of data available for each STP location, site, and survey unit, the evaluation of survey reliability could be substantially more comprehensive, more precise, less error prone, and easier to accomplish.

We also suggest that installation managers determine which kinds of sites are particularly important to discover through survey, which kinds are probably under- or overrepresented by survey, and what sample fraction will be necessary to ensure that enough sites are discovered, according to site type. As stated above, survey will never discover every site in a survey area, and attempting to do so is not necessarily an effective use of survey. Instead, installations must decide the kinds of sites for which they have a large enough sample and about which they do not know enough. Installation managers could then use this information to determine the reliability of survey for particular site types and installation areas.

CHAPTER 3

Site Location

A fundamental assumption of DoD CRM programs is that the locations of archaeological sites in an installation's database are accurate. All management decisions involving cultural resources depend on site-location information. Site locations affect where the installation conducts training activities or development projects, how much time and effort are required to comply with environmental laws, and what restrictions are placed on military activities. The credibility of the installation also depends heavily on the ability to provide accurate and reliable information on cultural resources to stakeholders, such as Native American tribes and SHPOs. Changes in the numbers of resources and their locations do great harm to the consultation process. It is essential, then, that the installations do everything in their power to get these fundamental and seemingly simple facts correct.

Obtaining site-location information is expensive and time-consuming. Not surprisingly, installation managers have the expectation that, with so much invested in fieldwork, the information is sound. Because sites consist of inanimate, physical objects, they should stay in one place and not change in size or character. Yet, installation archaeologists continually have to tell managers that some sites have apparently disappeared, others have grown, and still others have changed from insignificant, sparse artifact scatters to highly important residential camps. These archaeologists often are forced to request new money to survey areas that have already been surveyed because they cannot demonstrate that the old data are accurate.

Two distinct points about site locations are important for managers and archaeologists at installations to understand. The first is why the location of a site can be problematic. After all, a GIS can plot a site's location to an incredible degree of precision. Yet when we get to the field, the extent to which the mapped locations fail to correspond to the "true" locations of sites can be substantial. Some of this error stems from the recording method. Before about 2000, site locations were generally determined by means of topographic maps. Misplotting by archaeologists was common. Even when sites were plotted correctly, the scale of the maps introduced substantial error in determining the exact Universal Transverse Mercator (UTM) coordinates of the site.

Another source of error includes natural and anthropogenic processes that change the distribution of surface indicators of past human activity. In particularly dynamic environments, such as alluvial plains or dune fields, artifacts and features are commonly exposed, buried, or moved. The surface expression of sites, therefore, can vary over time, creating the impression that the site location and its boundaries are actually moving. Ultimately, apparent change in site location through time leads to uncertainty regarding the accuracy or validity of previous observations. Such circumstances also lead to the potentially erroneous conclusion that the lack of a surface expression in a particular location corresponds to the lack of a site in that location.

The second point is that we are not helpless in the face of these errors. Differences between "true" and recorded site locations are what statisticians call *uncontrolled variation* in these variables. Probability theory assumes that these errors exist and that they are random. To the extent that these assumptions are valid, we can use multiple recordings of the same site to create a more precise estimate of a site's location.

A composite measure of site location flies in the face of current CRM practice. Currently, most installations assume that the last recording is the most accurate. Consequently, they delete all previous recordings of site location, lest they become sources of confusion in the future. But there are good reasons why sites change location, and the last recording is not, necessarily, fundamentally better than previous ones.

By conceptualizing each recording episode as an individual case, we can create a "mean" location associated with a degree of precision that indicates the amount of confidence we can have in the estimate.

In this chapter, we first discuss one of the major factors affecting site location on installations: instrument error. We focus on this topic because instrumentation, particularly GPS, has become an essential part of archaeological inventory. Many CRM managers assume that these units are relatively free of error. But unless the factors involved in GPS are understood, new sources of error can be introduced into site locations. Next, we examine the history of site location from one installation, the UTTR, which kept historical site-location records. We show what these records suggest about the accuracy of site locations on the UTTR and why it is in the best interest of other installations to keep similar records. We end the chapter with a series of recommendations.

Instrument Error and Its Effect on the Location of Archaeological Sites

CRM has traversed the period from paper records to digital files. Nowhere has this transition been more keenly felt than in fieldwork and, in particular, how we locate ourselves in space. Finding our location has changed from visually matching landmarks with topographic maps to pushing buttons on a GPS unit. Most DoD installations have site-location data that belong to both technological ages. Installation staff have grappled with the differing levels of precision in several ways. Generally, it is assumed that paper records are less accurate than GPS locations. Some choose, however, to assume that visual map locations are as accurate as GPS locations until proved wrong. Others revisit sites, rerecording locations at tremendous cost and discarding old data. In this section, we explore the issue of Legacy site locations. We begin with a short discussion of errors associated with GPS and older surveying instruments.

Instrument error can be defined as the error resulting from use of a specific instrument to record the location of a point in geodetic space. Before the wide availability of consumer- or professional-grade GPS units, many archaeologists established site locations through the use of paper maps, handheld compasses, visual reckoning, and, less often, optical survey instruments, such as theodolites. In the United States, many archaeologists used U.S. Geological Survey (USGS) 1:24,000 quad maps, compasses, and visual estimation to locate sites during survey. In some circumstances, archaeologists may have had engineering maps at finer scales to establish site locations or may have used a combination of optical survey equipment and survey markers to measure them. These more accurate techniques, however, were probably restricted to sites that were intensively studied or were judged particularly important. Even in these cases, Wandsnider and Dore (1995:1) noted that "it is possible for site UTM coordinates to err by at least 24 m due to USGS error" and that "10 percent of the [USGS] points can have errors greater than 24 m."

Without instruments, the situation deteriorates. At sites where archaeologists can locate themselves accurately on a map, the scale at which geographic information is mapped is coarse enough to introduce uncertainty on the scale of tens to hundreds of meters. According to a study conducted by Wandsnider and Dore (1995), site locations recorded using paper maps typically are considerably more than 100 m off their "true" locations and can be as much as 350 m off, even for site locations for which field crew were highly confident of their accuracy.

In contrast to earlier methods for establishing site location, GPS instruments can offer much more accurate measurements but do not eliminate the problem of error in site location. When GPS instruments first became available to archaeologists, some measurements could be differentially corrected with data collected at base stations and with specialized correction software. In such cases, differentially corrected data were often accurate within one or a few meters. More-recent professional-grade GPS units with differential correction can provide measurements within centimeters of the true geodetic location. Where differentially corrected data are not available, as is true for many consumer-grade GPS units, errors in site location could be substantially larger. Unfortunately, we are aware of few formal studies that empirically estimated the error in uncorrected GPS measurements.

Our best estimates are that before selective availability was turned off in May 2000, uncorrected GPS measurements of site location were typically within 15 to 25 m of their true location but could be as much as 100 m off in some cases. After selective availability was turned off, many GPS units produced measurements that had a reported accuracy of within 12–15 m of the true location. With differential correction, some of these units could produce measurements within 3–5 m of the true location.

Another source of error in site location is user error. One of the most common problems with use of GPS is a misunderstanding or lack of knowledge of datums and projections. GPS instruments are sometimes used to collect UTM coordinates for site location, but the corresponding datum for those measurements is not recorded or the wrong datum is used. As a result, the data are later stored in GIS with the wrong datum; the result may be a substantial offset from the true location. A common problem is for measurements to be taken with a North American Datum 1983 (NAD 83) datum but then to be projected as if the measurements were collected with a NAD 27 datum. The resulting error in site location is a standard offset in site location, typically around 200 m, in a specific direction (Figure 3.1). The exact size and direction of the offset depend on the area of the country where the GPS coordinates were collected.

All instruments are associated with error. Errors in site location were presumed to have decreased with the use of GPS. Is this true? If so, must we revisit all sites on installations, or can we determine ways of utilizing Legacy data? We explore these questions with a case study from the UTTR.

The Utah Test and Training Range (UTTR), Hill AFB

The UTTR is located in the eastern Great Basin, within the Bonneville Basin of the Great Salt Lake Desert (GSLD) in northwestern Utah (Figures 3.2 and 3.3). It is divided into the North and South Ranges. Cultural resource investigations in the UTTR have identified sites from throughout prehistory; the Bonneville and Wendover period assemblages dominate much of the archaeological record (Young 2008:13–14). These sites are associated with the extinct Old River drainage system, which was characterized, in the early Holocene, by productive wetlands that covered some of what is now UTTR land. Significant sites from more-recent periods are found in the dune and spring environments in and around the GSLD.

The CRM office of Hill AFB Environmental Management (CEV) furnished SRI with GIS data and metadata for all sites presently known from the North and South Ranges of the UTTR. These included point data for site location (site datums) and polygon data for the extent of sites (site boundaries, site areas). The data did not initially track information about site location and extent for sites that had been visited two or more times (and had locational data collected during each site visit). Nevertheless, Kate Stratford, a GIS specialist with Hill AFB CEV, fortunately was able to reconstruct this information from UTTR site records and old GIS files archived in the Hill AFB CEV office (Stratford 2008a, 2008b).

Here is Stratford's (2008a) verbatim account of her UTTR site-data reconstruction:

> I have been working with the CRM data at HAFB Environmental Management GIS for the past 3 years. During that time I have consolidated all site GIS data into a master personal geodatabase, with separate feature classes (layers) for the different feature types (site datums, site areas, inventory areas, isolated artifacts, site artifact points, etc). Prior to that time site location information was scattered among several files, or combined with other feature types, and some site locations were missing. As I discovered missing properties, I updated the missing locations using UTM coordinates and report location maps (1:24K scale, USGS 7.5' topographic maps) if no other digital data could be found.
>
> For the current project I attempted to determine the source of each individual site record existing in the working CRM database. Because of the incomplete and fragmented condition

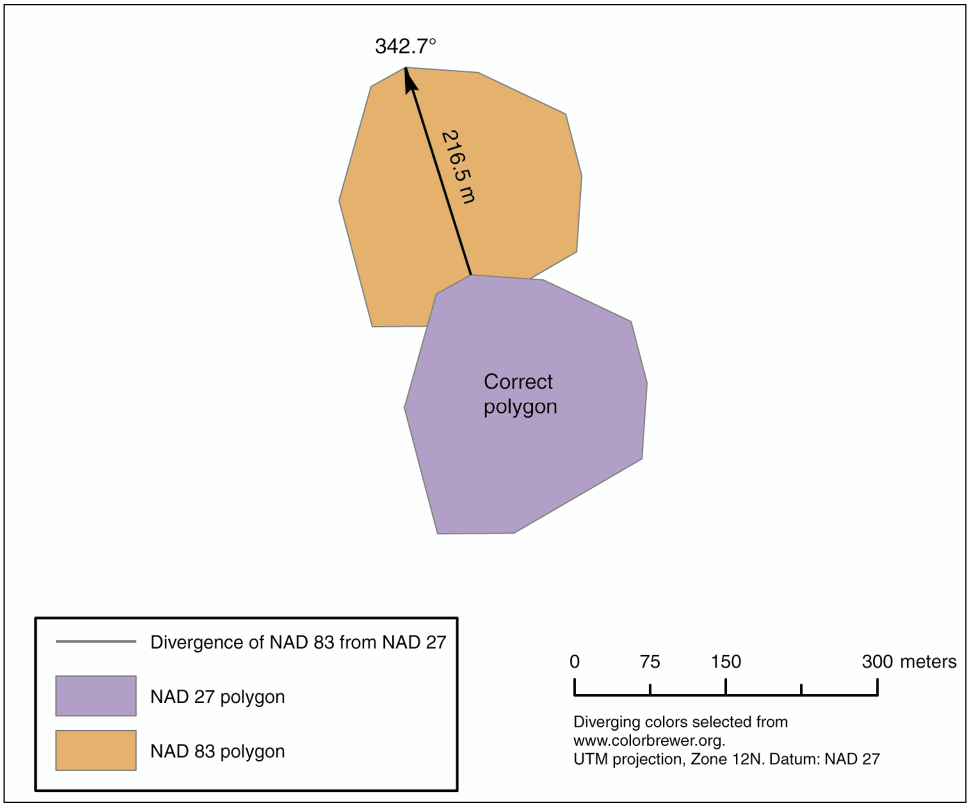

Figure 3.1. GPS measurements taken with the NAD 83 datum but then projected as if the measurements were collected with the NAD 27 datum (see text for explanation). The resulting error in site location is a standard offset in site location, typically around 200 m.

Figure 3.2. Landforms in the Bonneville Basin, northeastern Utah (photo courtesy of Hill AFB).

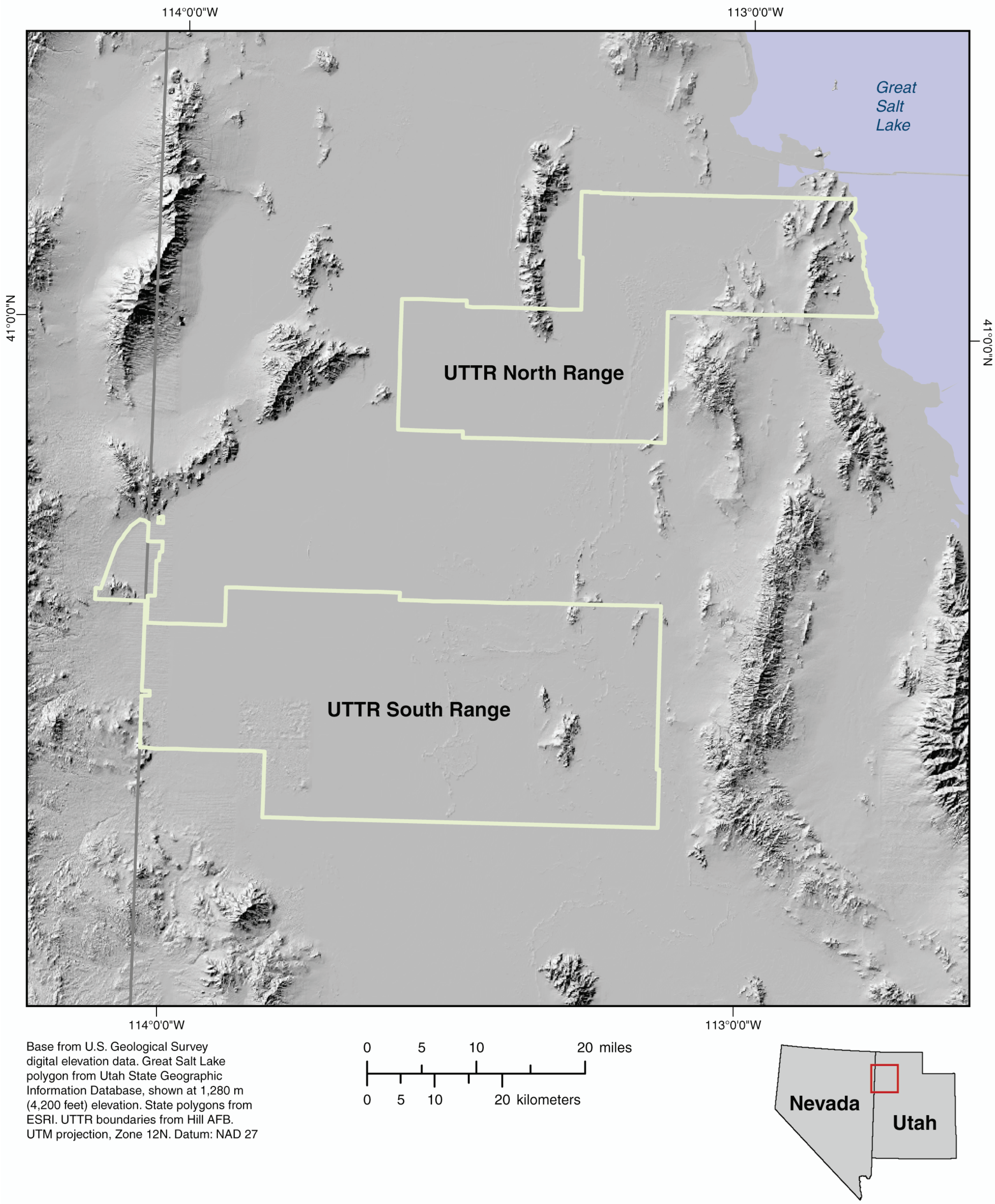

Figure 3.3. Map of the North and South Ranges of the Utah Test and Training Range (UTTR), Hill AFB, Utah.

of the information, it was sometimes unclear as to the source of individual records. For some records I knew that they were collected via GPS because they were included with files delivered with final reports or had a GPS file reference in the attribute table. For other records the locations did not correspond to the UTM location recorded on individual site forms and were obviously created with a different source method. After a little investigating I discovered that most of the known sites recorded prior to 1997 were visited by the HAFB staff in the summer of 1997 and a GPS location was taken for those sites (it did not appear that the sites west of Wildcat Mountain were visited and GPS recorded at this time). I discussed the possible sources for the remaining site locations with the Base Archeologist Jaynie Hirschi and we concluded that sites documented after 1998 were probably recorded using GPS and, with certainty, that sites recorded after 2004 were GPSed.

> HAFB maintains an Archeological Site (MS Access) database containing information recorded on the IMACS (Intermountain Antiquities Computer System) site forms for all archeological properties. Using the UTM northing and easting data I added the x, y for all sites to an Excel table, plotted the site locations in ArcMap v 9.2 (command: Tools > Add X, Y), and created a new shapefile. I compared these UTM locations to the existing locations contained in the HAFB CRM program's GIS database.

> Site locations were also compared to 1:24K USGS 7.5' topographic location maps found within the inventory reports. Often the recorded UTM location was inaccurate and didn't correspond to the existing location information, so the report map was used to determine the original or intended recorded location.

> When location discrepancies were identified, I would refer to the inventory report, site form, site sketch (if available), and location maps, and attempt to reconcile the difference. If a site had been revisited, re-recorded, or tested I attempted to document the datum/site location and boundary for each revisit. I checked through the inventory reports and identified sites that were re-recorded or tested and checked any updated site forms and site sketch maps, and plotted the UTM datum points.

> Another resource I used was Google Earth. I exported the shapefile (.shp) to a file (.kml) that could be viewed in Google Earth. This was especially useful to reconcile site locations for rockshelters and to provide a much larger overview of the installation area.

> As mentioned above, many sites were revisited during the summer of 1997 by the then CRM Program Manager and another HAFB GIS/GPS staff member. Their objective was an initial effort to collect site locations and boundaries for all recorded archaeological sites on UTTR property with GPS equipment in order to build a GIS database for the CRM program. Because neither individual had formal archaeological training, it is uncertain what criteria were used to define the site location or boundaries. For example, did they attempt to relocate and GPS existing site datums, or did they establish new ones? We believe the sites were relocated using the recorded UTMs and most likely the report location maps, but it is still unclear how the boundaries were determined—whether by referencing the site sketch maps or by their own observations on the ground.

This is what Stratford (2008a) was able to learn about site-boundary definition:

> Site sketches for sites recorded from 1991–1996 are considered to be inadequate. Many of the boundaries for these sites were digitized based on the report location maps, so they represent only generalized portrayals of these site areas.

> For those sites that were recorded multiple times (during a re-inventory or testing project) I referred to the site sketches and reports to digitize the boundary recorded for each visit. Recently recorded boundaries (>2005) were collected with GPS, so updates were made to the originally recorded boundary. These were in general modest changes resulting from testing activities. I digitized the changes from the site sketch as best I could (not georeferenced) and these can be considered approximations of the boundary adjustment.

> Not all sites are represented in the site area file. In some cases the sites did not have a defined site area when the site sketch was inspected (no boundary drawn, just features) and were identified solely by a point on the location map. Other sites were so small in size that it was not practical to add the boundary. Therefore, many sites still have no boundary recorded in the GIS database (just a location point in the site datum file).
>
> For all sites recorded in 1997 by the CRM program, I attempted to identify the originally recorded site boundary from the inventory reports and site sketches. If the boundary was significantly different than what was GPSed during the 1997 site visit I would digitize in the original boundary from the documentation (again an approximation).

The foregoing account illustrates not only many of the difficulties Stratford encountered in reconstructing older site-recording episodes but also some of the shortcomings in the UTTR data set. Such deficiencies are not unique to the UTTR data but are probably common to all large installations. These and other factors—sources of uncontrolled variation—condition the analyses presented below. They include the following:

- Some site data records are incomplete and fragmented condition.
- Not all sites have information on extent or area.
- The recording technique for some sites is either unknown or not precisely known.
- The "source" of location data for most of the earlier sites is just the most proximate one and does not actually identify the field methods used to collect them. In other words, "UTM coordinates" or "site sketch" refers to information on a site form, not to field techniques such as triangulation or estimating position using a topographic map. It may be possible in the future to extract this information from survey or site reports.
- Some site revisits could not be dated.
- Fifteen individuals and organizations have been responsible for recording the data over time. Individuals may have differed in their knowledge of, or training in, using location-recording equipment, such as total stations or GPS receivers.
- The models of GPS units employed are unknown. Again, it may be possible in the future to extract this information from survey or site reports or by interviewing installation staff or contractors.
- Older GPS units were not as accurate as newer models are. In addition, the removal of the deliberate "selective availability" error from the U.S. GPS system after May 2, 2000, has led to increased accuracy of all GPS receivers across the board after that time.

Despite the list of caveats just noted, there is no question of the worth of the UTTR data for data-quality studies. Stratford's work has proved invaluable for the objectives under investigation here. Without it, we would not have been able to analyze changes in site location and extent over time on the basis of temporal variation in recording techniques. It should be noted, also, that the benefits of Stratford's efforts were not all one-sided—her reconstructions of multiple site visits also turned up many problems in the UTTR database that required correction, thereby increasing the overall accuracy and quality of the installation's archaeological data.

Site Datums

The UTTR site-datums GIS file contains 643 points, representing 491 sites, recorded over 31 years from 1975 to 2006 (Table 3.1). Of the 643 records, 41 do not have any recording date assigned. The site datums are grouped into three site classes: Historic, Prehistoric, and Prehistoric/Historic; Prehistoric sites far outnumber the other two classes (Table 3.2). Eight sites were not classified to site class. As noted above, 15 individuals and organizations have been responsible for recording the site-datum data over time. Three CRM

Table 3.1. Counts of UTTR Site Datums, by Year Recorded

Year Recorded	Total
1975	1
1986	5
1991	95
1992	8
1993	24
1994	12
1995	26
1996	31
1997	67
1998	60
1999	6
2000	66
2001	12
2003	2
2004	32
2005	93
2006	62
Blank	41
Total	643

Table 3.2. Counts of UTTR Site Datums, by Site Class

Site Class	Total
Historic	54
Prehistoric	578
Prehistoric/Historic	3
Blank	8
Grand Total	643

companies account for most (497) of these. The remaining 146 sites were recorded by 5 individuals and 6 CRM companies. Forty site-datum records do not note a recorder.

Table 3.3 lists the "source" categories for site-datum data developed by Stratford (2008b) during her reconstruction of site records for sites that had been visited two or more times and had locational data collected during each site visit. It also shows general classes of higher- or lower-accuracy field measurement techniques matched to the data "sources." Table 3.4 shows the options for field measurement techniques available within each accuracy category.

Methods of Analysis

Investigation of the UTTR site-datum data began by creating polylines that joined an older site datum to a newer site datum for the same site. Sites with one revisit (two site-datum records) thus had only one line associated with them; sites with two revisits (three site-datum records) had two lines associated with them, and so forth. The length of each line connecting two site-datum records (the distance between the two points) was then calculated. Subsequent analysis focused on the distances between revised site datums, the sources of data or recording techniques employed, and the years in which the site-datum records were created.

This part of the study is underlain by the assumption that, all other things being equal, the point locations being measured for each site over time (the site datums) occupy the same physical location on the surface of the Earth. New measurements of a site datum over time are not creating new site datums. Although this assumption may not always hold true for the data set under investigation, it is a necessary starting point for guiding the analysis.

No significance testing of relationships was attempted, because the work is exploratory. Nevertheless, most (if not all) of the relationships identified or comparisons made throughout the discussion are obvious.

ArcGIS 9.2, ET GeoWizards 9.8, and Microsoft Excel 2003 were used to edit, extract, and otherwise manipulate data. In particular, the Point To Polyline Wizard (found in the freeware version of ET GeoWizards 9.8 for ArcGIS from ET SpatialTechniques) was used to create the polylines between site-datum points. ArcGIS 9.2 was used to generate GIS-based figures, and Golden Software's Grapher was used to construct graphs.

Results

Table 3.5 shows site-datum source or recording technique according to the year recorded, for all 643 UTTR site-datum records. Site-datum source is ordered in the table approximately from high to low accuracy. If one were to draw a diagonal line across Table 3.5 from the upper left to the lower right, it would almost perfectly bisect the table entries into two groups (excluding the Blank column): (1) a group on the lower left, dominated by site-datum-location recording methods that are of lower accuracy, are paper based, and use topographic maps; and (2) a group on the upper right, consisting mostly of higher-accuracy, GPS-based measurements. Site-datum-location recording methods that use topographic maps, once the mainstay of archaeological survey recording, began to drop off quickly after 1996, as GPS-based measurement supplanted them, certainly by 1997, and possibly a bit earlier.

A possible inference from Table 3.5 is that the measurement of a site datum's location becomes more accurate through time. If this inference is correct, then variation in measurements should decrease through time, ultimately converging on the manufacturer's instrument error.

Table 3.6 lists counts and average distances for the 110 site datums from 135 sites with multiple site-datum recording episodes for which year updated was noted. Statistics for the distance variable are shown at the bottom of the table. As a mean of displaying patterns, Figures 3.4 and 3.5 graphically depict the count and distance data, respectively. Figure 3.4 demonstrates that many site-datum measurements were taken in 1997; fewer measurements were taken in the other years studied. Figure 3.5 shows that 1996 is the year

Table 3.3. GIS File Attributes for Sources of Data on Site Datums (Stratford 2008b)

Source, by Field Measurement Technique	Explanation
Higher accuracy (GPS instruments)	
GPS	Datum location collected with GPS.
GPS?	Datum location probably collected with GPS, but not certain.
UTM/GPS	Datum plotted with UTM coordinates that were initially recorded/verified using GPS.
Lower accuracy (1:24,000 USGS 7.5-minute topographic map annotation)	
UTM/site sketch	Datum plotted by UTM coordinates (UTM NAD 27) recorded on site form (IMACS) with additional use of site sketch map to plot location.
UTM	Datum plotted by UTM coordinates (UTM NAD 27) recorded on site form (IMACS).
SHPO	Datum plotted by UTM coordinates (UTM NAD 27) recorded at SHPO.
Report topo	Datum plotted using report topographic location map; either site form UTM coordinates wrong or site form unavailable—location very imprecise.
Unknown	No information about data source.

Key: GIS = geographic information system; GPS = Global Positioning System; IMACS = Intermountain Antiquities Computer System; NAD = North American Datum; SHPO = State Historic Preservation Office; USGS = U.S. Geological Survey; UTM = Universal Transverse Mercator

Note: Sources are ordered approximately from highest to lowest accuracy. Please see Table 3.4 for more information on field techniques.

Table 3.4. Field Recording Techniques for Site-Datum Locations

Field Measurement Techniques	Options
Higher accuracy (GPS instruments)	• Consumer-grade GPS measurement before selective availability was removed in 2000 (±15 m?). • WAAS-capable consumer-grade GPS measurement after selective availability was removed in 2000 (±3 m?). • Professional-grade GPS with differential correction measurement after selective availability was removed in 2000 (±1 m down to ±0.01 m?).
Lower accuracy (1:24,000, 7.5-minute USGS topographic-map annotation)	Topographic matching or triangulation using 1:24,000 topographic map, visual estimations, and compass (±200 m?).[a]

Key: GPS = Global Positioning System; WAAS = Wide Area Augmentation System; USGS = U.S. Geological Survey.
[a] One way to get an idea of the accuracy limitations of annotating a 1:24,000 topographic map with a pencil is to keep in mind that an average pencil line on the map represents about 12–15 m on the ground.

Table 3.5. Site-Datum Source or Recording Technique by Year Recorded

Site-Datum Source	Year Recorded																		Total
	1975	1986	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2003	2004	2005	2006	Blank	
GPS									61	58	4	40		2	25	93	61	2	346
GPS?						7		1										19	27
UTM/GPS										1	2	21	1		7		1		33
UTM/site sketch	1	5	46	6									9						67
UTM			47	2	24	5	26	29	6	1		3	2						145
SHPO																		1	1
Report topo			2					1				2						3	8
Unknown																		16	16
Total	1	5	95	8	24	12	26	31	67	60	6	66	12	2	32	93	62	41	643

Key: GPS = Global Positioning System; SHPO = State Historic Preservation Office; UTM = Universal Transverse Mercator
Note: All UTTR site-datum records are included. Site-datum source is ordered approximately from high to low accuracy.

Table 3.6. Site-Datum Counts and Average Distances for Sites with Multiple Site-Datum Recording Episodes, by Year Updated

Year Updated	Number	Distance Mean (m)
1996	5	222.56
1997	67	55.80
1999	2	0.79
2000	16	94.40
2001	1	13.10
2003	2	121.68
2004	8	138.61
2005	7	143.22
2006	2	61.85
Blank	37	
Total/Overall mean	147	80.50[a]

Note: Distance variables are as follows: n = 110, minimum = 0.00, maximum = 607.08, range = 0.00–607.08, median = 52.02. The overall mean is the mean of all individual cases, not the grand mean of the sub-means by year updated.

[a]Sites having no update year recorded ("Blank") are not included in any of the distance calculations or statistics.

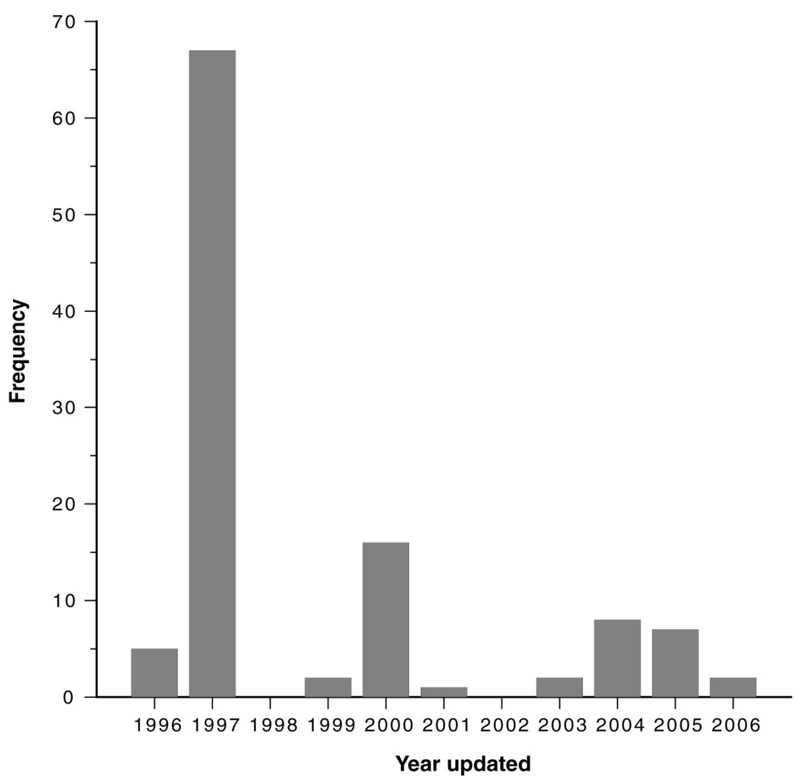

Figure 3.4. Site-datum counts for sites with multiple site-datum recording episodes (n = 110), according to year updated. Please refer to Table 3.6 for calculations and sample statistics.

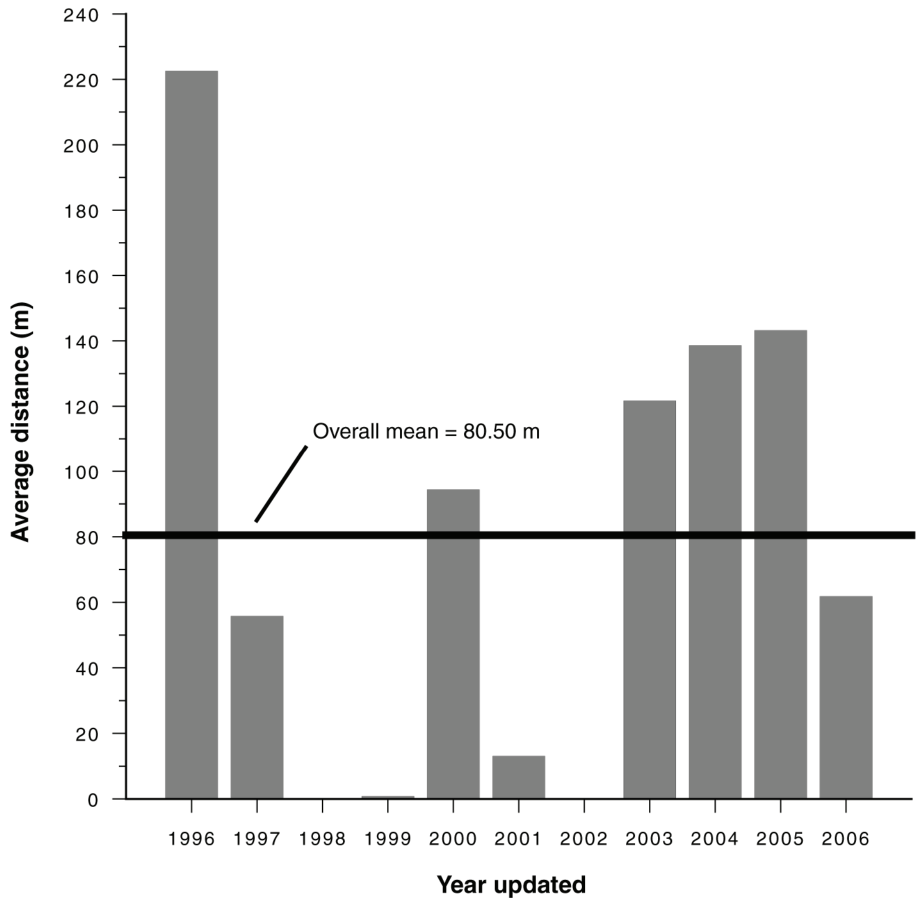

Figure 3.5. Average distance of site-datum-location changes for sites with multiple site-datum recording episodes (n = 110), according to year updated. The overall mean is the mean of all individual cases, not the grand mean of the sub-means by year updated. Please refer to Table 3.6 for calculations and sample statistics.

with by far the greatest average change in site-datum location, amounting to approximately 223 m. The average changes in other years are all less than 143 m, approximately. If there is any pattern at all in average change in site-datum location over time as shown in Figure 3.5, it is the apparent rise in average change in location between 1997 and 2005 (apart from 1999 and 2001; there are no values for 1998 and 2002). Except for 1997 and 2000, the number of observations for each of the years graphed is relatively small. We suspect that the rise in average change in site-datum location is a reflection of Hill AFB personnel's focus on sites with known or presumed locational problems.

The overall mean for all site-datum-location changes is 80.50 m. The most plausible interpretation is that the mean measures the improvement in accuracy from the set of earlier site datums (almost all of which range in age from 1991 to 2000, with one observation in 2004) to the set of later site datums that replaced them. In addition, if we had better data on the types of locational measurement instruments and recording techniques that were actually used to measure site-datum positions in the field, we might be able to begin assigning plus-or-minus ranges of uncertainty, like those listed in Table 3.4, to qualify the precision of measurements.

Discussion

Should UTTR managers be concerned with an average error rate of 80.50 m? Given the size of the installation, one might think that managers could easily deal with such a small distance during planning activities. For example, why not just buffer each site location by 80 or 100 m? In many cases, this solution might work well. There are cases, however, when it is important that we have a better understanding of exactly where a site is located.

One such situation is when the resource may be a highly significant scientific or traditional property. On the UTTR, cave sites tend to be one such resource. These sites often have little surface exposure because most of the deposit is actually “inside” a topographic feature, such as a hill or mountain.

Figure 3.6 shows site-datum locations at three different points in time for site 42BO0683, a prehistoric cave site. This site has no site-extent (i.e., site-area) polygon(s) associated with the site-datum points. The fact that the 1991 site location is more than 100 m south of the 2000 GPS recording is understandable in light of the switch from visual placement on 1:24,000 USGS topographic maps to increasingly higher-resolution GPS equipment. The more interesting move is between the 1997 and 2000 GPS site recordings, suggesting that early GPS site locations should be viewed with considerable skepticism. Regardless, the military needs to know where this site is located both to ensure that activities do not adversely impact the deposits and to steward the resource for future generations.

A second situation involves large sites, which are often to difficult to map because different archaeologists make different judgment calls on site boundaries. Site 42BO0909, for instance, is a large lithic scatter, which has two site-extent polygons associated with five site-datum points (Figure 3.7). It is possible that some of the putative datums are subdatums, because four were recorded in the same year (1996). The GIS file identifies the four older points only as “old datums.” Given the poor documentation and the radical shift in site size and location, our confidence in the accuracy of the last recording of the site is tenuous.

A final concern is illustrated by site 42BO0682 (Figure 3.8). This lithic scatter is associated with three site-datum locations measured in 1991, 1997, and 2000, respectively. It is curious that, as the site datum moved to the east, the site boundary did not change. Either the site datum was mapped inaccurately relative to the rest of the site, or the site boundaries were not reexamined during subsequent visits to the site.

The last two examples indicate that studying site datums alone is not sufficient to assess the quality of site-location data; we also must examine changes in the recording of site boundaries. Chapter 4 treats this subject.

Centroids

The UTTR site-areas GIS file contains 437 records, representing 386 sites, recorded over 15 years from 1991 to 2006 (Table 3.7). Of the 437 records, 80 do not have any recording date assigned. Table 3.8 lists the “source” categories for site-area data developed by Stratford (2008b) during her reconstruction of site records for sites that had been visited more than one time and had locational data collected during each site visit. It also shows general classes of higher- or lower-accuracy field measurement techniques matched to the data “sources.” Table 3.9 shows the options for field measurement techniques available within each accuracy category.

Methods of Analysis

Investigation of the UTTR site-area data began by creating centroids for each site-area polygon. A centroid is the geometric center of a feature; for a polygon, it is its center of area. Centroids of site-area polygons are useful in exploring the role of instrument error in the location of site areas because the centroid is a good way of representing the location of that site area in space by just a single point.

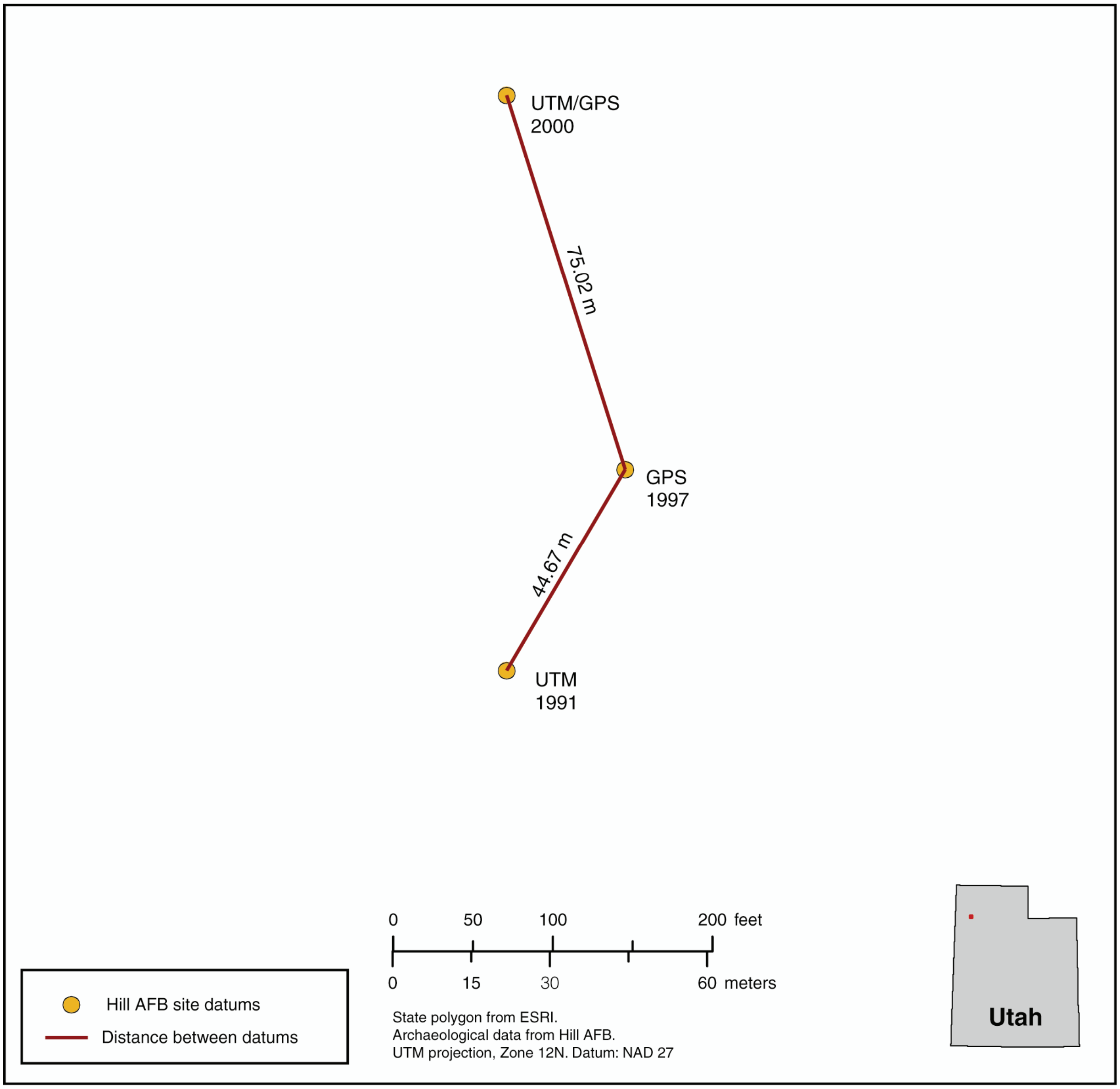

Figure 3.6. Example of site-datum change over time: site 42BO0683. This site has no site-extent (i.e., site-area) polygon(s) associated with the site-datum points.

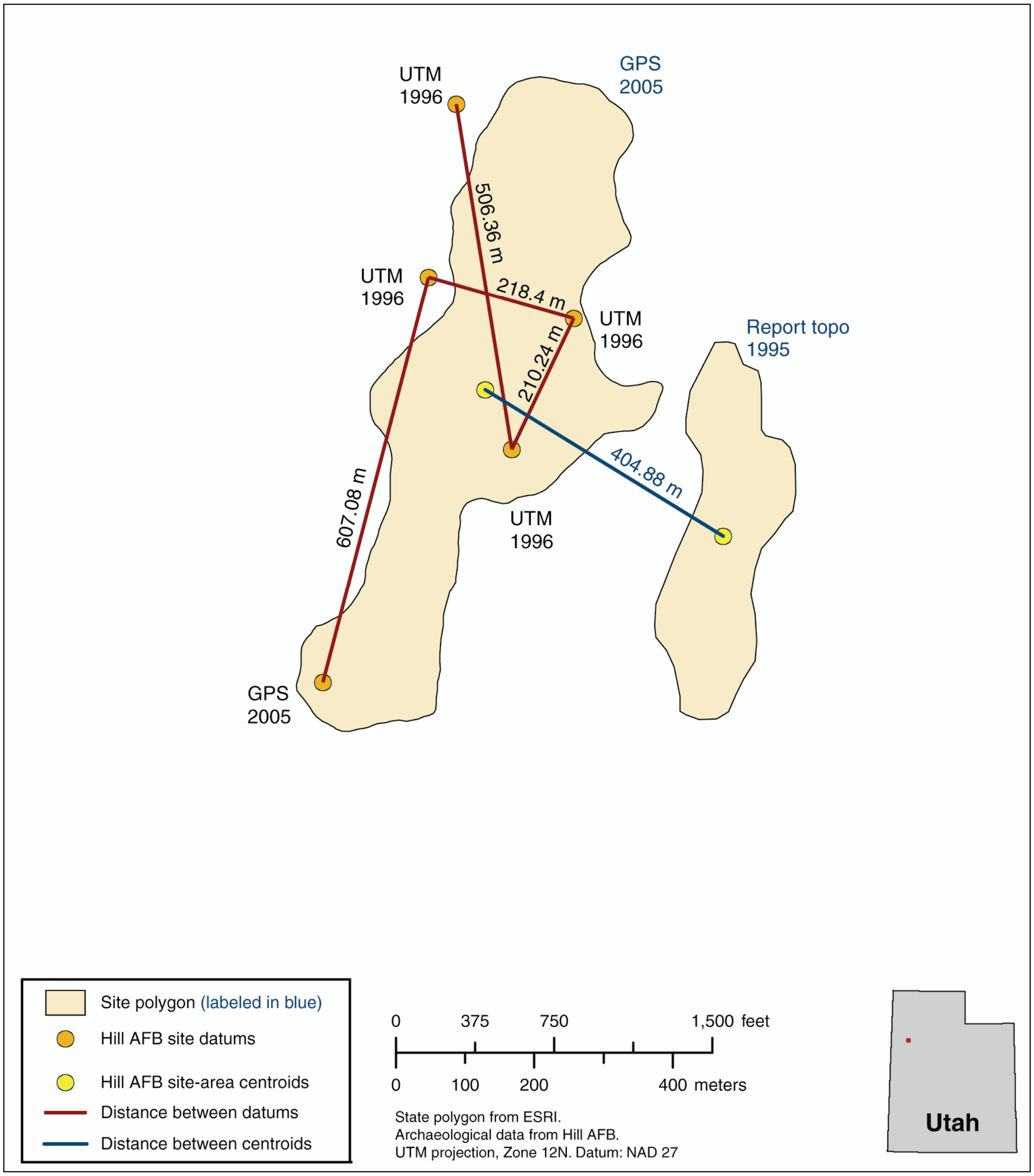

Figure 3.7. Example of extreme site-datum change over time: site 42BO0909. This site has two site-extent polygons associated with the site-datum points.

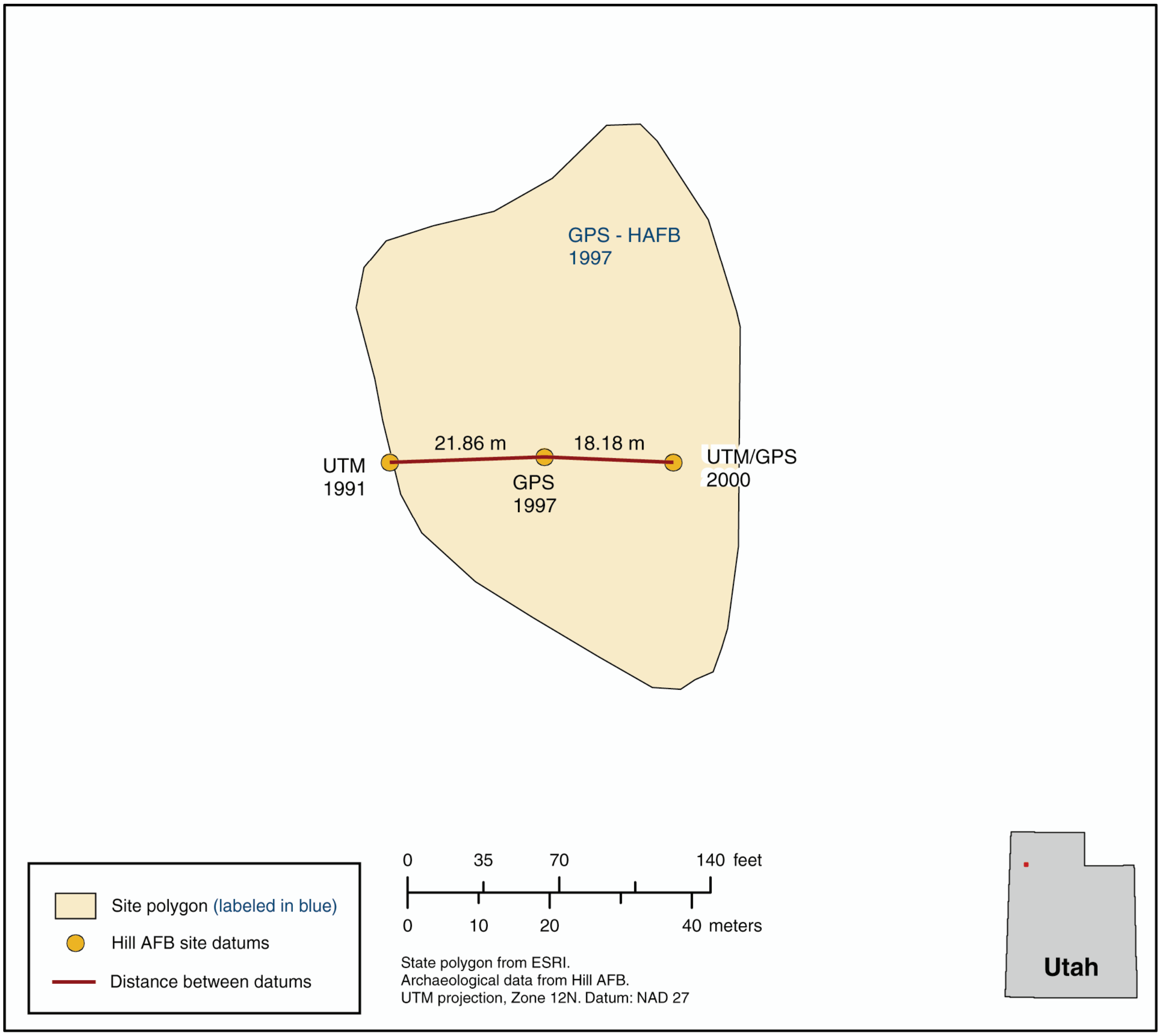

Figure 3.8. Example of site-datum change over time: site 42BO0682. This site has only one site-extent polygon associated with the site-datum points.

Table 3.7. Counts of UTTR Site Areas, by Year Recorded

Year Recorded	Total
1991	46
1992	4
1993	15
1995	13
1996	11
1998	58
1999	2
2000	33
2001	10
2004	21
2005	86
2006	58
Blank	80
Total	437

Table 3.8. GIS File Attributes for Sources of Data on Site Areas (Stratford 2008b)

Source, by Field Measurement Technique	Explanation
Higher accuracy (GPS instruments, total stations)	
GPS	Site boundary collected with a GPS during a site-recording project.
GPS-HAFB	Site boundary collected with a GPS during a site visit and not during a site-recording project.
GPS?	Site boundary probably collected with GPS, but not certain.
GPS/site sketch	Original site boundary collected with a GPS during a site-recording project, with later updated site sketch map changes (usually subsurface testing project resulting in boundary adjustments) digitized.
Georeferenced sketch	Site boundary digitized from a georeferenced site sketch (digital JPEG), tied into the site datum, that was brought into ArcView.
Lower accuracy (1:24,000, USGS 7.5-minute topographic map annotation, rangefinder, compass-and-pace)	
Site sketch update CEV GIS	Site boundary updated by HAFB CEV GIS staff from updated site sketch map (usually subsurface testing project resulting in boundary adjustments)—updated boundaries are approximations.
Site sketch–IMACS	Site boundary digitized from site sketch map—measurements are taken from datum location and are an approximation of the boundary.
Site sketch	Site boundary digitized from site sketch map in report only—boundaries are approximations.
Report topo/site sketch	Site boundary digitized using a combination of the report topographic location map and site sketch map—boundaries are approximations.
Report topo	Site boundary digitized using the report topographic location map only (no site sketch available or too unintelligible to use).

Key: CEV = Environmental Management; GIS = geographic information system; GPS = Global Positioning System; HAFB = Hill Air Force Base; IMACS = Intermountain Antiquities Computer System; JPEG = Joint Photographic Experts Group; USGS = U.S. Geological Survey; UTM = Universal Transverse Mercator

Note: Sources are ordered approximately from highest to lowest accuracy. Please see Table 3.9 for more information on field techniques.

Table 3.9. Field Recording Techniques for Site Areas

Field Measurement Techniques, by Level of Accuracy	Options
Higher accuracy	
GPS instruments	• Consumer-grade GPS measurement before selective availability was removed in 2000 (±15 m?). • WAAS-capable consumer-grade GPS measurement after selective availability was removed in 2000 (±3 m?). • Professional-grade GPS with differential correction measurement after selective availability was removed in 2000 (±1 m down to ±0.01 m?).
Surveying instruments	Total station (±0.01 m?)
Lower accuracy	
1:24,000, 7.5-minute USGS topographic-map annotation	Topographic matching or triangulation using 1:24,000 topographic map, visual estimations, and compass (±200 m?).[a]
Compass methods	• Compass-and-rangefinder (±10 percent?). • Compass-and-pace (±25 percent?).

Key: GPS = Global Positioning System; WAAS = Wide Area Augmentation System; USGS = U.S. Geological Survey.

[a] One way to get an idea of the accuracy limitations of annotating a 1:24,000 topographic map with a pencil is to keep in mind that an average pencil line on the map represents about 12–15 m on the ground.

Polylines that join the centroid of an older site area to the centroid of a newer one for the same site were then constructed. Because there are no more than two site-area records for any site in the GIS file, only a single line was associated with each site. The length of each line connecting two site-area centroids (the distance between the two points) was calculated. Subsequent analysis focused on the distances between the centroids of revised site areas, the sources of data or recording techniques employed, and the years in which the site-datum records were created or updated.

Again, no significance testing of relationships noted was attempted. Nevertheless, most (if not all) of the relationships identified or comparisons made throughout the discussion are apparent.

ArcGIS 9.2, ET GeoWizards 9.8, and Microsoft Excel 2003 was used to edit, extract, and otherwise manipulate data. In particular, the Point To Polyline Wizard (found in the freeware version of ET GeoWizards 9.8 for ArcGIS from ET SpatialTechniques) was employed to create the polylines between site-area centroids. ArcGIS 9.2 was used to generate GIS-based figures and Golden Software's Grapher was used to construct graphs.

Results

For all 437 UTTR site-area records, Table 3.10 shows site-area source or recording technique according to the year recorded. Site-area source is ordered in the table approximately from high to low accuracy. As is true for the corresponding table in the site-datum analysis, if one were to draw a line across Table 3.10 on the diagonal from the upper left to the lower right, it would almost perfectly bisect the table entries into two groups (excluding the Blank column): (1) a group on the lower left, dominated by site-area recording methods that are of lower accuracy, are paper based, and use topographic maps and perhaps compasses; and (2) a group on the upper right, consisting mostly of methods that are of higher accuracy and use measurements based on GPS and possibly other survey instruments. The use of lower-accuracy site-area recording methods began to drop off quickly after 1996, as instruments with higher-accuracy measurement supplanted them, beginning in 1998.

Table 3.10 suggests that additional measurement of a site area or boundary in a year following the year of a site's original recording should reflect better accuracy in determining the locations of site areas. This inference should be especially true for site areas determined by GPS or other high-accuracy surveying instruments. If sites are being affected by erosion or are located in a region of active sand dunes, however, the inference will not completely hold because, in these situations, site boundaries may be independently changing in size or shape over time as a result of natural or cultural processes.

Table 3.11 lists counts and average centroid distances for the 102 site-area records from 51 sites with two site-area recording episodes, according to year updated. Statistics for the distance variable are shown at the bottom of the table. Figures 3.9 and 3.10 graphically depict the count and distance data, respectively. Figure 3.9 highlights the small sample size on which the analysis is based. Figure 3.10 is similar to the graph of average site-datum-location changes (see Figure 3.5) in that the average distance between centroids of site-area polygons in 1997 is below the overall mean, whereas the average distances in 2004 and 2005 are well above the overall mean and more than twice the distance in 1997.

The overall mean for all site-area-location changes, as measured by the distance between centroids of site-area polygons, is 80.37 m. We suspect that this number is a reasonable measure of the improvement in accuracy from the set of earlier site-area polygons (almost all of which range in age from 1991 to 1998, with just three observations in 2005) to the set of later site-area polygons that replaced them. In addition (as was mentioned above in the section on site datums), if we had better data on the types of locational measurement instruments and recording techniques that were actually used to measure site-area polygons in the field, we might be able to begin assigning plus-or-minus ranges of uncertainty, like those listed in Table 3.9, to qualify the precision of measurements.

Table 3.10. Site-Area Source by Year Recorded

Site-Area Source	Year Recorded													Total
	1991	1992	1993	1995	1996	1998	1999	2000	2001	2004	2005	2006	Blank	
GPS										21	86	52	5	164
GPS-HAFB													36	36
GPS?													1	1
GPS/site sketch												6	4	10
Georeferenced sketch						58	2	32					4	96
Site sketch update CEV GIS													15	15
Site sketch–IMACS	22	1		3										26
Site sketch	1			2									1	4
Report topo/site sketch									9				2	11
Report topo	23	3	15	8	11								11	71
Unknown								1	1				1	3
Total	46	4	15	13	11	58	2	33	10	21	86	58	80	437

Key: CEV = Environmental Management; GPS = Global Positioning System; HAFB = Hill Air Force Base; IMACS = Intermountain Antiquities Computer System

Note: All UTTR site-area records are included. Site-area source is ordered approximately from high to low accuracy.

Table 3.11. Site-Area Counts and Average Centroid Distances for Sites with Multiple Site-Area Recording Episodes, by Year Updated

Year Updated	Number	Distance Mean (m)
1997	16	70.28
2000	1	69.89
2001	1	57.19
2003	15	21.91
2004	8	165.14
2005	7	150.19
2006	3	48.66
Total/Overall mean	51	80.37

Note: Distance variables are as follows: n = 51, minimum = 2.22, maximum = 404.88, range = 2.22–408.88, median = 60.26. The overall mean is the mean of all individual cases, not the grand mean of the submeans by year updated.

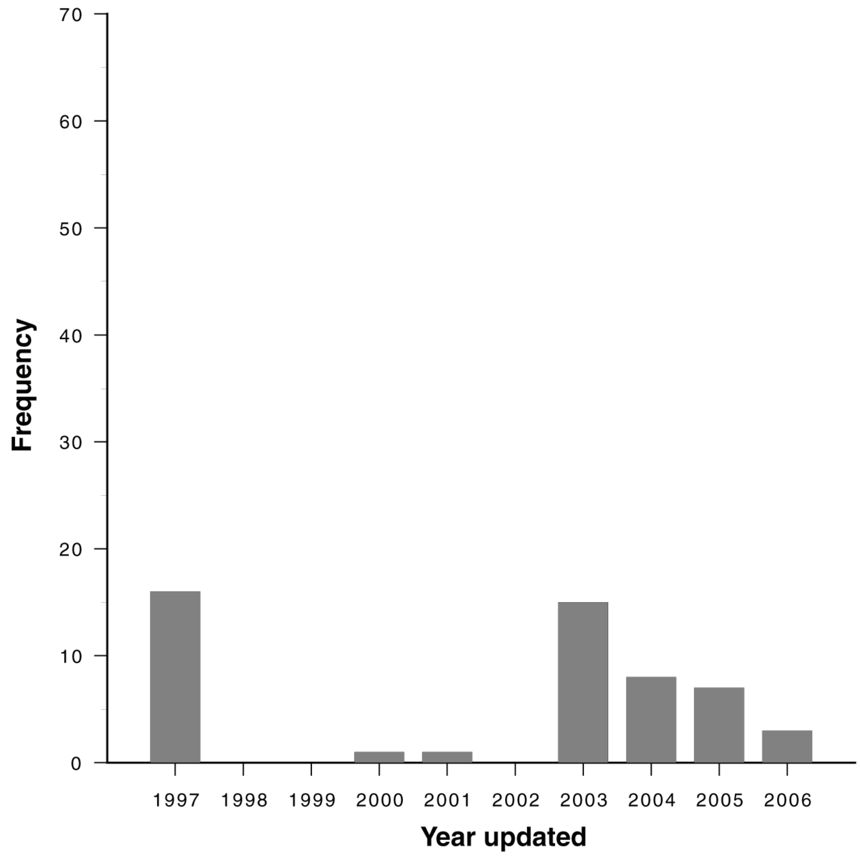

Figure 3.9. Site-area counts for sites with two site-area recording episodes (n = 51), according to year updated. Please refer to Table 3.11 for calculations and sample statistics.

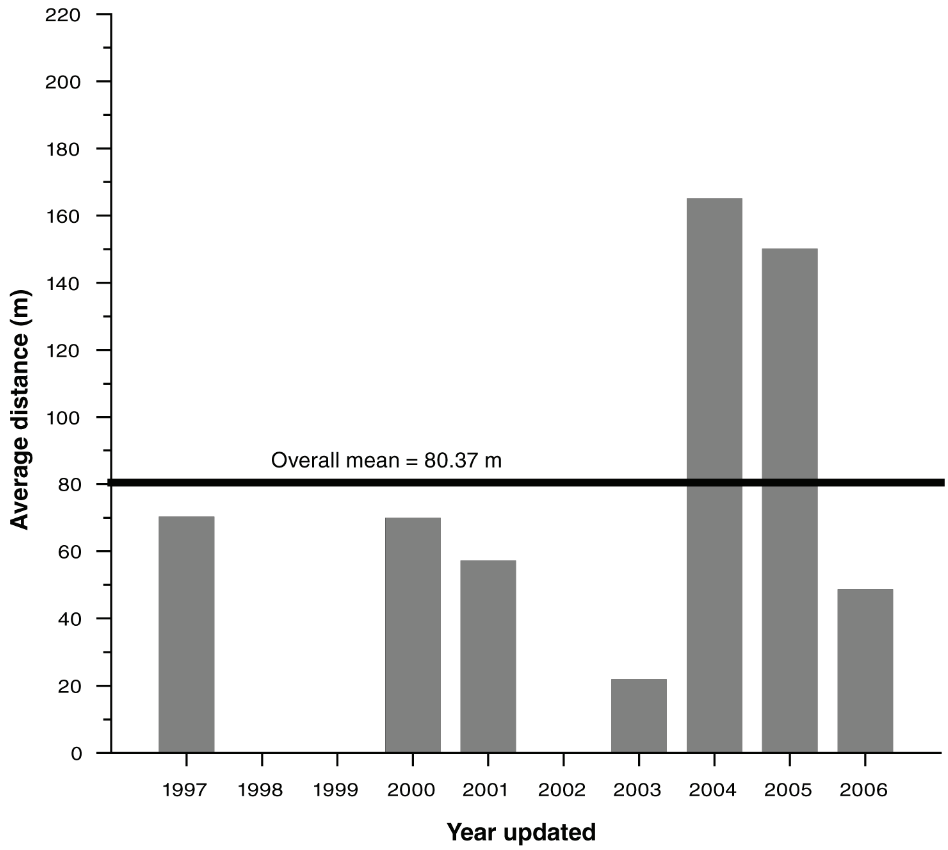

Figure 3.10. Average centroid distance of site-area changes for sites with two site-area recording episodes (n = 51), according to year updated. The overall mean is the mean of all individual cases, not the grand mean of the sub-means by year updated. Please refer to Table 3.11 for calculations and sample statistics.

Discussion

The overall mean of the change in site-datum location (80.50 m) and the overall mean of the change in centroid distances for site-area polygons (80.37 m) are very similar. One explanation for the similarity between the means follows from the fact that site boundaries were based on measurements taken from a site datum. Site centroids computed from polygons are thus indirectly derived from datum measurements. The small difference in accuracy between datums and centroids may simply suggest that the two measures are geodetically accurate to roughly the same degree. Another explanation, which is not mutually exclusive of the first, is that when archaeologists revisited sites, they simply measured the site datum in space and then reentered the new coordinates into the installation's GIS site database. The old polygon delimiting site boundaries was then adjusted to the new site datum.

We examined the case studies used in the centroid analysis to determine if either interpretation was consistent with the data. Although we discuss the issues of site boundaries in the next chapter, we highlight below three situations that seemed typical of the various recording methods used on the UTTR.

The first situation is one in which the site datum does not change, but the site area shifts location. A good example of this problem is site 42BO0810, a prehistoric quarry (Figure 3.11). This site has two site-datum points associated with two extent or area polygons. The two site-datum points have identical coordinates, however, and neither lies within the boundary of the older polygon. A line connects the centroids of the two site-area polygons. The site-area polygons partially overlap, and the newer site-area polygon is

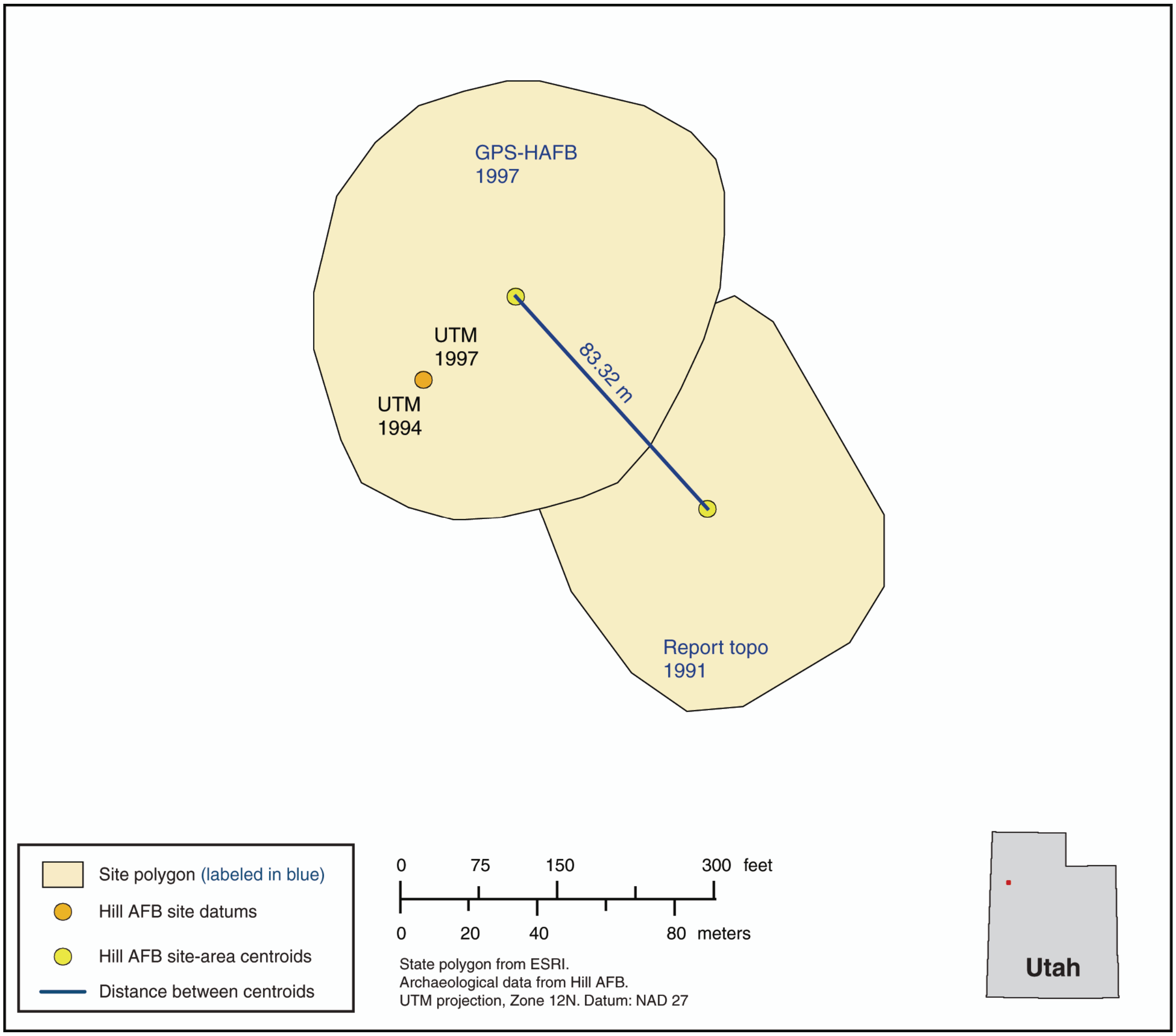

Figure 3.11. Example of site-area change over time: site 42BO0810. This site has two site-datum points associated with two site-extent polygons. The two site-datum points have identical coordinates, however.

larger than the older one. In this case, it is clear that the archaeologists responsible for recording the site at different times had different ideas about the size and shape of the site but not about the location of the site datum. The placement of the site datum may have been shaped by local conditions, such as a prominent landmark.

A second situation involves a change in size, shape, and location. In this case, there is no overlap between site polygons or site datums: indeed, the entire site seems to move. Site 42BO0911, a prehistoric lithic scatter, has two site-datum points associated with two extent or area polygons (Figure 3.12). Although the newer site datum lies within the newer site-area polygon, the older site datum is not within the boundaries of either site-area polygon. The site-area polygons do not overlap, and the newer site-area polygon is larger than the older one.

The third situation is when the site appears to grow in size. In this case, any change in site location is much less problematic than the dramatic increase in site size over time. A good example of this situation is site 42BO0674, a prehistoric lithic scatter/habitation/processing site (Figure 3.13). This site has two site-datum points associated with two extent or area polygons. The older site datum lies within the older site-area polygon, and the newer site datum lies within the newer site-area polygon. The site-area polygons also overlap slightly. But the real story here is that the newer site-area polygon is many times larger than the older one.

Our examination suggests little doubt that archaeologists revisiting sites rerecorded both their location and their boundaries. Although we cannot explain why the measurements of site-datum and centroid change are so similar, it is important not to lose sight of the fact that both the exact location (site datum) and the center of the site (centroid) are quite variable. Ideally, the recording of these locations is becoming more accurate as technology improves. But the variation is still substantial, and the question becomes whether we can use this variability to our advantage.

A Spatial Statistical Approach for Improving Accuracy of Site Datums

One method to gradually improve the precision of archaeological site-datum locations over time is presented below. Instead of just storing the latest coordinates (or the ones identified as the "best" or "most reliable," for example) for a site datum in GIS files, an installation could store a history of site-datum determinations for each site. Gradually, over time, the number of site-datum location records would accumulate to the point at which spatial statistics could be applied to the collection of points. One could then apply the spatial-data equivalents of well-known tabular statistics, such as the mean and standard deviation, with the objective of summarizing the variation in a collection of point records to a single, more precise average value, while maintaining and displaying the variation inherent in that value.

This approach can be demonstrated by creating a fictitious or hypothetical set of points to represent, for example, GPS measurements of a single site datum, accumulated over time. In Figure 3.14, 16 hypothetical site-datum points are mapped, together with the positions of the spatial statistics of Mean Center and Standard Deviational Ellipses for 1, 2, and 3 standard deviations.

The Mean Center is the average *x*- and *y*-coordinate of all the points shown in Figure 3.14 (ESRI 2008a, 2008b). Just like the traditional mean applied to a column of numbers, however, the Mean Center is sensitive to low or high outliers in any direction in a data set. Therefore, in Figure 3.14, the effect of the lowest and highest points (measured on the north–south axis) is to elongate the Standard Deviational Ellipses along this axis. Figure 3.15 illustrates how the shapes of the Standard Deviational Ellipses change to a more circular appearance when the northernmost and southernmost points are removed from the data set.

A Standard Deviational Ellipse, created by the Directional Distribution tool in ArcGIS (ESRI 2008c, 2008d), is derived by calculating the standard distance separately in the *x* and *y* directions. These two measures define the axes of an ellipse encompassing the distribution of features. The term "Standard Deviational Ellipse" refers to the calculation of the standard deviation of the *x*- and *y*-coordinates from the mean center

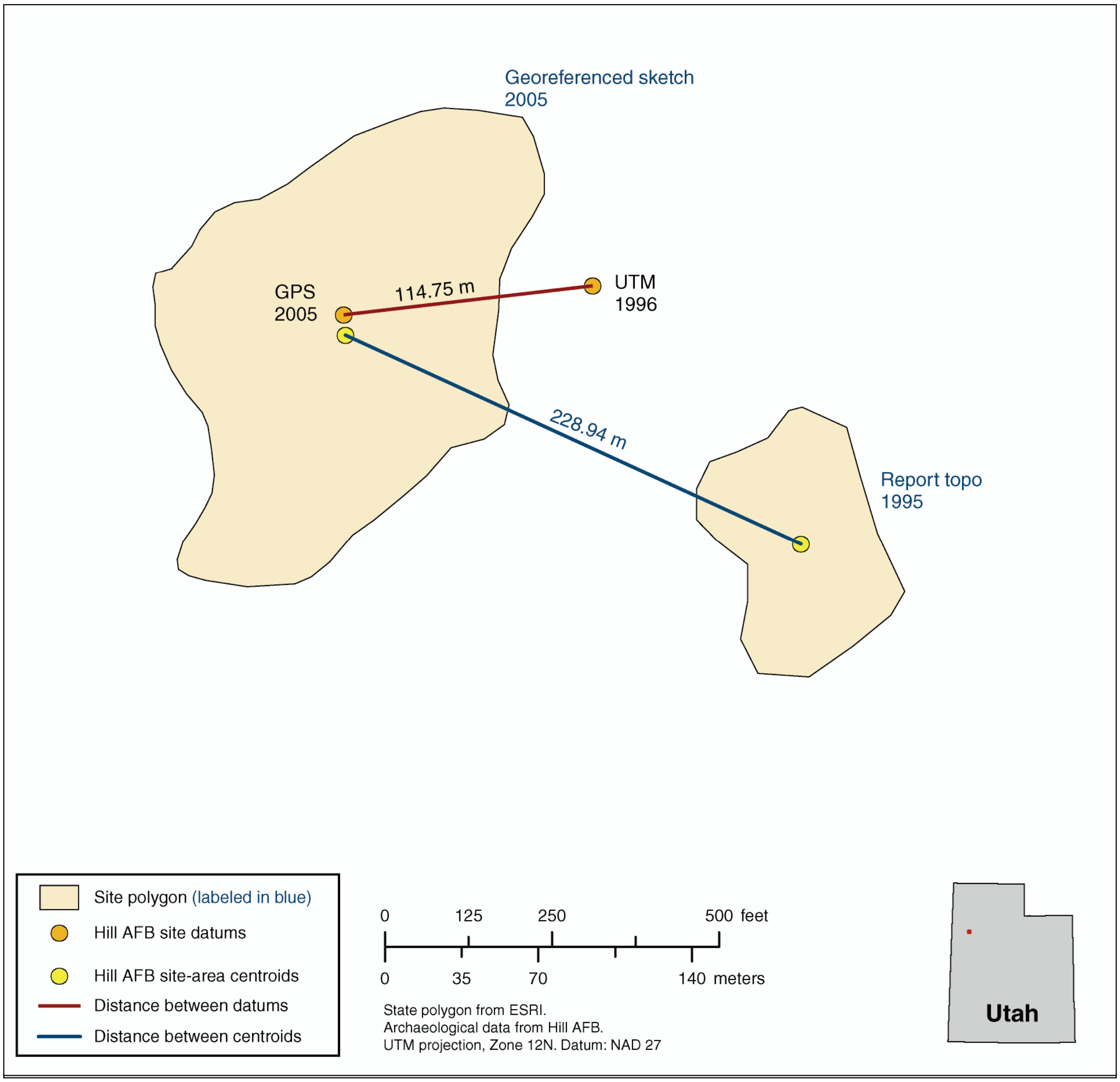

Figure 3.12. Example of site-area change over time: site 42BO0911. This site has two site-datum points associated with two site-extent polygons.

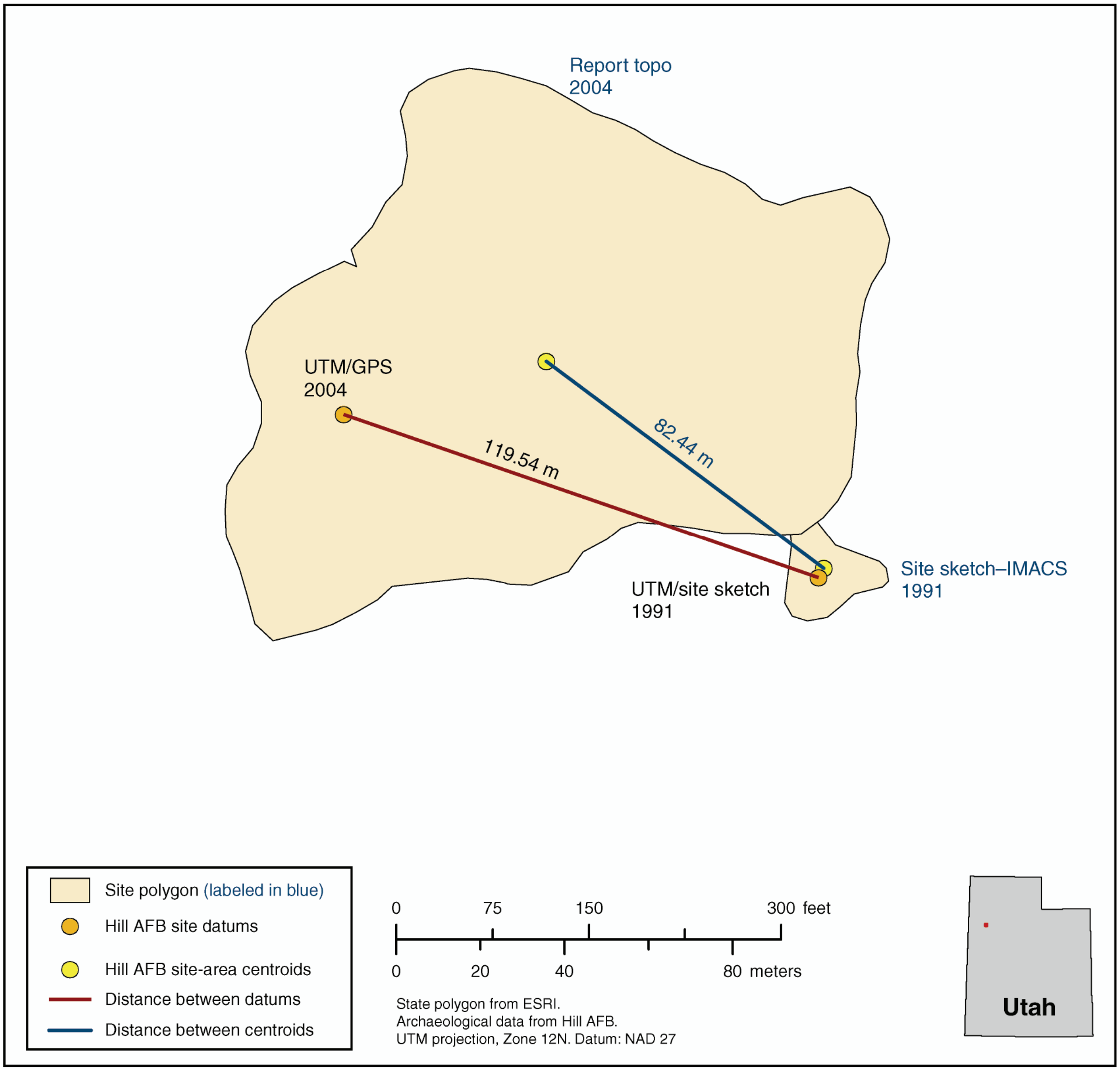

Figure 3.13. Example of site-area change over time: site 42BO0674.
This site has two site-datum points associated with two site-extent polygons.

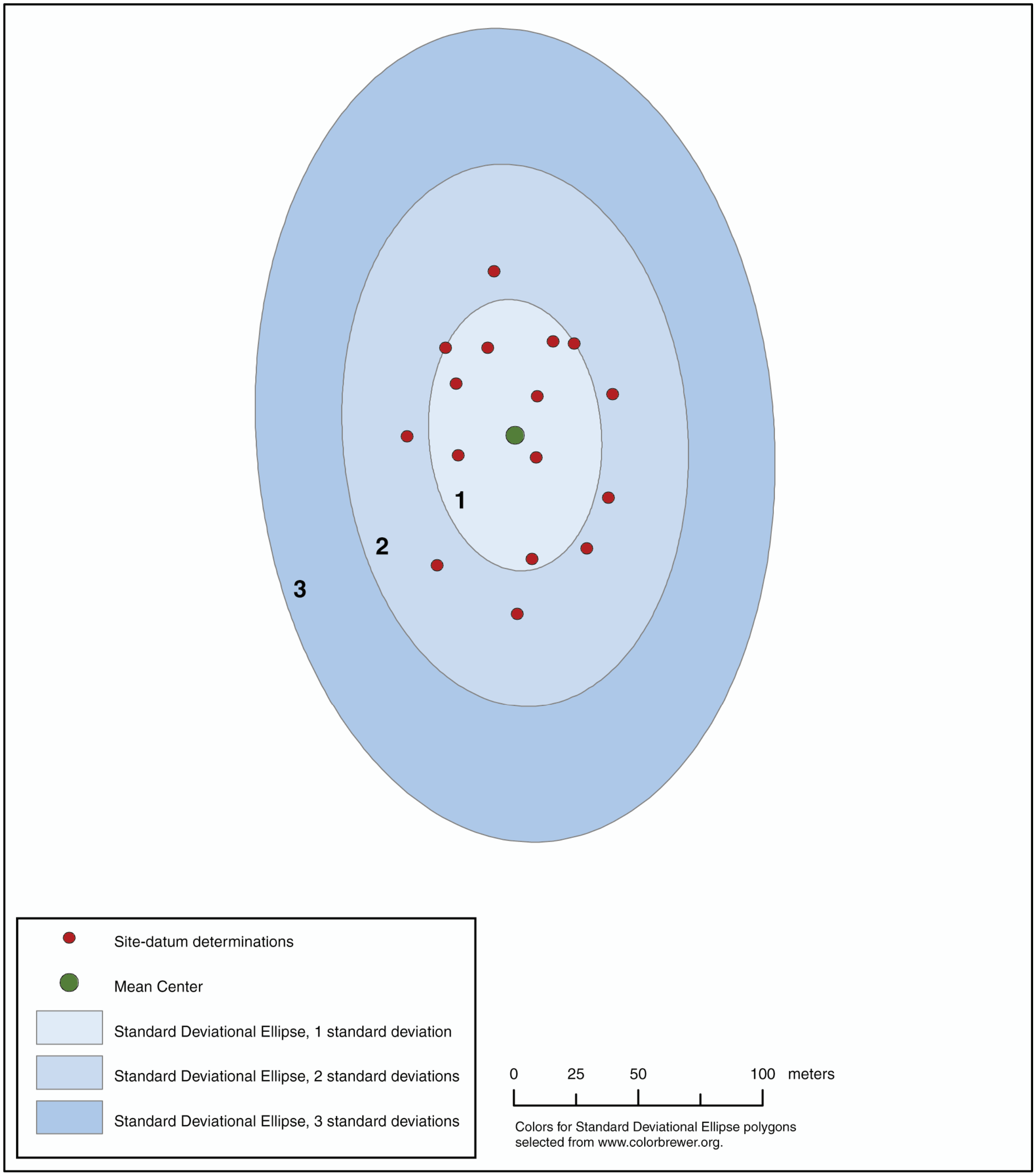

Figure 3.14. Hypothetical site-datum points (including outliers), determined over time by GPS, showing positions of Mean Center and Standard Deviational Ellipses for 1, 2, and 3 standard deviations.

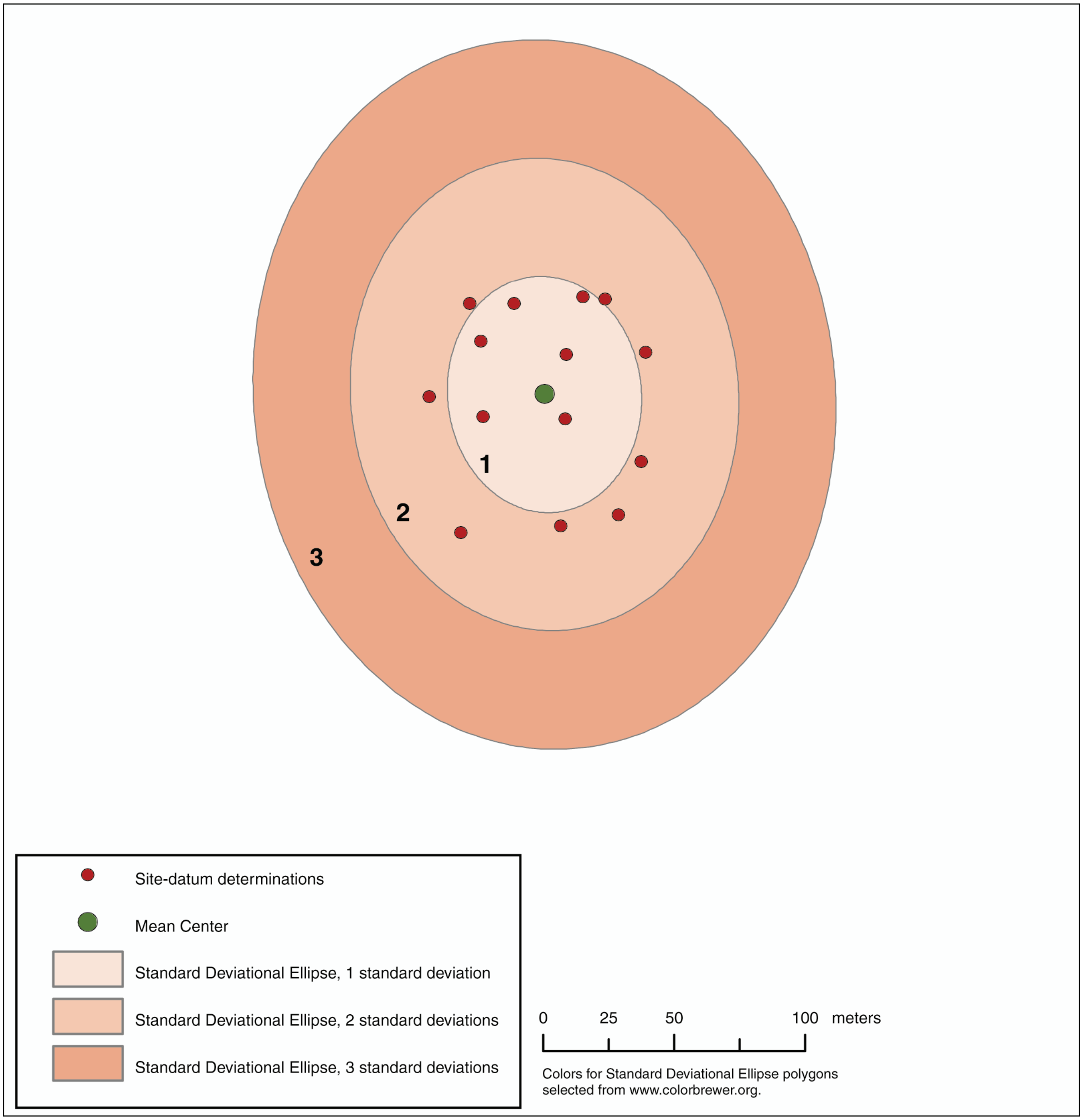

Figure 3.15. Hypothetical site-datum points (excluding outliers), determined over time by GPS, showing positions of Mean Center and Standard Deviational Ellipses for 1, 2, and 3 standard deviations.

to define the axes of the ellipse. The ellipse allows the observer to see if the distribution of features is elongated and hence has a particular orientation. If the underlying pattern in the data is normally distributed about the mean, the polygon for the 1-standard-deviation ellipse will cover approximately 68 percent of the features in the cluster. Two standard deviations will contain approximately 95 percent and 3 standard deviations approximately 99 percent of the features in the cluster. Calculations are based on Euclidean distance for the type of point data discussed here and require projected data to accurately measure distances.

Another spatial statistic, complementing the Standard Deviational Ellipse, is the Standard Distance (ESRI 2008e, 2008f; Figure 3.16). The Standard Distance measures the degree to which features are concentrated or dispersed around their geometric mean or median center (compactness). The Standard Distance is a useful statistic, as it provides a single summary measure of a point distribution around their center (similar to how a standard deviation measures the distribution of data values around the statistical mean). It provides a single value; therefore, compactness can be represented on a map by drawing a circle with the radius equal to the value. As with the Standard Deviational Ellipse, the Standard Distance is sensitive to low or high outliers in any direction in a data set. Therefore, in Figure 3.16, the effect of the lowest and highest points (measured on the north–south axis) is to increase the radii of the Standard Distance circles. Figure 3.17 illustrates how the circles of the Standard Distance Ellipses become smaller when the northernmost and southernmost points are removed from the data set. Again, if the underlying pattern in the data is normally distributed about the mean, the Standard Distance polygon will cover approximately 68 percent of the features in the cluster for 1 standard deviation, approximately 95 percent for 2 standard deviations, and approximately 99 percent for 3 standard deviations. Calculations are based on Euclidean distance, which requires projected data to accurately measure distances.

Summary and Recommendation

Archaeologists often find the relocation of previously recorded sites difficult, particularly if sites are relatively nondescript in their attributes, or if the surface expression of a site has changed through time. Archaeologists sometimes determine that observed boundaries of a relocated site differ from those previously recorded. A common response to such situations is to rerecord the site and then to update site records with new boundary information. Unfortunately, when such updates occur, obsolete data are rarely maintained, making it difficult or impossible to reconstruct the history of site location and thus, to determine what factors probably contributed to changes in recorded site location through time. The lack of historical information on site location also prevents the determination of which, if any, site polygon is most accurate.

We cannot directly measure the accuracy of UTTR site-datum locations because we do not have precise information about the kinds of recording instruments and recording techniques used in the field to obtain them. These data may be found in individual site or survey reports, or in the institutional memories of installation or CRM company personnel. If we did have knowledge of the instruments and techniques used, we could gather or construct estimates of their accuracy, apply them to the measurements (see the work of Wandsnider and Dore [1995] for an example using GPS measurements), and then perhaps buffer the points and polygons in GIS according to their error values to provide a visual sense of the overall degree of accuracy or inaccuracy of these types of archaeological data. Nevertheless, we do have data from the UTTR that reflect a general dichotomy of accuracy, such as "lower" (for example, "annotating on 24K topographic maps") and "higher" (for example, "GPS measurement"), together with site-data records that document two or more revisits to sites to update site-datum and/or site-area measurements.

By analyzing the information from site revisits at the UTTR, we have been able to come to the following conclusions:

- The use of lower-accuracy, topographic map site-datum and site-area recording methods, once the mainstay of archaeological survey recording in the second half of the twentieth century, began to drop off

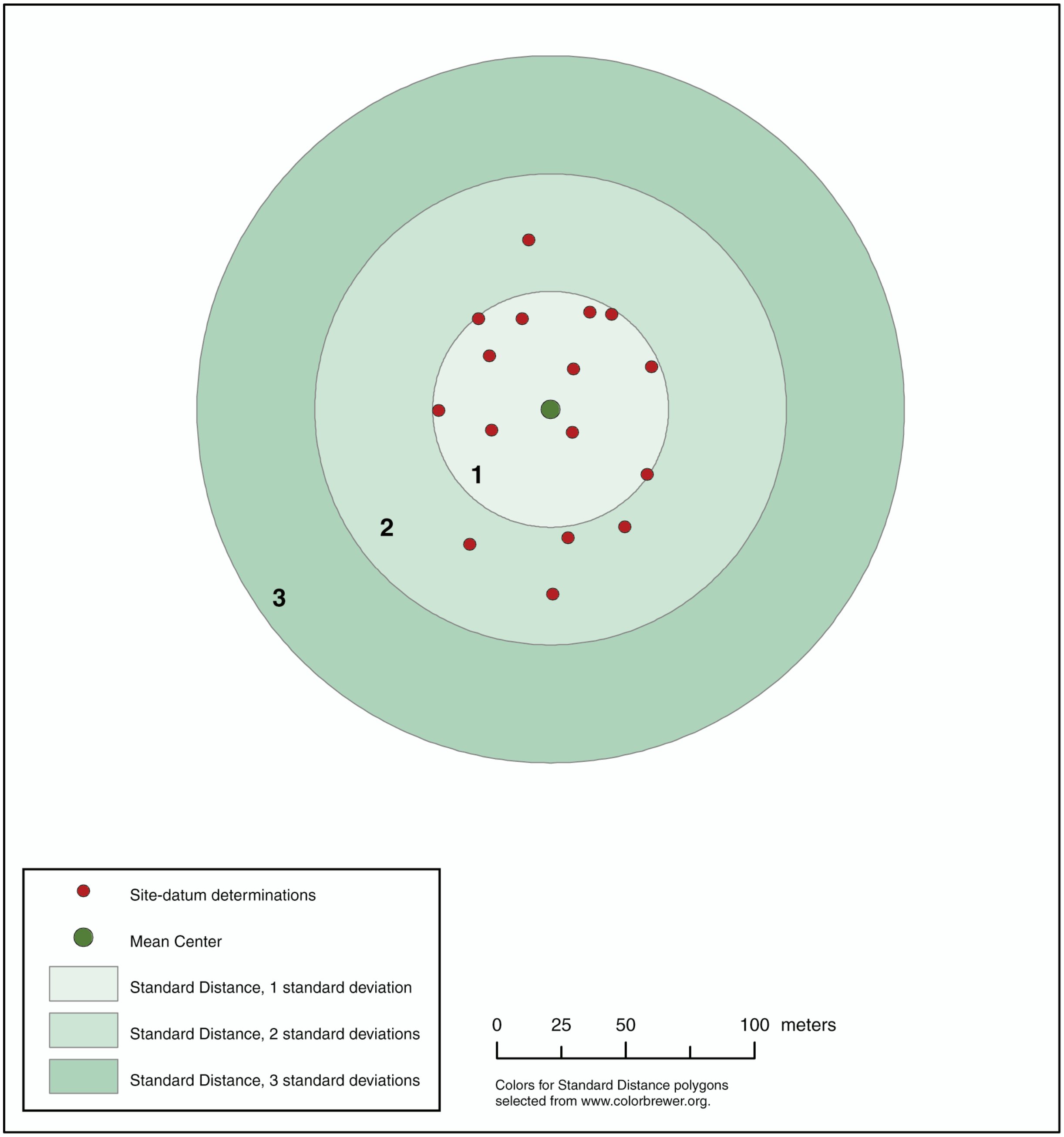

Figure 3.16. The Standard Distance statistic (a circular polygon) for the hypothetical site-datum points (including outliers), describing the compactness of the distribution. Polygons for 1, 2, and 3 standard deviations are shown.

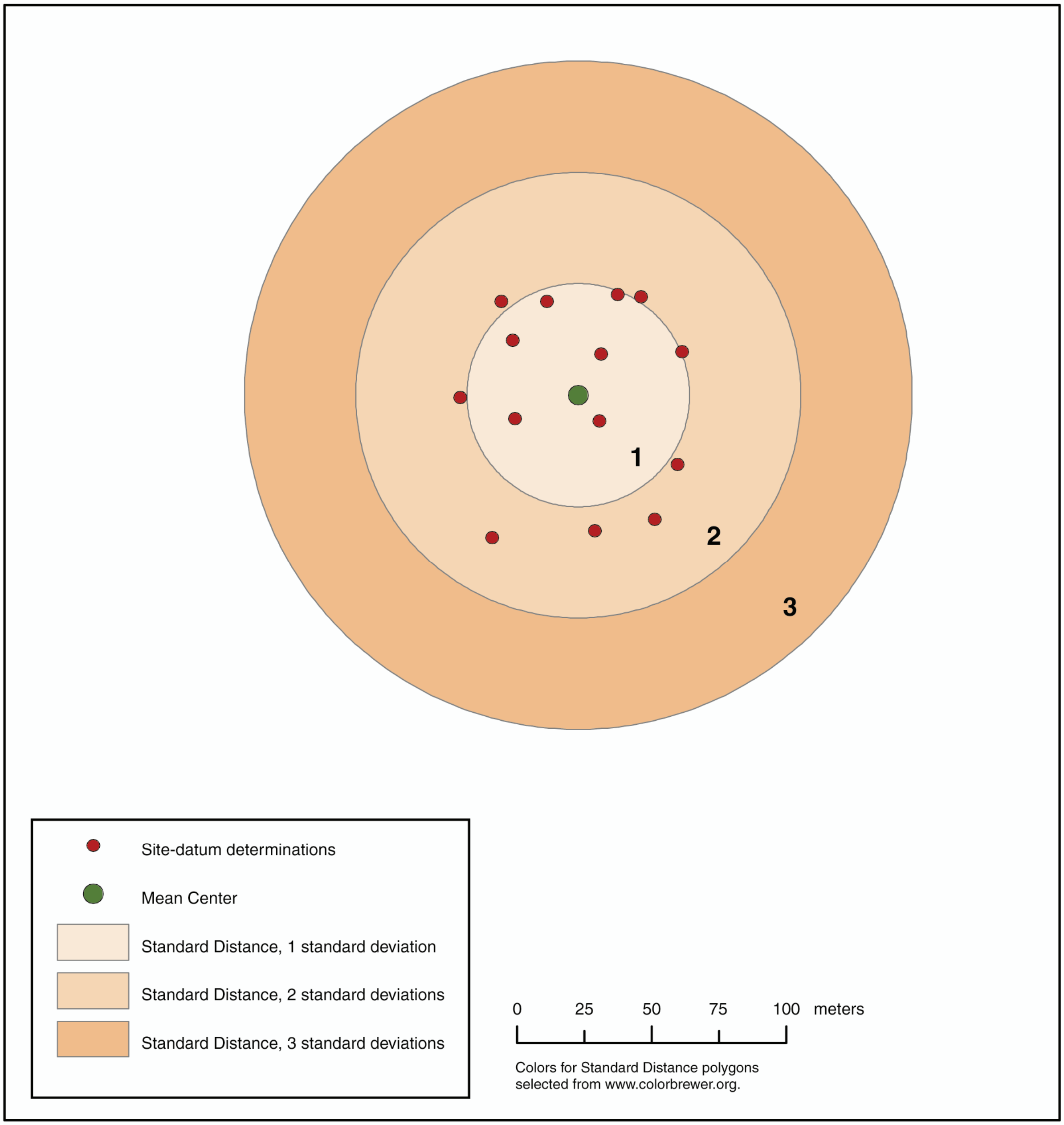

Figure 3.17. The Standard Distance statistic (a circular polygon) for the hypothetical site-datum points (excluding outliers), describing the compactness of the distribution. Polygons for 1, 2, and 3 standard deviations are shown.

quickly at the UTTR after 1996, as higher-accuracy, GPS-based measurement supplanted them, beginning certainly in 1997, and possibly earlier.

- The overall mean of the distances for all site-datum location changes is 80.50 m. This value may indicate the improvement in accuracy from the set of earlier site datums to the set of later site datums that replaced them.
- The overall mean for all site-area-location changes, as measured by the distance between centroids of site-area polygons, is 80.37 m. This value may reflect the improvement in accuracy from the set of earlier site-area polygons to the set of later site-area polygons that replaced them.
- The overall means of site-datum distances and the overall mean of centroid distances for site-area polygons are very similar. The convergence in the overall means for datums and centroids seems to indicate a close correlation between the two measures. Datums were used to measure boundaries, which were made into polygons, which we then used in GIS to calculate centroids. If the two measures are indeed correlated, then site centroids could be used as an alternative point estimate of site location in cases where site boundaries were measured from site datums. In cases where site boundaries were recorded independently using a GPS unit, site centroids could be considered an independent measure of site location, because boundaries were not measured from a site datum. In either case, site datums and centroids could be used to cross-check each other as point estimates of site location.

The quality of site-location data on the UTTR has increased with the advent of GPS technology. Sources of error in site location still exist, but these have less to do with mapping errors than with observational error. For the most part, archaeologists appear to record sites in similar ways. Yet there are problematic situations that managers should watch for. Sites whose sizes increase or decrease tremendously should be viewed cautiously, for they are the best evidence of differences in site recording among archaeologists. Also, unless the installation maintains strong control of records and insists that changes in site location and size are documented, the utility of these data will be offset by the investment needed to check discrepancies in site recording. One might argue that the last point suggests that installations would do well to delete historical records. We disagree. These records, particularly the convergence of site location and size between events, are the best evidence that an installation can use to demonstrate that their records are accurate. Additionally, the data collected from each recording event can be combined to create precise estimates of site-datum or centroid locations. Keeping of historical records should be encouraged, not dismissed.

The analysis of site location leads to two observations that affect DoD in general, as well as land-management decisions and inventory requirements specifically:

- Technological advances, particularly the advent of GPS, have greatly increased the accuracy of plotting site locations. Many installations require the use of GPS in inventory efforts. These procedures should be extended throughout DoD. Installations should require, at a minimum, that their archaeologists record manufacturing information on the GPS unit used to take each measurement and that they use a standard datum (i.e., NAD 83 or World Geodetic System [WGS] 84) for all GPS measurements taken in the field. Installations managers should also require that electronic copies of all site-location data be provided in GIS in a specific format, using a standard datum, units, and projection. These requirements should be specified programmatically. Installations may also wish to require that all GPS data are differentially corrected using specific base stations and that the electronic files recorded in the field with the GPS unit be maintained for the purposes of differential correction and maintaining historical records.
- Site locations plotted without GPS should be considered suspect. Installations would be well advised to maintain relatively large buffers around site locations during land-use planning and military activities. GPS locations plotted before May 2000 should be considered of higher quality but still subject to various degrees of error.

Finally, we can assume that improvements to GPS and other similar technologies will continue. In light of this expectation, we offer two suggestions concerning archaeological and GIS data management from the analysis of site-location data quality:

- Because of the value of historical site-locational data for studies of DoD archaeological data quality, we suggest that DoD installations develop ways for routinely incorporating storage of dated changes to site records in their archaeological and GIS databases.
- We also suggest that DoD installations take “snapshots,” or system saves, of the archaeological data management and GIS systems at regular time intervals, or before major data changes or database restructuring occurs, to archive their information. The existence of system archives may make investigation of presently unknown or unforeseen research or management topics possible in the future.

Note:

It is possible to differentially correct historical GPS data using base-station data from the Continuously Operating Reference Stations (CORS) network of GPS base stations. CORS continuously take measurements of geographic position, time, and satellite information from a fixed position. GPS users can use CORS data along with files of similar data generated by their GPS unit to correct GPS measurements taken in the field. CORS data are available for select locations at given dates as early as 1994, but the original files generated by the GPS unit are required in order to correct the original data. In order to differentially correct historical GPS data using CORS data, cultural resource programs would also need to know what particular GPS measurements correspond to, once corrected, given that GPS IDs generated in the field will not necessarily correspond closely to official site numbers assigned after fieldwork is complete. Because archived GPS files will probably be rare or nonexistent for many installations and because their correspondence to recorded site locations could be very difficult to understand, we feel that differential correction using CORS data probably is possible only under limited circumstances.

Site-Boundary Definition

The third data quality issue addressed in this report is site boundaries. Installation land managers often express frustration with the changing nature of CRM data. In Chapter 2, we showed how and why the numbers of sites change in inventoried areas. Next, we looked at the changing location of seemingly stationary sites. In this chapter, we explore how inanimate sites appear to grow or shrink in size.

We originally proposed to examine the subject of changing site boundaries with a series of before-and-after case studies. We wanted to compare site characteristics, including function and time of occupation recorded when the site was discovered, with those same characteristics after the site had been fully exposed and data recovery had been performed. Spectacular examples of sites changing from artifact scatters to complex villages are rampant in U.S. CRM, particularly in the transportation field.

After examining the installations chosen for this project, we found the data better suited to evaluating changes in site boundaries. The definition of site boundaries is no trivial issue. Archaeologists have debated the concept of the site, including how to define sites and how to delimit them in space (Banning 2002; Binford 1992; Dewar and McBride 1992; Dunnell 1992; Dunnell and Dancey 1983; Ebert 1992; Lucas 2001; Rossignol 1992). Beyond academic interest, definition of site boundaries is critical to management and planning on military installations. Put succinctly, site boundaries distinguish areas that will be managed for archaeological resources from areas free of such constraints. The site concept is central to Section 106 compliance, which is the driving force behind most DoD CRM programs.

We feel that several factors that contribute to variation in site-boundary definition. Although instrument error as discussed in Chapter 3 can affect boundary definition, we feel that the major contributing factors to variation in site boundaries are (1) methodological variation among recording organizations, projects, or phases of work; and (2) cultural or environmental processes that cause variation in the surface or near-surface exposure of sites.

In this chapter, we illustrate how site boundaries change between recording episodes, discuss the factors that contribute to change in site boundaries, and present methods for evaluating change in site boundaries. To do so, we evaluate case studies from Fort Bliss, the UTTR, Eglin AFB, and Fort Drum.

Methods for Evaluating Site-Boundary Data

Archaeologists often revise site boundaries on the basis on field checks. The common protocol for installations is to maintain the more recent boundary polygon, under the assumption that it is the most accurate and up-to-date representation of the site boundary. Although this approach is understandable, there are compelling reasons to maintain data on previously recorded site boundaries that are considered obsolete. Where a site surface is differentially exposed through time, for example, the union of multiple boundaries may represent a site boundary more accurately than the latest polygon. In addition, previous site-boundary data can be used to assess variation through time in site-boundary definition. These data can in turn be used to assess the degree to which site boundaries change and the risk that site boundaries will change with future site recordings.

When site boundaries change, site size probably changes as well. If a boundary does not change in size or shape, but its location does, then the problem may actually be one of site location (Chapter 3), rather than of site-boundary definition. We suggest a few simple metrics that can be used to assess change in site boundaries, all of which are based on the proxy variable of site size:

- Absolute change in site area
- Percent change in site area
- Ordinal change in site area
- Net change in site area
- Relative frequency of site-area change

We calculate the above variables using the size of a site before and after site-boundary revision. Absolute change in site area (Δa) is simply the previous site area (a_p) subtracted from the most recent site area (a_r). Absolute change in site area can be positive or negative, depending on whether a site has increased or decreased in size. Percent change in site area is $100(\Delta a/a_p)$.

To demonstrate ordinal change in site area (Δo), we define arbitrary size classes in the following section. Calculating Δo involves placing sites into arbitrary ordinal size classes, labeled 1 through *n*. For instance, ordinal size class 1 could be all sites between 0 and 500 m^2; size class 2 all sites between 500 and 1,000 m^2; size class 3 all sites between 1,000 and 2,000 m^2, and so on. The important point about ordinal size classes is that they should correspond to how installation archaeologists define site types. In this sense, ordinal change in size classes potentially indicates a change in site type, in addition to site size. Like Δa, Δo is calculated by subtracting the previous ordinal site-area value (o_p) from the most recent ordinal site-area value (o_r). Below, we define arbitrary ordinal size classes for sites on Fort Bliss and the UTTR to demonstrate how change in ordinal size can be calculated and evaluated. The ordinal size classes we assign are arbitrary. They are intended for illustrative purposes only and have no necessary correspondence to how sites are defined on either installation. Fort Bliss, for instance, does not assign site types according to ordinal size classes.

Net change in site area refers to the absolute and percent change in site area for a group of sites evaluated according to the supplied metrics. For instance, if each of five sites was 100 m^2 in size, two were revised down by 50 m^2, and three were revised up by 50 m^2, the net Δa for the group of sites would be +50 m^2. The net percent Δa would be $50 \text{ m}^2 \div (5 \times 100 \text{ m}^2)$, or +10 percent.

The last metric is relative frequency of site-area change. Relative frequency of site-area change can be calculated for a sample of sites for which site-boundary revision is evaluated. When representative of particular installation contexts, the measure can be considered to indicate the risk that a site boundary will change if rerecorded. Relative frequency of site-area change is calculated by dividing the number of site-area changes by the total number of cases, and by multiplying that number by 100. Relative frequency of site-area change can be calculated for multiple variables, including Δa and Δo, and can also be calculated separately for positive and negative changes, should the directionality of change be a concern to installation managers.

In the following sections, we apply the above metrics where possible to instances of site-boundary change on four installations: Fort Bliss, the UTTR, Eglin AFB, and Fort Drum. Fort Bliss is presented first, as the available data allow us to apply all the metrics discussed above. Not all the metrics can be applied without qualification to the other three installations.

Fort Bliss

Fort Bliss has a long history of sponsoring archaeological research, largely to comply with federal mandates. The installation covers a range of environments and has a rich and complex culture history. Thousands of sites have been recorded by surveys covering tens of thousands of acres, performed by a multitude of contractors. CRM managers must feel comfortable with the quality of archaeological data to make critical decisions affecting resources and military activities. Not surprisingly, Fort Bliss has been a leader in developing standards and practices to ensure the collection of comparable data.

CRM managers have long recognized that change in site boundaries as a result of resurvey is a recurrent problem on some areas of Fort Bliss where sites are found on active surfaces in dune environments (see Appendix A:Figures A.1 and A.2). In active dune environments, ongoing environmental processes can cause sites to change in apparent size, shape, or composition. In other areas of the installation, where stabilized land surfaces are more common, the effect of environmental processes on site boundaries or other characteristics may be less frequent or dramatic. In addition to the effects of environmental conditions, Fort Bliss CRM personnel and archaeological contractors are fully aware that change through time in how sites are recorded or defined has also led to change in site boundaries. Leckman et al. (2008a) argued, for instance, that previous definition rules led to more restrictive definitions of site area, resulting in many small sites separated by dunes and other landforms that bury cultural materials. More recent definitions are more inclusive and tend to include areas that may have previously been recorded as multiple small sites and intervening areas, where cultural materials are likely to be buried. How sites were discovered in the past probably has also had a substantial effect on site-boundary definitions. Previous definitions were dependent in part on the discovery of sites using fairly wide and variable transect intervals (e.g., 40–60 m), whereas current survey methods require 15-m transect intervals as the standard. Using earlier methods for discovering sites, substantial portions of a site could have been missed and thus not included in boundary definitions.

In this section, we examine change in site boundaries for two recently completed testing and evaluation projects conducted on Fort Bliss (Leckman et al. 2008a, 2008b). For both projects, SRI relocated previously recorded sites to document the current distribution of artifacts and features, to test features and subsurface deposits, and to provide recommendations with respect to eligibility for listing in the National Register of Historic Places (NRHP). For the 31-site project, 15 of the sites were originally identified in 1976 and 1977 by the Western Hueco Bolson Survey (Whalen 1978); 12 of these had not been revisited and documented since originally recorded. Five sites were identified during survey and testing work conducted by the University of Texas at El Paso (UTEP) in 1986 in preparation for the construction of the Loop 375 expressway (O'Laughlin 1987). The remaining 11 sites were recorded in 1990 as part of a Fort Bliss Directorate of Public Works–Environmental Division laboratory project (Project 90-05). The 34-site project tested and evaluated sites originally recorded during a survey of Maneuver Areas 3–8 conducted by UTEP (Carmichael 1986). Thirty-two of the sites were revisited by Lone Mountain Archaeological Services in 1997 and 1998 (Sale 1999), and 2 sites were tested by Geo-Marine, Inc., in 1994 (Mbutu 1997).

For both projects, SRI recorded boundaries of relocated sites using a Trimble GeoXH GPS unit capable of sub-foot to sub-meter accuracy (approximately 10–50 cm) when data are differentially corrected. Subsurface tests consisted of unscreened trowel tests, shovel tests, and 1-by-1-m test units. Subsurface tests were used to identify and characterize subsurface deposits, to evaluate features, and to describe the geomorphic context of each site. The limited number and extent of subsurface tests were insufficient to evaluate the subsurface extent of the site. Instead, site boundaries were recorded based primarily on the distribution of surface artifacts and features. Site boundaries for the 34-site project were recorded according to the same methods used in the 31-site project, but Leckman et al. (2008b:19) noted for the 34-site project that

> observed cultural material in the vicinity of a site was generally considered part of the site as long as it was not farther than 30 m from the rest of the site's surface assemblage. In some cases, however, previously defined sites were so close to each other that their redefined boundaries fell within this 30-m threshold. In these instances, sites were treated as separate entities unless their surface assemblages were clearly continuous.

Causes of Change in Site Area

Methodological factors played a major role in site-boundary change for the two projects. The relocated sites were initially recorded many years earlier, and all were recorded on the basis of methods different from those used currently. One major factor that Leckman et al. (2008a) identified as contributing to variation in site boundaries is change over time in criteria for defining sites. The current programmatic agreement under which Fort Bliss complies with Section 106 specifies that

> 30 m is the maximum distance within which cultural manifestations should be considered part of the same site. This relatively large distance potentially allows material exposed across several small blowouts to be incorporated into a single site, creating a unit of analysis that is more resistant to aeolian change and is also, presumably, a more robust representation of related prehistoric cultural activity. Many of the sites within the present project area, however, were defined using more-restrictive criteria that frequently limited a "site" to cultural material exposed within a single blowout. As a result, many sites lie within 30 m or less of each other, increasing the difficulty of meaningfully distinguishing them from one another [Leckman et al. 2008a:108].

Environmental factors also played a role in site-boundary change for the two projects. Most of the sites relocated and rerecorded by SRI for these two projects were originally identified in deflated areas in dune environments. In such environments, aeolian activity differentially buries and exposes archaeological sites. Leckman et al. (2008a:102) noted that

> aeolian environments pose numerous challenges for field personnel trying to relocate and identify specific site boundaries. On many occasions, increased deflation enlarges exposed blowout areas, exposing new material and enlarging the areal distribution of surface artifacts and features (e.g., Lukowski et al. 2006; Vierra et al. 1999). But aeolian processes can also bury previously visible archaeological manifestations, reduce the apparent area of surface sites, and concentrate dispersed archaeological materials (Abbott et al. 1996:91–95).

Thus, aeolian activity can expand or diminish site boundaries in multiple ways that are not easily predicted.

For the 31-site project, only around one-third of previously identified features mapped within sites could be relocated. Sites for the 34-site project also proved difficult to relocate; relocation of those sites included careful examination of aerial photos and field maps, and intensive survey within a 100-m radius of the site's plotted location. Although problems with mapping accuracy probably contributed to the difficulty in relocating sites and features, Leckman et al. (2008a) attribute some of the difficulty to aeolian erosion and dune migration.

Absolute, Percent, and Net Change in Site Area

All sites relocated for the 34-site project changed in size as a result of being rerecorded. One site (FB 3996) was removed from the analysis because of uncertainty about whether the site had been relocated. Of the 27 relocated sites, 11 became smaller, and 16 increased in size. Sites that decreased in size typically decreased by about 30–50 percent. Six of the 17 sites that increased in size increased by 400 percent or more. One became almost 45 times larger. Altogether, change in site size represents a net increase of 13.2 percent in site area (Table 4.1).

For the 31-site project, all relocated sites whose previous size was known changed in size. More than half of the sites decreased in size, ranging from 17 to 88 percent. The remainder increased in size. Although

Table 4.1. Change in Site Size for the 34-Site Project on Fort Bliss (Leckman et al. 2008b)

FB No.	Relocated?	Original Size Category	Original Site Size (m^2)	Revised Size Category	Revised Site Size (m^2)	Change in Site Size (m^2)	Change in Site Size (%)	Ordinal Change in Site Size
3527	yes	5	6,750	5	6,745	–5	–0.1	0
3988	yes	4	2,184	4	3,006	822	37.6	0
3990	yes	4	4,700	4	3,167	–1,533	–32.6	0
3991	no	1	100	—	—	—	—	—
3993	yes	4	4,550	4	1,851	–2,699	–59.3	0
3994	yes	6	14,574	7	15,072	498	3.4	1
3995	yes	4	1,332	5	7,637	6,305	473.3	1
3997	yes	5	7,081	4	3,022	–4,059	–57.3	–1
3998	yes	1	100	2	165	65	65.0	1
3999	yes	3	1,000	4	1,799	799	79.9	1
4000	yes	4	2,592	4	4,659	2,067	79.7	0
4002	yes	1	100	3	786	686	686.0	2
4003	yes	4	3,432	5	6,457	3,025	88.1	1
4005	yes	4	1,550	6	12,054	10,504	677.7	2
4012	no	1	100	—	—	—	—	—
4017	yes	3	532	3	874	342	64.3	0
4018	yes	1	100	3	927	827	827.0	2
4019	no	1	100	—	—	—	—	—
4022	yes	7	17,640	5	7,373	–10,267	–58.2	–2
4023	yes	2	330	2	158	–172	–52.1	0
4026	yes	4	2,139	4	1,448	–691	–32.3	0
4027	yes	2	360	3	849	489	135.8	1
4150	yes	1	100	4	1,403	1,303	1303.0	3
4151	yes	3	768	4	1,747	979	127.5	1
4152	yes	1	100	4	4,588	4,488	4488.0	3
4154	no	1	100	—	—	—	—	—
4155	yes	4	3,869	4	1,648	–2,221	–57.4	0
4157	yes	1	100	1	23	–77	–77.0	0
4158	yes	4	1,900	4	1,032	–868	–45.7	0
4161	no	4	1,170	—	—	—	—	—
4163	no	1	100	—	—	—	—	—
4164	yes	1	100	2	126	26	26.0	1
4595	yes	2	195	1	80	–115	–59.0	–1
Total	27	—	79,848	—	88,696	10,518	13.2	16

sites evaluated for this project more often decreased in size, sites that increased in size typically increased by 100 percent or more. The result is a net increase of 48 percent in site area (Table 4.2).

Ordinal Change in Site Area

Archaeologists often group sites into particular size classes in order to evaluate them, under the assumption that different size classes correspond to different levels of cultural phenomena. In order to track this kind of ordinal change in site size, we grouped each site into a size class based on the following size ranges: (1) 0–100 m^2, (2) 100–500 m^2, (3) 500–1,000 m^2, (4) 1,000–5,000 m^2, (5) 5,000–10,000 m^2, (6) 10,000–15,000 m^2, and (7) 15,000–20,000 m^2. The results of this analysis show that for the 34-site project, ordinal size class changed for 16 of 27 sites for which we have before and after data (see Table 4.1). In other words, sites changed to an altogether different size class for around 60 percent of sites. A similar result was obtained for the 31-site project (see Table 4.2). Of 23 sites for which we had before and after data, 11 (48 percent) changed in ordinal size class. The two projects differed in the magnitude of the change. For the 31-site project, sites typically went up or down one size class, if they changed in size. Only 1 site increased by two size classes. By contrast, 6 out of 28 sites for the 34-site project increased by two or three size classes. One site decreased by two size classes.

Relative Frequency of Change in Site Area

Ultimately, the results of both projects suggest that we can expect nearly a 100 percent chance that site boundaries will change as a result of rerecording and a 40–60 percent chance that sites will change in size classes as a result of site-boundary changes for similar kinds of projects and environmental settings. For installations that organize sites into size classes for evaluation or modeling purposes, sites could be classed incorrectly for a large percentage of cases. Installations facing this problem might need to use broader size classes or to calculate the chance that sites occupy smaller or larger size classes than recorded. Fort Bliss, for example, does not define sites or site types on the basis of size. Instead, site classes are defined by sets of artifacts and features inferred to correspond to behaviorally meaningful analytical units.

Net change in site area also increased for both projects (site areas increased more than they decreased overall). Ultimately, net change in site area may be the result of revisions in rules for site-boundary definition, which tend toward defining larger sites.

Fort Bliss's Approach to Problems in Site-Boundary Definition

As noted above, there is a wide variety of methodological factors and environmental and cultural processes that can affect change in site boundaries. As archaeologists, we cannot control for all the factors that affect variation in site boundaries among recording episodes. Using before-and-after data, however, we can estimate the frequency and magnitude of site-boundary changes for particular circumstances and can use these estimates to predict change for similar circumstances. We can also factor the potential for change in site boundaries into planning efforts.

Fort Bliss has addressed the problem of site-boundary definition by developing the Transect Recording Unit (TRU) survey system. The TRU survey system is designed to maintain consistency in survey and site-recording methods and to provide a more objective approach toward site definition. If site-definitional criteria change, the survey method also allows for sites to be redefined using the same original survey data. According to the Fort Bliss Integrated Cultural Resources Management Plan 2008–2012 (U.S. Army 2008):

Table 4.2. Change in Site Size for the 31-Site Project on Fort Bliss (Leckman et al. 2008a)

FB No.	Relocated?	Original Size Category	Original Site Size (m^2)	Revised Size Category	Revised Site Size (m^2)	Change in Site Size (m^2)	Change in Site Size (%)	Ordinal Change in Site Size
7635	yes	2	400	3	791	391	97.8	1
7636	yes	3	700	3	582	–118	–16.9	0
7640	yes	2	300	2	223	–77	–25.7	0
7641	yes	2	400	2	255	–145	–36.3	0
7642	yes	3	600	2	368	–232	–38.7	–1
7644	yes	2	400	2	424	24	6.0	0
7645	yes	2	300	2	250	–50	–16.7	0
7646	yes	2	400	2	294	–106	–26.5	0
7647	yes	2	400	2	252	–148	–37.0	0
7659	yes	3	600	4	3,161	2,561	426.8	1
7660	yes	2	300	3	654	354	118.0	1
7661	yes	3	700	2	450	–250	–35.7	–1
7662	mechanically excavated prior to evaluation	4	1,400	—	—	—	—	—
7667	yes	2	300	3	686	386	128.7	1
7677	yes	3	800	4	2,199	1,399	174.9	1
9749	no	—	—	—	—	—	—	—
9752	yes	—	unknown	2	211	—	—	—
9754	yes	—	unknown	4	3,552	—	—	—
9839	yes	—	unknown	1	31	—	—	—
9842	yes	—	unknown	2	198	—	—	—
11221	yes	1	< 100	1	40	–60	–60.0	0
11230	yes	1	< 100	2	378	278	278.0	1
11231	yes	3	900	2	451	–449	–49.9	–1
11232	yes	1	< 100	1	57	–43	–43.0	0
11234	yes	2	200	4	1,284	1,084	542.0	2
11242	yes	1	< 100	1	30	–70	–70.0	0
11243	yes	1	< 100	1	12	–88	–88.0	0
11244	yes	1	< 100	1	16	–84	–84.0	0
11259	no	1	< 100	—	—	—	—	—
11260	yes	1	< 100	2	277	177	177.0	1
12098	no	—	—	—	—	—	—	—
Total	27	—	9,900	—	17,126	4,734	47.8	6

The TRU method uses a grid system configured to line up with the UTMs (NAD 83) in the area for recording materials found on survey. The survey area is divided into 15- by-15-meter cells. All cultural materials are recorded within each cell and an approved threshold is established to organize positive cells into sites based on the current Fort Bliss site criteria. All TRU survey data are collected digitally and locational data are collected using high-accuracy GPS units. Hand-held computers (i.e., PDAs, Pocket PCs, etc.) are used as field data collection units and the surveyors will develop appropriate field data collection forms and software.

Before fieldwork, each crew member is assigned specific 15-by-15-m cells that are their responsibility to survey according to the TRU system. Assigned cell locations are loaded onto handheld computers, which crew members use along with GPS units to identify their location on the grid and to enter field data for each assigned grid cell. In this way, areas requiring survey cannot be as easily missed or overlooked. A further benefit of the system is that sites are defined on the basis of what is recorded in each grid cell. If site definitions change, then site boundaries can be redefined objectively using the same field data, rather than through resurvey or subjective interpretation of less reliable field maps.

It would be very useful to assess change in site boundaries between recording episodes for sites recorded and rerecorded using the same TRU system. In such cases, methodological change—including change in site-definitional criteria—could be held constant. Change in site boundaries could then be interpreted as resulting from other factors, such as change in visibility, erosion, deposition, or disturbance.

Other installations may be interested in utilizing the TRU system. Before such a move, however, it would be useful to demonstrate that the method produces comparable and reliable results. Installation archaeologists at Fort Bliss indicate that there are very few cases where a site recorded originally in the TRU system has been rerecorded, because current survey efforts focus on inventory in unsurveyed areas or resurvey in areas where previous survey no longer meets inventory requirements. Clearly, such a pilot study would be well worth the effort.

The Utah Test and Training Range (UTTR)

For the UTTR, we analyzed a series of 52 sites for which boundaries were revised as a result of rerecording efforts (Table 4.3). These data were collected for the purpose of assessing change in site location and were described in Chapter 3. Unfortunately, we were unable to interpret unambiguously the cases where a site was rerecorded in some fashion (e.g., the site datum was rerecorded) but the site boundary was not revised. We were unable to clearly interpret whether, in those cases, a site boundary was not changed because (1) it was considered the same as previously recorded or (2) the issue of boundaries was not addressed by rerecording efforts. As a result, we cannot unequivocally estimate the relative frequency of site-boundary change or assess the risk that the boundary of a rerecorded site at the UTTR will vary from a previously recorded boundary. Nonetheless, given what we know about the situation at the UTTR, the chance that a rerecorded site will require a boundary revision is probably near 100 percent.

We evaluated how often site-boundary revisions resulted in smaller or larger sites and estimated the relative change in site area as a result of boundary revisions. For some boundary revisions, site size increased or decreased by only a small amount, suggesting that site-boundary revision occasionally had only a minor effect on site size. However, substantial revision of site boundaries was the norm when boundaries were revised. Moreover, boundary revision predominantly resulted in larger sites. An increase of 50 percent or more in site area occurred for half of all boundary revisions; several other revisions had a markedly negative effect on site size. Site-boundary revision resulted in an increase in site area for 80 percent of cases.

Because the range of site sizes was large for the UTTR in comparison to the Fort Bliss data, we used a slightly different scheme for assigning ordinal values to site sizes: (1) 0–100 m^2, (2) 100–500 m^2, (3) 500–1,000 m^2, (4) 1,000–5,000 m^2, (5) 5,000–10,000 m^2, (6) 10,000–50,000 m^2, (7) 50,000–100,000 m^2, and (8)

Table 4.3. Change in Site Size for UTTR Sites

Site ID	Initial Recording Year	Original Size Category	Original Site Size (m^2)	Revision Year	Revised Size Category	Revised Site Size (m^2)	Change in Site Size (m^2)	Change in Site Size (%)	Ordinal Change in Site Size
42BO0677	1991	4	3,528	1997	4	2,850	-678	-19	0
42BO0684	1991	6	14,634	1997	5	6,374	-8,260	-56	-1
42BO0692	1991	4	1,249	1997	4	1,147	-102	-8	0
42BO0810	1991	5	8,739	1997	6	12,190	3,451	39	1
42BO0811	1991	4	4,336	1997	6	12,994	8,658	200	2
42TO0137	1991	4	4,509	2005	6	11,144	6,635	147	2
42TO0173	1991	6	18,127	1997	6	32,233	14,106	78	0
42TO0659	1991	3	825	2004	2	449	-376	-46	-1
42TO0665	1991	2	181	2004	5	8,282	8,101	4,476	3
42TO0667	1991	3	593	2004	3	925	332	56	0
42TO0668/0669/0670	1991	3	786	2004	7	56,460	55,674	7,083	4
42TO0674	1991	2	388	2004	6	14,030	13,642	3,516	4
42TO0675	1991	6	24,885	2004	7	78,972	54,087	217	1
42TO0690	1991	6	18,401	2001	7	57,879	39,478	215	1
42TO0691	1991	4	1,192	1997	4	2,174	982	82	0
42TO0694	1991	5	7,849	1997	4	3,497	-4,352	-55	-1
42TO0695	1991	4	3,808	1997	6	11,018	7,210	189	2
42TO0696	1991	2	489	1997	4	1,585	1,096	224	2
42TO0699	1991	3	696	1997	3	683	-13	-2	0
42TO0700	1991	4	4,321	1997	6	10,659	6,338	147	2
42TO0703	1991	5	8,970	1997	6	18,550	9,580	107	1
42TO0704	1991	6	17,244	1997	6	32,408	15,164	88	0

Site ID	Initial Recording Year	Original Size Category	Original Site Size (m^2)	Revision Year	Revised Size Category	Revised Site Size (m^2)	Change in Site Size (m^2)	Change in Site Size (%)	Ordinal Change in Site Size
42TO0705	1991	4	1,341	1997	5	5,691	4,350	324	1
42TO0706	1991	4	1,292	1997	4	2,419	1,127	87	0
42TO0906	1995	4	1,441	2004	6	10,053	8,612	598	2
42TO0907	1995	3	651	2004	5	8,267	7,616	1,170	2
42TO0908	1995	3	714	2005	5	5,343	4,629	648	2
42TO0909	1995	7	62,621	2005	8	346,719	284,098	454	1
42TO0911	1995	5	7,327	2005	6	25,672	18,345	250	1
42TO0912	1995	3	998	2005	3	677	-321	-32	0
42TO0913	1995	4	1,760	2005	6	13,892	12,132	689	2
42TO0922/0923	1996	6	16,755	2005	8	118,404	101,649	607	2
42TO0924	1996	6	33,176	2005	6	11,338	-21,838	-66	0
42TO0926	1996	6	12,426	2000	6	17,093	4,667	38	0
42TO1008	1998	5	5,543	2003	5	7,805	2,262	41	0
42TO1012	1998	7	60,097	2003	7	78,347	18,250	30	0
42TO1013	1998	7	95,780	2003	7	98,606	2,826	3	0
42TO1016	1998	6	38,956	2003	7	54,468	15,512	40	1
42TO1020	1998	6	20,032	2003	6	21,899	1,867	9	0
42TO1021	1998	6	47,731	2003	7	93,684	45,953	96	1
42TO1025	1998	6	18,045	2003	6	22,216	4,171	23	0
42TO1028	1998	6	11,391	2003	6	12,629	1,238	11	0
42TO1029	1998	7	57,747	2003	7	64,444	6,697	12	0
42TO1030	1998	6	25,550	2003	6	26,923	1,373	5	0

continued on next page

Site ID	Initial Recording Year	Original Size Category	Original Site Size (m^2)	Revision Year	Revised Size Category	Revised Site Size (m^2)	Change in Site Size (m^2)	Change in Site Size (%)	Ordinal Change in Site Size
42TO1037	1998	5	5,834	2003	4	4,772	-1,062	-18	-1
42TO1042	1998	6	38,316	2003	6	48,576	10,260	27	0
42TO1046	1998	6	49,002	2003	7	63,676	14,674	30	1
42TO1053	1998	6	27,089	2003	6	33,896	6,807	25	0
42TO1054	1998	5	9,096	2003	6	12,664	3,568	39	1
42TO2712	2005	6	46,198	2006	6	49,327	3,129	7	0
42TO2726	2005	7	57,551	2006	7	54,754	-2,797	-5	0
42TO2733	2005	4	3,872	2006	6	34,973	31,101	803	2
Total	—	—	904,082	—	—	1,725,730	821,648	91	40

Note: Data provided by Hill AFB.

100,000–500,000 m^2. As for Fort Bliss, the arbitrary size classes we used were assigned to illustrate the potential for change in size classes for the benefit of installations that use such criteria in assigning site types. The size classes assigned do not necessarily correspond to any site types used for defining resources on the UTTR. Using this scheme, UTTR site size increased in ordinal value for almost half of boundary revisions and increased by two or more size classes for more than one-quarter of boundary revisions. Ordinal size decreased for only four sites.

Altogether, these data suggest that we can expect most previously recorded sites on the UTTR to increase substantially in size when rerecorded. As at Fort Bliss, the net change in site area as a result of boundary revision was positive, but the magnitude of the increase was substantially greater for the UTTR. Net change in site area was +821,647 m^2 or +91 percent. The larger net increase could result from differences between the two installations in typical site sizes. The UTTR sites were much larger on average (33,187 ± 14,730 m^2) than the Fort Bliss sites used in this study (1,968 ± 813 m^2), for instance. Other factors could have also played a role, such as variation between installations in the characteristic sizes of landforms or differences in definitional criteria. For the Fort Bliss data we analyzed, dune landforms might have constrained the size of sites to a greater degree than at the UTTR, or Fort Bliss site-definition criteria might be more conservative than those at the UTTR. The point here is not to isolate the ultimate causes of change in site-boundary definition but to demonstrate ways to document and estimate the effect. Case studies for both installations suggest that under similar circumstances we can expect site boundaries to change and for site sizes to increase more than they decrease as a result of boundary revision.

Case Studies from Eastern Installations: Eglin AFB and Fort Drum

The site-boundary problem faced by installations where the surface is obscured is much different from the one posed at installations where surface archaeological materials are visible. In areas obscured by vegetation, installations rely on small subsurface probes, usually in the form of shovel tests or small test pits (hereafter referred to as shovel-test pits [STPs], as in Chapter 2), whereas in areas of good surface visibility, pedestrian survey is a reasonable site-detection technique. Although, as we have documented, there is still substantial variation in site-boundary definition as a result of environmental processes or change in site-definition rules, pedestrian surveys have the advantage of actually observing relatively large surface areas. Surveys based on subsurface probes may provide better observations (screened fill of subsurface contexts) but cover exceedingly small fractions of the survey universe. We have already discussed the effect of small subsurface probes on site detection (Chapter 2); here, we shift to their effect on defining site boundaries.

During survey, the interval between STPs is generally too large to define site boundaries. Installations typically address this problem by excavating additional STPs near positive ones. Although samples are more closely spaced than survey units, the sample size for site-boundary-definition efforts is still relatively small. Usually, a significant chance remains that additional STPs excavated within sites will be negative, particularly when artifact densities are low. In some cases, where sites were originally defined as small and sparse scatters, later military actions have required reassessments, which have resulted in substantial changes in the definition of site size and character. We were able to glean a few of these cases from installation personnel to be used for illustrative purposes. We note, however, that archaeologists are perhaps more apt to remember the extraordinary cases when results differed markedly from prior expectations, than to remember less remarkable cases. Also, because we originally requested data for the purpose of evaluating the correspondence between surface and subsurface components, installation archaeologists may have focused on sites that had a recognizable surface component and a subsurface component. We sense there are more cases worthy of examination than those described below, particularly those cases in which little was found beyond the original positive discovery unit.

We present five cases of site-boundary definition from Eglin AFB followed by two cases from Fort Drum (Table 4.4). We stress that these cases should not be considered representative of the full range of

Table 4.4. Change in Site Size for Example Sites on Eglin AFB and Fort Drum

Site No.	Initial Discovery Situation	Original Site Size (m^2)	Revised Site Size (m^2)	Change in Site Size (m^2)	Change in Site Size (%)
FDP 1093/1094	surface	2,000	40,469	38,469	1,923
FDP 1170	surface	8,094	52,609	44,515	550
8OK241	surface	20	6,898	6,878	34,390
8OK427	surface	600	11,390	10,790	1,798
8OK428	surface	400	9,208	8,808	2,202
8OK433	surface	200	1,500	1,300	650
8OK2114	surface	750	750	0	0
Total		12,064	122,824	110,760	918

Note: Fort Drum data (FDP sites) provided by Dr. Laurie Rush, Cultural Resources Manager, Fort Drum, New York; Eglin AFB data (8OK sites) provided by Prentice Thomas & Associates on behalf of Eglin AFB, Florida.

outcomes of efforts to define site boundaries, but merely illustrative of cases in which relatively small, discrete single-component sites are revealed through further investigation to be much larger and more complex than originally inferred. The problem that these cases illustrate is that installations do not necessarily know when or where such cases will emerge. This problem underscores the need to develop methods for identifying archaeologically sensitive areas and determining the risk of discovering significant sites in particular zones.

Eglin AFB

Eglin AFB provided information on five sites that could be used to evaluate the potential for change in site boundaries. Of particular note is that all five sites were discovered as a result of surface finds. Preliminary testing in all cases suggested that the sites were relatively small. On the basis of initial discoveries, several of these sites were recommended not eligible for listing in the NRHP. In four of the five cases, however, subsequent testing revealed the sites to be substantially larger and more complex than originally inferred. Further, the one site that was considered likely to have significant subsurface deposits (8OK2114) had no intact subsurface deposits worthy of further investigation. In that case, no changes were made in the site boundary. In the other four cases, redefined site boundaries resulted in sites that were 6.5–350 times their originally recorded size.

Site 8OK241

Site 8OK241 was first reported in 1982 when a small surface scatter of oyster and conch shell was discovered during the survey of Unit X-32. Investigations were limited to surface collection and the excavation of a few 30-cm STPs. The area of the site was estimated to be about 20 m^2.

The site was revisited in 2003 when additional examination and the excavation of 50-cm STPs revealed a number of subsurface shell concentrations, indicating that the site was used as a resource camp on one or more occasions. The site boundary was revised to include an area of about 6,898 m^2. Testing was recommended (Campbell, Morehead, and Thomas 2008).

Site 8OK427

Site 8OK427 was originally identified during survey of Unit X-194 in 1984 when five prehistoric ceramics were recovered on the surface of a road and around a borrow pit. Five recording STPs were excavated around the surface finds, but no in situ remains were recovered. The site was evaluated as not eligible for the NRHP because of extensive damage from the borrow pit, construction and maintenance of the road, and erosion. The site was estimated to be about 40 by 15 m (600 m^2) in size. Further work was not recommended.

The site was relocated in 2007 (Campbell, Morehead, and Mathews 2008; Morehead et al. 2008) (see Appendix A:Figure A.3) while Unit X-886 was being surveyed; artifacts were observed near the borrow pit and along the road, in the same location where the previous work had focused. In addition, three STPs recovered ceramic and lithic artifacts to a depth of 90 cm in an area depicted as inundated at the time of the 1980s investigations. Sixteen additional STPs were excavated nearby, but all were negative.

Subsequent to the above tests, an additional 103 STPs were placed at 10–20-m intervals around the positive STPs in order to define site boundaries and characterize the site. Twenty-four of the additional STPs were positive. The combined survey and recording effort expanded the site's boundaries to cover an irregular area of approximately 11,390 m^2. In addition, one component of the site—the Late Woodland Weeden Island component—appears to cover the whole site to a depth of 60 cm. Site 8OK427 was subsequently tested, and portions were determined eligible for listing in the NRHP.

Site 8OK428

Site 8OK428 was first identified during the survey of Unit X-194 in 1984 when prehistoric ceramics, lithics, and historical-period artifacts were recovered on the surface in a 20-by-20-m area on either side of the road. Five recording STPs were excavated around the surface finds, but no in situ remains were recovered. The size of the site was estimated at 400 m^2. Road construction and erosion were noted to have severely disturbed the site. Site 8OK428 was evaluated as not eligible for listing in the NRHP. Archival work was suggested; however, further fieldwork was not recommended.

The site was relocated in 2007 (Campbell, Morehead, and Mathews 2008; Morehead et al. 2008) (see Appendix A:Figure A.4) when buried prehistoric and historical-period artifacts were discovered in nearby STPs as well as on the surface of a road embankment. Because the artifacts were found outside the original site boundaries, the site extent was expanded. An additional 40 STPs, 19 of which were positive, were excavated, and these tests, along with tests from previous delineation efforts, were used to redefine site boundaries. Site 8OK428 was determined to extend over an area of approximately 9,208 m^2. The site evaluation was revised on the basis of shovel testing; the site was recommended to be potentially eligible for listing in the NRHP, and testing was recommended.

Site 8OK433

Site 8OK433 was originally discovered in 1984 when a piece of debitage was found on the surface. Surface inspection, one discovery STP, and four recording STPs resulted in the discovery of one additional artifact. Site area was estimated at about 200 m^2, and the site was considered not eligible for listing in the NRHP. No further work was recommended.

The site was relocated in 2007 during survey of Unit X-885 (Campbell, Morehead, and Mathews 2008; Morehead et al. 2005; Morehead et al. 2008) (see Appendix A:Figure A.5). Fifteen STPs were excavated in the area of the initial discovery. Some were located at 15- and 20-m intervals in cardinal directions and on diagonals, whereas others were placed judgmentally, as dictated by landform and site orientation. Eight of the recording STPs were positive, expanding the site boundaries to approximately 70 by 45 m or approximately 1,500 m^2 (as calculated from GIS polygon). The site boundary to the west continues onto private property where Prentice Thomas & Associates (PTA) located another site, with contemporaneous deposits (8OK1033), suggesting that 8OK433 may be substantially larger than recorded. Prehistoric cultural deposits were located at depths from 10 to 100 cm, with a sterile zone between 30 and 40 cm, indicating potential for stratigraphic separation of multiple site components.

Site 8OK2114

Site 8OK2114 was first identified in 1982 (Campbell, Morehead, and Thomas 2008), during survey of Unit X-666. The site is located on a barrier island and contains prehistoric ceramics, oyster shell, and rangia shell on the surface. Recording efforts involved the excavation of 11 50-by-50-cm STPs placed judgmentally in and around the shell concentration. The area occupied by the site was determined to be 20 by 35 m, or approximately 700 m^2.

Subsequent test excavations in 2003, designed to investigate the potential for subsurface deposits at the site, included nine 50-by-50-cm STPs and three 1-by-1-m STPs. All STPs were negative. The three 1-by-1-m STPs produced only shallow shell deposits. Despite surface evidence, testing revealed no indication of significant subsurface deposits. As a result, the area occupied by the site was left unchanged.

Fort Drum

Fort Drum provided two case studies relevant to the discussion of site-boundary definition. Both reveal the complexity and underscore the difficulty in defining clear site boundaries using subsurface tests, particularly in areas of high archaeological sensitivity. Both cases also illustrate the problematic use of the site concept in areas where cultural remains are concentrated. They underscore the use of sensitivity zones as a better means of delimiting areas where cultural materials are likely to exist, based on the known distribution of positive tests or the density of cultural materials.

Site FDP 1093/1094

Sites FDP 1093 and FDP 1094 were discovered in 1997 during the cultural resource inventory of Training Area 4A. Located in open, sandy sediments (Ahr 1998:37–38) (see Appendix A:Figures A.6 and A.7), the sites were found during a pedestrian survey. Site boundaries were determined via a 100 percent survey of bare areas. In addition to surface artifacts, each of two STPs at FDP 1093 contained one flake (Ahr 1998:88–92). No STP in FDP 1094 was positive; all cultural materials at FDP 1094 were discovered on the surface. Both sites were interpreted as relatively small, discrete pottery and lithic scatters, located approximately 150 m from each other. Historical-period artifacts discovered in the vicinity of FDP 1093 and FDP 1094 were recorded as separate sites: FDH 1120 and FDH 1121.

During the course of annual monitoring of protected sites in 1999, the field crew discovered tire tracks throughout FDP 1093 (Fort Drum Cultural Resources Program 1999:1). Tire-track patterns indicated that the site area was being used as a short cut between two portions of a training area. Fort Drum decided to test the site further to (1) create a culturally sterile right-of-way where traffic could be channeled, (2) evaluate site significance, and (3) post the site area as off-limits.

To address these goals, 5-by-5-m STPs were excavated in areas considered most at risk of disturbance, and others were placed near the loci where diagnostic artifacts were found during Phase IB surface collections. Fourteen STPs (10 5-by-5 m, 1 1-by-2 m, and 3 1-by-1 m) were excavated in 1999. STPs established the presence of intact features and deposits but also indicated the site was being severely damaged by natural processes. Discoveries indicated that the site probably had significant research potential.

In 2001, further evaluation of FDP 1093 was undertaken because the site occupied land strategically important to the military for training exercises. By determining the boundaries of the site, areas of importance to military operations could be cleared through data recovery, and other parts of the site could be protected and preserved for future study (Fort Drum Cultural Resources Program 2001:59).

Thirty-seven 5-by-5-m STPs were excavated (Fort Drum Cultural Resources Program 2001:60). Excavation revealed additional components of the site and provided additional information on site structure (Fort Drum Cultural Resources Program 2001:79–84). The southern and eastern areas of the site produced few signs of cultural activity. In contrast, subsurface features and deposits were considered most likely to be found in the western portion of the site, given the presence of stabilized dunes.

Site FDP 1093/1094 illustrates a case where two relatively small sites were determined through additional work to be loci of a much larger and more complex site. Originally two small, discrete sites on the order of 2,000 m^2, the area is now considered one large site that is perhaps 20 times the size (40,469 m^2) of either site as originally recorded. The site keeps growing. Southern and eastern boundaries of the site were established, as well as a culturally sterile right-of-way where traffic could be channeled across the site, but other portions of the site boundary remain untested. In short, the site is much larger and more complex than originally expected, and the boundaries of the site have yet to be clearly defined.

Site FDP 1170

Site FDP 1170 lies in the north-central portion of Training Area 8B (Fort Drum Cultural Resources Program 2000:1–10) (see Appendix A:Figure A.8). The site was first discovered during a pedestrian survey in February 2000. The surrounding environment consists of a relatively flat mixed hardwood and pine forest, with large interspersed areas of open sand and moss surrounded by waterways. Originally discovered in a sandy blowout approximately 2 acres in size, FDP 1170 might be part of an extensive corridor of sites that comprise approximately 358 acres.

In June 2000, 205 5-by-5-m squares were surveyed and recorded within the site to determine site boundaries and investigate the distribution of cultural materials. Seven units were shovel scraped; additional units were excavated in promising areas and three test trenches were dug. Despite the relatively large level of effort, it proved impossible to establish a precise site boundary. Every attempt to open a unit to determine a boundary for the site resulted in the discovery of more cultural material and features. The site, initially thought to be limited to 2 acres in size, was clearly much larger and possessed greater potential than originally thought.

In 2002, a field crew returned to the site, excavating 25 2-by-2-m STPs. Prehistoric cultural material was found in all of them (Fort Drum Cultural Resources Program 2002:61–98). The excavations expanded the area of the site into multiple nearby blowouts and other undisturbed areas, increasing the site's extent to at least 13 acres; the exact boundaries of the site have yet to be established.

Site FDP 1170 is situated in a 358-acre archaeologically sensitive area. The area is best defined by streams that surround it, as opposed to distinct areas of archaeological material separated by areas devoid of such material. It may prove difficult to establish boundaries for FDP 1170 because our common concept of a site does not apply very well in this situation. Given that a low density of archaeological material is likely throughout the entire 358 acres, it is probably best to consider the entire area an activity zone, of which FDP 1170 represents but one or more loci. Currently, FDP 1170 is about 6.5 times larger than originally expected, but this statement misses the point that the original designation covers an area several hundred times smaller than the one of concern for military managers.

Summary and Conclusions

In this chapter, we addressed the issue of assessing variation in site-boundary data on military installations. After discussing the relevance of the issue to management concerns, we presented several metrics that can be used to assess change in site-boundary data. We then applied these metrics to data from Fort Bliss and the UTTR, and in a more selective manner to data from Eglin AFB and Fort Drum. There are several lessons to be learned.

One lesson is that site boundaries frequently change, for a variety of reasons. Particularly in environments where surfaces are active, such as dune fields or alluvial floodplains, site boundaries determined from surface observations are apt to change frequently as a result of natural processes. In some cases, change in site boundaries can be expected in 100 percent of situations when sites are rerecorded. In the absence of directional processes that always expand or contract site boundaries, we can expect that natural processes should have a more-or-less random effect on site boundaries. That is, site boundaries should increase or decrease site size in relatively equal proportions through time, and the net change in site area should be relatively close to zero. Installation managers should keep in mind, however, that the changes in site size can be extreme and that to account for random change may require periodic observations over long time spans.

Changes in boundary-definition rules also must be accounted for in assessing site-boundary data. In the case of Fort Bliss, change in survey methods and rules for site-boundary definition had the effect of a net increase in site area when measured as a percent change and as a change in ordinal size. To Leckman et al. (2008a), the effect of change in site-boundary-definition rules on Fort Bliss could mean that many previously recorded small sites could be subsumed into fewer larger sites if new boundary-definition rules were applied. We should not expect that a net increase will always occur when boundary-definition rules change, however. No change or a net decrease is also possible for installations that have boundary-definition rules that either have no consistent effect on site size or have a tendency to create smaller sites with more restrictive boundaries. Understanding the causes and directionality of change in site-boundary data is important to understanding how to interpret and correct for change in evaluating site distributions.

Another lesson is that change in site boundaries not only affects change in the areas that need to be managed on military installations but also can affect how particular sites are interpreted and evaluated. If sites are organized into size classes, site-boundary revision can result in the redefinition of a site into a new class. This redefinition can be a contributing factor in determining NRHP eligibility and, ultimately, site management and treatment.

A third lesson that is particularly important on installations that rely on shovel tests concerns the number of STPs or STP locations that are necessary to delineate site boundaries, particularly when artifact densities are low. Simply expanding outward from a positive test may not be an adequate method of testing site boundaries. There is a significant chance that tests within the site itself will be negative, particularly when STPs are small. Cases from both Fort Drum and Eglin AFB indicate that sites originally interpreted as small and insignificant—because of limited numbers of positive tests—turned out to be much larger, significant sites when more intensive testing was initiated near previously recorded sites.

A final lesson to be learned is that the designation “site” can sometimes impede, as opposed to assist, CRM compliance and management. In areas of high archaeological resource density, clear site boundaries are nearly impossible to establish. The delineation of archaeologically sensitive areas based on the density of cultural materials or positive STPs sidesteps this issue. Instead of focusing on defining a moving target (i.e., site boundaries), installation CRM programs can focus on characterizing zones within archaeological landscapes according to the density and distribution of cultural materials. Sites can certainly be defined within these zones, but the areas identified for particular management concerns would not be isolated, individual sites, but larger, more inclusive landscape zones that are considered more or less sensitive according to various management criteria. Rather than in the revision of site boundaries, management efforts could be placed in the development and revision of sensitivity zones. Sensitivity zones could be periodically tested through survey and updated to incorporate new archaeological data, modeling techniques, or changes in management guidelines.

Recommendations

A useful method for assessing change in site-boundary data is to calculate a series of metrics that track change in site area as a result of boundary revision. Although we were able to use data from Fort Bliss and the UTTR to illustrate use of these metrics, such data are not easy to obtain. We suggest that installations collect and maintain historical data on change in site boundaries, including change in site size and extent, reasons for site-boundary revision, and changes in rules for defining sites. We also suggest that installation CRM programs require explicit information from contractors regarding how site boundaries were defined and what specific instruments were used in their definition.

CHAPTER 5

Conclusions

Military operations require extensive planning and consideration of multiple factors. Many questions must be asked about the action itself: What is required for the training exercise? Where is the most effective place to test a missile? How much land is needed as a buffer zone? Beyond military contingencies, environmental and archaeological concerns must also be considered.

Environmental concerns are generally assessed as part of complying with the National Environmental Policy Act (NEPA). NEPA requires federal agencies to evaluate their actions in terms of the impact on the "human environment." The goal of a NEPA evaluation is for a federal agency to reach an informed decision about how to proceed with an action that considers the environmental, social, and cultural consequences of various alternatives. Although cultural and historical resources are included in the NEPA process, the law does not require federal agencies to find all cultural resources scheduled to be impacted or to have comprehensive data on all alternatives. What it does require is that the agencies gather and evaluate sufficient data in a scientifically valid manner.

Most agencies comply with the cultural resource aspect of NEPA by fulfilling their obligation under Section 106 of the National Historic Preservation Act (NHPA). Section 106 of NHPA requires that agencies identify historic properties that could be affected by undertakings that they fund or approve. Understanding that no one can guarantee that all historical and cultural resources can be identified, federal agencies are required to put forth "a reasonable and good faith effort" to identify historic properties (36 CFR 800.4[b]1).

Although historic properties include historical, traditional, and archaeological resources, this project focused solely on the last of the three. In the 40 years during which federal agencies have complied with NEPA and NHPA, vast amounts of time and money have been spent on inventorying, evaluating, and treating archaeological sites. Early on, it became clear that U.S. archaeology had not developed a standard methodology for characterizing the archaeological record. The definition of a site and the proper ways of recording sites varied among archaeologists. Such variation, though expected and acceptable as a matter of academic discourse, was problematic with respect to legal compliance.

Over time, federal and state agencies, together with archaeologists, Native Americans, and other interested parties, developed acceptable definitions of archaeological resources and methods of identifying and evaluating them. Definitions and standard practices emerged at the state level. In the East, identification became synonymous with shovel-test surveys, whereas in the West, pedestrian surveys with regularly spaced transects became the rule. Site definitions varied greatly, from a single artifact in Nevada to 30 artifacts in an area 15 m in diameter in Arizona.

Military installations generally adopted state definitions and practices. With standardization came the expectation that data collected at different times, by different contractors, would be equivalent in quality. Essentially, data were data, and all were sound. This assumption still goes unquestioned at many installations. But its uncritical acceptance has led to an increasing number of cases in which archaeological sites are either mischaracterized (e.g., small artifact scatters later found to be large village sites) or missed entirely. These cases have an impact that goes well beyond the individual undertaking. Stakeholders become reluctant to trust the military. Installations are required to "prove" that they have not missed anything. Considerable time and money have been spent resurveying areas that have already been inventoried, revisiting sites that have already been recorded, and performing data recovery on sites that no longer exist, have no significant subsurface components, or are dramatically larger and more complex than expected.

Few, if any, archaeologists consider archaeological data free of error. Yet, in CRM, the number of studies devoted to assessing the types of error and their severity is surprisingly small. This lack of interest is unfortunate, for error is not only inherent in observation, it is essential in assessing the confidence we should have in observations we have already made.

Goals of the Project

The goals of this pilot project—to answer the following two questions—were simple: How good are CRM data on military installations? And does the quality of the data affect decisions critical to military activities? Our interest was less academic than practical. Survey reliability is important, for example, less because better estimates of site density will illuminate the past than because the difference between a region containing 3 sites and one containing 10 sites per 100 acres has direct impacts on military activities. At best, sites found after plans have been put in motion only add costs to the project; at worst, they constrain activities to a point that diminishes military objectives.

The pilot project focused on the "low-hanging fruit" of CRM data: standardized inventory and recording practices. There is little disagreement, for instance, among archaeologists on what constitutes a site datum. Thus, we can be assured that in plotting the same site datum on different maps, archaeologists are trying to locate the same point in space. Our task was simply to determine, using objective and replicable procedures, how well they have done.

Survey Reliability

Survey reliability is one of the best studied, yet least understood, methodological issues in U.S. archaeology. Numerous studies have been performed on various aspects, including sample fraction, transect size and shape, edge effect, and shovel-test size and spacing. All these studies have shown that discovery of archaeological sites is far from perfect and confirm Judge's (1981:128) statement that "the more time spent in the field looking for sites, the more sites will be found."

Even in the face of this voluminous evidence, managers continue to work under the assumption that a 100 percent Class III intensive survey will find 100 percent of the archaeological sites. How one might reach this conclusion is certainly understandable. After all, doesn't 100 percent mean 100 percent? The answer is that there is a difference between walking or probing 100 percent of the area and finding 100 percent of the target subjects—in this case, archaeological sites. As Chapter 2 demonstrated, all surveys are not the same. Walking transects or placing shovel-test pits (STPs) at 15-m intervals, for example, will find significantly more sites than observations spaced at 30- or 50-m intervals. Without a doubt, the most important factor for DoD in finding sites is visibility. Systematic pedestrian surveys in the arid West, where vegetation is sparse and ground visibility is good, will almost always outperform shovel-test surveys in vegetated areas in the East. More sites will be found at lower cost at the BMGR, Fort Bliss, the UTTR, or China Lake than at Fort Drum or Eglin AFB.

Even if surveys do not find every site, do they find enough sites? The answer depends on your objective: Do you need to know the location of every site to make a decision? Or does your required action need only enough data from which you can estimate the locations and types of sites in a project area? To evaluate whether survey reliability is adequate, we must know the purposes for which installations use these data.

Installation CRM programs have two distinct responsibilities. The first is to ensure that the military missions are not impeded or constrained by cultural resources. To meet this objective, CRM programs *react* to military decisions through the Section 106 process. Underlying this process is the assumption that survey methods find either all or a very large proportion of significant archaeological sites (i.e., sites that are listed or eligible for listing in the NRHP). Many installations meet the "reasonable and good faith" threshold for their identification efforts by preparing a survey or inventory design document that is then reviewed by the SHPO, Native American groups, and other stakeholders. In the ensuing agreement document, the parties explicitly acknowledge the possibility of missed sites in a "discovery clause" that stipulates how sites that are discovered during the undertaking will be handled. In short, Section 106 assumes that at some point, either before or during the undertaking, all significant sites can be found. It is to the installation's benefit that this knowledge is gathered earlier, rather than later, in the process.

The second responsibility of installation CRM programs is to preserve and protect the cultural and historical values embedded in the installation's historic properties. Often viewed through the lens of complying with Section 110 of the NHPA, this objective requires the installation to *manage* both the cultural resources that have been identified and those that are still to be discovered. To manage effectively, the installation must know where recorded sites are located and where heretofore undiscovered sites will be found. Installations must be able to answer the following questions: How many sites have been found? How many have been missed? How many more can we expect? In other words, installations also must be able to provide reasonable estimates of the types of sites that have been found, the expected site types that have yet to be recorded, and the expected total number of sites on the installation, according to site type. These questions require installations to understand methods used in discovering sites so that estimates and confidence intervals can be correctly calculated. Most installations employ rigorous and systematic survey methods that are designed exactly for these purposes. Yet few installations actually complete this process by developing trustworthy estimates using available techniques and information.

Instead, there is a tendency to manage cultural resources at an installation proceeding from the false assumption that survey identifies all the resources that have to be managed within a survey area. By linking all the survey efforts together and combining the identified sites into one cohesive database, installation managers then create a "measles" map that shows points or polygons of various sizes and shapes spread throughout the installation. Managing cultural resources, then, consists of surveying areas that have not been inventoried and avoiding or, if necessary, treating the "dots."

The measles map approach relies on the "site" as the basic unit for identifying and managing cultural resources. Although archaeologists and managers use the same term, their perceptions of the term "site" are quite different. Focused on their compliance obligations, managers look to the law for guidance. NHPA requires federal agencies to identify historic properties or historic resources, which are defined as "any historic district, site, building, structure, or object included in, or eligible for inclusion on the National Register" (16 U.S.C. 470w [5]). Although largely implicit in the law, a site as envisioned by NHPA has empirical dimensions that correspond to a cohesive set of cultural and historical characteristics. A site can be located in space with boundaries placed around it such that the historical and cultural values worthy of consideration lie solely inside those boundaries. It follows that, outside site boundaries, significant cultural resources are absent.

Archaeologists know such assumptions to be false and therefore do not make them part of their definition. To an archaeologist, a site is largely a bookkeeping device by which spatial clusters of archaeological material are distinguished from other spatial clusters. From a research perspective, there is nothing inherently meaningful about this designation. Sites take on scientific value only when the archaeological record can be linked to a cohesive set of human behaviors performed at that location or in comparison to other locations.

It is generally this latter step of linking behavior to the archaeological record that is missing in CRM. Sites are defined according to standardized definitions and practice. Managing the "dots" is more about protecting artifacts and features than stewarding the past. To actually meet the legal requirements of NHPA, installations should be less concerned with identifying locations of archaeological material than with identifying the environmental, cultural, and social influences that shaped past human use of the region and are encoded in the installation's archaeological record.

Shifting managers' focus from identifying sites as discrete locations in space to identifying regional patterns in cultural materials has distinct advantages. First, it moves CRM programs away from the quixotic task of “complete” identification of sites. Second, it places a strong emphasis on pattern recognition and on estimation that is based on a realistic understanding of survey reliability. Third, it allows installations to “think big” in terms of discerning which resources are significant and strategizing how to evaluate or manage significant resources. Significance should be attributed to patterns and relationships, not to locations. Instead of constantly returning to the same frustrating question—What will we learn if we dig this particular site?—we can ask another: What site(s) should we dig to elucidate this cultural pattern or to understand this relationship?

This pilot project has clearly demonstrated the faulty premise under which most installation CRM programs operate: 100 percent survey does not mean 100 percent site discovery. How many resources, and of what type, are found and how many are missed can be known and should be incorporated into an installation’s Integrated Cultural Resource Management Plans (ICRMPs) and similar documents. Improving survey reliability, however, should not be an end in itself. We must move beyond managing the “dots.” We must use inventory data to identify larger patterns in the distribution of cultural resources and use those understandings to develop and address important questions about the past. Using this approach, we can better meet the requirement of a making a reasonable and good faith effort in identifying significant resources.

Site Location

Once a site is found, it must be recorded. We must first figure out where we are on an installation and then define the size and shape of the archaeological site. Finding where a site is on a map seems simple and straightforward. Archaeologists in the field establish a site datum: a reference point for all site measurements. The datum is marked by a permanent fixture, such as a rebar post. Subsequently, then, other archaeologists can find the exact same spot. In some cases, archaeologists visiting a previously recorded site rerecord the site datum’s spatial coordinates. By comparing the various recorded locations, we can assess the variability in measuring the location of the same physical marker.

Ideally, there should be little to no variation in the recordings. Military activities are planned under the assumption that recorded archaeological sites truly are in their mapped locations. Yet we know that each measurement is subject to various sources of error, including instrument and observer error. Archaeology abounds with tales of sites being located on the wrong ridge, on the wrong side of a river, or simply hundreds of meters from the recorded location. The question we posed, then, was simple: Are sites located where they are plotted?

The answer is no. Using the UTTR as an example, we demonstrated that over the last two decades, sites have “moved” on average about 80 m. We presume, however, that most of this movement has been toward more-accurate recordings. Technological advances, particularly GPS, have greatly reduced site-location error. Although measurements made in recent years may be considered more reliable than previous ones, many sites on military installations are still subject to considerable site-datum error.

Probability theory, the cornerstone upon which statistical logic is based, can be used in this situation. Although it is commonly assumed that probability theory is helpful in determining the accuracy of any particular sample (e.g., the “true” location of a site datum), such is not the case. Probability theory, however, is useful in determining how that sample compares to all the possible samples that could have been chosen. Thus, although we may not be in a position to assess the accuracy of a particular measurement of site location, we can use that measurement together with other similar measurements to provide an estimate of the site’s location with a known degree of confidence. By combining measurements, we provided several approaches to using GIS to calculate reliable and precise locations of site datums.

Site Boundaries

Whereas every archaeologist records and rerecords a site from the same or a related datum, the definition of a site's boundary involves more than plotting points in space. Central to the definition are decisions about how to characterize the archaeological record. For example, clusters of surface artifacts distributed on two ridges on opposite sides of a small wash could be recorded as one large site, one site with two distinct loci, or two separate sites. There is nothing inherently right or wrong about any of these definitions. Another major factor is visibility. On dynamic geomorphic surfaces, artifacts are buried, are exposed, and move. Observations at different times record the changing surface distribution of archaeological materials as different sizes and shapes. Using surface data from the UTTR and Fort Bliss, we demonstrated that changes in site size are the norm, not the exception. Sites increase and decrease in size, often dramatically. Case studies from Eglin AFB and Fort Drum showed that the conclusions about shovel-test inventories extend to site recording: small sample fractions make it very difficult to accurately judge the subsurface archaeological record. Defining sites based on shovel tests often requires large levels of effort, which are rarely used in CRM.

Although archaeologists differ on what constitutes a site and how to draw its boundaries, there is general agreement on what is meant by archaeological materials (e.g., artifacts, features, and trash mounds). Most archaeologists would recognize the same material as cultural and, under specified guidelines, identify similar phenomena as "sites." Because they would differ on other aspects, it makes good sense to view site boundaries as approximations: that is, there is a significant cultural resource in the general vicinity, but our definition of its exact size, shape, and character is imprecise.

Beyond differences of opinion, there are other factors which must be borne in mind when evaluating site boundaries. As with survey, the more time spent on a site, the more features and artifacts will be found. Usually, the time spent on recording sites during survey is relatively short. Field crews focus on establishing a site datum, recording that datum in space, drawing a field map, recording information on artifacts and features, and creating a photographic record of the site and its setting. Currently, the expectation on the part of the installation and contractors at the BMGR is that field crews should be able to survey about 20 acres per person per day at intervals spaced at 15-m intervals, including the discovery and recording of three sites. Time spent recording a site should be on the order of 1–1.5 hours. When the same site later becomes a subject of interest (for instance, when a road through it is planned), a new recording team may revisit the site precisely to make sure that boundaries are adequately mapped and that the site has been accurately characterized. More time may be spent at the site during rerecording, resulting in a much more detailed site map and a greater understanding of the site's surface assemblage. Changes in site boundaries, then, should be expected not simply because of changes in the environmental conditions, but because of the military's need to have more-detailed information for planning purposes.

Given the expected change in site boundaries, how should the military use archaeological inventory data during planning? One approach is to buffer site areas when considering possible undertakings. For example, on the UTTR, it would be reasonable to add 80 m in all directions from the mapped site polygon. Once the footprint of a planned activity is set, installation archaeologists could revisit the site to refine the boundaries as much as possible. It is important to point out, however, that even if archaeologists accurately portray the surface distribution of archaeological material, there is no guarantee that subsurface deposits will be found in the same area. The relationship between the surface distribution of artifacts and features and the types and location of subsurface deposits is poorly understood. Without adequate subsurface testing there is simply no way to ensure that archaeological site boundaries are accurate. This caveat applies to all survey methods but is especially true where surface visibility is low and site boundaries are determined through relatively few widely spaced shovel tests.

Another approach is to augment military planning based on known archaeological site locations with sensitivity zones. Predictive models of archaeological site location are developed by correlating locations with and without archaeological material with environmental features and then generalizing these relationships to areas of the installation that have not been surveyed. The result is a map that provides the probability that any particular location will or will not contain a site. These probabilities are generally classed into high-, medium-, and low-sensitivity areas. For purposes of planning military activities, knowing the general likelihood that significant archaeological sites will be encountered can be helpful and may be sufficient for NEPA compliance. Predictive models would still have to be followed by inventory if the undertaking proceeds, but the level of effort expended to locate sites could be tailored to the likelihood that such a site would be encountered.

Modeling also should inform on the level of effort needed to define site boundaries, particularly in areas where surface visibility is low. As the probability of finding an artifact in a shovel test decreases, the number of shovel tests needed to accurately define its boundaries increases. If predictions yielded probabilities for finding particular site types, each with a specific discovery probability, then the number of shovel tests needed to adequately define and characterize that site could be calculated. For example, we need five times the number of shovel tests to define a deposit with an average of one artifact per square meter than a similar sized site with a deposit that has five artifacts per square meter. On most installations, however, the spacing and size of shovel tests used to record sites is set to a uniform standard in the survey design. The result is that the size and shape of most sites of low and medium artifact density are grossly misrepresented.

The pilot project has shown that advances in technology have led to improvements in mapping points in space, but that such improvements have not materially affected our ability to define site boundaries. The factors affecting site-boundary definition are complex. Instrument error and changes in environmental conditions may be controlled or at least understood. More challenging is the subjective nature of the site concept and the different roles it plays in archaeological research and land planning.

Recommendations

We have provided a number of specific recommendations in the preceding chapters regarding how to address problems in archaeological data quality for specific installations. These will not be repeated. In this section, we instead forward three general recommendations to DoD CRM programs.

Legacy Data

It is ironic that installation archaeologists—people who spend their careers protecting and preserving past human behavior—appear to have little interest in curating records of their own behavior. In general, installations keep information on the last site visit and discard earlier data on site location and site boundaries. This practice is based on the tenuous assumptions that the most recent visit is the most accurate assessment of the site and that multiple datasets will lead to confusion in the future. They also assume that many surveys are equivalent and do not curate explicit data on how survey was conducted or how sites were recorded. The result is that exactly those data needed to evaluate data quality are not available. It is telling that all the installations involved in this project were extremely interested in the subject but for the most part were not able to provide all the data required to perform the analyses.

We suggest that installations keep detailed records on the history of site discovery and recording. These should include, but not be limited to, technical information on mapping instruments, GPS units, and GIS technology. Additionally, data should be kept on survey intervals, transect size and shapes, shovel-test design and actual shovel tests dug, screen mesh size, and site definition. Environmental data on erosion and visibility (including season, lighting, and weather conditions) should also be kept on any recording or rerecording effort.

Metrics

Much of this pilot project was devoted to developing objective and replicable measures for three aspects of data quality: survey reliability, site location, and site boundaries. Our purpose was not simply to assess data quality on specific installations. We also wanted to show how other installations could use their own data to create more-reliable estimates of the resource base (e.g., total number of sites, sites by site type) and to evaluate the relationship between their specific methods and survey results (e.g., edge effect, survey intensity, sites missed). The metrics developed for this analysis are but a small sample of similar metrics applicable to this type of study.

Site Concept

The site concept is the cornerstone of DoD archaeological CRM programs. An installation's goal is to find, map, and avoid sites. This goal might be achievable if archaeologists could remove the ambiguity surrounding the site concept. It is not simply that the definition of a site changes between jurisdictions, but that what constitutes a site changes between archaeologists and even between visits to a location by the same archaeologist. Managers must accept that sites are theoretically and empirically "fuzzy." Sites may remain in the same general location, but their sizes, shapes, depths, and importance change in time with environmental conditions and academic debate. We have suggested two ways to manage these moving targets. First, buffer sites with an area large enough to minimize surprises. Second, focus less on specific locations and more on regional patterns. This latter approach includes using predictive modeling of where sites of specific types are more or less likely to be discovered, as well as using survey reliability to model where sites of specific types are more or less likely to have been missed.

Future Research

Although survey reliability, site location, and site-boundary definition are important topics in archaeological data quality, they are by no means the only areas in which data of questionable or misinterpreted quality are used to make DoD decisions. Two other subjects of great interest are in-field analysis and site exposure. In-field analysis is used primarily to classify sites according to period, culture, and site function. Site exposure relates to the correlation between the site's characteristics as inferred through the study of surface materials and those same characteristics as defined by subsurface excavation.

The classification of archaeological sites found during survey was traditionally based on detailed analyses of surface-collected artifacts in the laboratory. Combining laboratory analysis with field descriptions of feature types, archaeologists assigned sites to a period, site function, and archaeological culture. Disagreements about classifications were usually settled by reanalyzing the collections.

Today, there is a general tendency not to collect artifacts during a survey in the western United States. Period and site function are assigned on the basis of in-field analysis of artifacts and features. Field crews identify time-sensitive diagnostic artifacts and characterize surface artifacts in terms of raw-material type and artifact class or tool type. Surface features are mapped and classified according to descriptive and/or functional type. In-field analysis saves time and money; there are no curation costs, and the time spent in the field analyzing artifacts is substantially less than what would have been spent in the laboratory.

There is little question that the quality of data produced by in-field analyses is poorer than that generated by laboratory analysis. Observations and analyses performed in the field are done by field supervisors or by field technicians who are often not trained for these tasks. The field also offers less-than-ideal conditions; artifacts are not washed, lighting is not controlled, and field staff often have to contend with heat, cold, rain, or wind. In contrast, laboratory analyses are performed by trained specialists, such as ceramicists or lithicists, under controlled conditions. Most important, when analyzed only in the field, artifacts are not available for others to reanalyze, thereby precluding independent assessments of the classification and interpretation.

The situation has become so problematic that some federal agencies discount in-field analytical results in making determinations of significance. The Tonto National Forest, for example, no longer gives any credence to contractors' classification of pottery types (Michael Sullivan, personal communication 2007), relying instead only on their own staff's assessment. Such a move requires that all sites discovered by contractors must be revisited by U.S. Forest Service personnel.

Assessing the quality of in-field analysis and interpretation would require a much more elaborate experimental design than we established for this project. Many archaeological interpretations, such as site function, are based on subjective criteria that are not explicitly stated and that change among archaeologists. We shied away from these thornier issues and focused on topics that could be addressed with data that are more readily accessible and with more-objective and more-replicable measures.

Just because we did not include the problem of in-field analysis in our study, however, does not mean that the topic is not amenable to study. For example, counting surface artifacts by class, such as ceramic, lithic, and faunal, is a task of similar complexity to locating a point in space. Problems we identified with site datums, in terms of observation and instrument error, may have counterparts in these simple observations. By limiting the comparison to sites where all surface artifacts were analyzed and environmental conditions were held constant (e.g., in-field analyses conducted on the same day), we could ensure that the same surface assemblage was available to each recording crew. A somewhat more abstract problem involves typing diagnostic artifacts, such as painted ceramics, or classifying features into feature types (e.g., classifying a rock feature as a rock pile, hearth, roasting pit, or rock ring). In addition to observational error, these decisions include a definitional component, which is akin in some respects to the issues involved in the definition of site boundaries.

We could address variation in artifact classification by comparing different in-field analyses of the same site. Ideally, we would add a third component to the study by actually collecting the artifacts analyzed in the field and then reanalyzing them in the laboratory.

The issue of site exposure is of great interest to DoD. It is one thing for an artifact scatter to change size and shape; it is quite another for the scatter to change into a buried village, replete with house features, extramural pits, and cemetery areas. There is also the problem of unanticipated discoveries: finding undiscovered sites during construction in previously surveyed areas.

Investigating site exposure requires data from all levels of archaeological research. We must compare the site as drawn and described on the basis of survey data with the site as defined through subsurface testing or geophysical investigations, and the entire site as defined through complete data recovery.

Traditionally, field methods used to define archaeological sites could be divided neatly between survey and excavation. In the last few decades, advances in geophysical research have applied new instruments—such as magnetometers, ground-penetrating radar, and resistivity probes—to the detection of subsurface archaeological deposits. The subsurface extent and composition of some archaeological sites can now be defined without large amounts of subsurface testing. Geophysical techniques, however, still requires substantial research to demonstrate their utility as an effective archaeological tool. Each technique works under a specific set of environmental conditions for only certain types of archaeological deposits. Many techniques, for instance, are much better at identifying rock walls surrounded by fine sediment than they are at identifying features like pits, whose borders and fill consist of similar materials. Unfortunately for DoD, the most problematic types of sites for geophysical techniques—sites consisting of trash deposits, earthen house pits, and small extramural pit features—are ubiquitous on military installations. An exploration of site exposure, therefore, must include sites defined by both traditional and geophysical methods.

Ideally, we would find sufficient examples for investigating site exposure from individual installations to perform quantitative comparisons. Although we proposed such a study for this project, we found than none of the installations had more than a few case studies and that data of the kinds we required were difficult to come by. The problem was not always a lack of excavations. As with other Legacy datasets, we could not obtain all the needed data to perform the desired analysis. One approach to circumvent the problem posed by insufficient cases would be to augment an installation's dataset with cases from surrounding areas. For example, we could combine the data from installations in the eastern California desert: Fort Irwin, China Lake, Twentynine Palms Marine Corps Air Ground Combat Center (MCAGCC), Edwards AFB, and the Chocolate Mountain Range. These installations have similar archaeological records and face similar site-exposure issues. Another approach would be to combine an installation's data with those obtained from other adjacent or nearby federal agencies. In particular, state Departments of Transportation (DoTs) have sponsored many data recoveries, which could be combined with data from the installation. In that manner, we would have a good chance of obtaining data on each site class represented on the installation.

Final Comment

In this report, we examined the quality of the data used by military installations to make decisions about archaeological resources. For the most part, we found that the data had been collected by competent archaeologists adhering to professional standards of research. We were concerned less by errors in the data than by the fact that installations had gone to considerable time and expense to collect data in systematic and rigorous ways only to fail to put the data to their best use: to produce trustworthy and comparable estimates of archaeological resources. By failing to do so, military CRM programs are squandering hard-won information that is central to meeting their responsibilities. We hope that our project has shown not only that this type of information is readily developed using work that has already been completed, but also that it is well worth the effort.

REFERENCES CITED

Abbott, J., T. R. Mauldin, P. E. Patterson, W. N. Treirweiller, R. J. Hard, and C. R. Lintz
1996 *Significance Standards for Prehistoric Archaeological Sites at Fort Bliss: A Design for Further Research and the Management of Cultural Resources*. U.S. Army Corps of Engineers, Fort Worth District, Fort Worth, Texas.

Ahr, Steven W.
1998 *Cultural Resource Investigations at Fort Drum, New York, FY 1997: Volume I: Results*. Technical report prepared for the Cultural Resources Program, Environmental Division, Public Works, Fort Drum. Center for Ecological Management of Military Lands, Colorado State University, Fort Collins, Colorado.

Altschul, Jeffrey H.
2000 Programmatic Approach to the Inventory and Mitigation of Sparse Lithic Scatters at Camp Pendleton. Manuscript on file, Statistical Research, Tucson, Arizona.

Altschul, Jeffrey H., and Christopher L. Nagle
1988 Collecting New Data for the Purpose of Model Development. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling,* edited by W. J. Judge and L. Sebastian, pp. 257–299. USDI Bureau of Land Management Service Center, Denver, Colorado.

Altschul, Jeffrey H., and Adrianne G. Rankin
2008 Introduction. In *Fragile Patterns: The Archaeology of the Western Papaguería*, edited by Jeffrey H. Altschul and Adrianne G. Rankin, pp. 1–27. SRI Press, Tucson, Arizona.

Altschul, Jeffrey H., Lynne Sebastian, and Kurt Heidelberg
2004 *Predictive Modeling in the Military: Similar Goals, Divergent Paths.* Preservation Research Series 1. SRI Foundation, Rio Rancho, New Mexico.

Altschul, Jeffrey H., and Ronald H. Towner
1994 Prehistoric Settlement in the Intermontane Regions of the Sand Tank Mountains. In *Intermontane Settlement Trends in the Eastern Papaguería: Cultural Resources Sample Survey in the Northeastern Barry M. Goldwater Range, Maricopa County, Arizona,* by Jeffrey A. Homburg, Jeffrey H. Altschul, and Rein Vanderpot, pp. 285–322. Technical Series 37. Statistical Research, Tucson, Arizona.

Banning, E. B.
2002 *Archaeological Survey*. Kluwer Academic/Plenum Press, New York.

Binford, Lewis R.
1962 Archaeology as Anthropology. *American Antiquity* 28:217–225.

1965 Archaeological Systematics and the Study of Culture Process. *American Antiquity* 31:203–210.

1992 Seeing the Present and Interpreting the Past—and Keeping Things Straight. In *Space, Time, and Archaeological Landscapes*, edited by Jacqueline Rossignol and LuAnn Wandsnider, pp. 43–59. Plenum Press, New York.

Burger, Oskar, Lawrence C. Todd, Paul Burnett, Tomas J. Stohlgren, and Doug Stephens
2002–2004 Multi-Scale and Nested-Intensity Sampling Techniques for Archaeological Survey. *Journal of Field Archaeology* 29(3/4):409–423.

Campbell, L. Janice, James R. Morehead, and James H. Mathews
2008 *Test and Evaluation of 8OK427, 8OK433, 8OK434, 8OK435, 8OK2621, 8OK2627, 8OK2631 & 8OK2632 (Task Order CR-08-0042), Cultural Resources Management Support, Eglin Air Force Base, Okaloosa, Santa Rosa and Walton Counties, Florida*. Report of Investigations No. 1147. Prentice Thomas & Associates, Mary Esther, Florida.

Campbell, L. Janice, James R. Morehead, and Prentice M. Thomas
2008 *Test and Evaluation of Fourteen Sites on Santa Rosa Island (Task Order CR-03-0008), Cultural Resources Management Support, Eglin Air Force Base, Okaloosa, Santa Rosa, and Walton Counties, Florida*. Report of Investigations No. 853. Prentice Thomas & Associates, Mary Esther, Florida.

Carmichael, D. L.
1986 *Archaeological Survey in the Southern Tularosa Basin of New Mexico*. Historic and Natural Resources Report No. 3. Environmental Management Office, Directorate of Engineering and Housing, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas.

Dean, Jeffrey S.
1978 Independent Dating in Archaeological Analysis. In *Advances in Archaeological Method and Theory*, Vol. 1, edited by Michael B. Schiffer, pp. 223–255. Academic Press, New York.

Dewar, Robert E., and Kevin A. McBride
1992 Remnant Settlement Patterns. In *Space, Time, and Archaeological Landscapes*, edited by Jacqueline Rossignol and LuAnn Wandsnider, pp. 227–255. Plenum Press, New York.

Dunnell, Robert C.
1971 *Systematics in Prehistory.* Free Press, New York.

1992 The Notion Site. In *Space, Time, and Archaeological Landscapes*, edited by Jacqueline Rossignol and LuAnn Wandsnider, pp. 21–40. Plenum Press, New York.

Dunnell, Robert C., and W. S. Dancey
1983 The Siteless Survey: A Regional Scale Data Collection Strategy. In *Advances in Archaeological Method and Theory*, Vol. 6, edited by Michael B. Schiffer, pp. 267–287. Academic Press, New York.

Ebert, James I.
1992 *Distributional Archaeology.* University of New Mexico Press, Albuquerque.

ESRI
2008a ArcGIS Desktop Help 9.3: Mean Center (Spatial Statistics). Electronic document, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Mean_Center_(Spatial_Statistics), accessed November 3, 2008.

2008b ArcGIS Desktop Help 9.3: How Mean Center (Spatial Statistics) Works. Electronic document, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20 Mean%20Center% 20(Spatial%20Statistics)%20works, accessed November 3, 2008.

2008c ArcGIS Desktop Help 9.3: Directional Distribution (Standard Deviational Ellipse) (Spatial Statistics). Electronic document, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Directional_Distribution_(Standard_Deviational_Ellipse)_(Spatial_Statistics), accessed November 3, 2008.

2008d ArcGIS Desktop Help 9.3: How Directional Distribution: Standard Deviational Ellipse (Spatial Statistics) Works. Electronic document, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20Directional%20Distribution:%20Standard%20Deviational%20Ellipse%20(Spatial%20Statistics)%20works, accessed November 3, 2008.

2008e ArcGIS Desktop Help 9.3: Standard Distance (Spatial Statistics). Electronic document, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Standard_Distance_(Spatial_Statistics), accessed November 3, 2008.

2008f ArcGIS Desktop Help 9.3: How Standard Distance (Spatial Statistic) works. Electronic document, http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=How%20Standard%20Distance%20(Spatial%20Statistic)%20works, accessed November 3, 2008.

Fish, Suzanne K., and Stephen A. Kowalewski (editors)

1990 *The Archaeology of Regions: A Case for Full Coverage*. Smithsonian Institution Press, Washington, D.C.

Fort Drum Cultural Resources Program

1999 *Cultural Resource Investigations at Fort Drum, New York, FY 1999: 1999 Report.* Cultural Resources Program, Environmental Division, Public Works, Fort Drum, New York.

2000 *Cultural Resource Investigations at Fort Drum, New York, FY 2000: 2000 Report*. Cultural Resources Program, Environmental Division, Public Works, Fort Drum, New York.

2001 *Cultural Resource Investigations at Fort Drum, New York, FY 2001: 2001 Report*. Cultural Resources Program, Environmental Division, Public Works, Fort Drum, New York.

2002 *Cultural Resource Investigations at Fort Drum, New York, FY 2002: 2002 Report*. Cultural Resources Program, Environmental Division, Public Works, Fort Drum, New York.

Homburg, Jeffrey A., Jeffrey H. Altschul, and Rein Vanderpot

1994 *Intermontane Settlement Trends in the Eastern Papaguería: Cultural Resources Sample Survey in the Northeastern Barry M. Goldwater Range, Maricopa County, Arizona.* 2nd. ed. Technical Series 37. Statistical Research, Tucson, Arizona.

Judge, W. James

1981 Transect Sampling in Chaco Canyon—Evaluation of a Survey Technique. In *Archaeological Surveys of Chaco Canyon, New Mexico*, edited by A. W., Hayes, D. M. Brugge, and W. James Judge, pp. 101–137. USDI National Park Service, Washington, D.C.

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.

Kintigh, Keith
1988 The Effectiveness of Subsurface Testing: A Simulation Approach. *American Antiquity* 53:686–707.

Krakker, James J., Michael J. Shott, and Paul D. Welch
1983 Design and Evaluation of Shovel-Test Sampling in Regional Archaeological Survey. *Journal of Field Archaeology* 10(4):469–480.

Leckman, Phillip O., Bradley J. Vierra, and Monica Murrell
2008a *Archaeological Testing and National Register of Historic Places Evaluation of 31 Prehistoric Sites in Maneuver Area 1B, outside Loop 375, El Paso County, Texas*. Technical Report No. 07-38. Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas. Prepared by Statistical Research, El Paso, Texas.

2008b *Archaeological Testing and National Register of Historic Places Evaluation of 34 Prehistoric Sites in Maneuver Area 8, Fort Bliss, Otero County, New Mexico*. Technical Report No. 07-39. Environmental Division, Conservation Branch, U.S. Army Air Defense Artillery Center, Fort Bliss, Texas. Prepared by Statistical Research, El Paso, Texas.

Lucas, Gavin
2001 *Critical Approaches to Fieldwork: Contemporary and Historical Archaeological Practice*. Routledge, London.

Lukowski, P., M. A. Flowers, and E. Perez
2006 *Thirty Years Later: Results of a 2,045 Acre Survey to Assess the Conditions and Boundaries of LA 37179 (The McGregor Site), LA 37180, LA 145197, and LA 150612, Otero County, New Mexico*. Fort Bliss Cultural Resources Report No. 0518. TRC Environmental, El Paso, Texas.

Mbutu, S. K.
1997 *Results of Phase II Investigations and Recommendations for Phase III Data Recovery at 37 Archeological Sites in Selected Areas of Maneuver Areas 1, 2, and 8, Fort Bliss, Texas.* Miscellaneous Report of Investigations No. 88. Geo-Marine, Plano, Texas.

Morehead, James R., L. Janice Campbell, William R. Mallory, and Thomas Jennings
2008 *Cultural Resources Survey of X-885 & X-886 (Task Orders Cr-07-0036, Cr-07-0037), Cultural Resources Management Support, Eglin Air Force Base, Okaloosa, Santa Rosa, and Walton Counties, Florida.* Report of Investigations No. 1030. Prentice Thomas & Associates, Mary Esther, Florida.

Morehead, James R., L. Janice Campbell, and Frank Servello
2005 *Test and Evaluation at 8WL1404, 8WL1543, 8WL1727, 8WL1759 and 8WL1863 (Task Order CR-03-0014), Cultural Resources Investigations, Eglin Air Force Base, Okaloosa, Santa Rosa, and Walton Counties, Florida.* Report of Investigations No. 804. Prentice Thomas & Associates, Mary Esther, Florida.

Mueller, James W. (editor)
1975 *Sampling in Archaeology*. University of Arizona Press, Tucson.

Nance, Jack D., and Bruce F. Ball
1986 No Surprises? The Reliability and Validity of Test Pit Sampling. *American Antiquity* 51:457–483.

O'Laughlin, T. C.
1987 An Interim Report for Phase I Archaeological Investigations, Loop 375, Fort Bliss Maneuver Area I, El Paso County, Texas. Manuscript on file, Conservation Branch, Directorate of Environment, Fort Bliss, Texas.

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.

Pollard, A. Mark, and Carl Heron
1996 *Archaeological Chemistry.* The Royal Society of Chemistry, Cambridge, Great Britain.

Rossignol, Jacqueline
1992 Concepts, Methods, and Theory Building: A Landscape Approach. In *Space, Time, and Archaeological Landscapes*, edited by Jacqueline Rossignol and LuAnn Wandsnider, pp. 3–19. Plenum Press, New York.

Rossignol, Jacqueline, and LuAnn Wandsnider (editors)
1992 *Space, Time, and Archaeological Landscapes*. Plenum Press, New York.

Sale, M.
1999 *National Register Eligibility Assessment of Prehistoric Archaeological Sites in Maneuver Area 8, Fort Bliss, Texas.* Report No. 506. Lone Mountain Archaeological Services, El Paso, Texas.

Savinskii, Igor D.
1965 *Probability Tables for Locating Elliptical Underground Masses with a Rectangular Grid.* Consultants Bureau, New York.

Schiffer, Michael B.
1976 *Behavioral Archaeology*. Academic Press, New York.

1987 *Formation Processes of the Archaeological Record.* University of New Mexico Press, Albuquerque.

Schiffer, Michael B., Alan P. Sullivan, and Timothy C. Klinger
1978 The Design of Archaeological Surveys. *World Archaeology* 10(1):1–28.

Schiffer, Michael B., and S. Wells
1982 Archaeological Surveys: Past and Future. In *Hohokam and Patayan: Prehistory of Southwestern Arizona,* edited by Randall H. McGuire and Michael B. Schiffer, pp. 345–383. Academic Press, New York.

Sebastian, Lynne, Jeffrey H. Altschul, Chris M. Rohe, Scott Thompson, and William E. Hayden
2005 *Adaptive Management & Planning Models for Cultural Resources in Oil & Gas Fields: New Mexico Pump III Project.* Submitted to the U.S. Department of Energy National Energy Technology Laboratory, Contract No. DE-FC26-02NT15445. Copies available from Gnomon, Carson City, Nevada, and the SRI Foundation, Rio Rancho, New Mexico.

Shennan, Stephen
1985 *Experiments in the Collection and Analysis of Archaeological Survey Data: The East Hampshire Survey*. Department of Prehistory and Archaeology, University of Sheffield, Sheffield, Great Britain.

Shipman, Pat
1981 *Life History of a Fossil: An Introduction to Taphonomy and Paleoecology.* Harvard University Press, Cambridge, Massachusetts.

Shott, Michael J.
1985 Shovel-Test Sampling as a Site Discovery Technique: A Case Study from Michigan. *Journal of Field Archaeology* 12(4):457–468.

1989 On Tool-Class Use Lives and the Formation of Archaeological Assemblages. *American Antiquity* 54:9–30.

1992 Commerce or Service: Models of Practice in Archaeology. In *Quandaries and Quests: Visions of Archaeology's Future*, edited by LuAnn Wandsnider, pp. 9–24. Occasional Paper No. 20. Center for Archaeological Investigations, Southern Illinois University, Carbondale.

Stratford, Kate
2008a HAFB/UTTR Archeological Site Data: Explanation of Data Updates for SRI Legacy Data Quality Project. Draft document, Hill AFB Environmental Management GIS, Ogden, Utah.

2008b HAFB/UTTR Cultural Resource Data: Site Datums and Site Areas Source Explanation. Draft document, Hill AFB Environmental Management GIS, Ogden, Utah.

Tagg, Martyn D., and Michael P. Heilen
2008 *STAC 2003: Intensive Archaeological Survey of 4,945 Acres on the South Tactical Range, Barry M. Goldwater Range East, Arizona*. Barry M. Goldwater Range East Cultural Resource Management Program, Cultural Resource Studies in the Western Papaguería 16. Statistical Research, Tucson, Arizona.

Tainter, J. A.
1983 Settlement Behavior and the Archaeological Record: Concepts for the Definition of "Archaeological Site." *Contract Abstracts and CRM Archaeology* 3(2):130–133.

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. University of Arizona Press, Tucson.

Thomas, Prentice M., Jr., and L. Janice Campbell (editors)
1993 *Eglin Air Force Base Historic Preservation Plan: Technical Synthesis of Cultural Resources Investigations at Eglin, Santa Rosa, Okaloosa and Walton Counties, Florida.* Report of Investigations No. 192. New World Research, Mary Esther, Florida.

Trigger, Bruce
1984 Alternative Archaeologies: Nationalist, Colonialist, Imperialist. *Man* 19(3):355–370.

1989 *A History of Archaeological Thought*. Cambridge University Press, New York.

U. S. Army
2008 *Integrated Cultural Resources Management Plan (ICRMP), Fort Bliss, Texas, 2008–2012*. Directorate of Public Works, Environmental Division, Fort Bliss, Texas.

Vanderpot, Rein, Jeffrey H. Altschul, Michael K. Lerch, and Michael P. Heilen
2008 Nonriverine Native-Plant Exploitation in the Sonoran Desert: A View from the Western Papaguería. In *Fragile Patterns: The Archaeology of the Western Papaguería*, edited by Jeffrey H. Altschul and Adrianne G. Rankin, pp. 179–213. SRI Press, Tucson, Arizona.

Vierra, Bradley J., P. Lukowski, G. Smith, S. Sitton, B. Boeke, V. Provencio, M. Sechrist, A. Sullivan, J. Johnson, S. Wilcox, and M. Yduarte
1999 *Assessment of 431 Archeological Sites in Doña Ana Range Firing Groups C, D, G, H, I, and J, Fort Bliss, Texas and New Mexico*. TRC Mariah Associates, El Paso, Texas.

Wandsnider, LuAnn, and E. Camilli
1992 The Character of Surface Archaeological Deposits and Its Influence in Survey Accuracy. *Journal of Field Archaeology* 19:169–188.

Wandsnider, LuAnn, and Christopher Dore
1995 Creating Cultural Resource Data Layers: Experiences from the Nebraska Cultural Resources GIS Project. Part II: Combining GPS and GIS Data. *SAA Bulletin* 13(5):16–18.

Whalen, M. E.
1978 *Settlement Patterns of the Western Hueco Bolson*. Publications in Anthropology No. 6. El Paso Centennial Museum, University of Texas, El Paso.

White, Leslie A.
1959 *The Evolution of Culture; The Development of Civilization to the Fall of Rome*. McGraw-Hill, New York.

Willey, Gordon R., and Philip Phillips
1958 *Method and Theory in American Archaeology*. University of Chicago Press, Chicago.

Young, D. Craig
2008 *The Archaeology of Shifting Environments in the Great Salt Lake Desert: A Geoarchaeological Sensitivity Model and Relative Chronology for the Cultural Resources of the US Air Force Utah Test and Training Range*. Far Western Anthropological Research Group, Davis, California.

Zeidler, James A.
1995 *Main Text*. Archaeological Inventory Survey Standards and Cost-Estimation Guidelines for the Department of Defense, vol. 1. Special Report 96/40, CRRC Series. U. S. Army Construction Engineering Research Laboratories (USACERL), Champaign, Illinois.

2001 *Dynamic Modeling of Landscape Evolution and Archaeological Site Distributions: A Three-Dimensional Approach* (SERDP Project CS-1130). Technical Publication Series 01-8. Center for Environmental Management of Military Lands, Colorado State University, Fort Collins, Colorado.