Predictive Modeling in the Military
Similar Goals, Divergent Paths

by Jeffrey H. Altschul, Lynne Sebastian, and Kurt Heidelberg

Prepared through a grant from the Legacy Program to Air Force Materiel Command,
Martyn D. Tagg, Investigator

Submitted to:
Headquarters Air Force Materiel Command
4225 Logistics Avenue Room A-128
Wright-Patterson AFB, Ohio 45433

U.S. Army Corps of Engineers, Fort Worth District
819 Taylor Street, P.O. Box 17800
Fort Worth, Texas 76102-1012

Prewitt & Associates, Inc.
7701 N. Lamar, Suite 104
Austin, Texas 78752-1012

Preservation Research Series 1

A Legacy Project

SRI Foundation
Advancing historic preservation through education, training, and research

Rio Rancho, New Mexico

2004

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Introduction

Archaeology, as a discipline, addresses basic questions of time and space: How did things change or stay the same through time? How were things different or the same within a geographical area? Archaeologists examine the material remains left behind by human societies and attempt to identify temporal and spatial patterning. Based on observations of historical and contemporary societies and on theoretical constructs about how societies of different types operate, archaeologists assign meaning to those patterns and combine them into descriptions and explanations of life in the past.

From its beginnings as an antiquarian pursuit, through the first 100 years or so as a scholarly discipline, American archaeology was largely descriptive and historical. Beginning in the 1960s, however, archaeologists began adopting an explicitly "scientific" approach to the study of the past—hypothesis testing, statistical sampling and analysis, and an emphasis on explanation. Given the importance of patterns of distribution, association, and relationship to the field of archaeology, it is not surprising that archaeologists during this period became interested in the possibilities of quantitative models as a pattern recognition tool.

Because most of the theoretical perspectives that were in vogue during the 1960s and 1970s were ecologically based, most of the models that were developed focused on identifying correlations between quantifiable environmental variables and the location of archaeological sites. Depending on the predilections of the modeler, these correlations were then used either directly to "explain" why the sites were located where they were or indirectly to test theoretically-based hypotheses about what factors might have conditioned the locations of human activities in the past. The potential for using these models to predict where archaeological sites should be located was also clear to researchers, who used predictive success as a means of testing and refining their models.

Predictive Modeling and Archaeology

Predictive models are tools by which trends are characterized and anticipated. The principles of predictive modeling extend back almost as far as the use of mathematics, though their application was somewhat less formalized for activities such as gambling, military tactics, and hunting. In the context of archaeology, a predictive model is a "simplified set of testable hypotheses, based either on behavioral assumptions or on empirical correlations, which at a minimum attempts to predict the loci of past human activities resulting in the deposition of artifacts or alteration of the landscape" (Kohler 1988:33).

Most archaeological predictive models attempt to answer one basic question: "Where are sites located?" Whether they recognize it or not, all archaeologists have developed "predictive models" that answer this question. Most of these models are intuitive statements, such as "sites are found on flat land near water," "rock shelters facing east contain cultural material," or "elevated knolls in a swamp are midden sites." Many of our intuitive models propose correlations between site locations and one or more environmental features. Unless we provide explicit measures of these variables, however, we can never know how good our models are.

Archaeological predictive modeling has its roots in the settlement pattern studies of the 1950s and 1960s. Perhaps the most influential of these studies was Willey's (1953) classic analysis of settlement

patterns in the Virù Valley of Peru. Most of these studies were descriptive; surveys were completed and patterns of correlation were sought between settlement locations and environmental features, leading to ad hoc and largely untested conclusions about cultural evolution.

Quantitative methods emerged as an important component of archaeological studies in the 1960s. Issues of sampling, economic models, and spatial analysis were keenly discussed. One outgrowth of particular relevance to predictive modeling was the Southwest Anthropological Research Group, or SARG. The SARG organization was dedicated to investigating why archaeological sites are located where they are (Plog and Hill 1977). One of the lasting legacies of SARG is a strong interest among southwestern archaeologist in developing objective measures of environmental variables.

Green (1973) advanced predictive modeling in a study of prehistoric Mayan sites in northern British Honduras. Previously, archaeologists had related site location to environmental variables on a bivariate basis. Green used multiple linear regressions to incorporate a series of environmental variables to predict the probability that a particular location would contain a site.

With Green's study laying the foundation, predictive modeling in cultural resource management began in earnest in the late 1970s. Many of these studies were funded by large land-managing agencies attempting to assess the potential effects of enormous development projects on archaeological sites. Full-scale survey was considered economically prohibitive, and predictive models, wherein a small fraction of the project area was surveyed with the results then generalized to the rest of the affected area, seemed to offer tremendous promise. Not surprisingly, the discipline was flooded with predictive models (see Thoms 1988). Because of the substantial variability in quality among these models, the Bureau of Land Management (BLM) sponsored development of a comprehensive text on predictive modeling that remains a seminal study of the subject (Judge and Sebastian 1988).

Almost as soon as the BLM text was published, the field was revolutionized by the advent of geographic information system (GIS) technology. A GIS is essentially a set of related databases with one major addition: the means to store, manipulate, and display geographically referenced data. These applications allow models to combine geographic data from different sources, some of which may have different projections, formats, and scales, such that individual "layers" can be viewed and analyzed in conjunction with one another. The added visual component provides a means of creating maps from which the archaeologist may develop hypotheses for later quantitative testing. By viewing assorted spatial components in conjunction with one another, a researcher may easily observe apparent covariation among environmental attributes and cultural behavior.

With the advent of GIS, the days of tediously measuring environmental variables by hand and then manually entering the data into databases were over. GIS-based models have become relatively common in the last decade. From the standpoint of research, the aggregation of data under one common structure provides a platform for regional analysis. All archaeological data can be observed in the context of environmental and spatial data, thereby allowing the investigation of questions regarding land use, trade, mobility, and subsistence. The standards inherent in the design of the system assure that the data structure remains constant. In this way, analyses are cumulative, regardless of current research questions, researchers, or biases. In addition, the intuitive, graphical platform of GIS provides an interface through which ideas may easily be communicated with nonspecialists.

One of the main problems that has emerged with GIS-based models is, in fact, the ease with which they can be developed. The difficult theoretical and methodological issues discussed in the Judge and Sebastian (1988) volume—e.g., regional sampling, measurement of environmental variables, and statistical applications—and those discussed below are not resolved simply by using GIS technology. Although innovations continue (see Kvamme 1995, 1999), many GIS-based models are flawed or at best provide simple visual images with no analytical power.

Theoretical and Methodological Issues in Archaeological Predictive Modeling

Issues considered in the development of predictive models are similar to those inherent in all archaeological research. For one thing, modelers must adopt either a deductive or an inductive approach. Models based on theoretical propositions are more compelling, for they not only predict the location of prehistoric sites, they also explain why sites are located where they are. Theoretical models, however, tend to have limited predictive power. Human behaviors are complex, and behaviors related to the placement of activities are difficult to intuit. Consequently, theoretical variables related to settlement are hard to define and measure, which means that many theoretical models yield “soft” or “fuzzy” statements, such as “base camps will be found at the junction of ecotones.” Such predictions make good intuitive sense, but do not lend themselves to precise locational predictions. Even quantitatively based theoretical models, such as optimal foraging models, generally result in predictions covering rather large environmental strata (e.g., the pinyon-juniper zone).

Some have argued that prediction without explanation is an empty endeavor. This opinion, however, masks some fundamental truths. As Binford (2001) points out, archaeological explanations that derive from anthropological propositions, such as ethnographic analogy, are by nature limiting. Such explanations can never advance beyond the founding propositions, and thus do not add new information about human behavior. In contrast, pattern recognition offers a means to compare classes of archaeological data so that relationships between classes of phenomena can be discerned. Patterns can prompt questions about why sites are found in some locations and not others, questions that can only be explained by new theories.

Predictive modeling in archaeology has been greatly enhanced by GIS technology. Archaeologists can now "collect" regionally specific data on vegetation, soils, elevation, slope, aspect, hydrology, and climate against which to examine proxy measures of human behavior regarding settlement and land use. Because these environmental data come from a single source, rather than being collected piecemeal as a result of separate projects using different methods or definitions, these "layers" offer a less biased metric representation of the variables.

There is a constant trade-off between accuracy and precision with predictive models. Accuracy refers to the success of the prediction—the number or proportion of “hits” as opposed to “misses.” Precision refers to the statistical confidence we have in the predictions. To illustrate these concepts, imagine a predictive model defining three "sensitivity" areas, high, medium and low, reflecting the likelihood of encountering archaeological sites in each area. If 85 of 100 sites fall within the medium and high sensitivity areas, the model may be said to have an accuracy of 85 percent. On the surface, this may appear to be a good model. If the medium and high sensitivity areas constitute 85 percent of the study area, however, then the only statement that can be made is that 85 percent of the sites fall within 85 percent of the study area—a model as effective as a dart throw. Clearly, the goal is to hone the precision of the model such that the medium and high sensitivity areas represent considerably less than 85 percent of the entire study area, while maintaining the same level of predictive accuracy.

Many variables can be used in building a predictive model. Examples of environmental variables include elevation, slope, aspect, vegetation, and proximity to water. Each environmental variable is measured on an interval, ordinal, or categorical scale. Traditionally, archaeologists categorized the scores on each variable into sensitivity classes that represented relative probabilities of finding sites. For example, assume that in a specified area, most archaeological sites are found on slopes that range between 0 and 30°; these areas could be defined as “high sensitivity.” Similarly, we may find that a higher than expected number of sites are found within 500 m of a water source. Accordingly, these areas would be defined as high sensitivity for the variable “proximity to water.” By overlaying the variables, we create a mosaic in which areas are classified as high, medium, or low sensitivity, based on the combined sensitivity scores for all or most variables. This process is called the *intersection method* (BRW 1996).

The main problem with the intersection method is that it treats all environmental variables equally. On theoretical and empirical grounds, however, there is good reason to believe that some environmental variables are more strongly correlated with human settlement behavior than others are. For example, for

hunter-gatherers of the Great Basin making decisions about where to carry out activities, proximity to water was probably more important than aspect. A solution to this problem is to weight the values. Each environmental variable is categorized, and graduated values are assigned. For example, proximity to water may be divided into four categories: 0–160 m, 161–430 m, 431–800 m, and >800 m, with weighted values of 4, 3, 2, and 1 being assigned to these categories, respectively. For slope, 0–5° may be given a value of 3, 6–15° a value of 2, and so on, in the same incremental fashion. Each variable can then be assigned a weight (mathematically, a factor) by which it is multiplied.

The weighted value method is an improvement over the intersection method because it allows variables to have differing impacts on the model. It introduces a new problem, however, by restricting the model to a structure where weights are scaled by factors. This imposes linear covariability among the subcategories of each variable as they are scaled, which may not be a desirable constraint as the model is being refined. For example, if the slope variable described above were to "weigh" more than other variables, one could simply multiply all of the slope values by 2 or some other factor. It may be, however, that when slopes fall within a certain range, say, between 6–15°, then slope is less of a determining factor in the overall scheme. The weighted value, then, is over-representative in some cases. The two modeling systems can be combined to take advantage of the simplicity of intersection and the sophistication of weighted values; the end product, however, becomes very complex and beyond the bounds of intuitive understanding.

The problem with using intersection, whether weighted or not, is that *all* values for *all* variables come into play as the variable themes overlap. This results in a complexity that increases at a geometric rate, regardless of the benefit (or lack thereof) to the end product. More sophisticated mathematical modeling techniques can break down this complexity. Many of these techniques follow a general linear regression model and include, among others, multilinear regression, discriminant function analysis, principal components analysis, and logistic regression (see Rose and Altschul 1988).

All multivariate regression techniques follow similar approaches. The covariation among the independent variables (in the case of archaeological predictive models, these are the environmental variables) is evaluated. Only those environmental variables that independently explain sufficient variability in the dependent variable (that is, site location) enter into the regression analysis. The analysis results in one or more equations that are used to calculate the probability that a case (usually a cell of specified area) will contain an archaeological site. The cells are then placed in a two-dimensional grid with the resulting probability scores being used to create a three-dimensional isopleth map.

Regression models have become extremely popular among archaeological predictive modelers. Many of these models are quite powerful, combining precision and accuracy. For practical applications, however, these models tend to be ignored in favor of the much less accurate intersection models. Although this trend appears counterintuitive, it is simple to explain. Multivariate regression models are statistically complex. For people without considerable training in statistics, these models are intimidating and not "user friendly." Intersection models, on the other hand, though not elegant or powerful, are easy to understand; more important they make intuitive "sense." The best archaeological models for practical applications are those that combine the intuitive nature of intersection models with the power of multivariate statistics.

Predictive Modeling and Cultural Resource Management

As noted above, although archaeologists were excited about the research and explanatory potential of modeling, it was the predictive potential of archaeological models that caught the attention of federal land managers. With the passage of the National Historic Preservation Act (NHPA) in 1966 and the National Environmental Policy Act (NEPA) in 1969, federal agencies were required to consider impacts to historic and cultural resources in their planning for development projects. The compliance process for Section 106 of NHPA required that agencies identify historic properties, including archaeological sites, that could be affected by undertakings that they funded or approved.

Some cultural resource staff and agency managers, faced with the substantial costs of large, intensive archaeological surveys in anticipation of major energy develop projects or broad-scale military training exercises, seized upon predictive models as a cost-effective alternative. The logic underlying this line of thought was that the agency would spend money up front to create an objective and verifiable model whose predictions would then substitute for large-scale survey. This view represented a fundamental misconception about what predictive models could and could not do, and in some cases, a fundamental misunderstanding of the requirements of good resource stewardship under NHPA.

The archaeological community and, in some cases the Section 106 reviewing agencies, spoke out against the misuse of predictive models as a means of "identifying" archaeological sites that would be affected by federal undertakings. The critics of this approach pointed out that the result of a predictive model was a set of probabilities for the likelihood that an archaeological site would or would not be found at a particular location. What the Section 106 process requires is that the agency make a reasonable and good faith effort to determine what effects its undertaking will have on actual properties listed on or eligible for listing on the National Register of Historic Places—something that cannot be discovered from a sensitivity map or other representation of the results of a modeling effort

Not only does the statistical or probability-based nature of the modeling *results* lead to problems in using modeling to substitute for survey-based identification, the very nature of the modeling *process* also argues against using predictive models in this way. For one thing, as Judge and Martin (1988:580) point out, "modeling is a cyclical process of ongoing refinement, rather than a one-time event, and thus models cannot be developed by outsiders and then simply 'turned over' to agency field office archaeologists for 'application.'"

Another frequently raised issue was the fact that models were being developed for artificially bounded areas—the area to be affected by a specific federal undertaking, for example, or the area within the boundaries of particular federal installation. As Kincaid (1988:552) notes:

> Observed site patterning in the study area may be responding to factors that are "uncontrolled" in the terms of the model because they are a response to forces or events located outside the study area. The chances of developing an accurate and interpretable model are greatly reduced by this circumstance.

Additionally, American Indians often ascribe special values to archaeological sites, values that are encoded not in the natural environment but in esoteric knowledge and oral traditions. Because archaeological predictive models are generally based on environmental variables, sites with traditional cultural values could be underrepresented or overlooked entirely by the model.

In reaction to these and other problems, full survey coverage of specific project areas became the standard tool for inventory (e.g., Fish and Kowalewski 1990). This conservative approach appeased preservation-minded archaeologists and the Section 106 reviewing agencies, but failed to provide a greater understanding of the past, primarily because no mechanism emerged to synthesize the various projects. Moreover, as Judge (1981) pointed out, the conclusion that full survey coverage finds all archaeological sites in a project area is naïve and wrong. No data have ever been produced to counter Judge's (1981:128) dictum that, "the more time spent in the field looking for sites, the more sites will be found."

Because predictive models were not a panacea that could be substituted for high-cost intensive surveys, many land-managers abandoned the whole concept of archaeological models. A few agencies, however, continued to see value in them. The agencies that did continue to develop and use predictive models, many of them military installations, realized that even though modeling could not be substituted successfully for archaeological survey, there are valid and important uses for models in the Section 106 process and in federal land management in general. Indeed, as Kincaid (1988:554) has pointed out,

> Perhaps the most cost-effective context for model development is within the framework of general planning by a land-managing agency or a local government. These programs can develop and sustain long-term approaches that are funded incrementally and result in cumulative and refined

data bases. Such databases, and the models based on them, may take years to develop and test. The end result, however is a powerful and effective management tool.

Among the most effective and appropriate uses of predictive modeling in the Section 106 process are those related to project planning and project design. If undertakings are designed from the beginning to avoid making impacts on critical areas for cultural resources, the cost-savings from avoided re-designs and mitigation can be substantial. Once the design of an undertaking is complete, the predictive model can also provide an excellent mechanism for structuring the Section 106 compliance survey. Use of sensitivity maps can allow managers to target proportionately more of their survey dollars in those areas most likely to contain the greatest number of the sites, while surveying areas less likely to contain sites at a lower intensity.

In addition to these Section 106-related uses, however, agencies that see the value in predictive models realize that there is more to cultural resource management than Section 106. Section 110 of the NHPA, for example, requires that federal agencies establish a program to identify and evaluate historic properties under their jurisdiction and that they manage and maintain National Register eligible properties under their jurisdiction in a way that preserves their historic qualities. The information from a good predictive model enables land managers to meet their Section 110 obligations by making informed decisions about land uses and by making the best use of scarce dollars available for resource identification, stabilization, and protection.

The National Environmental Policy Act (NEPA) requires that federal agencies evaluate their actions in terms of their impact on the "human environment," including cultural and historic resources. All federal projects require some level of NEPA assessment; those that have the potential to have or are known to have a major impact require formal evaluation of effects and alternatives through the Environmental Assess or Environmental Impact Statement processes. The purpose of these NEPA assessments is to enable federal agency managers to make informed decisions that consider, in an even-handed way, the environmental, cultural, and social consequences of alternative agency actions.

The NEPA process does not require that agencies identify all the specific resources that will suffer impacts or that they have comprehensive, detailed data about all the alternatives that are being considered. What it does require is that agencies have and/or gather sufficient data to evaluate the alternatives effectively and that those data be gathered and evaluated in a scientifically valid process. Predictive modeling, perhaps combined with minimal ground truthing, is an excellent approach to NEPA compliance for archaeology. A military installation or other land-managing agency with a well-designed, thoroughly tested and refined predictive model is in a good position to evaluate a wide variety of alternatives and assess their relative impacts in an impartial, scientifically sound fashion.

After a decade or more of being somewhat out of fashion among land-managing agencies, predictive models are making a comeback. Agencies with existing models are looking at updating and improving them; agencies that let their models languish on a shelf after learning that they could not simply substitute models for survey are rethinking that decision.

There are several reasons for this resurgence. For one thing, after more than a quarter of a century of cultural resource management work, we are still struggling with the issue of synthesis. Section 106, which is the driving force behind most of the work that has been done, is totally case-specific. We have no mechanism for taking what we learn from each individual survey or excavation project, combining that information with all the other surveys and excavations in the area, and using the synthesized data to inform future management decisions.

There are areas where literally hundreds of surveys have been completed; yet we know very little more about the nature and meaning of the archaeological record than we did in the beginning. Archaeological modeling provides us with a framework in which to combine all of those existing data and examine them for patterns of correlation or evaluate them against theory-based predictions. Archaeological models offer us the opportunity to interpret and explain human behavior in the past and to develop more sophisticated research and management strategies for the future.

One of the most critical needs in cultural resource management today is to develop better mechanisms for assessing archaeological significance. Decisions must be made every day about the potential of archaeological sites to "yield important information" about the past; this is one of the criteria of the National Register of Historic Places that determines whether sites will be afforded consideration under Section 106. Archaeological modeling can assist with this process by providing a mechanism for synthesizing existing data and characterizing what is common and what is rare in the archaeological record, but even more important, archaeological models can help us identify the things that require explanation.

A great deal of emphasis is place on the predictive accuracy of models, but this tends to obscure the fact that predictive power does not equate one-to-one with archaeological importance. For example, it is conceivable that, based on a cursory review of a topographic map, archaeologists could give fair, best-guess estimates of where archaeological sites are likely to occur with an accuracy rate of more than 50 percent. Accuracy of greater than 80 percent is not out of the question. These high success rates may lead one to question the value of formal predictive models that are intended to better that accuracy by something less than the remaining 20 percent.

In all likelihood, however, it is the remaining 20 percent of sites that have the greatest potential to teach us important things about the past. These "red flags" (Altschul 1989, 1990) are sites that do not fit expected distributional patterns; they represent aspects of human behavior that are outside the range of our predictive power. Most importantly for managers, these sites, by definition, are likely to yield important information on settlement behavior and, thus, are probably eligible for listing in the National Register of Historic Places. This is one of the important reasons why modeling cannot simply be substituted for inventory.

Predictive models will, by definition, not identify these sites, and archaeological inventory alone will not indicate their significance. A combination of modeling to identify the common, environmentally predictable patterns and inventory to identify the exceptions, however, can provide us with a powerful means to identify those sites that should be the focus of our research, mitigation, and site protection efforts.

A third reason why predictive modeling is receiving renewed interest is the current government-wide emphasis on environmental streamlining. One of the most effective ways to streamline environmental review is to move decision-making to a point earlier in the planning process. By a happy coincidence, this is also the most effective thing that can be done to increase protection and preservation for archaeological sites and other kinds of cultural resources; the largest number of avoidance and mitigation alternatives are available at the earliest stages of planning.

In order to make effective, well-informed cultural resource management decisions at the planning stage, long before Section 106-driven resource identification takes place, however, land-managers need a mechanism for synthesizing, manipulating, and evaluating all initially available data in a scientifically sound fashion. Archaeological modeling, combined with limited identification and testing, has great potential to contribute to environmentally responsible streamlining efforts.

CHAPTER TWO

Project History

Background

In June of 2000, one of the authors (Altschul) attended a workshop sponsored by the Legacy Program and the Strategic Environmental Research Development Program (SERDP) in Lexington, Maryland, titled "Archaeology in the Department of Defense (DoD)." The main goal of the workshop was to identify topics critical to DoD's management of cultural resources that could benefit from research and development funds. Many of the topics discussed permeated American archaeology at the turn of the century: how to find sites through technology, addressing the crisis in curation, and developing procedures for government-to-government consultation with American Indians, to name a few. Throughout the workshop, managers and archaeologists continually returned to the fundamental issue of identification. How does an installation know when it has found enough sites and surveyed enough acres?

For those who had been involved in cultural resource management for several decades, these questions had a familiar ring. We had heard them before, and many at the workshop had tried to answer them. As described in the previous section, in the 1980s and 1990s many installations had created predictive models of site location. These models, however, tended to be short lived. Interest waned when it became clear that the installation would still have to complete pedestrian surveys for most undertakings. The reports that produced these models were put on the shelf and largely ignored.

Many of the participants representing military installations in the Legacy/SERDP workshop had heard of predictive models but were not using them. No clear reason was given, although there was an underlying impression that predictive models had been tried and failed.

But had they failed? No one really knew. Although inventory had continued unabated, in some cases for more than two decades after model development, none of the installations had used these data to evaluate whether the models on which they had spent so much money actually worked. Fred Limp of the University of Arkansas urged Martyn Tagg of the Air Force Materiel Command and Jeff Altschul of Statistical Research, Inc., and the SRI Foundation to conduct such a study. After meeting with Legacy Program staff, Tagg and Altschul prepared a grant proposal the next week.

Their proposal outlined a project with the objective of answering four specific questions:

- Do predictive models that have been created for military installations work?
- Can they be refined to work better?
- Are they sufficiently accurate so that land managers and State Historic Preservation Officers can use them in evaluating management decisions about installation resources?
- Can a predictive model be integrated into a more dynamic operational model that would be useful across the DoD to increase cost efficiency of cultural resource management at large installations?

The proposal defined a three-step process for answering these questions. The first step was to determine the pervasiveness of predictive modeling in the military through a questionnaire sent to installations representing all branches of the service. Although not intended to be a complete canvassing of the military use of predictive models, the objective of the questionnaire was to achieve a reasonable sample from

which inferences could be drawn. The second step was to choose models from four of the responding installations for an in-depth evaluation of their technical quality, accuracy, and general utility as a management tool. This evaluation was to include suggestions about ways to improve the utility of each model. The proposed third step was take the results of the evaluation and design a follow-up study that could directly contribute to more effective use of predictive modeling by military installations.

Below we describe the questionnaire and the selection process. The next section of this report provides an evaluation of each of the four selected predictive models. In the final section of the report, we suggest what we would propose as a useful follow-up study.

Participant Selection

Twenty-three military installations were sent a questionnaire (see Appendix to this report) that inquired about the nature and use of predictive models in their cultural resource management program. The installations were chosen for geographical diversity, size, and known use of predictive models. We also attempted to cover all the services; of the 23 installations, 7 were Air Force facilities, 12 were Army reservations, and 2 each represented the Marine Corps and Navy (Table 1).

Table 1. Military installations contacted and their responses

Service	Installation	Existing model?	Date developed	Still in use?
Air Force	Avon Park Air Force Range, FL	yes	1985	yes
Air Force	Barry M. Goldwater Range, AZ	yes	1995	no
Air Force	Dover AFB, DE	no response	–	–
Air Force	Edwards AFB, CA	yes	1995	yes
Air Force	Eglin AFB, FL	yes	1982	yes
Air Force	Patrick AFB, FL	no response	–	–
Air Force	Vandenburg AFB, CA	no	–	–
Army	Fort Benning, GA	no response	–	–
Army	Fort Bliss, TX/NM	yes	1995	yes
Army	Fort Carson, CO	yes	1985	yes
Army	Fort Drum, NY	yes	1998	yes
Army	Fort Hood, TX	yes	1989	yes
Army	Fort Huachuca, AZ	yes	1990	yes
Army	Fort Irwin, CA	yes	2002	yes
Army	Fort Lewis, WA	no	–	–
Army	Fort Stewart, GA	yes	1985	yes
Army	Fort Leonard Wood, MO	no response	–	–
Army	White Sands Missile Range, NM	yes	2001	yes
Army	Yakima Training Center, WA	no response	–	–
Navy	China Lake Naval Air Weapons Station, CA	yes	1996	no
Navy	Fallon Naval Air Station, NV	no response	–	–
USMC	29 Palms Marine Corps Air-Ground Combat Center, CA	no response	–	–
USMC	Camp Pendleton, CA	no	–	–

The questionnaire asked 10 questions that focused on when the installation's predictive model was created, who created it, how many sites were used in the development of the model, whether it was still in use, what platform it used, and how many sites had been found subsequent to the model's creation. Thirteen installations responded that they had developed predictive models, and that they were willing to participate in the study. Seven installations did not respond to the questionnaire and only three replied that they had not created a predictive model.

The 13 positive responses provided insight into the history and use of predictive models by the military. Contrary to our assumption that predictive modeling had waned after the initial enthusiasm of the 1980s, the responses indicate that predictive modeling has continued at a relatively steady rate for the past two decades. Five of the predictive models were created in the 1980, six were completed in the 1990s, and two were developed within the past two years. With the exception of two created in the mid 1990s, all models are still in use.

For the most part, models were created by outside contractors. Among the exceptions are the predictive models for Fort Drum and White Sands Missile Range, both of which were created by base personnel within the last five years. Two other recent models, those for Fort Bliss and Fort Irwin, were created by the Army's environmental research laboratory (ERDC-CERL). These four models may indicate a trend toward the use of in-house expertise, particularly in the area of GIS. All of these models use Earth Scientific Research Institute (ESRI) software, particularly ArcView. ESRI software is ideal for creating intersection models (see previous section), but less useful for creating quantitative analytical models. ESRI software has the advantage of being widely available and relatively user friendly. This trend has both positive and negative implications. On the positive side, predictive models are being used by installation archaeologists and managers precisely because they are intuitively based, easy to create, and on computer platforms that are accessible. The downside is that the models created today are less powerful than those developed 20 years ago.

The move toward ESRI products is not limited to recently created models. Of the 13 installations, 8 use ESRI software for their predictive model platforms. Included in this count are several models developed in the 1980s that have been migrated to current versions of ArcInfo, ArcView, or ArcGIS. Those models that have not been migrated remain static; several exist only as paper maps. These latter models have lost all ability to be upgraded or to serve as dynamic management tools; their utility, therefore, is continually decreasing. The models vary significantly in the size of the region they cover, the number of sites used to construct them, and the number of sites found subsequent to their creation. Some models, such as the one for Eglin Air Force Base, were created early in the development of the installation's cultural resources management program with the intent of using the model to guide subsequent inventory and management decisions. Others, such as the White Sands Missile Range model, were created after decades of inventory, with the objective being both to manage an enormous data set and to make analytical and predictive statements about site location.

Given the wide range of models, selecting four installations for in-depth analysis was not an easy task. We first developed six criteria for selection. These were: the age of the model, the amount of inventory subsequent to model development, availability of documentation, the size of the datasets, participant installation's interest in being part of the study, and installation staff's familiarity with the structure of the data set.

Age of the Model: To maximize variability among the models, we sought a range of model ages. Older models have a longer experience record, which makes it easier to evaluate their management and analytical utility. The trade-off for age is that the older the model, generally the more questionable the integrity and availability of the data.

Availability of New Data: A fundamental task in this project was to determine whether models "held up" when tested using data collected subsequent to model development. For this evaluation it was necessary that cultural and environmental data not used in model creation be available.

Availability of Documentation: Models were often developed by individuals who are no longer available to provide explanations of model heuristics, or who simply cannot remember the details of model development. Consequently, availability of good documentation was an important criterion.

Size of Dataset: Larger datasets generally allow for more comprehensive models. We focused on larger installations with small, but adequate, survey and on smaller installations with a high percentage of survey.

Willingness of Participants: Clearly, those installations without models and those that did not respond to the questionnaire were eliminated from further consideration. In some cases, installations initially expressed a desire to participate but subsequently were reluctant or unable to provide data for evaluation. These latter installations were also removed from the pool of potential candidates.

Control over the Data: To ensure timeliness in the evaluation, it was imperative that installations be able to deliver data in suitable formats on schedule. Although perhaps obvious, this criterion was enlightening, for several installations with models and digital data could produce neither in a timely fashion.

Based on these criteria, we selected four installations for analysis: Eglin Air Force Base, Fort Bliss, Fort Drum, and Fort Stewart (Figure1; Table 2).

Figure 1. Locations of the four installations selected for study.

The models from the four selected installations varied widely in design and operation. Fort Drum's model is based on a visual, intuitive method that capitalizes on the tendency for sites to be located along ravines. An ArcView algorithm was used to model ravines, with the resulting topography being associated mechanically with areas likely to contain sites and those that should not. Given the environmental setting, Fort Drum recognizes that surface distribution of artifacts is a poor reflection of the location of archaeological sites. Model testing in this environment requires some form of subsurface survey, in this case, shovel tests.

Fort Stewart's model correlates site location with soils. Paper maps serve as the presentation medium, with ArcView software used solely to maintain site datums and survey blocks.

Eglin AFB's model uses proximity to water, elevation relative to water, and regional topography as the independent variables predicting site location. Association with prehistoric site location is based heavily on prior knowledge. Site boundaries and survey blocks are maintained on paper topographic maps and are in the process of being transferred to digital format using Geomedia Professional by Intergraph.

Table 2. Installations selected for study

Installation	Total acres on installation	Acres surveyed Prior to original model	Sites in original model	Acres surveyed since model development	Sites identified since model development	Platform of model
Fort Bliss, Texas/New Mexico (McGregor Range)	700,000	224,000	901	148,800	approx. 2,000	GRASS and ArcView
Fort Drum, New York	107,000	22,000	104	14,000	approx. 3,000	ArcView, ArcInfo, and Access
Eglin AFB, Florida	464,000	55,000	511	130,000	approx. 1,100	Paper maps and Geomedia Professional (in process)
Fort Stewart Georgia	279,000	35,582	409	135,000	approx. 1,600	Paper maps ArcView

Finally, the Fort Bliss model also uses multiple environmental variables to predict the location of archaeological sites. In this case, soils, aspect, landform, and elevation are the independent variables; they are derived from secondary themes in a GIS. The model was developed in GRASS 4.0, which is no longer maintained by the installation. The Fort Bliss model currently exists only in paper form.

All of the models generate “sensitivity maps,” where high and low sensitivity areas are defined. For some models, these regions were defined using statistical thresholds; for others, simpler approaches were taken. To evaluate the effectiveness of the models, SRI Foundation developed a standardized performance measure (see discussion in the section on Fort Stewart, below). We applied this measure to all models except that for Fort Drum. The predictions yielded by the Fort Drum model are not compatible with this performance measure.

Because of the substantial differences in how the models were designed and developed, certain assumptions within some of the models had to be changed to permit comparisons with the other models. For example, sites at Fort Stewart are designated as points in the installation’s GIS. To make our evaluation of the Fort Stewart model comparable with those for the other models, Fort Stewart sites were arbitrarily defined as including a 50-m buffer zone around each point. The Eglin AFB model posed a different problem; the base could not furnish all environmental data in digital form, so our evaluation of the model was based on the one available variable, proximity to water, and on a proxy measure of distance from the coast. In the following discussions, we explicitly identify the assumptions that had to be made in the course of evaluating each model.

CHAPTER THREE

Results of Model Evaluation

This section provides our assessments of the predictive models from Fort Drum, Fort Stewart, Fort Bliss, and Eglin AFB. The discussion proceeds from the simplest to the most complex of these predictive models. For each installation, we provide background information on the model, including its geographic range, data inputs, and the methods used to create the model. Next, we present the results of the evaluation, followed by a discussion of the model's strengths and weaknesses and our recommendations.

Fort Drum

The cultural resource management program at Fort Drum, which is located in upstate New York, was established in 1985. The installation covers more than 107,000 acres; approximately 36,000 acres have been inventoried for cultural resources. More than 3,100 sites have been recorded within the boundaries of the fort.

Fort Drum is representative of installations that have developed "informal" predictive models, that is, models whose assumptions are not explicitly stated and whose predictions take the form of intuitive statements of presence or absence of cultural material. The current sensitivity model replaced an earlier predictive model that utilized proximity to water, soil type, and slope. Archaeologists were concerned that the correlative model was not accurately representing the archaeological record. Because archaeological surveyors routinely bypassed ravines and ignored areas with substantial slope, installation personnel felt that the data were skewed, resulting in a poor sensitivity map. In 1998, the Fort Drum Cultural Resources Program embarked on a different course that combined intuitive knowledge of the archaeology with GIS technology.

Site Representation/Sampling Strategy

Given the poor surface visibility in upstate New York, archaeological survey at the installation takes the form of shovel tests spaced at 20-m intervals. The result is a map (Figure 2) that shows locations where archaeological materials have been recovered, but not site boundaries in the traditional sense. Prior to 1988, actual archaeological sites were defined, but those sites are plotted as points. It is not clear whether the plotted sites represent simply the presence of archaeological material or more extensive deposits.

In 1999–2000, Fort Drum completed a survey designed to test their model. Shovel test were placed in areas thought to be of the highest sensitivity and the lowest sensitivity. More than 50,000 shovel-tests were dug, yielding almost 3,000 positive cultural hits.

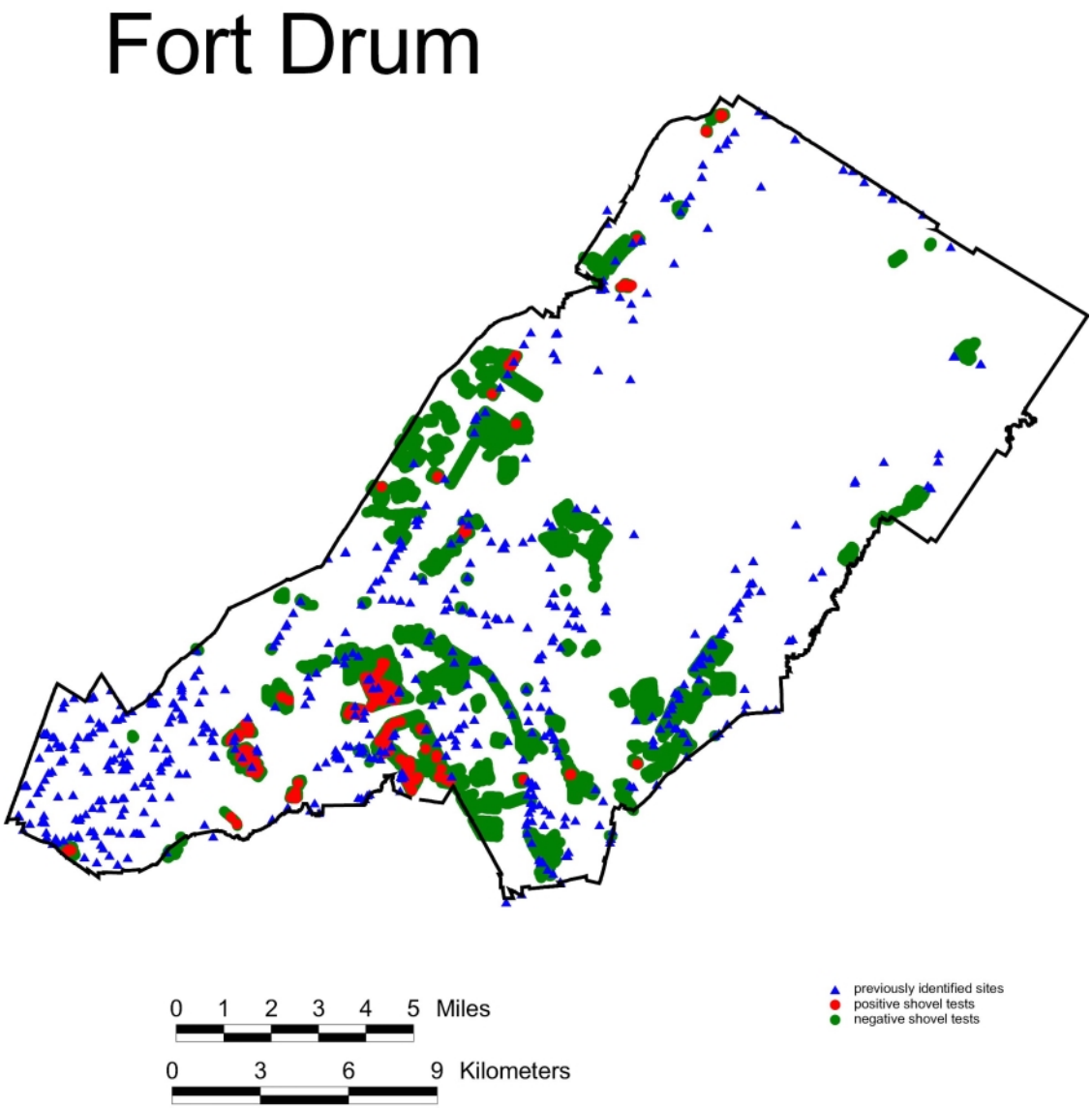

Figure 2. Shovel tests and sites at Fort Drum.

Prediction Method

The Fort Drum model is based primarily on a generated hillshade model. A sun angle and azimuth are entered into an algorithm that generates a visual "shadow and light" rendering of the topography. Shadow effects highlight ravines, along which archaeological sites tend to be located. Using approximately 120 known site locations, environmental zones were evaluated to determine whether they exhibited differential densities of sites. Sensitivity areas are not formally defined, but areas roughly within 150 m of the ravine edges were found to be higher than average in site density across the installation. In addition, ancient shorelines were modeled using elevation data, and additional high site density areas were found at

levels of 700–740 feet above current mean sea level. The lowest sensitivity areas were defined as comprising the ancient lakebeds and the northwestern plains area of the installation, where little water exists.

The Fort Drum predictive model is unusual in that it is not a visual map, but instead a set of mechanical statements linked through Boolean logic. High sensitivity zones are those areas that are within about 150 m of a ravine edge OR located between 700 and 740 feet above mean sea level. Low sensitivity zones are areas not located along ravines OR areas inside ancient lakebeds OR areas where little water exists. For any given location on the military reservation, the base archaeologist applies the logical statements above and assigns a sensitivity value. These logical statements are kept purposely fuzzy. Terms such as "within about 150 feet of a ravine edge" allow the base archaeologists to use their own judgment in defining the sensitivity of any particular location. Because variables are not explicitly defined so that the measurement is the same for every investigator, however, the model is not replicable.

Post Model Development

Subsequent to developing the model, the installation engaged in a survey designed specifically to test it. According the Fort Drum staff, the tests of the model were productive. They validated the developers' expectations: high sensitivity areas had higher site density, whereas the low sensitivity areas did not contain sites. The modelers also observed that additional environmental variables such as soils and proximity to water could play a valuable role in strengthening the model.

Evaluation Results

Although the installation staff is to be commended for having explicitly tested their model, it is important to point out that there are no statistics to support the staff's finding that the test supports the model. We do not know how many of the 3,000 hits or how many of the 50,000 shovel pits were in low sensitivity areas as opposed to high sensitivity areas. Because the model is based on intuitive statements that cannot be mapped in a replicable manner, we could not independently assign the 50,000 locations to high or low sensitivity zones, and thus, we are not able to mathematically evaluate the results of the survey or the efficacy of the model.

Strengths and Weaknesses of the Model

The strength of the Fort Drum model is that it is intuitive in nature and easy to understand. The model attempts to mimic implicit archaeological knowledge about the location of cultural resources. It creatively uses topographical features, which is rare in most other archaeological predictive models owing to the complexity involved in digitally representing landforms. Substantial effort has been spent to test the model. Whereas most models rely heavily on collections of old, disparate datasets, the Fort Drum model was created concurrently with the development of a survey strategy and recording scheme that produces consistent data.

The model's greatest weakness is its informality. There is simply no way of knowing how well the model works. Statements such as "more sites are located in the high sensitivity zone than in the low sensitivity zone" are the strongest prediction the model can make. Because the model is not based on a probabilistic approach, we can never interpret its predictions in any measurable way. For example, statements such as "this location has a 90 percent chance of having a site" have no basis in this type of model. The model exists simply as a means of combining visual inspection and the base archaeologists' experience and knowledge. As such, it tends to reify what we already know; that sites are located in certain types of settings. It does not, however, offer an approach for learning more about these sites or their distribution.

In addition, the model's reliance on a generated hillshade to serve as a proxy for topography is not a strong solution. Although hillshade algorithms are visually pleasing, they can often be misleading. Large shadows obscure much of the topography, and the shaded regions all appear to be the same, when most

often they are not. For example, a shadow cast by a mountain may look as dark as that cast by a ravine, but the two represent completely different topographical phenomena.

Recommendations

The goal at Fort Drum, as elsewhere, is to capitalize on prior information about archaeological site locations. Unfortunately, at Fort Drum this information is not formalized, but instead is held by a few people who have gained it through personal experience. The task at Fort Drum, as with many installations, is to transform this intuitive knowledge into formal predictive statements that are replicable and measurable.

As it stands, the Fort Drum Archaeological Predictive Model appears to accurately designate areas of high archaeological sensitivity. The problem is that we cannot assess the model's accuracy. The installation can use this model as a heuristic device, but without replicable measures, few agencies or interested parties are likely to accept the results as the basis for management decisions.

We have two suggestions to improve the model.

1. Use of more and better-defined independent variables

   Hillshade algorithms are one of many heuristics that use Digital Elevation Models (DEM) to render secondary themes. Other algorithms using DEMs can render slope, aspect, terrain roughness, topographical features, and transportation cost analyses. Availability of other kinds of spatially rendered environmental data varies greatly from region to region, but generally information on soils, geology, and hydrology can be found for most areas. Sometimes vegetation, faunal zones, geomorphology, and paleoenvironmental themes are available as well.

2. Formal models

   We recommend that Fort Drum continue to use the hillshade theme model but that, in addition, the installation create a second model based on replicable and measurable variables. Specifically, we suggest that, in the second model, hillshade be replaced with a topographic feature theme generated from the DEM, perhaps in conjunction with slope data. Discrete topological proximity zones should be defined and added to the model. At the same time, the predictive potential of some other DEM-derived themes, such as soils and hydrology, should be explored. Those that make a statistically significant contribution to the model's effectiveness should be added.

   The second predictive model should use at least the intersection method of modeling, or a general linear regression model, if possible. Indeed, once the data are formatted correctly, all types of modeling procedures could, and should, be tried. We suspect that at first the hillshade theme model will work better than the intersection or regression model, because it most closely mimics prior knowledge. As the second model is refined with more and better-defined variables, however, we would expect it to begin outperforming the hillshade theme model. By knowing how the predictions are made, base archaeologists can determine why the model works in some situations and not others and for some site types and not others. By examining cases of "wrong" predictions – sites found in low sensitivity zones or high sensitivity zones without sites – base personnel can refine the model by adding or subtracting variables or deriving new proxies of environmental conditions. Most importantly, by understanding why the model works, base archaeologists can also gain better insight to prehistoric human behavior, leading to ever more precise predictions, better defined statements of significance, and better management decisions.

Fort Stewart

Fort Stewart, Georgia, includes some 279,000 acres, approximately 170,000 of which have been surveyed for cultural resources. The cultural resource management program at the fort, which was established as an in-house program in 1995, maintains information on more than 2,000 archaeological sites. Fort Stewart is representative of installations that have built relatively simple statistical predictive models. The Fort Stewart model is based on the correlation between one environmental variable—soils—and one cultural variable—the location of archaeological sites. Although soils are statistically related to archaeological site locations, the relationship is weak. When the model was developed in 1985, the high sensitivity zone represented just over 25 percent of the surveyed areas, and included 38.75 percent of the archaeological sites. Subsequently, the staff at Fort Stewart has been working to refine the sensitivity zones and increase the predictive precision of the model.

Site Representation/Sampling Strategy

The model is based on 409 sites that had been recorded on 35,582 acres within the installation boundary prior to 1985. In the model, sites are represented by their datums, that is, as points. Sites are not classified into site types, but the model excludes historical period sites.

The model was never formally tested; no survey was designed specifically to ensure that all soil types were surveyed proportionally or that all environmental zones had been sampled. Even so, given the relatively large number of acres surveyed at the installation, and based on a visual inspection of the distribution of these surveys, we are confident that sufficient data exist to model site location accurately.

Prediction Method

The model used soil type, exclusively, and ranked the categorized soils with sensitivity values of high, low, and indeterminate. The modelers experimented with another independent variable – proximity to water – but had little success in achieving a strong correlation with site location. Most likely, this was a result not of the prehistoric inhabitants being indifferent to the location of potable water, but rather of archaeologists being unable to develop proxy measures that successfully capture this relationship.

The sensitivity values were calculated by comparing each soil's areal representation within the survey areas to the number of sites found within that soil type. An observed/expected (O/E) ratio was then calculated for each soil. All soils with an O/E value greater than 1.0 were considered high sensitivity, while all those falling below 1.0 were considered low sensitivity (Figure 3). In cases where survey results appeared to be anomalous, that is, where site density was considered to be too low and to misrepresent the true sensitivity of some regions, some low sensitivity soils were recoded as high sensitivity.

Sites and sensitivity zones were plotted on paper topographic maps. Although survey and site locations are now digitally recorded in an ArcView database, the model still exists only as paper copies.

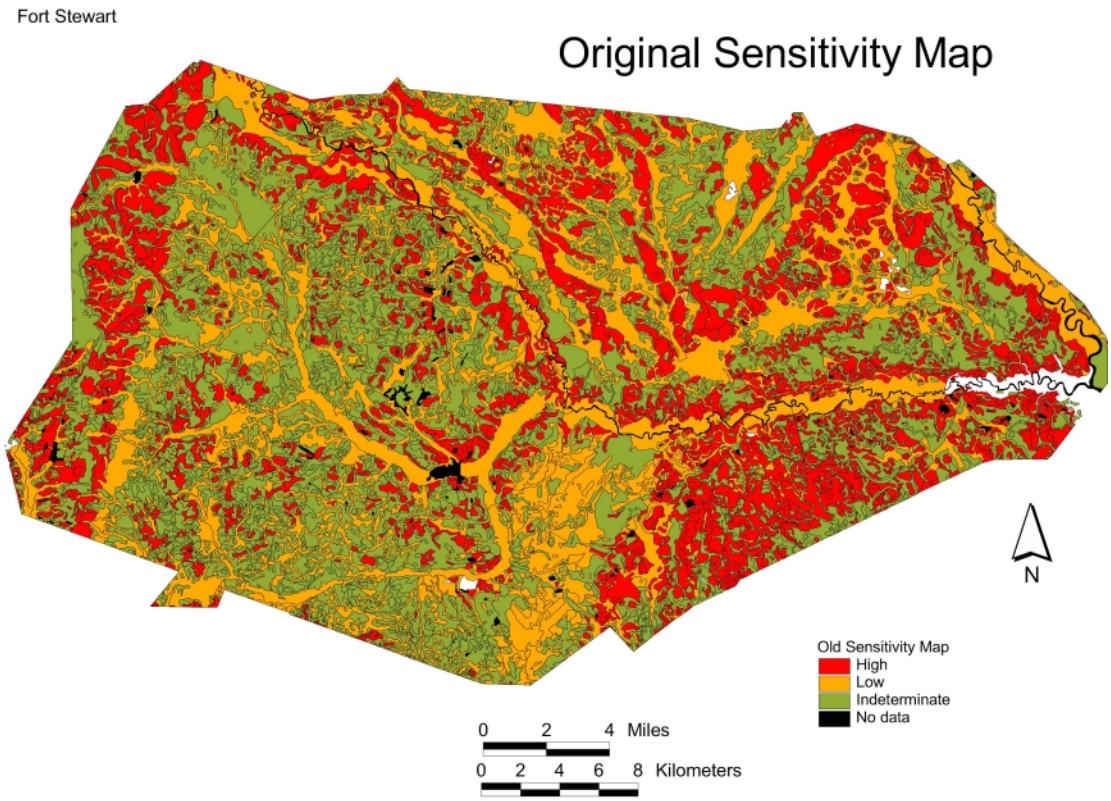

Figure 3. Sensitivity map from 1985 Fort Stewart model.

Post Model Development

Since 1985, more than 1,600 additional sites have been recorded during more than 135,000 acres of additional survey. As more survey was performed, indeterminate soils, or those for which insufficient data existed to establish a reliable understanding of the relationship between soil type and site location, were eliminated. The sensitivity map was changed to reflect the new data (Figure 4).

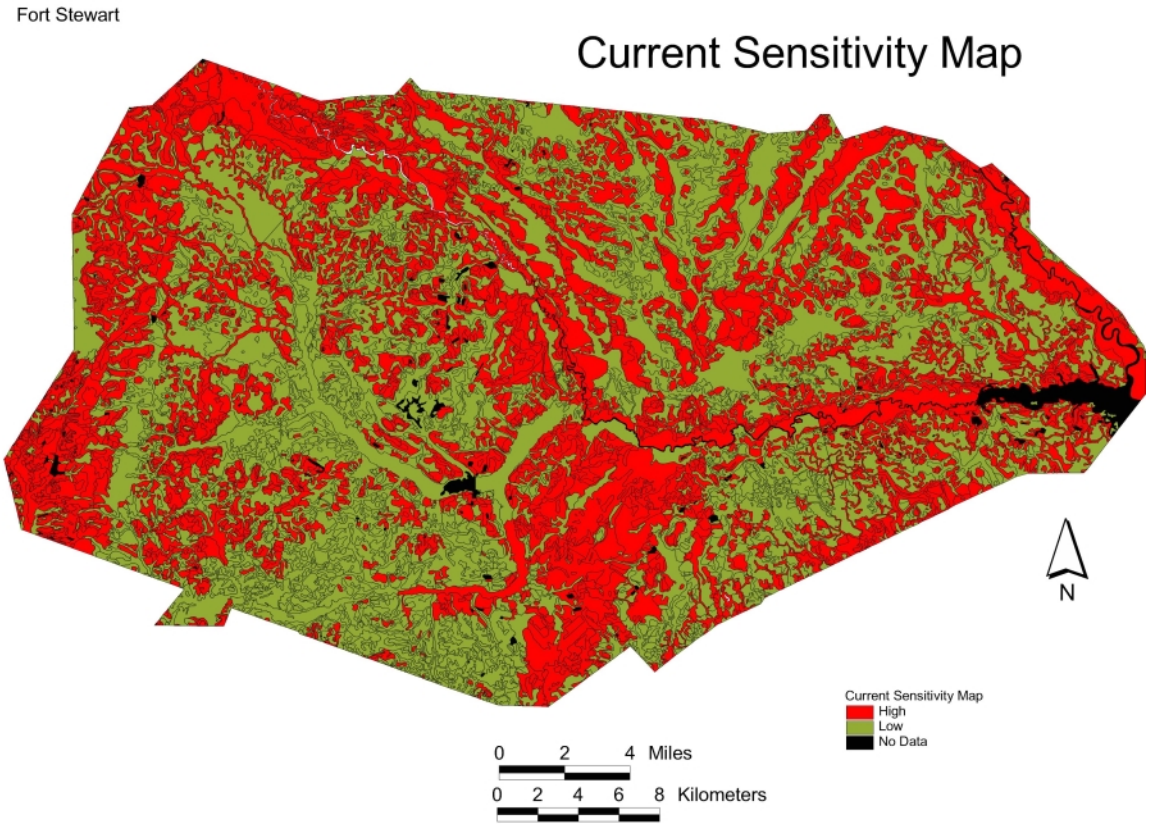

Figure 4. Current Fort Stewart sensitivity map.

Evaluation Results

Using the logic and assumptions embedded in the Fort Stewart model, SRI Foundation reconstructed the model in a GIS using ArcView software and soils, survey, and site data provided by Fort Stewart archaeologists. To correct the unrealistic assumption that sites exist as points (see below), we arbitrarily defined site boundaries as an area within a 50-m radius of the site datum.

For prehistoric sites, 67.38 percent of all sites were located within the high sensitivity zones, which constitute 44.94 percent of the surveyed area. This result differs from that of the original 1985 model wherein the high sensitivity areas constituted 25.07 percent of the survey area and contained 38.75 percent of the prehistoric sites. This difference is a result of the relatively high proportion of indeterminate soils in the original model.

As a measure of the performance of the model, we compared the proportion of sensitivity zone surveyed with the proportion of the sites located in that sensitivity zone. For sensitivity zone (i), this measure (S) can be defined mathematically, as

$$S_i=(a_I)/(b_i),$$

where a_i is the proportion of sensitivity zone (i) surveyed to the total area surveyed, and b_i is the proportion of the total number of sites that are found in sensitivity zone (i). Assuming that some sites have been recorded in each sensitivity zone, S varies between zero and infinity. Below 1.0, S indicates that the proportion of sites in a sensitivity zone is larger than the proportional size of that zone relative to the total

area encompassed by the model. The closer S is to zero, the greater the sensitivity. Above 1.0, the proportion of surveyed land in the sensitivity zone is greater than the proportion of total sites in that zone. This situation reflects lots of space devoid of sites. The larger S becomes the lower the sensitivity of the zone.

The high sensitivity zones of both the 1985 and the 2003 Fort Stewart models yield a performance measure of about 0.65 ($S_{(1985)}$=64.70; $S_{(2003)}$=66.70). One can argue, given their statistical equivalence, that the second model performs better because it encompasses a larger area. Neither model, however, can be termed a strong predictor. For example, many multivariate predictive models yield sensitivity zones in which half the sites are found in a quarter of the installation, or S has a score of 0.5 or lower (for example, see Fort Bliss below).

Strengths and Weaknesses of the Model

The model's greatest strength is its dynamic nature. Too often predictive models are developed and then left static, never changing even though new data are collected. Archaeologists at Fort Stewart have regularly updated and re-evaluated their model, creating a far better management tool than the original model.

The model suffers from three weaknesses. The most serious is the absence of site boundaries. Site location is plotted only by site datum. Representing site locations with points makes for a difficult modeling environment. For example, the datum may be on dry land that slopes to the northeast at 2° and is composed of some anthropogenic soil type. Without site boundary information, however, we cannot determine that the site includes a rock shelter, a stream, and outcrops of granite. This limits the effectiveness of many independent variables. In the absence of field maps, proxies for site boundaries can be developed by placing a buffer zone of arbitrary size around each site datum (as was done in this analysis). This allows the analysis to take into account the surrounding environment. The problem with this approach is that it introduces errors that cannot be quantified. All sites, large and small, are treated as if they were equal in size. Sites with less than 100-m diameters, therefore, are given too much weight, whereas the statistical influence of large sites, which are often the most significant sites, is diminished.

The second weakness is the poor predictive power of the model. This results from reliance on a single environmental variable. Proxies of other variables that influence human behavior, such as distance to water, type of water, slope, aspect, and vegetation, need to be developed, and these proxies must be measurable and replicable. Furthermore, formation processes that affect the archaeological record, such as erosion and alluviation, must be taken into account. Variables reflecting these processes can be derived and included in the statistical modeling effort, or separate models of site burial can be developed.

The model's third weakness is in presentation. Currently, the predictive model is maintained as a series of 7.5-minute U.S.G.S. maps. For this project, we replicated the model in a GIS format. Fort Stewart personnel should migrate the model from a paper to a digital medium.

Recommendations

Fort Stewart has a nearly 20-year commitment to predictive modeling. This commitment goes beyond simply using a model; it includes an ongoing process of refining the model. Creating a research and management atmosphere that supports *modeling* as opposed to static *models* is a critical step in successfully using and integrating predictive modeling into cultural resource management. Fort Stewart, then, has accomplished the hardest part of the job. What are left are technical tasks.

1. Site boundaries

   Fort Steward needs to add site boundaries to the ArcView GIS layer on cultural resources. This will provide a far more accurate representation of environmental components within the domain of each site, and will make using additional environmental variables more practical. If resources are not available to digitize the actual site boundaries, it is recommended that some means of assigning stan-

dard-sized site areas be added to the site database. This would allow archaeologists to variably buffer datums so that the size of the buffer better reflects the magnitude of the site.

2. Develop more independent and dependent variables

   The model is based on a simplistic statement that predicts the locations of all prehistoric sites, based on soil type, without regard to site type. Other environmental factors besides soil type clearly affected prehistoric behavior. Proxy measures of these factors can easily be created from source data such as DEMs. Additionally, the prehistoric sites could be subdivided by temporal and functional categories and then modeled separately. With clearly defined and measurable variables, modeling could include more robust statistical techniques based on regression (Rose and Altschul 1984:212).

3. Incorporate post-depositional processes into the models

   Geomorphic data on site burial and site destruction need to be included in the model. One approach is to create a separate GIS theme for these processes that could then be overlaid onto the sensitivity map.

Eglin AFB

During the 1980s and early 1990s, cultural resource management at Eglin AFB what handled by private contractors. In 1992, however, an in-house cultural resource management program was established for the installation. Eglin AFB comprises approximately 464,000 acres in northwestern Florida. Of that acreage, a total of 185,000 acres has been surveyed for cultural resources, and more than 1,600 archaeological sites have been recorded. The installation's interest in predictive modeling began in 1982 and continues today. To create its predictive model, Eglin AFB sponsored a 12 percent inventory of the installation that combined probabilistic sampling with judgmental survey. After the first year of survey was completed, the installation began experimenting with multivariate statistical modeling. Altschul (1993) used discriminant function analysis to distinguish between site and nonsite locations in the random sample survey areas. Variables included seven environmental features: slope, topography, type of nearest potable water source, horizontal distance to water, vertical distance or elevation above water, relative elevation, and the number of streams within 200 m of a location. The results were disappointing, in large part because of poorly defined measures of the environmental themes. Because the modeling effort predated GIS technology at Eglin AFB, the multivariate model was extremely labor intensive.

In 1983, Eglin AFB abandoned sophistical statistical modeling techniques in favor of simpler methods. Prior knowledge was relied on extensively to develop proxy measures of environmental variables associated with site location. The sensitivity map was based on results from all inventory efforts, and included 511 sites in 55,000 acres of survey area (Thomas and Campbell 1993). The sensitivity zones defined in the model were plotted on 7.5-minute USGS maps, the format in which the model continues to be used today. Currently, the sensitivity map is being converted, digitally, to a system based on Geomedia Professional by Intergraph.

Site Representation/Sampling Strategy

To obtain data to create the predictive model, Eglin AFB first sponsored a 6 percent stratified random sample survey. Upon completion of the initial model, another 29,405 acres, or 6.5 percent of the installation, were surveyed on a judgmental basis to test the model (Thomas and Campbell 1993:151). Additionally, 87 sites recorded prior to 1982 were re-recorded.

Field methods included sketch maps and the plotting of site boundaries on 7.5-minute USGS maps. The model is based on these maps, and thus sites are represented by actual site boundaries.

Prediction Method

The model is heavily based on proximity to potable water, elevation above the closest water supply, and proximity to coastlines and alluvial plains. High probability zones (Figure 5) are those areas less than 200 meters from potable water and less than 50 feet above water, coastal zones, and a 400-m band defining the floodplain/upland juncture. Low probability zones are those greater than 200 meters from water or over 50 feet above water. In the current version of the model, any areas that have been surveyed and found to have no sites are also included in the low probability zone, whereas areas within site boundaries are classified as high probability zones. Finally, wetland areas are regarded as indeterminate probability zones. Generally site densities are expected to be quite low in wetlands, but small knolls or seasonally inundated areas that are not visible on topographic maps may be rich in cultural material.

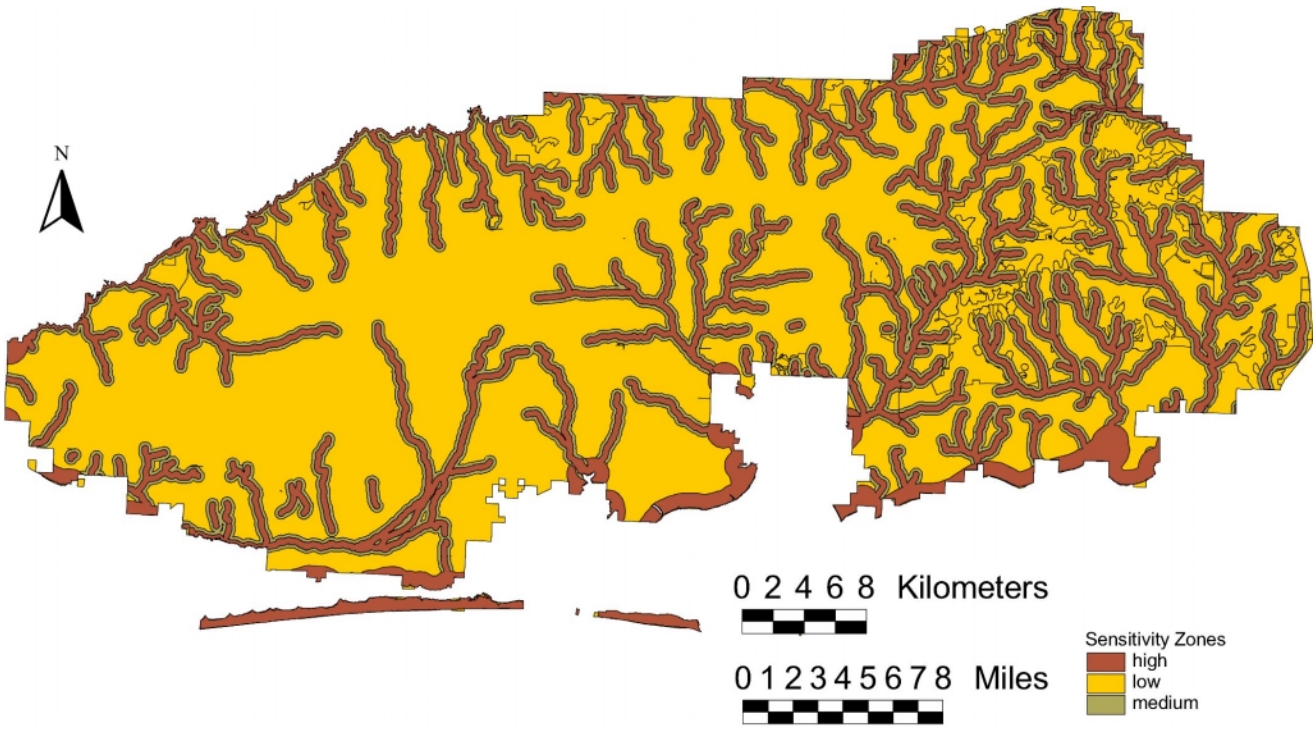

Figure 5. Eglin AFB sensitivity map.

Post Model Development

Approximately 1,100 sites have been discovered since development of the model as a result of approximately 130,000 acres of additional survey. The sensitivity map is regularly updated as new survey is performed, but unfortunately this update consists of marking all portions of newly surveyed areas that do not contain sites as “low sensitivity” and those that contain sites as “high sensitivity,” regardless of the environmental parameters. The rationale for this approach is that all resources in these areas are known and therefore these classifications reflect actual, as opposed to predicted, sensitivity.

Evaluation Results

The only environmental data for this model that were available digitally were those for the “distance from potable water” variable. SRI Foundation created a proxy of coastal areas by designing a GIS theme that

defined all areas within 1 km of the coast as a high sensitivity zone and all areas farther away as low sensitivity zones. A digital predictive model that used the logic underlying the Eglin paper map model was created by overlaying the two GIS themes and defining three sensitivity zones. The high sensitivity zone included all areas less than 200 m from potable water or within 1 km of the coast. Medium sensitivity zones were at least 1 km from the coast and between 200 and 400 m from potable water. Low sensitivity zones were also more than 1 km from the coast, but were more than 400 m from potable water.

The 1,100 sites recorded since the initial sample survey in 1982 follow the general trend predicted by the model (Table 3). High sensitivity zones have more than twice the number of sites per 1000 acres as either the medium or low sensitivity zones. Performance, as measured by the S statistic (see discussion in section on Fort Stewart above), also shows that, proportionally, sites tend to be concentrated in the high sensitivity zone. Somewhat surprisingly, the medium and low sensitivity zones have about the same predictive power, as measured by the S statistic.

The two variables included in the analysis, then, yield a relatively poor predictive model. The S score for the high sensitivity zone, although below 1.0, is still relatively high, indicating that about 60 percent of the sites are contained within approximately 40 percent of the surveyed area. This result is hardly what we would call a strong predictor, especially given that the installation is continually inflating the model's power by automatically redefining all site areas identified during survey as high sensitivity zones. The lack of significant differentiation between the medium and low sensitivity zones highlights the importance of the variables that we were unable to include in our GIS reconstruction—floodplain/upland ecotone locations and elevation above water. This lack of differentiation occurs despite the practice of adding areas known to contain no sites to increase the low sensitivity zone sizes.

Table 3. Predictive model performance evaluation for Eglin AFB

Sensitivity	Number of Sites	Acres of Survey	Density Sites Per 1000 Acres	Performance (S)
High	978	80,097	12.21	0.70
Medium	238	36,771	6.47	1.32
Low	407	73,658	5.53	1.54

Strengths and Weaknesses of the Model

The predictive model has a solid foundation, being based on a well-planned sample survey and on a multivariate prediction strategy. Archaeologists at Eglin update their sensitivity map regularly and continue to use the model.

Weaknesses in the model stem from its format and application. The sensitivity model is a paper map, which means that Eglin personnel are unable to take advantage of modern computing abilities. This weakness is presently being remedied by Eglin AFB archaeologists, who are digitizing the map. The act of digitizing, however, is creating its own problems. The sensitivity map being digitized is a product of a hand-drawn one, rather than one generated through GIS using algorithms that proxy the modeling rules. This point is important, as bias and oversight affect hand-drawn delineations. Furthermore, hand-drawn models cannot be as easily tested or improved upon. The model should drive the sensitivity map, not the reverse.

The model's utility is also being compromised by the mechanical approach of defining all surveyed areas as low-sensitivity areas and all sites as high sensitivity areas. Logically, all cultural resources are known within the surveyed areas, so presuming the resources are avoided, the military can use the area free of cultural resource restrictions. This means that, from the Air Force's perspective, the "sensitivity" of surveyed areas is known and can be easily managed.

Unfortunately, this also means that current sensitivity maps are the result of two fundamentally different sets of map rules. One set of rules is based on environmental features that have been logically and statistically associated with site location; the second rule set simply designates sensitivity on the basis of suitability for potential military activity. Thus, one cannot infer that humans avoided low sensitivity areas in the past. Instead, the best we can infer is that the region's prehistoric inhabitants avoided some of these areas, although without also knowing where surveys have been performed, we do not know which areas.

Current practices restrict our ability to learn more about past behavior and, accordingly, model it better. New data acquired through survey are the lifeblood of the modeling process. Newly discovered sites provide an ever-increasing dataset from which the environmental signature of archaeological sites can be more accurately determined. As an example of the problems caused by the conflicting rules that have generated the sensitivity maps, recent survey and testing projects at Eglin AFB have demonstrated that archaeological deposits are buried in aeolian beach and dune formations (Campbell et al. 2003). Although archaeologists have long suspected that humans favored these geomorphic features for some activities, they were always designated low sensitivity areas by virtue of the lack of surface expression of the archaeological sites. By categorizing all surveyed areas with negative results as low-sensitivity, Eglin AFB precludes the possibility of refining the predictive model to account for these geomorphic situations. Potentially valuable information that could further refine the model, or at the very least bolster our confidence in its predictions, is thereby ignored.

Recommendations

In light of the above discussion, we have the following recommendations for improving the Eglin AFB predictive model.

1. Generate a new version of the predictive model using GIS technology.

   We suggest that modern environmental datasets and GIS technology be used to regenerate the sensitivity map using the same environmental criteria that were used as independent variables in the current version of the model. This task should be relatively straightforward, using hydrology and digital elevation data provided by the United States Geological Survey. In this form, the model will be easier to refine and use.

2. Eliminate the current practice of defining all surveyed space with negative results as "low sensitivity" areas and all surveyed space containing sites as "high sensitivity" areas.

   This practice defeats the purpose of creating a replicable and objective model. Information about which areas have been surveyed and information about which areas have high or low site sensitivity are both extremely important for resource management, but the two cannot be combined and remain truly useful. Since the data on the original probability zones still exist as a GIS layer, the integrity of the model can be maintained by developing a separate, management-oriented "surveyed lands" layer. By overlaying the site location and surveyed lands layers, a layer showing "constraint-free zones" in which where there are no cultural resource constraints on mission activities could be created.

3. Incorporate geomorphology into the model

   One of the problems long noted by archaeologists working at Eglin AFB is that buried sites are not predicted well by the model. A geomorphological map of surfaces likely to contain sites could easily be added as a theme within a GIS-based predictive model.

Fort Bliss

The Fort Bliss cultural resource management program, which was established in 1978, has developed two Archaeological Predictive Models for the 700,000 acre McGregor Range in the Tularosa Basin of south-central New Mexico (Zeidler et al. 2002). Nearly 373,000 acres of the range have been surveyed, and more than 2,900 sites have been recorded as a result of those surveys. The McGregor Range models are among the most sophisticated in use in the military. They are based on multiple environmental variables that have been used to predict various types of archaeological manifestations. The models are based on 901 sites that were recorded prior to 1996. Approximately 2,000 additional sites have been recorded since model development.

Site Representation/Sampling Strategy

Site datums are used to represent site locations. Sites are ranked by size: small, medium, and large. Fairly comprehensive data about site characteristics are available in the database, including cultural affiliation, features, and artifact types. Eligibility for the National Register of Historic Places is also noted.

No survey has been performed specifically to test the model. Inspection of Figure 6, however, reveals that nearly 40 percent of the McGregor Range has been surveyed. Survey driven by Section 106 compliance has taken place in all parts of the range. We conclude that this coverage is adequate to create correlative models of site location based on environmental variables.

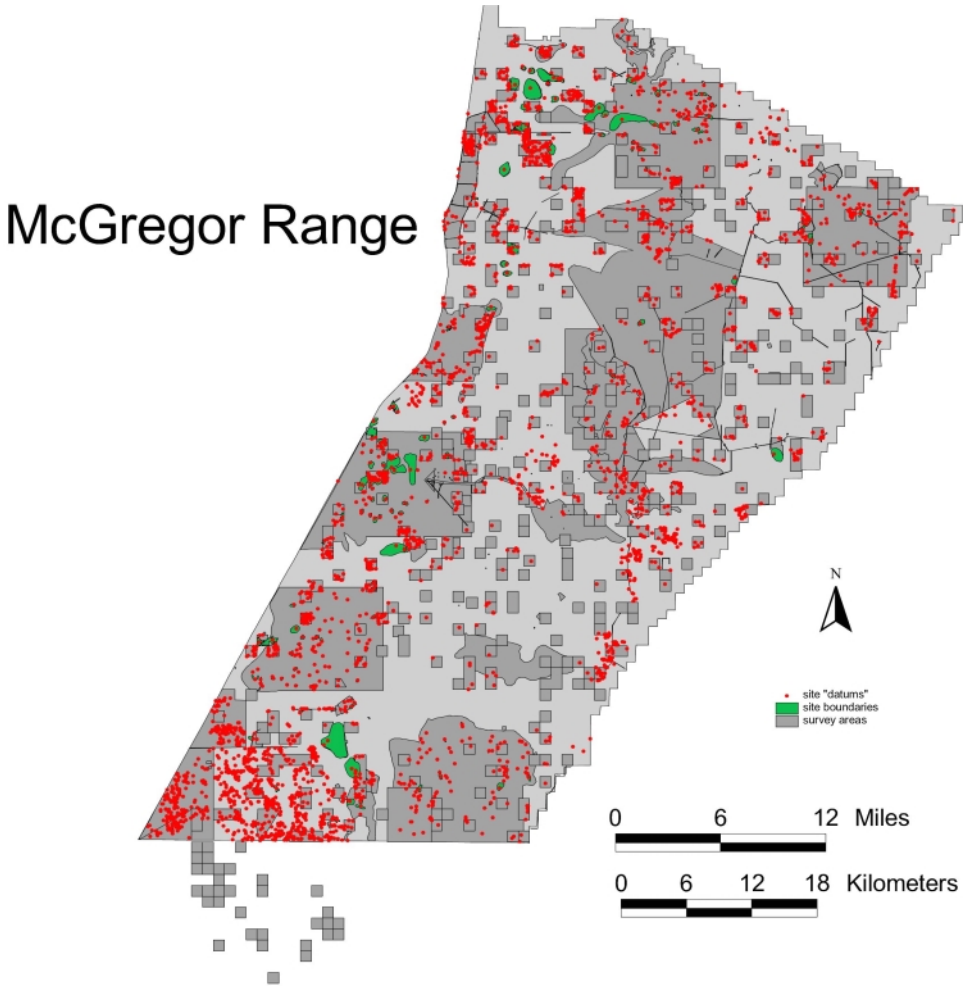

Figure 6. McGregor Range with sites and survey areas.

Prediction Method

For purpose of modeling, the McGregor Range was divided into 2,984,882 cells, each of which measured 30-by-30 m. Each cell was associated with a score on a series of environmental and cultural variables. Environmental variables were: slope, aspect, elevation, distance from streams, distance from playas, geomorphic features, soils, vegetation, and soil moisture. Archaeological sites were classified as limited activity, extended activity, and rock shelters. Associational statistics, including chi-squared, phi-squared, and Cramer's V tests, were calculated to determine covariance among environmental and archaeological data. Two predictive models were ultimately developed: one which modeled all prehistoric sites using soils, aspect, elevation, and landform, and another which modeled rock shelter sites (n = 32) using slope, aspect, landform, elevation, and distance to water. The first model created three sensitivity maps: one for all sites, one for limited activity sites, and one for extended activity sites.

To develop the sensitivity maps, each environmental variable used in the model was divided into three favorability scores (1 = favorable, 2 = neutral, 3 = unfavorable). The scores for all variables were summed for each cell. Cells were then grouped into six categories from lowest to highest site sensitivity. The probability of any cell in each category to contain a site was calculated, and the results were framed as probability ranges as follows: 0–20 percent, 21–50 percent, 51–70 percent, 71–80 percent, 81–90 percent, and 91–100 percent.

Post Model Development

The predictive model was developed in an older version of GRASS (4.0). Fort Bliss is no longer using a GRASS platform and, consequently, is not maintaining the model. The predictive model exists as a paper map that has not been updated since its creation. Although the model has not been kept current, archaeological survey has continued unabated. Since 1996, approximately 148,800 acres have been surveyed and more than 2,000 archaeological sites have been recorded.

Evaluation Results

To evaluate the model, SRI Foundation first had to transform the hard-copy map into a digital file. The hard-copy map exists only as an 8.5 by 11 inch, second-generation copy. The image is fuzzy and the sensitivity zones are depicted as gradations of gray. Although of poor quality, the map represents a model of fine-scale resolution. There are nearly 3,000,000 cells that vary among six probability categories represented on the paper copy. Hand digitizing the model was out of the question. The expense would have been enormous and the result of dubious accuracy, for there was no easy way to distinguish among probability categories represented as slight variations of gray. The only feasible approach was to scan the map and collapse the probability categories from six to three sensitivity zones: 0–20 percent (low), 21–90 percent (medium), and 91–100 percent (high).

To digitize the map, we first scanned the paper copy to create a digital file. Each pixel was assigned a value between 0 and 255 corresponding to the visible color spectrum ranging from white to black. We then utilized an image enhancer, termed a "mean digital filter," available in the IDRISI GIS software package, to create a new image in which each pixel's value is based on its own value and those of its immediate neighbors. The nature of this operation is determined by the values stored in a 7-by-7 template that is centered over each pixel as it is processed. The values of the 49 pixels are averaged, and the resulting value is assigned to the subject pixel.

By comparing the results with the paper copy, we determined that values 0–95 represented the high sensitivity zone, values 96–218 corresponded to the medium sensitivity zone, and values higher than 218 fell in the low sensitivity zone. A digitized sensitivity map was thereby created (Figure 7), and we were able to use that map to test the predictive model.

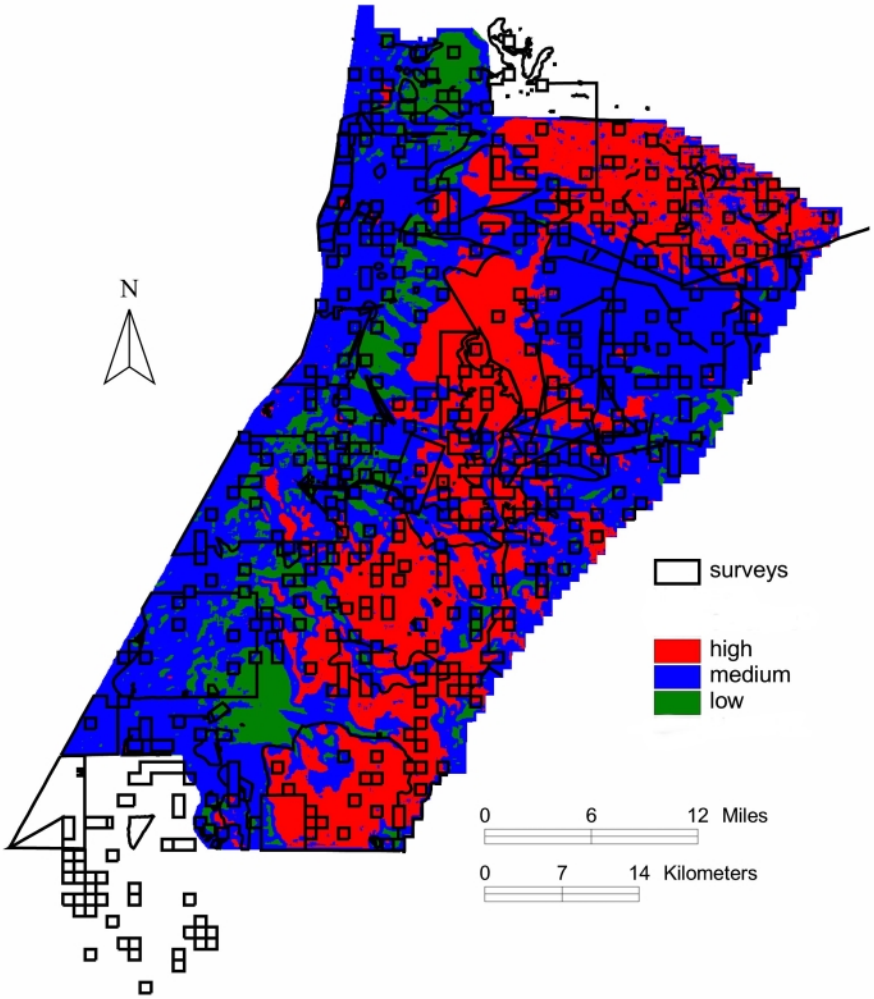

Figure 7. Fort Bliss McGregor Range sensitivity map generated by SRI Foundation.

Many of the sites in the McGregor Range model are represented as points rather than polygons. Of even greater concern, most sites are represented by more than one point. Staff at Fort Bliss explained that this is a result of multiple visits to the sites where an approximate "site center" was recorded by Global Positioning Station (GPS) each time (Figure 8). These practices have led to two major problems with evaluating the McGregor Range model. First, because site boundaries are not accurately portrayed, the covariation of environmental variables with archaeological site locations is imperfectly measured. Second, the practice of including repeated visits to the same site as separate records in the database has the effect of inflating the importance of heavily visited sites.

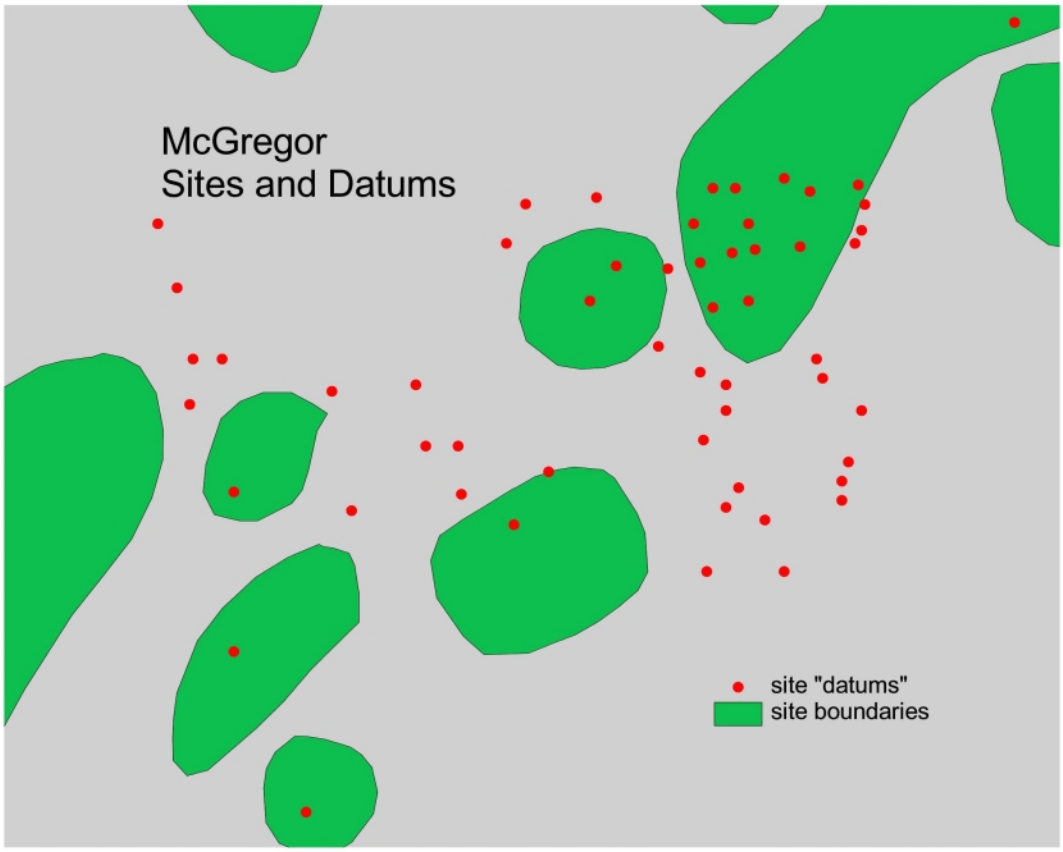

Figure 8. Examples of multiple site datums at McGregor Range.

To improve the McGregor predictive model, site boundaries need to be accurately portrayed in the GIS. This step is presently under way at Fort Bliss. In the meantime, the practice of creating multiple site datums can be used as a proxy measure of site size. We buffered all site datums creating approximately circular boundaries using data on site area (a) from the installation's Excel cultural resources data base. We then used the function

$$r = \sqrt{\frac{a}{\pi}}$$

to determine buffer radius (r) for each datum. The resulting layer was joined with the partially completed true site boundary layer to represent total site coverage for the analysis.

Many generalizations were made in the evaluation process. The sensitivity map was digitized from a very fine-scaled but low print-resolution image. The color values were then recategorized to make them manageable. The site coverage data were grossly approximated. It is important to point out that the McGregor Range model includes sensitivity maps for limited and extended activity sites, as well as the "all prehistoric sites" map. Owing to the scale and resolution issues in digitizing, however, the subtle differences among the three models could not be practically identified. For this reason, our evaluation focuses on the model for all prehistoric site locations.

Even with the introduction of these errors and biases, it is clear that the environmental signatures of sites are quite robust (Table 4). Sites are nearly 10 times as dense in the high sensitivity areas as in the low sensitivity areas. Performance scores (S), as defined in the section on Fort Stewart, demonstrate that high sensitivity areas are strongly associated with the location of archaeological sites, whereas for mid-range sensitivity areas, the proportion of sites is roughly equivalent to the amount of surveyed space. In low sensitivity areas we find mostly "empty" surveys, devoid of cultural remains.

Table 4. Predictive model performance evaluation for McGregor Range

Sensitivity	m^2 of site	m^2 of survey	Density sites per 1000 m^2	Performance (S)
High	2,393	15,805	15.14	0.19
Medium	10,526	385,373	2.73	1.05
Low	1,838	115,434	1.59	1.79

Strengths and Weaknesses of the Model

The McGregor Range predictive model works remarkably well. The model is well thought out and includes a wide range of environmental and cultural variables. Unlike many other examples, the modeling process for the McGregor Range subdivides the cultural dimension into behavior sets. This use of site types makes the model more effective as an analytical tool for understanding past human behavior.

There are a number of weaknesses in the model, but the most serious is its current static state. The model exists only in paper form; the operating system used to create the model is no longer supported at Fort Bliss, which precludes the ability to add new data.

Other weaknesses include the lack of site boundaries, reliance on the intersection method, and correspondingly, the lack of more rigorous analytical modeling procedures, such as regression-based modeling techniques. Although the model uses a variety of environmental variables, it does not incorporate proxy variables for the geomorphology. Hall (2002) has demonstrated that land surfaces in the region are quite dynamic, and that aeolian processes have buried archaeological remains to such an extent that surface expressions are limited and misleading. A more powerful model would include layers indicating landforms conducive to buried sites.

Another problematic area is the practice of placing multiple datums at sites. As mentioned in the previous section, this practice has the potential to inflate the importance of highly visited sites because each datum is treated as a separate site. This practice has not previously caused serious problems because the Fort Bliss model is static and could not be updated to include these additional datums.

Recommendations

We suggest that Fort Bliss perform a full conversion of the model from GRASS to an ArcGIS platform. Installation personnel are trained on ESRI software, so this conversion would enable Fort Bliss to develop a dynamic system that is regularly maintained.

With respect to design, a few recommendations follow:

1. Complete the site boundary layer

   Accurate site boundaries are the most immediate concern for the Fort Bliss model. Site buffering is, in the long run, a poor substitute for documented boundaries. Arbitrary buffers fail to characterize the shapes of sites, which, quite often, mimic the nature of environmental variables.

2. Condense sensitivity categories to facilitate resource management

   The model divides site sensitivity into too many categories to be helpful for resource management. Managers need categories with clearly identifiable management implications. Multiple categories of "gray" are more confusing than helpful. We recommend reclassifying sensitivity into three categories (low, medium, and high) that can be tied to specific management concerns and practices.

3. Use more of the available data to inform the model

   Substantial amounts of data are associated with the sites in the database. Since one of the main goals of predictive modeling is to assist in finding historic properties, it is important to identify the environmental signature of NRHP-eligible sites. Developing models for specific property types – those with and without thermal features, for example, or sites with ceramics vs. those without – would be a step in this direction. Identification of environmental variables that measure the potential for NRHP-eligible sites to be buried with little or no surface expression would also be useful.

4. Use more rigorous modeling techniques

5. The McGregor Range model uses rudimentary modeling techniques. Although the current model is useful, more powerful models could be created through the use of multivariate statistical techniques. These models would not only provide for more accurate predictions, but would also allow the modelers to determine the proportional value of each independent variable. Such an understanding could allow for greater insight into past human behavior.

Summary and Conclusion

There are a number of widely applicable lessons about the use of predictive modeling by military installations that can be drawn from our analysis of four specific models in this study. In many ways, for example, the history of archaeological predictive modeling at Eglin AFB could be viewed as a microcosm of the military's experience with modeling throughout the United States. Eglin AFB began with great enthusiasm, embracing predictive modeling as a tool to help understand the installation's archaeology and to support the best possible management decisions. The installation funded a complicated and expensive stratified random sample survey, leading to an even more complicated multivariate statistical predictive model. By the time the model was finished, however, the bloom was off the rose. Although the model's accuracy would have improved in time with more data, better proxy variables, and different statistical techniques, it was clear that even a new and improved model would serve neither the interests of archaeologist nor those of the installation managers. The model's predictions could not be intuitively grasped; there was no way to connect the results with human behavior that structured decisions about land use. Without such an understanding, there was little confidence in the results. The model, which required considerable effort to maintain, was dropped.

In its place arose a more traditional archaeological settlement model. Using their collective knowledge about where sites had been found, archaeologists "backed into" a predictive model. They defined where they thought sites should be located and then created measurable rules—so far from water, so high above water, so far from the coast, etc.—to capture these notions. The resulting model was presented as a "sensitivity" map. The term sensitivity was used to signify that these were qualitative assessments about prehistoric settlement, as opposed to the probability maps generated by the multivariate statistical techniques. Sensitivity also suggested a certain "fuzziness" about the zones, which meant that the installation could not use these maps to eliminate areas from survey. As more survey was performed, archaeologists became more confident that the model worked well in some areas and that it did not work well in others. In time, the reason for one category of errors became clear; post-depositional geomorphic processes obscured sites in some of the areas where the model did not appear to work well. Archaeological resources actually exist in the suspect landforms, but they have no surface expression and thus were not being found during surveys.

The Eglin AFB predictive model is a reasonable reflection of the confidence the archaeological community of northwest Florida has in their knowledge about regional prehistoric settlement. The predictions provide managers with a notion of what to expect, but given the "fuzziness" inherent in those predictions, they do not carry sufficient weight to support decisions eliminating survey in low sensitivity zones. The Air Force, however, is interested in "clearing" areas, that is, creating areas that can be used free of restrictions for military activities. Managers have increased the likelihood that low sensitivity areas will not contain a site by including all surveyed areas that are known not to contain sites in this zone. By "stacking the deck," they are creating a "model" that appears to perform remarkably well. Some managers may actually argue that low sensitivity zones really contain no sites. This appearance, however, is misleading, and by mixing management and archaeological criteria in sensitivity zone definition, it is not at all clear any more what exactly the model is predicting.

Eglin AFB's history with predictive models—the initial interest in research and development, the transformation of prior knowledge into a sensitivity map, and the bending of the map to incorporate management objectives—parallels the experiences of many military installations. In this final section, we explore the reasons for this history and how predictive models can be used more effectively in the future.

The Initial Blockage: Archaeology versus Management

The perilous course of predictive modeling in cultural resource management was set early by a lack of specific awareness of the competing objectives of research and management. To cultural resource managers, the only rationale for engaging in predictive modeling was that it would assist in complying with laws and regulations pertaining to cultural resources. More to the point, many installation commanders expected predictive models to eliminate the need for inventory in large areas, and initially some cultural resource managers shared a belief in this possibility. Archaeologists contributed to this belief by observing that correlation of site location and environmental attributes can sometimes lead to very accurate predictions. Generally, however, the archaeologists were much less interested in where sites were located and more intrigued by why sites were located where they are. Most archaeologists viewed predictive models not as end products, but rather as necessary first steps in understanding prehistoric decisions about where to live, where to collect and process resources, and how these decisions fit together into settlement systems.

These differing interests and expectations were rarely explicitly addressed. More often, cultural resource managers and archaeologists assumed that they were in agreement on the goals and objectives of predictive modeling. Funding sources in the military favored the development of predictive models in the 1980s. As a result of these unexamined differences in expectations and objectives, neither cultural re-

source managers nor consulting archaeologists informed those funding the projects that the likely outcome was not going to be the end of archaeological survey as we know it.

This historical review undoubtedly overstates the dichotomy between management and research objectives. Many cultural resource managers have been, and continue to be, trained as archaeologists, and most archaeologists working in CRM are fully versed in compliance issues. The blurring of the lines, however, does not diminish the main point. Cultural resource managers needed predictive models that were useful in meeting compliance objectives. Archaeologists wanted predictive models that could be used as a tool in their arsenal for exploring past behavior.

By and large, neither group has been satisfied with the results. Military installations have generally not been able to persuade State Historic Preservation Officers and other interested parties to accept the results of predictive models as a substitute for archaeological survey. Without some compliance-related payoff, installations have been reluctant to fund further research into archaeological settlement. What is so surprising, then, is that even though neither archaeologists nor cultural resource managers are really getting what they want, interest in predictive modeling has remained strong. In large part, the reason for this interest can be found in three words, or better yet, in one acronym: geographic information system or GIS.

GIS: A Brave New World

In 1988, the Bureau of Land Management published a watershed monograph on predictive modeling under the title, *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling* (Judge and Sebastian 1988). The volume remains the most comprehensive treatment of the subject. When the volume was conceived in 1985, the BLM contracted for individual chapters and selected the volume editors through requests for proposals. No proposal was requested for a chapter on GIS. It was only at the pre-work conference in Denver that the participants agreed that GIS was a technology looming on the horizon that could be important in future modeling projects. A chapter was hastily added and two of the authors who had been selected to write other chapters agreed to write a GIS chapter as well (Kvamme and Kohler 1988).

Fifteen years later, no one would conceive of a predictive modeling project that did not make use of GIS technology. In fact, many people conflate GIS and predictive modeling, which is the cause of much confusion and frustration. Many of the old arguments regarding predictive modeling have been repackaged to suit the new technology. GIS, so we are told, will allow us to know where previously recorded sites are located and where we can expect to find new sites. Ultimately, as the argument goes, the application of GIS should allow installations to conduct less archaeological survey.

There is a fundamental flaw in this argument. GIS is not an analytical tool. Rather, it is a system for organizing, relating, and formatting spatial data. The four installations discussed in this report gravitated to GIS technology, not because they were interested in analyzing spatial data, but because they had amassed huge amounts of information on a variety of topics. These installations needed to track observations on topics as disparate as endangered species, vegetation, roads, military activities, flight patterns, and hazardous waste. Instead of maintaining sets of maps in various offices, an installation can use GIS to centralize and standardize geo-referenced data. Cultural resources, then, are simply another set of observations, maintained as one or more themes, in a GIS.

As documented at Fort Stewart, Fort Drum, Fort Bliss, and Eglin AFB, information on cultural resources can be recorded in a GIS in many ways. Sites can be represented as pixels, points, polygons, and even lines (in the case of linear features like roads). As with any decision tree, choices made at the design phase about how information will be entered and displayed will greatly affect the purposes for which those data can be used and how those data can be analyzed. Once made, these decisions are not only difficult but also very costly to undo. Entering information into a GIS requires a large investment in person hours. It is extremely disappointing to find out, after making such an investment, that the format chosen does not allow the promise of GIS to be realized. At military installations, GIS specialists and not archae-

ologists have generally made data and formatting choices. The result, as the decision at Fort Stewart to represent all sites as points bears witness, is that cultural resources are often represented in ways that are unsuitable for spatial analysis.

As a formal database, GIS requires systematic and standardized data, a requirement that is sometimes problematic, given the vagaries of archaeological data collection. The problem of multiple site datums noted in the discussion of Fort Bliss is a good example. Archaeologists have long known that the plotting of site locations and site boundaries on maps is a frequent source of error. Global positioning stations (GPS) hold out the promise of reducing these recording errors, but even this technology cannot eliminate human error. An even bigger problem for coding GIS-based data is the lack of standardization in archaeological terms. Site types, feature types, raw materials, and even artifact types are assigned different names by different recorders.

Once decisions have been made regarding how spatial data will be represented in the GIS and how attributes will be defined and measured, the long and arduous process of entering data must be completed. Only then can installation archaeologists begin to ponder the structure of the archaeological record. Unless predictive modeling is one of the perceived uses of the GIS from the beginning, it is likely that the myriad of decisions made in defining the database structure will limit the subsequent utility of the GIS for this purpose. Additionally, the GIS software used by the installation will largely define the nature and type of predictive modeling that can be pursued.

At all the installations examined, with the exception of Fort Drum, the predictive models in this study utilize the intersection method. Environmental variables, such as slope, aspect, or elevation are divided into categories that are either strongly correlated with the presence of archaeological sites or strongly correlated with the absence of archaeological sites. The next step is to create a GIS layer for each variable; the layers map those areas of the installation where, based on that variable alone, archaeological sites should occur and those areas where archaeological sites should be absent. The layers are then overlaid on each other to create high (the intersection of all favorable zones), low (the intersection of all unfavorable zones), and medium (some favorable and some unfavorable) sensitivity zones.

The attraction of the intersection method is not that it mimics human behavior. Archaeologists have long known that human decisions about where to live, where to get food, and where to interact with the cosmos are very complicated. At its best the intersection method provides an accurate picture of where archaeological sites can be expected. But as the history of predictive modeling has shown, such a picture is of limited use. Regulatory agencies, American Indians, and other archaeologists have not been satisfied with the results to the point that further inventory in low sensitivity zones can be eliminated.

Why, then, is the intersection method so popular? The answer lies in the software. Once a GIS is populated, it is a relatively simple matter to create an intersection model. This statement is particularly true of the ESRI software packages, ArcView and ArcInfo, which dominate the GIS market, particularly within federal, state, and municipal governmental agencies. Alternative modeling techniques, such as logistical regression, viewshed, and least-cost transportation routes, are best conducted with different software packages, such as Idrisi or Tnt mips. Managers of natural and cultural resources in the military do not commonly use these packages, which are more analytical than display oriented.

Even though there is no assurance that the results of predictive models will reduce the need for inventory, backing for predictive modeling of cultural resources remains strong within the military. We believe there are two reasons for this tendency. First is the wide availability of GIS. There is no doubt that the military is committed to using GIS technology to manage resources and activities, and cultural resources are an integral part of any environmental program's responsibilities. The ease in which intersection models can be created once a GIS is in place suggests that modeling as an offshoot of GIS use will continue.

The ease with which models can be created is not the only attraction, however. As the four installations examined in this study amply demonstrate, almost every military installation has a very large cultural resources database. Hundreds, often thousands, of sites have been recorded and entered into the installation's GIS. Such databases are inconceivable in academic research, and the possibilities for using these data to gain insight into a wide range of human behaviors are unlimited.

Yet the potential of these databases is largely left untapped. In part this is because the military does not encourage outside use of their data, but this is not the entire story. Most installation archaeologists want to know why sites are located where they are. This is not simply a matter of intellectual curiosity, but flows from a desire to manage resources more effectively and efficiently. If management is to consist of something more than the rote statement that all cultural resource locations are off-limits for training and testing activities, then cultural resource managers must be able to distinguish sites that are significant from sites that are not and important cultural patterns from noise.

Predictive models are ideal tools for identifying anomalous sites or settlement patterns that should be the subject of more intense work. They also are powerful mechanisms for identifying robust patterns that are unlikely to change; in such cases, additional work will not increase our knowledge concerning where sites are located. With much time and effort already having been devoted to GIS development at most installations, the incremental cost of predictive modeling is low and the potential to create an important tool that can move compliance beyond rote procedures is great. The potential contribution of predictive models to more active management of cultural resources, then, is the second, and more compelling, factor leading to the resurgence of predictive modeling in the military.

Predictive Modeling: Future Prospects

We began this study by asking a simple question, “Did predictive models created by military installations in the last 20 years work?” As with many scientific endeavors, when we pursued this topic we learned that the question asked was not really the question we wanted answered. Humans take into account their environmental surroundings when making decisions about where to carry out their activities. Models that capitalize on this premise, like the four models chosen for this study, can be accurate predictors of the archaeological record. What was interesting in the analysis was not that the models worked, but that each model used a different approach to defining the relationship between archaeological sites and environmental variables.

In this diversity, we identified five recurring themes. First, with all the interest in predictive modeling in the military, we found it interesting that there was no centralized instruction. Each installation was left to surmount the difficulties associated with site recording, GIS development, and predictive modeling by itself. This approach certainly encouraged innovation and has led to a wide variety of models. The downside is that the potential of many of these models is restricted because of decisions made early in the process, and the installations could have profited from both one another’s miscalculations and successes.

Second, the models are rudimentary in nature. In many respects, predictive modeling has witnessed a loss of sophistication in the models developed in recent years. Most models are simple intersection models or simple correlation models. Few models are based on multivariate statistical techniques or theoretically based constructs, such as optimal foraging. Because of the simplistic nature of the models, Fort Drum, Fort Stewart, and Eglin AFB have added judgmental criteria into their models to increase their accuracy; even though by doing so they reduce their systematic and objective character.

Third, the models are restricted to predicting surface manifestations. Although Eglin AFB and Fort Drum recognize the importance and predictability of buried sites, geomorphology is not a component of their modeling efforts. Additionally, the models reviewed for this study do not make use of remote sensing techniques. The lack of satellite imagery is particularly noticeable. Such imagery can be a useful proxy for ground cover and land surfaces. The imagery exists in digital form that can easily be included as a separate theme in an installation’s GIS. Importantly, much of this imagery is available to the military at little or no cost.

Fourth, while much effort has gone in to creating models, little effort has been expended in refining them. Models are treated as final products rather than being viewed as a process that involves continual modification and improvement. Eglin AFB, for example, continues to assess survey results against the

model, but there has been no attempt to refine the model since 1984, even in light of results that demonstrate the model predicts site locations in dune settings poorly.

Fifth, the models are not integral to the compliance process. Decisions regarding level of inventory, determinations of eligibility, and resolution of adverse effects rarely include model predictions. Yet, this does not have to be the case. How many acres should we survey? Where should they be placed? How should we identify sites (e.g., shovel probes or pedestrian survey)? These are questions that predictive models can assist in answering. Determinations of eligibility require archaeologists to state why a site is significant, and what we may learn from it. Models could be used to highlight why a particular site's location is unusual or typical of a class of behaviors. Data recovery plans could incorporate model predictions about the type of site and the resources available to its residents.

It is the last theme that leads us to the question that we should have asked, "If the military is not incorporating predictive modeling into the compliance process, then why are they interested in modeling?" The answers are probably as diverse as the various services and installations. We believe that at least one common thread is the DoD-wide commitment to GIS, which offers installation archaeologists the potential to be both better managers and better archaeologists if they master the technology. But our study has shown that widespread use of GIS has not led to a common understanding about predictive modeling. Models vary from qualitative statements about environment and site location to rigorous statistical relationships to mixing and matching of the two. We have learned that although there is general agreement that GIS should allow installations to create useful predictive models, there is no consensus about what a predictive model is or what it can do. The result is a proliferation of models and modeling techniques with little cross-fertilization and, individually, with little power to influence compliance procedures.

Twenty years after the advent of predictive models, we believe it is time for the military to reach a consensus on how predictive modeling will be used to comply with cultural resource laws and regulations. We are not suggesting a top-down approach in which the Department of Defense in Washington issues another set of regulations. Indeed, in the case of predictive modeling, the expertise lies with the individual installations that have been struggling to realize the potential of this technique for the last several decades. In the course of this study, we have been profoundly impressed by the knowledge and creativity exhibited at the installation level. The absence of a mechanism for sharing this expertise beyond the installation is unfortunate.

Our suggestion for a follow-up to the current study involves creating a mechanism to capitalize on this expertise. We propose to work with the military to convene a workshop in which installation archaeologists and cultural resource managers share their collective knowledge and come to a consensus on how the military can more fully incorporate predictive modeling into cultural resource management programs. Four topics would need to be addressed at this workshop: database issues, modeling techniques, modeling and compliance, and the role of spatial analysis.

Database issues: We have found that many predictive models fail before they start. Decisions about how sites will be represented in the installation's GIS, how the environment will be characterized, and what cultural attributes will be coded greatly affect the type and usefulness of a predictive model. Many installations delegate these decisions to GIS specialists, with the result often being that sites of all types are lumped together and that the environment is too crudely represented to be of much use as a predictor. And sometimes even these data-structure decisions can be too late in the process to ensure a successful model. Failure to impose quality assurance standards on the collection of field data often results in site locations being incorrectly plotted, features and artifacts being misidentified, and sites being assigned to incorrect types or periods. Each installation has a history of addressing these problems, which can be of great benefit to others.

Modeling Techniques: The intersection method is currently the modeling technique of choice among military installation. This development is not necessarily a welcome one. The intersection method is easy and generally accurate. It does not, however, tell us much more than we already know. Intersection models have not been usefully integrated into compliance with historic preservation laws, largely because the results do not provide a compelling argument that we have learned anything of importance about past be-

havior. We need a thorough discussion of the range of modeling techniques available to installation, along with guidelines about the level of expertise needed to put them into operation.

Modeling and Compliance: The goal of the whole modeling process is not to create predictive models, but to provide a useful tool to assist installations to comply with laws and regulations more efficiently and more effectively. The emphasis should not be on models as end products, but on modeling as a process that assists with compliance. Although most installations have a long history of developing models, few have integrated the modeling process effectively into their CRM programs. We need to explore why this has been the case and how the military can better use predictive modeling in decisions regarding inventory, evaluations of eligibility, and resolution of adverse effects.

Spatial Analysis: At the risk of having installation commanders resort to Ronald Reagan's admonition, "There you go again," we raise the need for predictive models to be useful in understanding the past. Correlation is not explanation. Knowing that archaeological sites can be predicted by a set of environmental features is not the same as knowing why humans chose those areas. Until military installations are able to demonstrate that they have programs focused less on identifying and avoiding and more on predicting and understanding, the military will continue to meet resistance when they propose not to survey every acre or test every site. Finding the appropriate approach to integrating research with compliance continues to haunt each installation. Answers will undoubtedly differ depending on the region and the nature of the resources, but common ground may be found that will enable all installations to develop strategies for integrating management and research.

The outcome of the workshop should be a manual that presents various problems, solutions, and avenues for securing additional help. We are not suggesting the creation of a cookbook or text on predictive modeling. Rather we propose to facilitate sharing of expertise, strategies, and creative problem-solving. The great strength of the current program is its diversity. Archaeologists and managers have struggled with problems and crafted solutions that are peculiar to their installation. Instead of a how-to manual, we propose offering procedures, techniques, and approaches that recognize this diversity and the need for each installation to tailor predictive modeling to its own situation.

REFERENCES CITED

Altschul, Jeffrey H.

1989 Modeling as a Management Strategy. In *Man, Models, and Management: An Overview of the Archaeology of the Arizona Strip and the Management of its Cultural Resources*. Jeffrey H. Altschul, and Helen C. Fairley. pp. 273–310. U.S.D.A. Forest Service and U.S.D.I. Bureau of Land Management, Washington, D.C.

1990 Red Flag Models: The Use of Modeling in Management Contexts. In *Interpreting Space: GIS and Archaeology*, edited by Kathleen M.S. Allen, Stanton W. Green, and Ezra B.W. Zubrow, pp. 226–238. Taylor and Francis, London.

1993 Discriminant Function Analysis. In *Eglin Air Force Base Historic Preservation Plan: Technical Synthesis of Cultural Resources Investigations at Eglin, Santa Rosa, Okaloosa and Walton Counties, Florida: Appendix V*, edited by Prentice M. Thomas, Jr., and L. Janice Campbell, pp. 275–296. Report of Investigation No. 192, New World Research, Inc., Mary Esther, Florida.

Binford, Lewis R.

2001 Where Do Research Problems Come From? *American Antiquity* 66:669–678.

BRW, Inc.

1996 *Draft Research Design for the Development of a High Probability Predictive Model for Identifying Archaeological Sites*. Prepared for Minnesota Department of Transportation, Minneapolis.

Campbell, L. Janice, Prentice M. Thomas, Jr., and William R. Mallory

2003 *Survey of X-666 (Task Order CR-02-0063) Cultural Resources Investigations, Eglin Air Force Base, Okaloosa, Santa Rosa and Walton Counties, Florida*. Report of Investigation No. 757. Prentice Thomas & Associates, Inc., Mary Esther, Florida (Draft).

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

1990 *The Archaeology of Regions: The Case for Full-Coverage Survey*. Smithsonian Institution Press, Washington.

Green, Ernestine L.

1973 "Location Analysis of Prehistoric Maya Sites in Northern British Honduras." *American Antiquity* 38:279–293.

Hall, Stephen A.

2002 *Field Guide to the Geoarchaeology of the Mescalero Sands, Southeastern New Mexico.* State of New Mexico, Historic Preservation Division and New Mexico Bureau of Land Management, Santa Fe.

Judge, W. James
1981 Transect Sampling in Chaco Canyon—Evaluations of a Survey Technique. In *Archaeological Surveys of Chaco Canyon, New Mexico*, by Alden C. Hayes, David M. Brugge, and W. James Judge, pp. 107–137. Publications in Archaeology 18A. National Park Service, Washington, D.C.

Judge, W. James, and Daniel W. Martin
1988 An Appraisal. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. James Judge and Lynne Sebastian, pp. 571–580. USDI Bureau of Land Management, Denver.

Judge, W. James, and Lynne Sebastian, editors
1988 *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling.* USDI Bureau of Land Management, Denver.

Kincaid, Chris
1988 Predictive Modeling and its Relationship to Cultural Resource Management Applications. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. James Judge and Lynne Sebastian, pp. 549–569. USDI Bureau of Land Management, Denver.

Kohler, Timothy A.
1988 Predictive Locational Modeling: History and Current Practice. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. James Judge and Lynne Sebastian, pp. 19–59. USDI Bureau of Land Management, Denver.

Kvamme, Kenneth L.
1995 A View from Across the Water: The North American Experience in Archaeological GIS. In *Archaeology and Geographic Information Systems: A European Perspective*, edited by G. Lock and Z. Stancic, pp. 1–14. Taylor and Francis, Ltd., London.

1999 Recent Directions and Developments in Geographical Information Systems. *Journal of Archaeological Research* 7(2):153–201.

Kvamme, Kenneth L., and Timothy A. Kohler
1988 Geographic Information Systems: Technical Aids for Data Collection, Analysis, and Display. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. James Judge and Lynne Sebastian, pp. 493–547. USDI Bureau of Land Management, Denver.

Plog, Fred T., and James N. Hill
1977 Explaining Variability in the Distribution of Sites. In *The Distribution of Prehistoric Population Aggregates*, edited by George J. Gumerman, pp. 7–36. Prescott College Anthropological Report 1. Prescott College Press, Prescott, Arizona.

Rose, Martin R., and Jeffrey H. Altschul
1988 An Overview of Statistical Method and Theory for Quantitative Model Building. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeologi-*

cal Predictive Modeling, edited by W. James Judge and Lynne Sebastian, pp. 173–255. USDI Bureau of Land Management, Denver.

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 Investigation No. 192, New World Research, Inc., Mary Esther, Florida.

Thoms, Alston V.
1988 A Survey of Predictive Locational Models: Examples from the Late 1970s and Early 1980s. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, edited by W. James Judge and Lynne Sebastian, Appendix pp. 581–645. USDI Bureau of Land Management, Denver.

Willey, Gordon R.
1953 *Prehistoric Settlement in the Virù Valley, Peru*. Bureau of American Ethnology Bulletin 155. Washington. D.C.

Zeidler, James A., Michael L. Hargrave, and Daniel Hagg.
2002 *Predictive Locational Modeling of Archaeological Resources on McGregor Range, Southern Tularosa Basin, New Mexico.* Tri-Services Cultural Resources Research Center, US Army Construction Engineering Research Laboratories, Champaign, IL.

Questionnaire Sent to Potential Participating Installations

Dear Cultural Resource Point of Contact,

The SRI Foundation has received a Legacy grant to evaluate archaeological predictive models that have been developed for Department of Defense installations. The project will involve working with installations that are interested in having their models evaluated and updated using data collected since model development. We will also be looking at how the model has been used, how well it has worked, and whether we can help the installation make the model more effective as a cultural resource management tool.

If your installation has an archaeological predictive model and if you would like to be considered as a possible participant in this study, please respond with answers to the brief set of questions below. From the responses, we will be selecting a set of installations that are geographically and environmentally diverse and offer a range of model types for further evaluation.

Thank you for taking the time to consider this request. We look forward to partnering with a variety of Department of Defense installations and to providing the installations with a technical product that will truly assist them in managing the resources under their stewardship.

Lynne Sebastian, Ph.D., RPA
Co-Principal Investigator
SRI Foundation
advancing historic preservation through education, training, and research

lsebastian@srifoundation.org

Questionnaire:

If you would like to be considered for participation in this project, please provide very brief answers to the following questions:

1. Installation name:
2. When was your archaeological predictive model developed?
3. Who developed it?
4. Is it still in use?
5. Do you have any documentation?
6. What is the platform – hardware and software?
7. How many sites were used to develop the original model?
8. What is the acreage of the study area?
9. How many sites have been located since the model was developed?
10. Who should we contact for additional information if you are selected to participate in this study?