Department of Defense
Legacy Resource Management Program

PROJECT NUMBER (11-158)

Institutionalizing Protocols for Wide-Area Inventory of Archaeological Sites by the Analysis of Aerial and Satellite Imagery

Douglas C. Comer
Cultural Site Research and Management

24 January 2014

Table of Contents

Abstract **3**
Executive Summary and Project Description **4**
Test Areas 5
Types of Sites 6
Review of Applicable Legislation 7
Problem Sets Addressed by this Project 7
Comparison of Archaeological Predictive Models and the Direct Detection Model 10
Development of the DDM **11**
SERDP Research 13
NCPTT Research 15
NASA ROSES Research 18
Legacy Project **19**
First Legacy Results 24
Legacy Results Using Point Cloud Sampling Protocols 25
DDMs for the Fort Irwin Test Areas 27
Applications of the DMB to Problem Sets 27
Notes on Using the ROC Curves for Decision Making 29
Protocols for the Use of the DDM 34
Cost Avoidance and applicability to all DoD Landholdings 41
Summary and Conclusion 46
Next Steps 46
Acknowledgements 47
Bibliography 48
Factsheet 50

Figures

Fig.1: Comparison of structural and non-structural sites 4
Fig.2: The three study areas 5
Fig.3: Typical open occupation site 6
Fig.4: Open archaeological site disturbed by vehicular traffic 7
Fig.5: Open archaeological site disturbed by traffic and erosion 7
Fig.6: Model produced by site prediction protocols developed during SERP research 15
Fig.7: Image layers used in the signature development analysis 16
Fig.8: Detection model, Santa Catalina Island 17
Fig.9: Composite images of signal detection layers 19
Fig.10: Site error and posterior probability maps, Santa Catalina Island 20
Fig.11: Schematic of statistical framework 22
Fig.12: Annuli method as image processor 23
Fig.13: ROC curve for the 36 band direct detection model 24
Fig.14: Visualization of point cloud sampling method 26
Fig.15: Chance diagonal 30
Fig.16: Fort Irwin Eastern Swath Results 31
Fig.17: Fort Irwin Western Swath Results 32
Fig.18: China Lake Swath Results 33
Fig.19: Fort Irwin Eastern Swath with previously surveyed areas overlaid 35
Fig.20: Fort Irwin Eastern Swath detail showing the correspondence between vehicle tracks and returns from the model 37
Fig.21: Fort Irwin Eastern Swath detail showing correspondence between high-return areas and site boundary polygons 38
Fig.22: Venn diagram showing the subset of sites most likely to be eligible for the NRHP 39
Fig.23: Venn diagram showing the subset of sites most likely to have been disturbed 40
Fig.24: Venn diagram showing the subset of sites most likely to be listed as ineligible for the NRHP 40

Charts and Tables

Chart 1: Previous Research, Key Objectives, and Accomplishments 13
Chart 2: Gain Statistics for Original Direct Detection Model 16
Chart 3: Number and percentage (of total land area) of probable habitation site pixels detected 18
Table 1: Applications of the DDM to Problem Sets 27

Institutionalizing Protocols for Wide-Area Inventory and Evaluation of Archaeological Sites by the Analysis of Aerial and Satellite Imagery:

Direct Detection Model (DDM) as a Decision Support Tool for Strategic Archaeological Survey and Evaluation of Discovered Sites

ABSTRACT

Archeological sites are, in almost all cases, places where human activity has produced an enduring change at a specific location in the natural environment. By analyzing that change, scientists can gain information about both human and environmental history. We report the results of our Legacy project here, by which we have developed the statistical treatments and computational capacity required to analyze data collected with a variety of remote sensing devices in order to detect environmental change associated with human activities. The sensors, statistics, and computational network that we employ can detect even subtle change. Sensors today are extraordinarily sensitive to chemical, vegetative, and structural difference. We have adopted and modified new statistical protocols developed by engineers and medical scientists over the last two decades to our task. The network of computational facilities at The Johns Hopkins University, the NASA Goddard Space Center, and the NASA Ames Research Center that we have established can effectively process the enormous volume of data collected by these devices. Aircraft and satellites carry the remote sensing devices that we employ. These sensors can collect data very rapidly over wide areas. By the analysis of this data, we produce maps of anthropogenic micro-environmental change by comparing environmental conditions at archaeological sites with conditions that obtain in the surrounding landscape. The maps are representations of a Direct Detection Model (DDM).

EXECUTIVE SUMMARY

We present here a decision support tool that offers substantial and immediate cost avoidance to the military by minimizing activities required to comply with the National Historic Preservation Act of 1966 (as amended), specifically Sections 106 and 110. The tool described in this report will provide:

- A substantial reduction in the number and intensity of required Section 106 archaeological surveys
- A practical way to find, categorize by age and materials, and evaluate sites and the context of those sites without the use of traditional survey methods
- A way to greatly increase training areas without fear of damage to archaeological resources
- The tool by which to protect the truly important archaeological sites on DoD lands
- *De facto* Section 110 surveys
- The means by which to strategically plan Section 106 Surveys by merging DoD training and facility needs with information about distribution and significance of archaeological sites
- A rich source of information by which to review, interpret, and improve upon the results of previously conducted archaeological surveys without conducting additional fieldwork, allowing for fewer mission delays.

All of the above, and more, rely upon the essential capacity that we have developed to detect the presence of materials peculiar to the types of archaeological sites that are, by far, most likely to be found on DoD lands. Most archaeological sites on DoD lands are not large scale structures made of stone, such as those that might be found at Mayan sites in Central America or Roman sites in Europe. In the United States, we do have archaeological sites that display remarkable architecture (e.g., at Chaco Canyon and Mesa

Verde). There are also enormously impressive, monumental earthworks (e.g., Poverty Point and Cahokia), and there are other sites (Ancestral Pueblo, Hohokam, etc.) that are smaller in size, but that are constructed from stone or mud brick that has endured. For each of these, however, there are tens—perhaps hundreds-- of thousands of archaeological sites that are comprised only of soils and lithic materials that have altered by human occupation and use. The Direct Detection model (DDM) can be used to find sites such as Poverty Point, Chaco Canyon, and the Ancestral Pueblo and Hohokam sites, but it can also be used find the soil and lithic sites that are much more numerous, and for that reason constitute the bulk of the compliance challenge facing DoD.

Archaeologists frequently refer to the all of the types of sites described above as either "structural" or "non-structural," even while understanding that some non-structural sites might have once contained relatively small shelters that were constructed from organic materials that have since deteriorated so completely that very little evidence of their existence can be detected. Schematics of these types of sites can be seen in Figure 1. The boundaries of a structural site, such as the one seen on the left side Figure 1, are established by finding the remains of a formal enclosure. Typically, no such clear demarcation of a non-structural site, such as the one on the right side of Figure 1, can be detected. More often than not, non-structural sites have been formed by repeated occupations of locations that are especially useful to humans, for one purpose or another. These occupations sometimes overlap, and are larger or smaller, containing more or fewer overlapping or clustered occupations, depending upon how great the attraction of the location for a certain type of human activity. The distribution of these sites resembles an oil slick: oil spilled on the ocean accumulates to varying depths, or can be absent altogether in places, as does archaeological material deposited by nomadic groups. The DDM that we have developed detects this distribution of archaeological material. Concentrations of archaeological materials constitute sites. When examined on the ground, archaeologists often disagree about precise site boundaries, as archaeological materials are distributed in greater or lesser densities. Vegetation and the expertise of archaeologist on the ground often affect the determination of site boundaries on this way. The DDM provides and more objective and empirical way of determining site boundaries. Often, the site boundaries seem to lump together sites that were found in previous ground surveys.

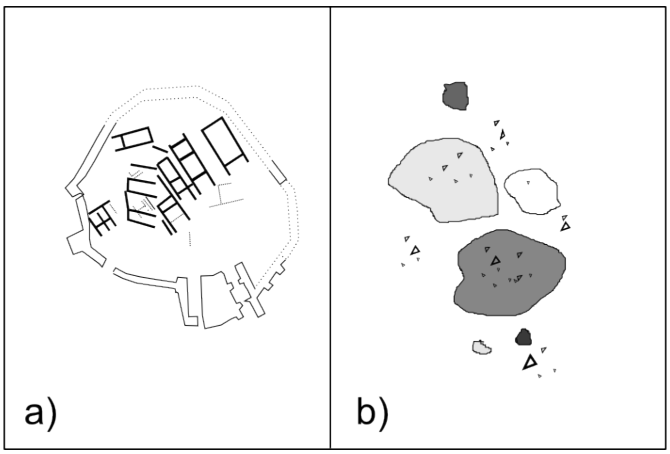


Figure 1: comparison between a) 'structural' and b) nonstructural sites. Non-structural sites, indicated by differences in soil chemistry, vegetation, and surface scatters of artifacts, represent the vast majority of sites in DoD lands.

Within Federal preservation law, Section 106 is treated as a compliance obligation, while Section 110 requirements are often unmet because of the extraordinary time and costs associated with a comprehensive, on-ground survey that, until now, has been the only means by which to comply with Section 110. Section 110 requires each Federal agency to establish "a preservation program for the identification, evaluation, and nomination to the National Register of Historic Places, and protection of historic properties." On the ground survey and evaluation in the service of this has proven prohibitively expensive. Because there are no clear timelines for completing Section 110 requirements, no Federal agency has completed survey and evaluation. From a preservation standpoint, this, we argue, is fortunate, in that the excavation that has been integral to most 110 evaluations is a destructive process. Yet in the absence of comprehensive survey and evaluation, important sites remain undocumented and unprotected, and there is no sound decision support prior to conducting 106 surveys for directing undertakings away from areas that are likely to contain important (significant, in terms of National Register criteria) sites. Archaeological predictive models (APMs) have in the half-century since they were first devised proven too unreliable for widespread adoption due to problematic assumptions and inappropriate statistical techniques. Also, they are incapable of differentiating between sites that are more likely to be evaluated as eligible for listing in the National Register from those that are not. In response to both the need for a wide area inventory and evaluation instrument and the deficiencies of traditional APMs, we present in this report a site detection protocol based on Bayesian statistical analysis of direct returns from a variety airborne and satellite remote sensors, which might include those that collect multispectral, hyperspectral, Lidar, and synthetic aperture radar data (here, we employ only multispectral and Lidar data). Images are used as data matrices. This direct-detection model (DDM) not only provides more rigorous probable site locations but, even more importantly, also allows for direct comparison of detected sites to sites known to be eligible for inclusion on the National Register of Historic Places.

Test Areas

The test areas for this project are located not only within China Lake Naval Air Weapon Station, which was originally identified as the test area, but also within Fort Irwin National Training Center. We established two test areas, Fort Irwin East and Fort Irwin West, within Fort Irwin National Training Center. In the adjacent China Lake Naval Air Weapon Station, we established one test area. The three test areas are seen in Figure 2. We began development of the DDM at Fort Irwin first because we anticipated challenges in developing a reliable and robust statistical treatment as we began adopting Bayesian statistics to this task, and additional challenges in constructing a computing network that could analyze extremely large volumes of data.

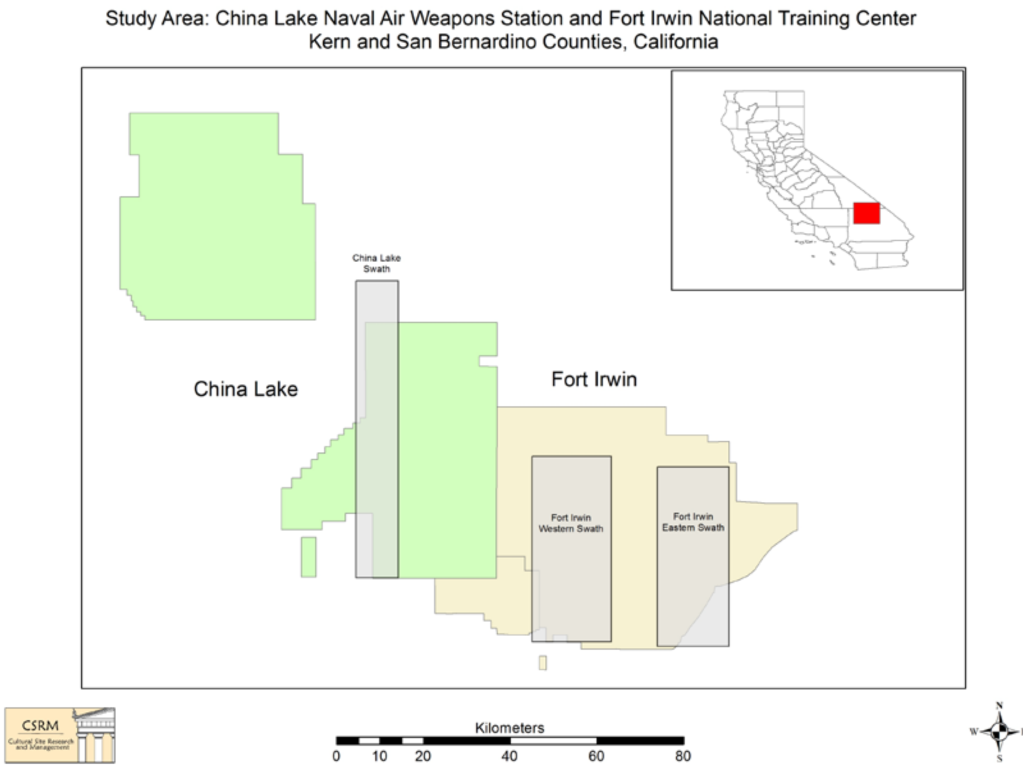


Figure 2: The three study areas.

We had relatively quick access to the data sets we required for Fort Irwin. In particular, these were site location data that we knew to be reliable, having collected it during previous Cultural Site Research and

Management, Inc. (CSRM) investigations at Fort Irwin. We also had already acquired precise Lidar DEM for Fort Irwin. It took considerable time before we could acquire what we could be sure were precise archaeological site locations for China Lake. And, although Lidar data had been collected over China Lake, we were never able to acquire it. Therefore, the DDM for the China Lake test area was ultimately developed without Lidar. As an important footnote, the sampling protocols that we ultimately devised for the statistical treatment utilized in the DDM do not demand that locations for archaeological sites that provide sample sensor returns be highly accurate. Nonetheless, we did not know this when we began investigations, and in any case we wanted to be sure that data we used to develop the DDM was as reliable as possible.

Types of Sites

Many different taxonomies have been used to categorize archaeological sites that are located within Fort Irwin and China Lake. Categories vary according to the theoretical orientation and research interests of the different groups of archaeologists who have conducted surveys at these two military landholdings. For example, large habitation sites have been divided into “Streamside Residential Bases” or “Lakeside Residential Bases;” they are essentially the same, but the archaeologists who used these terms were interested in how different sources of water might have influenced site distribution and the types of artifacts found at the sites. Places where humans might have camped briefly or even simply stopped to manufacture stone or other tools, gather plants or watch for animals have often simply been called “lithic scatters” or “camp sites.” The name refers to the flakes of stone that were left behind as tools were shaped. Occasionally, a tool broken during manufacture is found, or one that was used for some purpose and left behind is mixed in with these flakes of stone. Some site have been termed “detrital quarries,” when in fact there is often evidence that humans reoccupied such sites over hundreds or thousands of years. The detritus can contain lithic materials that have come from the site itself or from nearby locations and then shaped into tools. Therefore, while this type of site might be called a quarry, it might also very probably have been a site where humans stayed for days, weeks, or longer periodically. Stone circles seem to be a certain kind of occupation site, a type that displays, as the name implies, stones arranged in circles. All of the sites just mentioned are essentially open occupation sites. People probably lived at these sites for varying lengths of time.

Figure 3: Typical Open Occupation Site

In contrast, other sites are bedrock quarries, places where stone that is suitable for making tools was taken. Such sites, which are often away from water and on steeper slopes, might well have been occupied only briefly, during quarrying, and so are less like occupation sites. Other general types of sites at China Lake and Fort Irwin are rockshelters and caves, which are occupation sites but not open to the surface, and petroglyph sites. Petroglyph sites might or might not be found near occupations sites. They are places of ideological significance, where special activities, probably ritualistic in nature, were conducted.

In this project, we developed DDMs for open occupation sites. They are characterized by scatters and sometimes surfaces of lithic material, some of which is associated with ancient activities at the site. As mentioned just above, these have been identified by many different terms, but fall into two general categories, habitation sites and lithic scatters. Most of the sites at China Lake and Fort Irwin are open occupation sites. As importantly, these are the sites that are most affected by military activities, as most are found in relatively level terrain that is suitable for vehicular movement and development. Rock shelters and petroglyph sites are found for the most part in terrain with greater relief. While activities occur in such areas, the level of activity is much less, and vehicular movement and development poses less risk of damage to such sites.

Level of site disturbance varies from slight to profound. Figure 4 is a photo of a site that has been disturbed by vehicular traffic, but which retains a good amount of archaeological material. Figure 5 shows a site disturbed by vehicular traffic that has precipitated removal of archaeological materials by water erosion. Because the site is at a slightly lower elevation than the surrounding terrain, water has rushed through the area during rainstorms and carried archaeological materials away. Note the effect of this flooding on vegetation at the site, as well.

Figure 4: Open archaeological site disturbed by vehicular traffic.

Review of Applicable Legislation

A very brief review of the legislation and regulations will be used as a way to explain how the project proposed here can integrate the technologies and protocols developed by this investigation to bring DoD into full compliance with all applicable legislation, regulation, and policy relevant to archaeological resources quickly and with minimal expenditure of funds. It is also worth noting that, for reasons that will be described, DoD is in a unique position to provide a pathway to such compliance, which is both more effective and more cost-effective, for all other federal agencies

Figure 5: Open archaeological site disturbed by traffic and erosion.

The National Historic Preservation Act of 1966 (as amended) (NHPA) and the National Environmental Protection Act of 1969 (NEPA) were passed almost a half century ago. Section 110 of NHPA assigned responsibility to each federal agency to inventory lands it administered for the presence of archaeological and historic sites, and then to evaluate identified sites in order to determine if they were listed or eligible for listing in the National Register of Historic Places. Section 106 of NHPA requires a consultative process when an undertaking on federal lands or one with any federal involvement (for example, one that receives federal funding or required a permit from a federal agency) to avoid, minimize, or mitigate harm to properties listed on the National Register. An assessment of potential adverse effects is essential to this process. The location of sites eligible for listing on the National Register is, of course, necessary before this consultative process can begin.

Executive Order 11593 laid out the steps to be followed to implement NHPA in 1971, during the Nixon administration. These were further elaborated by 36CFR part 800. Several problems have come into clear focus in the almost half-century since NHPA became law:

Problem Set 1: The methods used to inventory and evaluate archaeological sites have proven to be too expensive, time-consuming, and unreliable to permit compliance with Section 110

1.1 No federal agency has fully complied with Section 110 of NHPA, which requires an inventory of archaeological sites and an evaluation of discovered sites to determine if they are eligible for listing on the National Register of Historic Places. The guidelines that have been established by each state were formulated before the development and common use of many powerful data collection and analytical technologies. Among these are geospatial positioning systems (GPS); aerial, satellite, and on-ground remote sensors; powerful personal computers; and sophisticated analytical software.

The exclusive reliance on field surveys has rendered the full completion of inventories of archaeological sites to be unfeasible in most cases. Approximately 30% of lands under the administration of DoD have been inventoried to date. In fact, most federal agencies have tacitly decided not fund Section 110 inventories, and to rely instead upon the Section 106 inventories that must be conducted prior to an undertaking. As mentioned in the introduction to this report, the DDM can serve as a Section 110 inventory.

1.2 Review of existing site inventories have revealed that the location of sites that were recorded without the use of GPS technology are almost invariably incorrect. Many sites are commonly found to be tens or hundreds of meters away from the location initially recorded.

1.3 This, in turn, reveals yet another, related problem: because sites are not located exactly where they were recorded to have been, it is difficult or impossible know if a site rediscovered is the recorded site for which one is searching, or a different site entirely. For example, if site X is not found at the location entered on a site inventory form, should a site found 60 meters away be labeled as site X? The challenge is compounded if yet another site is found less than 100 meters away from the location recorded on the site form.

1.4 The obvious approach to solving problem 1.3 is to re-examine the site to see if it corresponds to the site described on the site form. The quality and reliability of information on site forms varies, however. Site forms produced more recently generally contain more information. For the most part, this is because State Historic Preservation Offices have made the inclusion of more types of information mandatory over the years. Even so, line drawings of sites from a plan perspective can be sketchy or detailed, depending upon who filled out the site form. Photographs can be very helpful, although sites without very distinctive features (which are the majority of sites at China Lake and Fort Irwin) can still be difficult to identify.

1.5 Full documentation of archaeological sites discovered during inventories is therefore much more time-consuming than in previous decades, when forms were brief and did not require so much information. There is incentive, therefore, for surveys conducted through low-bid contracting to record as few sites as possible. This is not to say that sites are intentionally overlooked, only that in order not to lose money, field personnel must be very conscious of the need to move as quickly as possible. Low-bid contracting is a practice that produces what economists call a price-taking market. Ultimately, the only way to make a profit or simply stay in business is to reduce costs. Large blanket contracts, on which work orders can be written and costs negotiated, are not a solution to this reduction in quality. Large firms have large overheads, and large contracts granted to a firm that has established a network of other firms must bear the burden of overhead and profit taken by the organizing firm, which usually adds no value to the service provided.

1.6 The majority of sites found and documented on site forms during inventory surveys have not been evaluated in terms of their eligibility for listing on the National Register of Historic Places. There are good reasons for this. *Evaluation as it has been done has generally required excavation.* Excavation is time-consuming, labor intensive, and therefore expensive. Excavation is also a destructive process. Unless it is done with great care, much information will be lost. Done correctly, it often requires recording the locations of recovered artifacts *in situ*, taking a wide variety of samples that might prove to be informative (soil, pollen, charcoal, microfaunal, microfloral, etc.), rigorously analyzing recovered artifacts (e.g., identifying all lithic debitage as primary, secondary, or tertiary flakes, performing blood residue analysis, piecing together potsherds, etc.), analyzing all recovered samples, and preparing a report that carefully synthesizes all of this. *Even this does not fully compensate for the destruction of the site, because analytical tools in the future will be superior to the ones we use today*. Therefore, by removing archaeological materials from context, we lose all of the information that we might gain if excavation were to be done in the future. The tacit decision by many federal agencies not to evaluate by excavation until a site is actually threatened by development or some other agencies is therefore highly defensible.

1.7 Artifacts taken from the ground must be examined. They must be washed if washing will not remove materials such as blood residue that should be further analyzed, and examined again to be identified, labeled, catalogued, and then placed into storage. All of this is typically as time-consuming, labor intensive, and as expensive as the excavation of the material from the archaeological site. Storage must be in a climate-controlled environment, which is yet another substantial expanse, and one that continues. Storage requires not only construction, maintenance, and monitoring of the facility, but also personnel to do all of this in perpetuity

Problem Set 2: Archaeological Predictive Models (APMs) are unreliable as decision-support tools.

2.1 Archaeological Predictive Models (APMs) would seem to offer a way to plan, schedule, and predict costs for Section 110 and Section 106 survey and evaluation, but are seldom used for this because there is little confidence in their reliability. APMs have been formulated for fifty years. After initial enthusiasm for them on the part of both researchers and those with responsibilities for the protection and management of archaeological resources, interest waned because many, and probably most, proved to be so flawed as to be misleading. Many are unreliable because they are formulated from assumptions and statistical protocols that are inadequate. Data employed in development of models is frequently out of date or inaccurate, as discussed in Problem Set 1. APMs sometimes include unsupported assumptions about the relationships among environmental factors that have influenced site locations. Attempts to use statistics to establish such relationship more often than not employ inappropriate protocols (for evaluation of one such example, see Comer 2012). Even the rare use of more sophisticated tests of association between site location and environmental and social factors are plagued by suppression (Cohen and Cohen 1983:95–96), which is sometime due to redundancy (Cohen and Cohen 1983:115). Suppression occurs because some factors are more influential in site selection than are others. Depending upon the landscape and type of activity that took place at the site, the most important factor might be distance to water, soil type, elevation, or something else. Unless the influence of each factor is precisely calculated and then weighted correctly with all others, the influence of more important factors will not contribute as much as it should to the final model. This is difficult to do for many reasons, among them sampling error and the use of inappropriate statistical protocols. Redundancy might be an even more stubborn problem, because factors can be related in ways that are not apparent. There are often mediator variables (for example, soil water content can be influenced by distance to water, soil type, and other variables that might also be used in the model). Another continuing issue is that APM probability areas are sometimes tweaked to produce good results for a given installation or base by identifying high probability areas as those in which a substantially higher than average density of sites were found and low probability areas as those in which less than the average density of sites were found. While that might (or might not) provide a way to generalize about site location on the installation or base in question, the model has at that point been modified to fit only the specific region in question, and can therefore not be transferred to a different region with any certainty.

2.2 The mechanism used to evaluate reliability and productivity of APMs is likewise not informative enough to constitute a reliable decision-support tool. This mechanism has long been the gain statistic. A gain statistic is calculated as follows: 1 - % area / % sites; it is essentially a measure how much the density of sites in different areas deviates from what would be expected were all sites distributed randomly.

2.3 APMs do not provide the type of information of greatest use to those who must make decisions about placement of ground disturbing activities (e.g., certain types of training, construction, ordnance ranges). Placing a development or another type of ground disturbing activity in what in an APM appears as an area where there is low probability of encountering and archaeological sites does not mean that sites will not be encountered. When this occurs, as it will, during 106 surveys, the resources required to document, evaluate, and in some cases mitigate them can cause a cascade of unexpected delays not only in fieldwork, but also in planning, design, and construction, at ever-mounting expense.

Comparison of the Archaeological Predictive Models and the Direct Detection Model as Decision Support Tools

Before we compare the APM and DDM, we will provide definitions that are key to making the comparison, and then explain in more detail how the direct detection model works. Following the comparison, we will describe how the DDM can be used to address the problem sets described above, and solve many of the key problems in those sets.

Definitions

An Archeological Predictive Model (APM) ranks areas according to the likelihood that they will contain archaeological sites, based upon the association of environmental factors with the location of archaeological sites. Concisely: it *predicts where sites should be.*

A Direct Detection Model (DDM) is based on statistical analysis of signal returns from a variety of remote sensing devices carried by aerial and satellite platforms to identify the location of different types of archaeological sites. The manner in which the model is generated provides the means by which to assign probabilities that a site will be found at any particular location. Concisely, *it detects areas where sites are located.*

A ROC Curve is a Receiver Operating Characteristic graph. Such graphs were originally developed during the Second World War for detecting enemy objects on battlefields. It is now used in a wide variety of subjects and disciplines, including engineering, meteorology, psychology, and medical research. A ROC curve is used in medical research, for example, to provide an elegant display of the efficacy of a diagnostic test. It plots the relationship between actual positive results (called a True Positive Rate or TPR) and negative results (called False Positive Rates or FPR) for different tests, or of the same test using different variables or threshold (Metz 1978). At a time when resources are scarce and organizations are held ever more accountable for decisions, the ROC curve provides a means to assess the efficacy of procedures and tests.

A ROC curve runs from point 0, 0 where there are 0% TPR and 0% FPR to 1, 1 where there is 100% TPR but also 100% FPR. A diagonal line runs between these two points and this is called the chance diagonal. This line represents the expected random distribution of results. The closer a point or result is to this line, the more random it can be said to be. Conversely, if a point or points falls consistently well below the line, it is likely significant but the sensitivity and FPR need to be reversed. The efficacy of model productivity is measured by the Area Under the Curve (AUC).

DEVELOPMENT OF THE DDM

CSRM has been developing and improving a direct detection probabilistic model of archaeological sites with excellent results since 2004 (Comer and Blom 2007a, 2007b, 2007c; Comer 2008; Tilton, et al. 2012; Tilton and Comer 2013; Chen, et al. 2013). The essential elements of the approach that has yielded excellent results is as follows:

- Utilize a variety of remote sensing technologies, using each to characterize different aspects of the environment: e.g., soil chemistry, vegetative type and vigor, terrain, structure (broadly defined as regular geometric forms, clustering or pattering of objects of interest, surficial roughness and rugosity).

- Consider images that have been developed from data collected by aerial and satellite to be data matrices.

- Work backwards from pixel data in images to raw data that was collected by sensors. The products typically supplied by commercial vendors (as well as NASA) have been produced by manipulating raw data in order to produce visually appealing images.

- Correct for other factors that corrupt data sets (e.g., atmospheric conditions, radiometric changes produced by illumination and view angle, missing scan lines, sensor calibration, and other terrain effects) using the most effective of available protocols.

- Identify and modify as necessary statistical protocols that most effectively provide a probabilistic basis for differentiating between archaeological sites and non-sites, and more specifically, different types of archaeological sites and non-sites.

- Construct the framework within which this can be done with the greatest elegance: that is, with the fewest number of non-redundant statistical features.

- Model outputs in ways that can provide reliable decision support. This includes generating maps, charts, and tables.

In summary, our approach is to statistically compare return values collected by diverse sensors, thereby accessing very different kinds information about the microenvironment of the archaeological site and that of the surrounding environment. In our early work we used frequentist statistics to test for similarities and differences between archaeological site and surrounding environment. While outputs were outstanding in comparison to those provided by traditional archaeological predictive models, we identified drawbacks to frequentist approach. Foremost among them was the need to "bin" data, that is, to lump returns from sensors in to categories so that the frequency with which these categories were sensed at sites and non-sites. Binning data inevitably reduces precision. If data differences and similarities are subtle, binning might therefore produce errors at the classification stage of analysis. Also, the frequentist approach assumed a normal distribution for all sensed data, an assumption that is inherently problematic. Because of this, we now employ Bayesian statistics, which in our application does not require binning and does not assume a normal distribution. Individual archaeological sites are detected by the use of either frequentist or Bayesian statistics to degrees of certainly that can be stated, which is a vital consideration in the application of the detection model. Yet no matter how great the certainty, there is, of course, a corresponding level of uncertainty, and one relevant aspect of that uncertainty is that a location identified by the model as being the probable location of a site is in fact not a site. An additional enhancement to our protocols enables us to factor that uncertainty into the decision support tool: this is the use of the receiver operating characteristic (ROC) curve. As more fully explained below, the ROC curve track true positives, false positives, true negatives, and false negatives, and can therefore be used to estimate the probability that a certain number of sites should be found within a given area. Areas can range in size from a single pixel (2 meters square, in this case) to entire regions, to subareas within regions. This information can be used to develop time and cost estimates that can guide decisions as to where undertakings that would trigger 106 compliance should be placed, to most cost-effectively develop compliance strategies that will ultimately discharge 110 responsibilities, and to provide the basis for Memoranda of Understanding with State Historic Preservation officers that will minimize reporting obligations, facilitate negotiation of the most cost-effective survey and evaluation protocols, and in some cases to simply set aside areas where the most important archaeological resources are most probably located for protection until the time when much more informative non-destructive technologies can be used to investigate them.

Producing the Decision Support Model

CSRM, Inc., in collaboration with the NASA Jet Propulsion Laboratory at Caltech (JPL/NASA), NASA Goddard Space Flight Center, and Johns Hopkins University has been developing protocols for the practical use of aerial and satellite imagery in rapid, wide area archaeological survey since 2004. Our test areas have included San Clemente Island and Santa Catalina Island in the coastal waters of Southern California. The results that we report here used as test areas China Lake Naval Weapons Station and Fort Irwin National Training Center in the California Mojave Desert.

Our support has come from the Department of Defense Strategic Environmental Research and Development Program (SERDP), the National Center for Preservation Training and Technology (NCPTT), the NASA Research Opportunities in Solid Earth Science (ROSES) program, and the Department of Defense Legacy program.

Quite apart from the development of archaeological predictive models, there have been numerous attempts to detect archaeological sites by the use of remotely sensed data collected by aerial and satellite platforms. Some of these have been notably successful, yet all have been used to find sites that were monumentally large, and as importantly, discovery relied upon enhancing imagery in ways that allowed a viewer to see an archaeological site in the image. Our approach does not rely upon visual identification, but instead uses statistical analysis of pixel values to detect sites.

For example, Elizabeth Moore of the School of Oriental and African Studies at the University of London employed data from a 1994 NASA Spaceborne Imaging Radar C/X-Band Synthetic Aperture Radar (SIR-C/X-SAR) in in her studies of the Angkorian civilization in Cambodia. Using the various SAR bands and polarizations, Moore generated images of the locations of ancient Angkorian reservoirs and moats, many of which were not visible in optical imagery and had not been located on the ground. Since more recent reservoirs and moats were superimposed over older ones, she was able to rework a chronology of the Angkorian dynasties by associating reservoirs with the temple and palace complexes of succeeding dynasties (Moore 2000; Moore and Freeman 1996). William Saturno used multispectral imagery to identify areas where vegetation displayed less vigor in the jungles of the Peten, in Guatemala, and upon inspection, found that many such areas were the sites of Mayan temples (Saturno, et al. 2007). More recently, Arlen and Diane Chase used a high-density Lidar scan of Caracol, in Belize, to develop a comprehensive map of the ancient Mayan city (Chase et al. 2013). Yet, because these discoveries rely upon visual inspection, they admitted many false positives, although this is much less of a problem with the visual inspection of Lidar imagery. What can be reliably identified in Lidar images through visual inspection, however, are large structures and features, such as those found in a Mayan city. Lidar must be used quite differently to detect most of the archaeological sites that lie within the boundaries of the United States. These are non-structural, and are sites by virtue of artifacts scatters and soils below that contain archaeological materials.

Nonetheless, it was success in detecting large sites though visual identification by means of enhancing images generated by aerial and satellite remotely sensed data that provided the inspiration for the use of such data in statistical analysis to automate site detection. The chart below outlines steps in this process:

Chart 1: Previous Research, Key Objectives and Accomplishments

Project	Objectives	Accomplishments
San Clemente Island (DoD SERDP)	Develop protocols for wide area inventory of archaeological sites by statistical analysis of synthetic aperture radar (SAR) imagery	Statistical probabilistic approach developed that is much superior to image enhancement or classification approaches used in the past.
Santa Catalina Island (NCPTT)	Test transferability to very different environment and image types.	Transferability of probabilistic statistical protocols demonstrated.
Santa Catalina Island (NASA ROSES)	Automate protocols, reduce data analysis run time	Beta software reduces analysis run time from five hours to 10 seconds. Methods to “bin” data in ways that improve modeling developed.
China Lake / Fort Irwin (DoD Legacy)	Improve direct detection protocols by use of statistical learning method protocols in the analysis of images as data matrices. Make use of increasingly precise widely available Lidar and multi spectral imagery. Recommend actions that would mainstream the use of direct detection protocols.	Statistical learning method replaces frequency distribution test, which eliminates need to bin remotely sensed returns, thus increasing sensitivity of analysis. Use of ROC curve to identify most informative classifiers as the basis for elegant models.

SERDP Research

The Strategic Environmental Research and Development Program (SERDP) sponsored initial efforts to develop direct detection protocols (Comer and Blom 2007a, 2007b, 2007c, Comer 2008). San Clemente Island, located approximately 125 kilometers from the Southern Californian mainland, was the test area. Primary data sets (features) were taken from synthetic aperture radar (SAR) imagery. Data were harvested from the areas in images occupied by archaeological sites of different kinds and statistically tested for association with them. Later in the research, we also used multispectral imagery. Sampling for the statistical tests was from a small subset of the hundreds of archaeological sites for which very accurate locations had been established. Features could be individual bands of data collected by SAR or multispectral remote sensing devices, as well as bands of data generated by algorithms that combined other bands of data. As an example of a multiband feature, the four bands of multispectral data collected by the IKONOS satellite were used to calculate a normalized difference vegetation Index (NDVI), which is developed by subtracting the red (or similar) band from the infrared band from and then dividing by the near infrared plus red (or similar) band, or NDVI = (NIR — VIS)/(NIR + VIS), where VIS is generally red or red edge. The NDVI algorithm is a band difference ratio that, among other things, normalizes the data set produced. All spectral images used were corrected for atmospheric interference and to original radiance values prior to analysis.

In this and all later tests, bands provided data relevant to different types of site characteristics. These included the physical structure of soils, rocks, and vegetation at archaeological sites, which affected SAR returns, as well as the chemical composition of such materials, as characterized from multispectral returns. SAR data also varied with dielectric property (the capacity of materials encountered to conduct electricity) which was affected by soil moisture. Finally, SAR data could be used to characterize structure on a grand scale (as it did on the small and micro scales). A digital surface model was generated by interferometric analysis of certain SAR wavelengths. With this surface model, a slope model and view shed models were generated that provided additional features. SAR of three different wavelengths generated data. These were the C-band, with a wavelength of about 5 cm, L-band, about 25 cm in length, and the P-band, approximately 75 cm centimeters long. Each of these bands could be polarized at both transmission and reception. Depending upon the length of the band and how it was polarized, the band was scattered in ways that affected return values according to the size and orientation of materials encountered. Radar bands are scattered only by objects larger than 1/4 to 1/3 of their length, and are strongly affected by shape and orientation. An object, such as a blade of grass, which is oriented vertically, will produce a strong return for bands polarized vertically at both transmission and reception. Returns will be weak for grasses if bands are polarized horizontally at either transmission or reception. Returns from the set of bands used by AirSAR could differentiate among objects based on differences in size and orientation. This set of bands was complemented by yet another that was obtained by the Geo-SAR platform, which carried P-band and X-band, both polarized vertically and horizontally. X-band is about 3 cm in length. The familiarity and experience gained with SAR technology informs the project proposed here.

The Student's *t*-test was used to test for association with archaeological sites, as follows: Our null hypothesis was that there is no difference between the population of values that lie within site boundaries and the population of values that lie outside site boundaries:

$$H0: u^1=u^2$$

If the null hypothesis were upheld, there would be no significant statistical difference between pixel values from associated with archaeological sites and those associated with non-sites. If, on the other hand, the null hypothesis was disproven, pixel values associated with archaeological sites would be statistically different from those taken from non-sites.

We tested this with the formula $(\sum_n^1 \frac{x^1}{n}) - (\sum_n^1 \frac{x^2}{n}) < 1.96\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}$, which is to say that the difference of the means of the site and randomly selected samples (usually 15), or $(\sum_n^1 \frac{x^1}{n}) - (\sum_n^1 \frac{x^2}{n})$, will be less than 1.96 standard deviations apart, with the standard deviation of the difference of means being determined by the elementary formula: $\sigma_{x_1-x_2} = \sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{n}}$

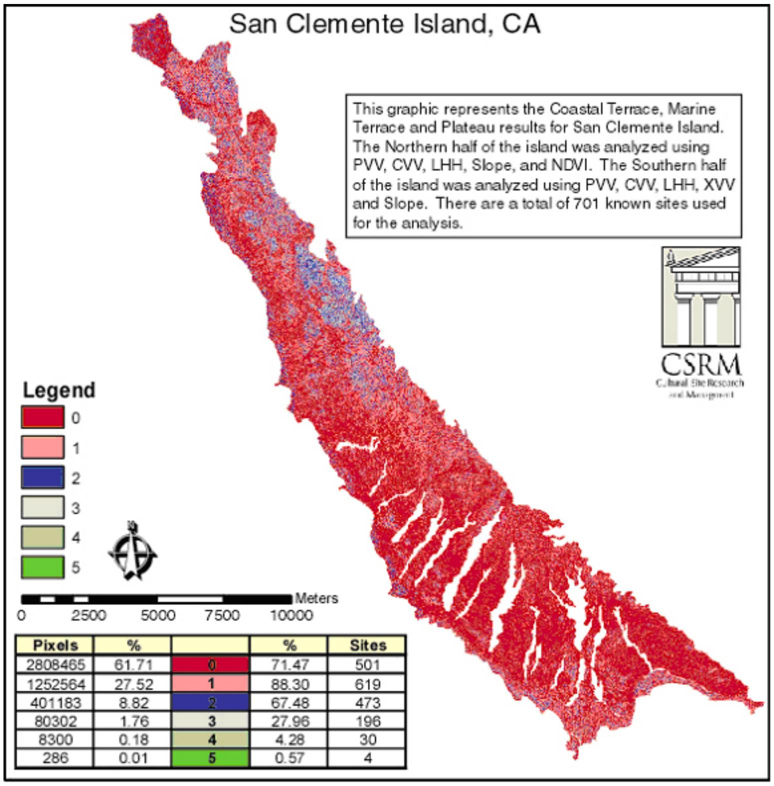


Figure 6: Model produced by site detection protocols developed during SERDP research.

In this way, we identified pixels containing feature values that were logically associated with archaeological sites, and the probabilistic parameters for the association. The greater the number of features statistically associated with a certain location, the stronger the overall statistical association. The final product was a map, or model, developed from numbers of features associated with sites (Figure 6). A test of model productivity employed hundreds of sites that had not been used to statistically test associations. The number of sites found was calculated and expressed as the percentage of the total universe of sites, then compared to the percentage of the survey area covered by pixels for which statistical associations with sample sites had been established. The ratio of the percentage of sites found and the percentage of the survey area covered by statistically associated pixels generated a *gain statistic*, calculated as 1 - % area / % sites (Kvamme 1988). Gain statistics varied according to how many features (data sets) were used to detect sites. Unsurprisingly, the fewer the features, the greater the percentage of sites detected, but the lower the gain statistic. This can be seen in the key to Fig. 6. Gain statistics for the direct detection model are shown in Chart 2.

NCPTT Research

Our first test of direct detection statistical protocol transferability was made with support from the National Center for Preservation Technology and Training (NCPTT). The test area for this was Santa Catalina Island, approximately 20 km northeast of San Clemente Island. Whereas 13 marine terraces and a central plateau, which provide large areas suitable for human occupation, are the main geomorphological features at San Clemente Island, ridges and valleys dominate the terrain of Santa Catalina Island. Despite the fact that the two islands are very near, Santa Catalina Island had a much larger population than San Clemente Island before European contact, as it receives twice as much precipitation, which feeds streams that flow into several large coves. Densely populated villages, organized around the habitations of chiefs and ceremonial areas, were located in coves. In contrast, no villages have been found on San Clemente Island.

Chart 2: Gain Statistics for Original Direct detection Model

Number of features in model	Percent of survey area	Percent of sites detected	Gain statistic
1	27.52	88.3	0.688335221
2	8.82	67.48	0.869294606
3	1.76	27.96	0.937052933
4	0.18	4.28	0.957943925
5	0.01	0.57	0.98245614

Features that provided data for the statistical analysis were formed from the red, green, blue, and near-infrared bands of the IKONOS satellite, and again red and near-infrared were used to generate a normalized difference vegetation Index (NDVI). NASA provided X-band data collected by the GeoSAR platform. The X-band was polarized vertically on transmission and reception and horizontally on transmission and reception. The X-band data were also analyzed interferometrically to produce a surface model, and a slope model was developed from the surface model.

Once more, the outcome was excellent as measured by the gain statistic (Comer 2007). When any two of the three features indicated the presence, the percentage of the survey area covered by the probabilistic model was 5.48%, while the percentage of sites detected was 91.6%, for a gain statistic of .94. When all three of three features indicated the presence of archaeological sites, these figures were 0.26% and 33%, yielding a gain statistic of .992. In this case, it is likely that the performance of the model benefited from the peculiarities of the Santa Catalina Island landscape; there is a high likelihood that any part of Santa Catalina Island with relatively level terrain would have been occupied at one time or another because most terrain is not level. Nonetheless, these results are also consistent with the general statistical principal that an elegant model is generally a better model; in this case, only three features produced an extremely productive model.

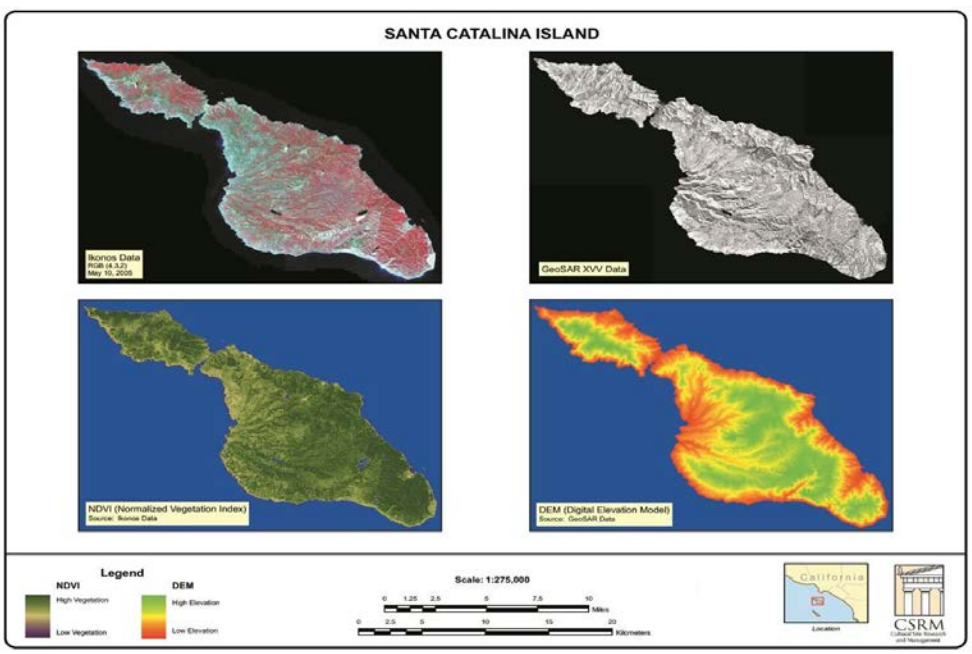


Figure 7: Image layers used in the signature development analysis. Upper left: IKONOS image in false-infrared layer stack (reds indicate vigorous vegetation). Lower left: NDVI analysis if IKONOS image (analytical statistics were run on this image). Upper right: the XVV image, created from XVV backscatter (analytical statistics were run on this image). Lower right: a digital elevation model (DEM) created through Interferometric analysis of XVV backscatter (analytical statistics were run on this image).

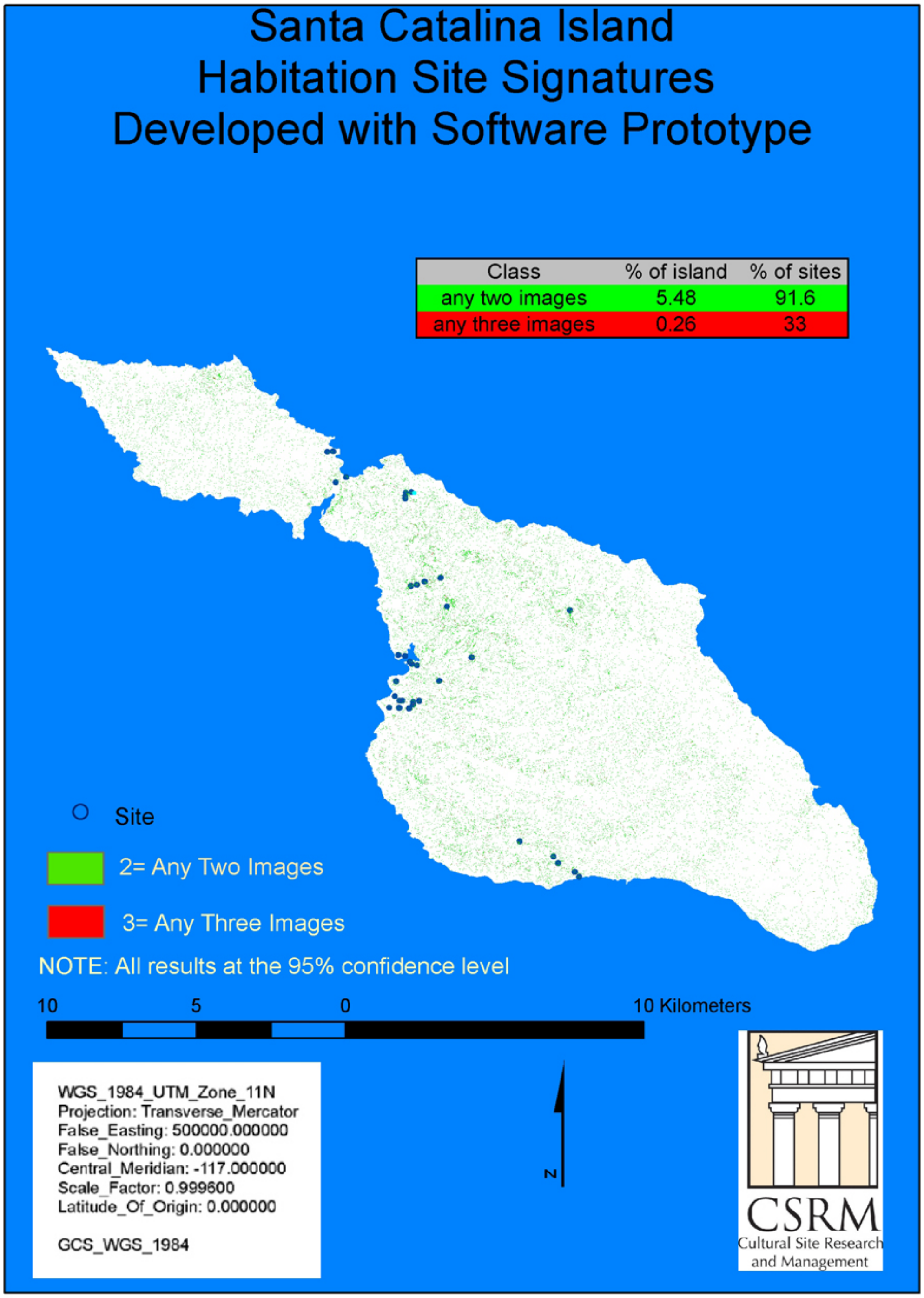


Figure 8: Detection model, Santa Catalina Island. See text for explanation.

NASA ROSES Research

A NASA Research Opportunities in Solid Earth Science (ROSES) grant was used to produce stand-alone software (written in C++) and to test a variety of data quantization schemes (Tilton and Comer 2013). The test area was again Santa Catalina Island. The software was designed to accommodate the analysis of imagery comprised of pixels that varied greatly in size, from the approximately 0.5 m pixels that make up IKONOS and Digital Globe satellite imagery to the 15 to 90 m pixels in Landsat and ASTER images. Statistical calculations by means of the software could be carried out in 10 to 15 seconds rather than the more than five hours required previously, which meant that many more tests could be conducted and that were used to better define sampling areas, delineate environmental zones from which sets of sites were sampled, refine statistical protocols, and develop quantization methods.

There are many types of data quantization approaches. Some of these are basic ones common to image analysis and enhancement software, such as histogram or linear stretches. Others are based on knowledge of the sensor characteristics. We used Recursive Hierarchical Segmentation (RHSEG) software developed by Co-PI Dr. James Tilton to identify discrete ranges of pixel values to identify classes that were used to compare population frequencies. Gain statistics realized with the new software equaled those realized during the NCPPT research. Results were even more impressive than those obtained for the San Clemente Island model. As seen below, 100% of known sites were found in the 9% of the survey area where one or more feature were statistically associated with sample sites, for a gain statistic of .91. The gain statistic rose with the number of features that were statistically associated with sample sites, from approximately .95 with two or more features, to a near 100% gain statistic with nine or more features (Tilton and Comer 2013):

Chart 3: Number and percentage (of total land area) of probable habitation site pixels detected and % known sites detected utilizing 192 level RHSeg based quantization plus proposed concentration and gain factors:

Number of detections	Number of pixels detected	% detection (out of 12,873,517 pixels) (%)	% Known sites detected (%)	Concentration factor[a]	Gain statistic[b]
Nine or more	691	0.0054	2.78	517.51	99.81
Eight or more	7,304	0.057	8.33	146.88	99.32
Seven or more	31,663	0.25	19.44	79.06	98.74
Six or more	79,499	0.62	36.11	58.48	98.29
Five or more	153,967	1.20	52.78	44.13	97.73
Four or more	260,199	2.02	66.67	32.98	96.97
Three or more	410,799	3.19	88.89	27.86	96.41
Two or more	639,730	4.97	97.22	19.56	94.89
One or more	1,162,841	9.03	100.00	11.07	90.97

[a]Concentration factor = % known sites detected/% detection

[b]Gain = 1 − (% detection/% known sites detected)

The site detection model image is seen below:

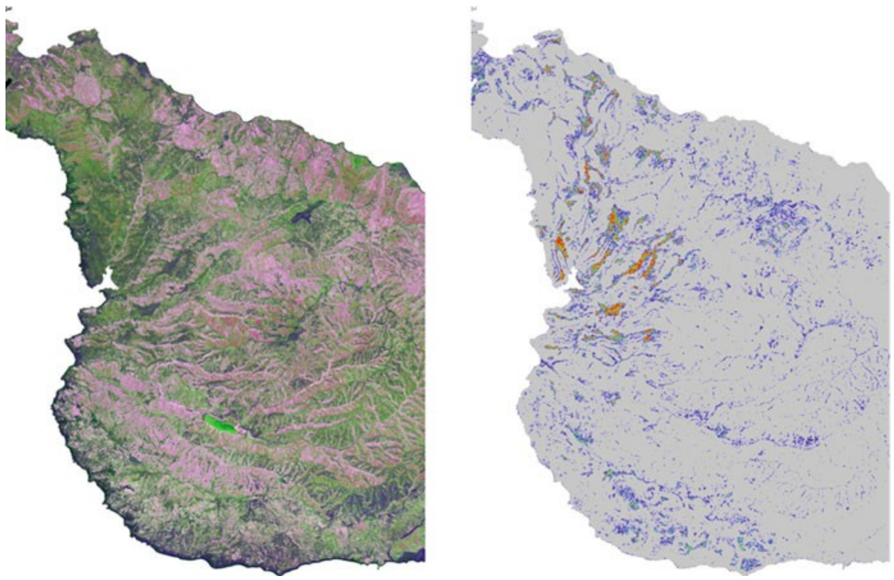


Figure 9: (a) An RGB representation of some of the analyzed image features: *Red* = TC , *Green* = TC , and *Blue* = NDVI. All are displayed with 255 level RHSeg quantization. (b) Colored coded display of the number of positive detections of probable habitation sites: 1 detection = *blue*, 2 detections = *magenta*, 3 detections = *cyan*, 4 detections = *green*, 5 detections = *yellow*, and 6 or more detections = *red*.

Also, the model developed by means of the new software in the RHSEG quantization appeared superior in several ways. Signature patterns were more consolidated. Therefore, the model was more easily understood; that is, it displayed greater interpretability. The model was tested during our current Legacy project, described below, and the results of these tests suggested important improvements to the direct detection protocols.

LEGACY Project

The objective of our current Department of Defense Legacy project is to put our modeling protocols into use for cultural resource management (CRM) purposes. From a CRM perspective, objectives go beyond detecting where sites are located. With DDM model, we also evaluate sites in terms of eligibility for listing on the National Register of Historic Places. Even so, future applications of our modeling are very obvious and numerous.

Improvements to model verification and interpretability began with an analysis of the Santa Catalina Island model. Figure 9 below displays a site error map (left) and a posterior map (right) based upon results of CSRM direct detection research at Santa Catalina Island. The site error map shows known correct, known wrong, random correct, and random wrong. To posterior probability is the outcome of the classification protocols. The site error map suggested the use of receiver operating characteristic (ROC) curves, which are generated from true positives, false positives, true negatives, and false negatives. Collecting these requires field verification. The ROC curve became integral to the improvement of protocols because it can be used to both test for the most informative classifiers and to generate posterior probability map tailored to compliance objectives.

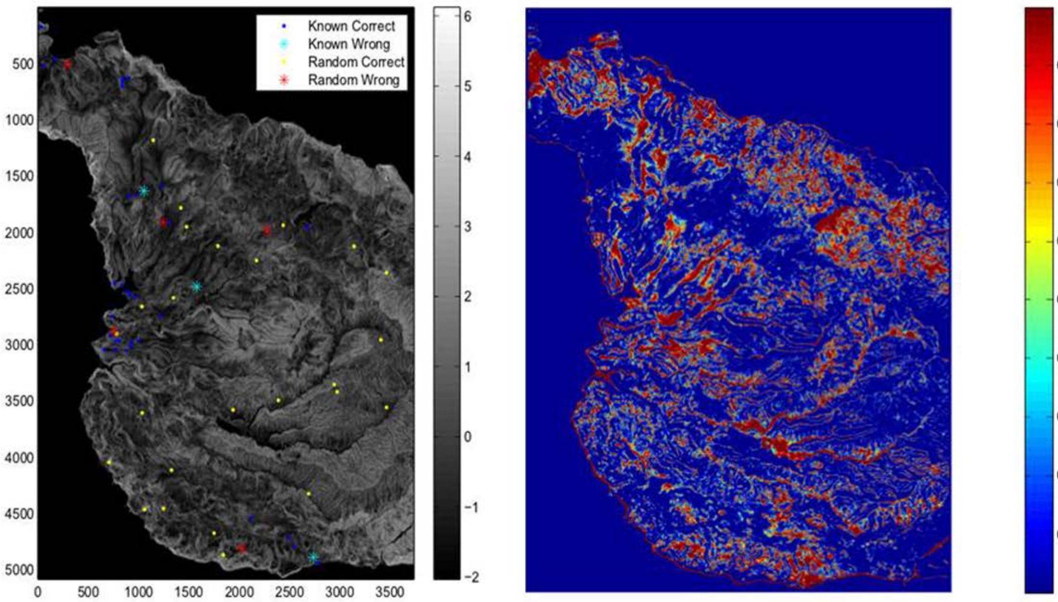


Figure 10: Site error and posterior probability maps from detection-based modeling results, Santa Catalina Island.

ROC curves

Receiver Operating Characteristic (ROC) Curves were developed during the Second World War for detecting enemy objects on battlefields. They are now used in a wide variety of subjects and disciplines, from psychology to medicine. A ROC curve can be used to assess the efficacy of a detection system (as it is in missile defense systems) or diagnostic test by plotting the relationship between actual positive results (called a True Positive Rate or TPR) and negative results (called False Positive Rates or FPR) (Metz, 1978). A ROC curve extends from point 0, 0 where there are 0% TPR and 0% FPR to 1, 1 where there is 100% TPR but also 100% FPR (see, for example, Figure 14). A diagonal line runs between these two points, which called the chance diagonal. This line represents the expected random distribution of results. The closer a point (result) is to this line, the more random it can be said to be. Conversely, if a point or points falls consistently above (or, more specifically, to the upper left) of the line, the productivity of the mechanism being tested is confirmed, the farther the points from the diagonal, the more productive. If points or points are well below the line, the FPR can be reversed to render the model productive.

In medicine, ROC curves explore the relationship between sensitivity (the ability of a test to diagnose correctly a patient with a disease) and specificity (the probability of a positive test result for a patient without a disease). Sensitivity can be understood as the TPR and specificity as FPR. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The area under the ROC curve (AUC) is a measure of how well a parameter can distinguish between two diagnostic groups (diseased/normal). Therefore, the overall productivity of the test remains constant, although the rate of true and false diagnoses can vary depending on the threshold set. This threshold defines whether a person has a specific condition or disease (Obuchowski 2005). As the threshold is lowered, there is a greater chance that the test will correctly diagnose patients (TPR); however, there is also a greater chance that people will be misdiagnosed (FPR). Conversely, if the threshold is raised, fewer people will be correctly diagnosed but there will also be fewer misdiagnoses.

In recent decades, ROC curve analysis has been used in meteorology to assess the predictive capability of forecasting tests (Mason and Graham 2002). Hit-rates (TPR) and false alarms (FPR) can be plotted on the ROC curve. Here, too, the efficacy of a test is defined not by the curve itself, but by the area under the curve. A value greater than .5 is deemed to signify a successful test, the greater the value, the more successful. Many other examples could be given; the ROC curve is used in many fields, but has not before been applied to archaeological modeling.

ROC curves in the Legacy project

Posterior probabilities are found in the area under the curve, which is in the lower right-hand portion of the chart. The development of a ROC curve for archaeological modeling requires field confirmation not only of sites, but also of non-sites. Fortunately, CSRM, in preparation for the Legacy project, had bid on and won a contract with Fort Irwin in the year before the Legacy project was to begin. Under the terms of this contract, CSRM was to examine 149 known sites and determine their eligibility for listing on the National Register of Historic Places. This provided the Legacy team with a very reliable data set showing precise locations of sites and non-sites. All 149 sites were revisited, locations for them were reestablished, and sites were classified based on first-hand examination. Fort Irwin is located adjacent to, just to the southeast of, our main test area, China Lake. To explain how ROC curves have been used to improve productivity of probabilistic site detection modeling and to increase interpretability of models, the protocols are outlined below.

Protocols Developed Under Legacy Award

Whereas the original statistical treatment was the Student's *t*-test, which compares relative frequencies, the protocols in our current Legacy project utilize statistical learning methods. There are four steps in the protocols: Image Pre-processing, Image Processing, Feature Extraction, and Classification. In mathematical notation, our classification scheme consists of three steps: image processing *f*, feature extraction *g* and classification *h*.

The images that we analyzed are for our purposes Z two-dimensional arrays. Each of these we call a feature image.

<u>*Image Pre-processing*</u>

Images are comprised of pixels that contain numbers related to sensor returns. As received, images contain data that are not optimal for image processing. We correct data for the influence of atmospheric conditions, sensor malfunctions, and previous binning of data done to develop images that are visually appealing. Image pre-processing must be done in order to generate a matrix of data that provides sensed returns uncorrupted by these various agencies.

<u>*Image Processing*</u>

Annuli are used to extract values from each component (or data set; in earlier research these sets were called features) of the model associated with site and non-site locations. For each site, 30 annuli are calculated, each with inner and outer radii. Annuli grow large as they move outward from the site center, and some overlap. For each annuli, we estimate the median and median absolute deviation (MAD) using the formulae: $\nu_i(x,\mathcal{B}) = \text{median}\{\mathcal{B}_{x'} : x' \in A_x(r_i^{(in)}, r_i^{(out)})\}$ and $\delta_i(x,\mathcal{B}) = \text{MAD}\{\mathcal{B}_{x'} : x' \in A_x(r_i^{(in)}, r_i^{(out)})\}$ where $A_s(r^{(in)}, r^{(out)}) = \{s' \in \mathcal{S} : r^{(in)} \leq \|s - s'\| < r^{(out)}\}$. This method is rotation invariant and its validity is held by the bias-variance tradeoff. It also has the highly beneficial effect of eliminating the need to quantize or otherwise bin sensed values, as would be done if frequency counts were the basis of statistical analysis. The statistical manipulation is extremely taxing on computing resources, however.

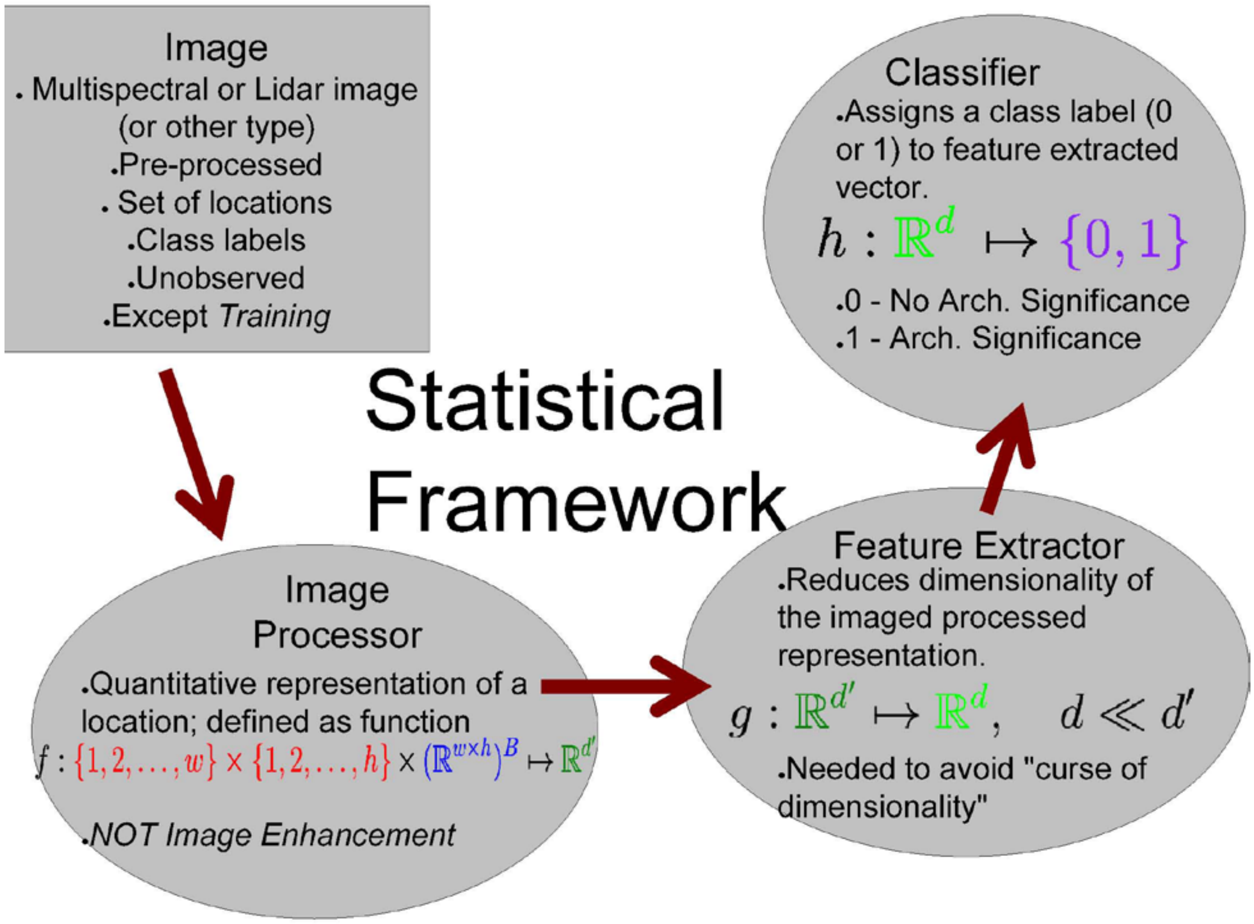


Figure 11: Schematic of Statistical Framework

The image band has 60 dimensions because of the annuli technique (30 *2), calculated as follows: Each annuli generates two statistics: median and MAD, and there are 30 annuli. There are as many has 36 bands so total dimensionality is 60x36, or 2160. For this reason, the annuli approach is extremely taxing on computational capacity. Therefore, the image processor now utilizes a point cloud approach, which generates much less dimensionality, described below.

Feature Extraction

Principal Component Analysis (PCA) is used to reduce the number of dimensions that must be considered for each site. For example, there are 33 surveyed and 100 unsurveyed lithic sites, for a total of 133. Dealing with enormous dimensionality of the direct detection image analysis is an example of what is commonly called the *curse of dimensionality*. In general, a feature extractor is a function such that $g : \mathbb{R}^{\tilde{d}} \to \mathbb{R} \ with \ d < \tilde{d}$, reducing the original dimension d to $\tilde{d}$. The extractor of the principal component analysis is defined by $g(x) = \tilde{U}^T x \in \mathbb{R}^d \text{ for all } x \in \mathbb{R}^{\tilde{d}}$, where $U \in \mathbb{R}^{\tilde{d} \times d}$ is a matrix whose columns are orthogonal eigenvectors associated with eigenvalues from the data covariance matrix.

We apply leave-one-cross validation to select the optimal PCA dimension corresponding to the minimum error rate for each surveyed site.

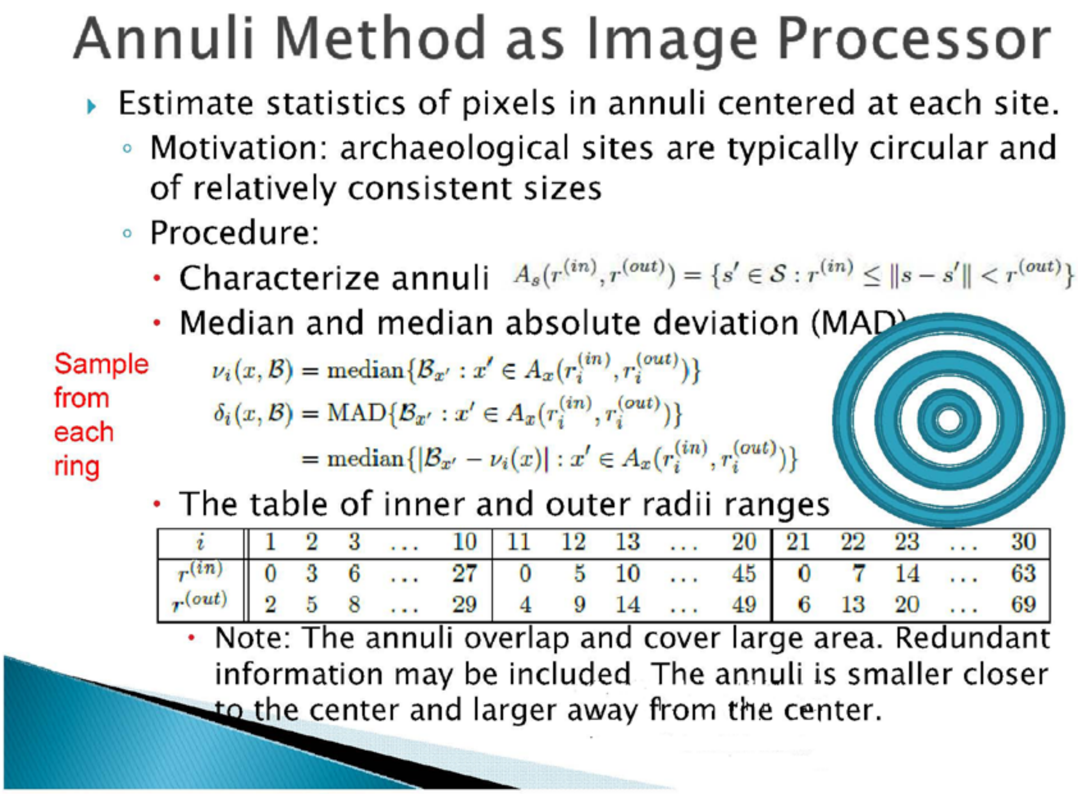


Figure 12: Annuli Method as Image Processor

Classification

Linear discriminate analysis is a supervised learning method, in this case, one that does not concern empirical probabilities. We estimate the mean and covariance by

$$\hat{\mu}_y = \frac{1}{n_y} \sum_{i:Y^{(i)}=y} X^{(i)} \quad \text{and} \quad \hat{\Sigma} = \frac{1}{n-2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X^{(i)} - \hat{\mu}_{Y^{(i)}})(X^{(j)} - \hat{\mu}_{Y^{(jT)}})^T$$

The posterior probability from Class 1 is defined as

$$\hat{\eta}(x) = \hat{\mathbb{P}}(Y = 1 | X = x) = \frac{\phi(x; \hat{\mu_1}, \hat{\Sigma})}{\frac{n_1}{n}\phi(x; \hat{\mu_1}, \hat{\Sigma}) + \frac{n_0}{n}\phi(x; \hat{\mu_0}, \hat{\Sigma})}$$

Thus, the classifier itself is defined as $h(x) = \mathbb{I}\{\hat{\eta}(x) > \tau\}$. It uses instead predicted, or conditional, probability. If a pixel displays a value (normalized) of greater than .5, the pixel tentatively represents an archaeological site. The optimality of the classifier is selected based on leave-one-out cross validation. Thus, the algorithm for composing the feature extraction with LDA is an inner and outer loop of leave-one-out cross-validation.

The ROC curve shows our algorithm performs well on the given surveyed sites in the western region of Fort Irwin, which was then used to train the classifier on the entire set of surveyed sites and test it on all of the western region. This procedure will generate a posterior probability map, where each pixel is assigned a probability score for being a site or not. Each pixel gives us 60x36 dimensions (=2160) using Worldview2 multispectral bands and slope data. The band images we analyzed consisted of 23,000 columns and 9,859 rows, presenting 226,757,000 pixels, each of which is loaded with 60x36, or 2160, calculations. For this reason, the map of the complete study area will be produced after the model has been improved as much as possible.

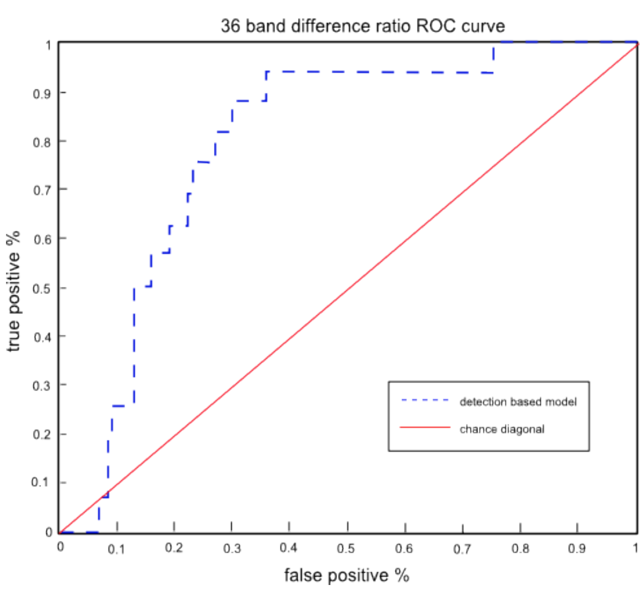
First Legacy Results

The first results obtained through this Legacy project were presented in Chen, et al. (2013). The machine learning algorithm just mentioned was used to generate many different sets of components, and many of these have been evaluated by means of ROC curves.

Data Sets

The components that we used for the first application of our direct detection protocols were obtained from high-resolution, multiband, multispectral WorldView-2 (WV-2) imagery and from a Lidar point cloud. The multispectral bands supplied by the WV-2 are: 400-450 nm (coastal), 450-510 nm (blue), 510-580 nm (green), 585-625 nm (yellow), 630-690 nm (red), 705–745 (red edge), 770–895 (near IR-1), and 860-900 nm (near IR-2). From the point cloud, we developed a high-precision surface model for Fort Irwin, and the surface model was then used to generate a slope model. A machine learning program was used to process the eight multi spectral bands and the slope model, using this algorithm to generate a band difference ratio: $F(bij) = [bi] - [bj] / [bi] + [bj]$ for all i > j N=n(n=1)/2

The most productive set of features were produced by generating 36 band difference ratios (BDRs) using the eight multispectral WV-2 bands, and slope (normalized values). The blue dashed line seen in Figure 12 shows performance of the resulting model. Given this performance, we may set a desirable threshold to gain a much lower false negative rate. By this threshold, the posterior probability map generated by training on all the surveyed sites will identify an area within which 90% of the sites detected will in fact be sites, and 30% of sites detected will not be real sites. Alternately, a posterior probability map can be generated that delineates an area within which almost 50% of detected sites will be true sites, and 15% of detected sites will not be real sites. These results were very significant, but computing time was very high.



A guiding principle in our project development is that adding components does not necessarily yield a higher performing model. Previous models used 12 features: the eight WV-2 multispectral bands, the slope, and three bands developed by the Kauth-Thomas Tasseled Cap Transformation. These bands are typically used to highlight wetness, greenness, and brightness in a landscape. All 12 bands were used to calculate band difference ratios, producing 66 of them.

In general, as in this example, adding more components risks masking of the effect of one variable (component) with that of another. This is called *suppression* (Cohen and Cohen 1983:95–96); *redundancy*, which might be collinearity or

Figure 13: ROC curve for the 36 band direct detection model.

multicollinearity, is also possible (Cohen and Cohen 1983:115). More generally, these phenomena can be termed *overfilling*. Larger numbers of data sets can include those that are informative and useful, but whether or not this is the case also introduce “noise." Therefore, an elegant model as a rule is superior to a more complicated one.

Model complexity also introduces *bias variance trade-off*. As a rule, as bias decreases, variance increases, and as variance increases, bias decreases. A statistical model that correctly identifies 100% of the objects that one would like to detect has virtually eliminated bias, but only for the set of data that was used to develop the model. In terms of archaeological sites, this would mean that the model would perform perfectly to detect all of the sites used to build a model, but would do less well when used to detect sites that differed even slightly from those used to produce the model or in a different environment. This limits the utility of the model in other areas.

As this applies to the Fort Irwin direct detection model, there are many more dimensions for the 66 BDR (3960) compared to the 36 BDR model (2160). Many generated dimension are redundant, which is indicated by the feature extraction step; in fact, PCA indicates that only a few combinations of the dimensions are useful. A classifier is not successful when too many features are added because they cause too much noise in the data. Moreover, because of the bias-variance tradeoff, if we find the optimal classifier for such noisy training data, the classifier cannot perform well for the testing data.

While it is standard to conduct radiance and atmospheric correction on multispectral remote sensing imagery as received from the vendor, this is not necessary with WV-2 multispectral imagery, which is delivered by Digital Globe after these corrections are made. This is not widely appreciated by archaeologists who have experimented with WV-2 images. Previous modeling attempts produced inferior ROC curves after two standard atmospheric correction protocols were employed on the WV-2 imagery; because they were redundant, performance was degraded. This is an example of how ROC curves are valuable in not only testing models, but also in evaluating data sets used in models.

Test areas

The dimensions of our Fort Irwin East Test Area were 41.1 x 16.5 km. The 21 sample sites used included 14 eligible sites and seven sites in good condition, but ineligible.

The dimensions of our Fort Irwin West Test Area were 42.6 x 18.2 km. The 27 sample sites used included eligible, 12 good but ineligible.

The dimensions of our China Lake Test Area were 63.3 km x 10 km. The 33 sample sites included 3 eligible, 30 untested, mixed lithic and habitation.

Legacy Results Using Point Cloud Sampling Protocols

Although the annuli sampling method produced excellent results, it proved to be very computationally expensive to generate post-posterior probability maps from these results. The annuli method required calculating the medians and MADs (median absolute deviation from the median) in the annuli surrounding each sampled archaeological site and non-site. Members of the project team from the Applied Mathematics and Statistics Department of The Johns Hopkins University Whiting School of Engineering suggested a point cloud sampling method. By this method, pixel values in the pre-processed images are sampled within a 161 pixel square, with the center of the square being the central location of known sites. However, this approach requires dealing with an incredibly high number of dimensions, specifically $161\times161\times(9) = 933{,}156$; far more than we can train a classifier for, given the amount of data we have. This necessitates the use of two dimension-reduction steps: an initial one, which gets rid of the grand majority of the features, since few of these will be useful for classification purposes, and a secondary one, which was the feature extraction step in the old approach. Through experimentation, we determined that the optimal PCA dimension for training the LDA classifier was very rarely greater than

30. Thus, our initial dimension-reduction step was to calculate the p-value of a two-sided Wilcoxon rank-sum test for each of the pixel values in the multiband image and retain only the pixels with the smallest 50 p-values. After this, as before, we used leave-one-out cross-validation to choose the optimal PCA dimension, then trained an LDA classifier on all of the data using this optimal dimension and used that classifier to create the map. Schematically, the new method is as follows:

$[0.\ 255]^{wxh} \longrightarrow$ initial dimension---feature---classification [0.1]

Or, $$S\ 3\ s\ 1 \rightarrow \{0, 1\}$$

$$h \circ g \circ rT\,(s) \in \{0, 1\},$$

where $r_T(s) = (s_{\pi(i)})_{i \in \{1,2,\ldots,50\}}$, and $s = (s_i)_{i \in \{1,2,\ldots,933156\}}$,

and $\pi : \{1, 2, \ldots, 933156\} \rightarrow \{1, 2, \ldots, 933156\}$ is a permutation of the numbers 1-933,156 in increasing order of p-values calculated from the training data ***T.***

As we can see, the image processing step in the old method, which required calculating medians and MADs for each of 30 annuli and 36 band difference ratio images, has been replaced with making a vector of just 50 coefficients from the band difference ratio image, drastically reducing computational demands.

As with the annuli method, which also sampled from a very large area around the site, the approach also compensates for inaccuracy in the location of sites used for training. Inaccuracies on the scale of 10 to even 30 meters will not degrade model performance in almost all cases, because the sampling area is 322 meters square. Such a large area also is appropriate to the types of site that are found in the test areas, and indeed in many other tracts of land owned by the military. Sites can be of an enormous size. Scatters of lithic material can continue for many hundreds of meters. Disagreement on proper site boundaries have been ongoing at both Fort Irwin and China Lake because from the perspective of a person on the ground it is very difficult to determine where a given site ends and the next begins. An area that one archaeologist might consider a single site with several loci, another might think to be several sites. From the perspective of the DDM, many areas appear to be a continuous scatter of archaeological material.

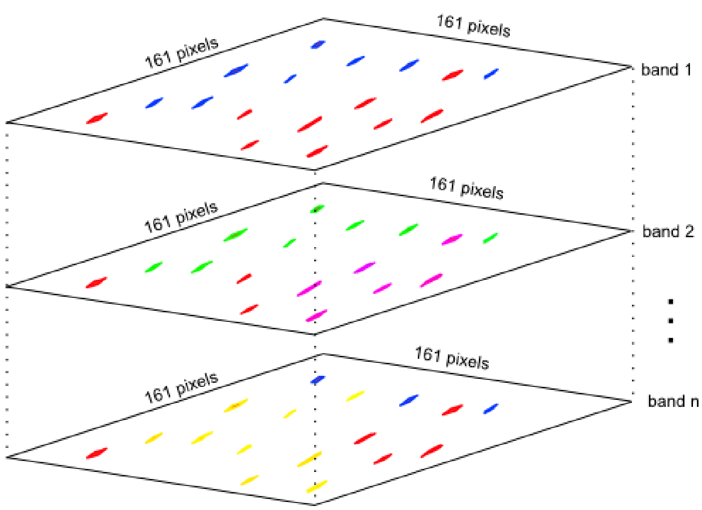


Figure 14: Visualization of point cloud sampling method.

DDMs for Fort Irwin Test Areas
Figure 16 (page 31) shows the DDM for the east test area in Fort Irwin. Symbols on the map represent locations of sites previously found and evaluated. These are, respectively, archaeological sites 1) eligible for listing on the National Register, 2) sites in relatively good (that is, relatively undisturbed) condition that were evaluated as being ineligible for listing on the National Register, 3) isolated finds (non-sites) that are often related to sites that have been disturbed, 4) sites that have been greatly disturbed, and 5) sites that have been completely destroyed.

The DDM is a probability map. Our statistical analysis indicates that the probability of finding an archaeological site in relatively good condition in red areas is from 30%-100%, of finding such a site in yellowish-orange areas is 17%-30%, and of finding such a site in green areas is 0%-17%. In the Fort Irwin East Test Area, we see that the great majority of eligible sites (white dots) fall within red areas. The great majority of all other sites fall within yellowish-orange and green areas, and of these more destroyed sites in green areas, and more disturbed sites and isolates in yellowish-orange areas. The large yellowish-orange area between the lower large red area and the upper large red alignment displays streaking, as if the terrain here has been greatly altered by the movement of vehicles. This might have obliterated any intact sites in this area, distributing site material in a thin layer over the landscape.

Figure 17 displays the DDM for the Fort Irwin West Test Area. Here, we see essentially the same pattern of eligible sites in red areas, and all others in green or yellowish-orange areas. There is a general pattern of sites in yellowish-orange areas being disturbed or destroyed. Sites in good condition that are in green areas are not eligible for listing on the National register of Historic Places. This is because these sites are generally small and thin scatters of lithic material, and would not provide scientific or historic information of any importance if they were investigated further.

Figure 18 is of the China Lake Test Area. The DDM for this test area was developed without benefit of Lidar data. The statistical protocols were run on only the eight bands WorldView-2 satellite image, and a ninth feature. This was a change detection feature: the change was detected by comparing an NDVI image generated from WV-2 data collected during the dry season in the Mojave Desert to and NDVI image generated from WV-2 returns collected during the wet season there.

Applications of the DDM To Problem Sets

PROBLEM SET 1: The methods used to inventory and evaluate archaeological sites have proven to be too expensive, time-consuming, and unreliable to permit compliance with Section 110.

1.1 No federal agency has fully complied with Section 110 of NHPA, which requires a full inventory of archaeological sites and an evaluation of discovered sites.

Direct detection models constitute *de facto* Section 110 surveys. They will serve both the military and archaeological resources better than on-ground survey and evaluation, which requires the controlled destruction of portions of archaeological sites by excavation. They also provide the basis for strategically planned inventory surveys and evaluation of discovered sites. As range boxes and training areas are moved, the model will identify areas that contain few sites. Few sites will be found in these areas, which will greatly reduce the time and expense of recording and evaluating sites.

The sites more likely to be eligible for listing on the National Register are evident in the direct detection model. This is because ancient human activity at these sites altered the environment to a greater degree at significant sites than at less important sites, and recent human activity has disturbed the remains less. Direct detection of sites will therefore guide archaeologists to the most important sites. The use of the direct detection model also provides information (e.g., context, precise location, aerial images, a landscape perspective) that will more quickly and accurately document the site and site conditions. Moreover, it will facilitate the negotiation of time saving fieldwork protocols for areas identified by the model as containing few sites. For example, SHPOs will often accept wider transect spacing when there is good reason to expect low site density. Finally, when all areas except those with very high site densities have been surveyed, managers can negotiate protected zones with the SHPO. This would eliminate the need for survey and evaluation in high density areas. Such areas are, of course, the most expensive to survey and evaluate, and might require mitigation, which could be several orders of magnitude more expensive.

1.2 Site locations recorded with dithered GPS signals (pre-2006) or with topographic triangulation are almost invariably incorrect.

The direct detection model identifies the location of detected or suspected sites with great accuracy, using 2 m resolution Worldview 2 Multispectral datasets and Lidar derived terrain models. As commercial data acquisition technologies improve, the model will become more accurate.

1.3 Because sites are not located exactly where they were recorded to have been, it can be difficult or impossible know if a site rediscovered is the site originally recorded.

The DDM can be used to obtain a synoptic view of a landscape and site distribution, which will aid in matching sites with those previously identified. Direct detection will also reveal locations that have been heavily disturbed, and so where sites may have been obliterated or drastically altered.

1.4 State site forms have required increasingly detailed descriptions for discovered sites.

The use of high resolution images inherent in the model provides the basis for more complete and accurate mapping and a context for photographs and notes obtained on the ground, while planning to avoid areas detected as high-density reduces the cost and accumulated paperwork of describing large numbers of sites.

1.5 The quality of archaeological work can suffer under low-bid award protocols that are frequently employed in contracting.

DDMs can be used to check the quality of on-ground surveys, and reduce survey and evaluation costs.

1.6 Most archaeological sites found by survey have yet to be evaluated for good reason: evaluation that involves excavation is expensive and time consuming.

The direct detection model provides information that can greatly reduce the time and cost associated with evaluation. Areas indicated by the DDM to have a high probability of containing sites are those where the richest, most undisturbed sites are located. The DDM also provides the context that bears greatly on significance or lack thereof; context includes the location of any given site in the distribution of all sites in the region and environmental zones. Context provides the basis for evaluating a site as unique, unusual, or commonplace.

Also, the detail visible in the DDM provides a way to identify areas that once probably contained rich archaeological deposits, but which have since been greatly disturbed. Figure 19 provides a good example of this.

1.7 Artifacts taken from the ground must be examined, washed (if washing will not remove materials such as blood residue that should be further analyzed) and examined again to be identified, labeled, catalogued, and then placed into storage. All of this is typically as time-consuming, labor intensive, and as expensive as the excavation of the material from the archaeological site in the first place. Storage must be in a climate-controlled environment, which is yet another substantial expense, and one that continues. Storage requires not only construction, maintenance, and monitoring of the facility, but also personnel to do all of this in perpetuity.

Because much less excavation is needed to evaluate sites, the model greatly reduces these costs, which are often enormous and continue indefinitely. The DDM provides a way to identify the richest and most intact sites that are most likely to be eligible for listing on the National Register of Historic Places. If sites are found in places that are not indicated to be highly likely to contain sites by the DDM, they are probably not eligible for inclusion on the

National Register.
1.8 Sites that have been or might be evaluated as eligible for inclusion on the National Register of Historic Places must be protected. The certainty with which sites can be located and identified and the precision with which they can be recorded, as well as the landscape perspective that is fundamental to the modeling protocols, provides guidance to determine effective buffer zones around sites, or around areas that might contain high concentrations of significant sites.

Notes on Using the ROC Curves for Decision Making

ROC curves not only evaluate the productivity of a model, but also are the basis for effectively applying the model, that is, rendering the model an effective decision support tool. ROC curves have been widely used in medicine to access the accuracy of diagnostic and screening tests. These tests explore the relationship between sensitivity (the ability of a test to correctly diagnose a patient with a disease) and specificity (the probability of a negative test result for a patient without a disease). Sensitivity can be understood as the TPR and specificity as the FPR. True and false diagnoses can vary depending on the threshold set. This threshold defines whether or not a person is sick or well (Obuchowski 2005).

Example: Setting a threshold for blindness.

By alternating this threshold, the results of a test can alter considerably between those who are totally blind and those with varying degrees of sight. This threshold will also change the sensitivity and specificity of the test. As the threshold is lowered, there is a greater chance that the test will correctly diagnose patients (TPR) however there is also a greater chance that people will be misdiagnosed (FPR). Conversely, if the threshold is raised, less people will be correctly diagnosed but there will also be less misdiagnosis.

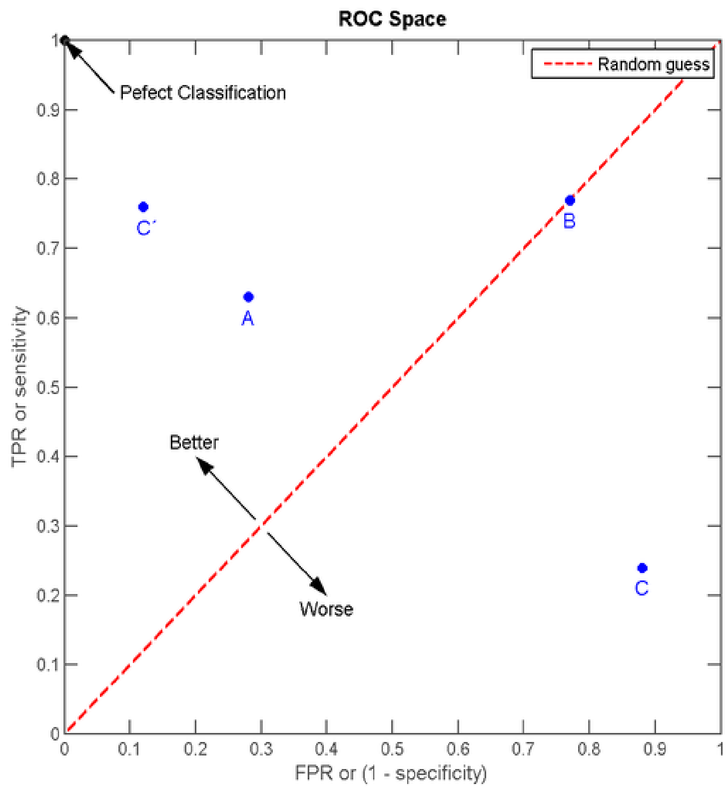
Deciding the most effective threshold can be difficult as it is necessary to weight the efficacy of a model in detection of a disease with the amount of false positive results. This is where the ROC curve can be used. Results from different thresholds can be plotted with the sensitivity or TPR (y) against the false positive rate or FPR (x). Points are then connected to generate a curve.

ROC curves have also been used in psychology, specifically in the field of signal detection theory. ROC analysis has been used to compare test which assess the ability of individuals to perceive sensory information like brightness, tone or pitch. Given the subjectivity of these senses the ROC curve allows psychologists to explore which tests are more effective than others. Similarly, the ability of people to perceive in less clear contexts (like over distance or through mist) can be assessed. This is referred to as discrimination acuity.

In recent decades, ROC curve analysis has been used in meteorology to assess the predictive capability of forecasting tests (Mason and Graham 2002). Hit-rates (TPR) and false-alarms (FPR) can be plotted on the ROC curve. In this case, the efficacy of a test is defined not by the curve itself, but by the area under the curve. A value greater than .5 is deemed to signify a successful test.

As mentioned above, a ROC curve runs from point 0, 0 where there are 0% TPR and 0% FPR to 1, 1 where there is 100% TPR but also 100% FPR.

A diagonal line runs between these two points and this is called the chance diagonal. This line represents the expected random distribution of results. The closer a point or result is to this line, the more random it can be said to be. Conversely, if a point or points falls consistently well below the line, it is likely significant but the sensitivity and FPR need to be reversed.



Why is the ROC curve a useful tool for testing Archaeological Predictive models?

A traditional APM identifies relevant variables using exploratory statistics like the chi-squared or Kolmogorov-Smirnof test, checks for overlapping using regression, and then compiles a series of overlapping layers identifying regions most likely to contain archaeological sites. Ideally, a model should be built using a sample from a larger population and then tested using the remaining sites. This testing is done using a gain statistic which explores the number of sites identified, and the quantity of land in the model. A good predictive model identifies a large number of sites in a small area.

More often, models are assessed for ‘accuracy’. The problem with this method is that it is possible for a test to be 98% effective purely by declaring that the entire landscape is void of archaeological remains. While they will be wrong in 2% of cases (where sites will be found), the model can still be considered highly accurate! It is therefore necessary to weight the accuracy or sensitivity (the number of true positive identifications) of a test against its specificity (number of false positive results). This is where the ROC curve can be used as it explores both the positive, but also the negative site identifications.

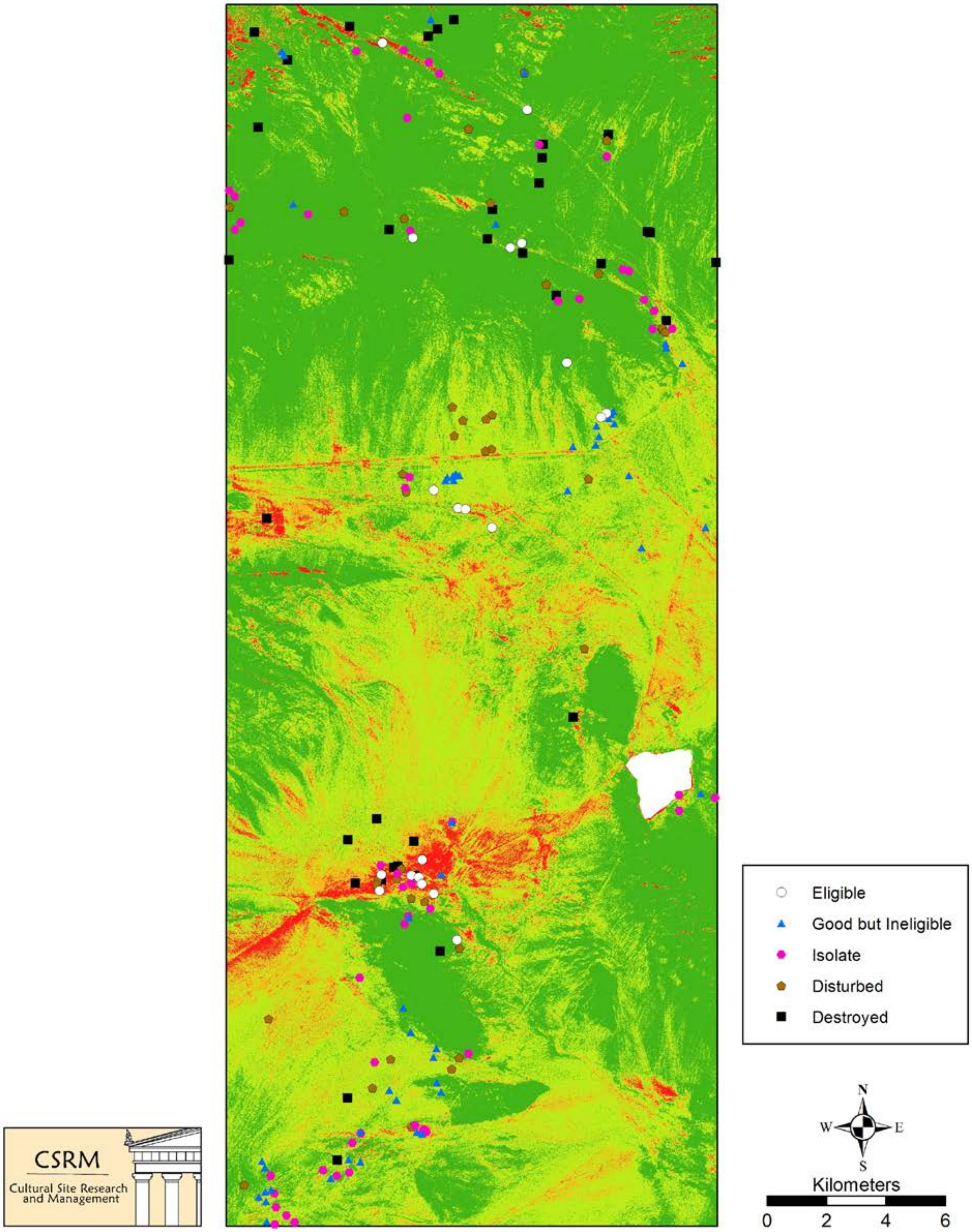


Figure 16: Fort Irwin Eastern Swath Results, sites organized by condition. Red is high probability of containing a significant site, green low.

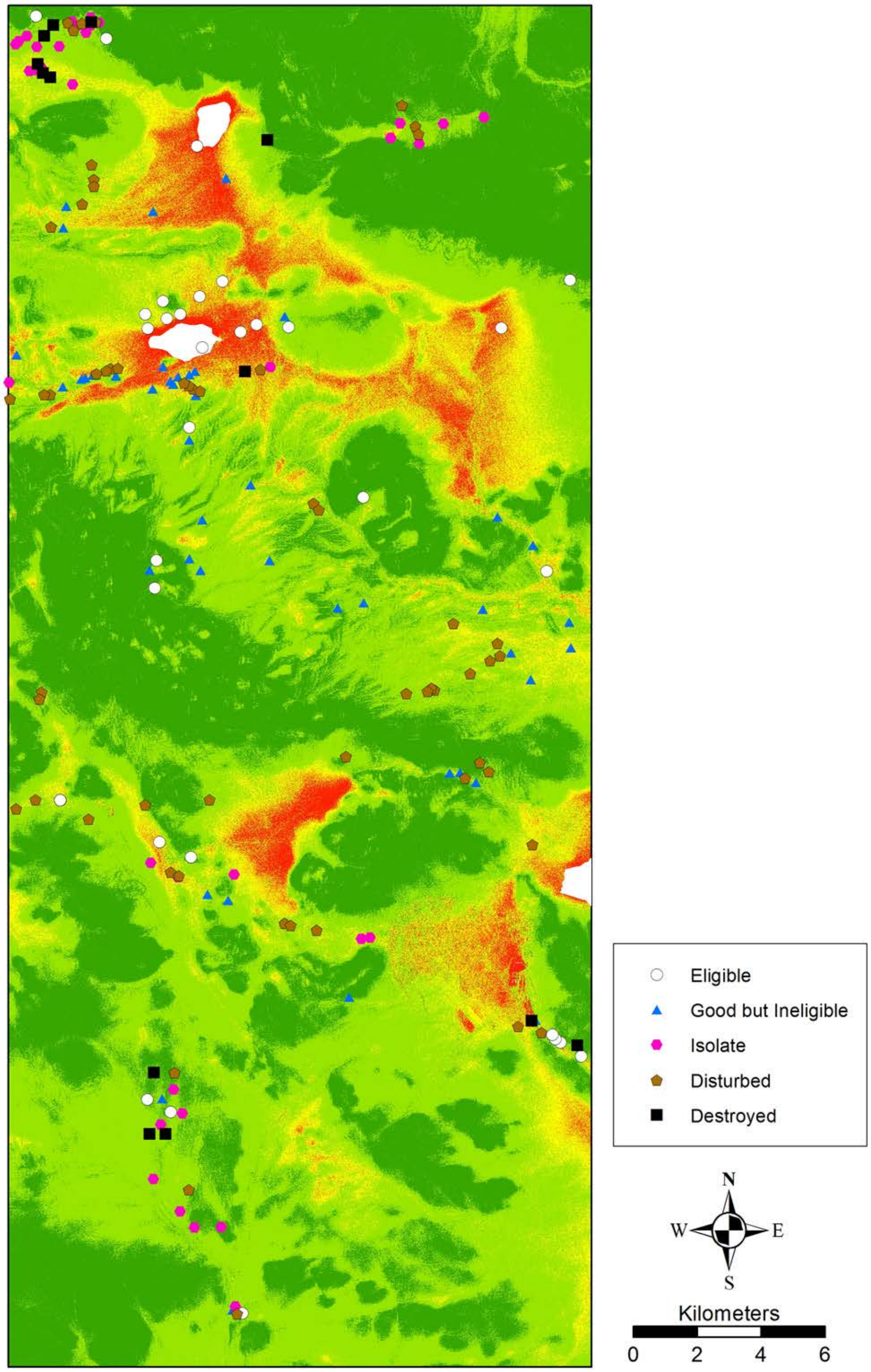


Figure 17: Fort Irwin Western Swath Results, sites organized by condition. Red is high probability of containing a significant site, green low.

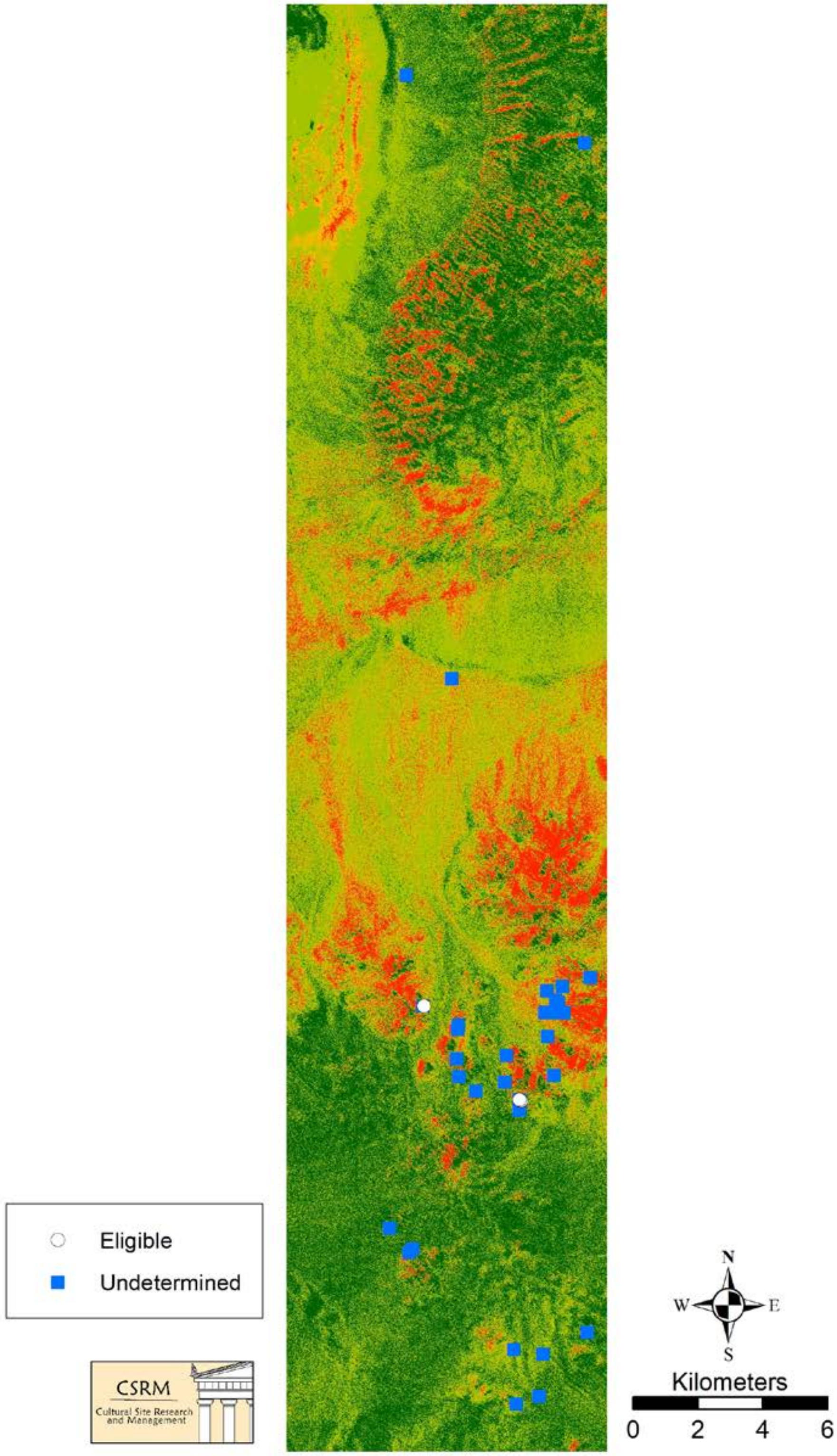


Figure 18: China Lake Swath Results, showing sites by known eligibility status. Red is high probability of containing a significant site, green is low. Note: the site data from China Lake did not permit fine-grained site condition analysis, and the majority of the sites in the study area have not yet been evaluated for eligibility for inclusion on the National Register of Historic Places.

In the direct detection model, actual site locations are tested against random points seeded according to a posterior probability index. This means that the non-site locations will share some similarity in location to the sites, further increasing the significance of the result.

Why is the ROC curve a useful tool for planning developments and activities?

The ROC curve makes it possible to determine probability that any given pixel will fall within:

1) the boundaries of a relatively rich and intact archaeological site that is more likely than others to be eligible for listing on the National Register, or

2) that it will fall within a site that has been disturbed, or

3) that it will fall within an area that does not contain archaeological sites or an area where sites are a thin scatter of artifacts or are completely disarrayed: such sites are unlikely to be eligible for the National Register.

Protocols for the use of the DDM

<u>1) Construct the DDM</u>

The direct detection model is most productive when a variety of remote sensing data sets are employed, and when each data set provides a different kind of information about the landscape in which archaeological sites are to be detected. This is demonstrated by the results of case studies at the two test areas of China Lake and Fort Irwin. Results at Fort Irwin are more useful because two types of remote sensing technology provided data to the DDM: the eight-band WorldView2 multispectral imagery and a digital surface model produced from a high-density Lidar scan of Fort Irwin. Lidar was not available for China Lake. Data sets that have proven valuable to DDM production elsewhere have been various bands and polarizations of synthetic aperture radar (SAR). This was described earlier in this report. Hyperspectral imagery would also be of great utility, however, it is not as commonly available as the three types of data just mentioned.

Site location data is also valuable. CSRM makes every attempt to utilize highly accurate site location data for developing DDMs. For this reason, we have frequently collected data ourselves. This has sometimes proven difficult on military lands where safely and security are an issue. In such situations, it is important to have access to data that has been vetted carefully by GIS professionals after it has been collected. The quality of data collected by contractors varies, for example, and data collected by field schools often requires careful review and correction. While accuracy is very important in developing the DDM, less accurate site locations seen against the backdrop of the DDM can often be instructive. The DDM reveals an archaeological landscape, and sites often fit into the pattern presented by the modeled landscape, as discussed just below.

The statistical treatments that are used to construct a DDM carry with them enormous advantages. By minimizing or eliminating the need to lump, or “bin,” sensor returns, we are able to take full advantage of sensor precision. We are also able to sample very large areas around sites, which means that if site location has not been recorded absolutely precisely, the protocols will still yield a reliable and informative model. These statistical treatments require extreme computational capacity, however. This capacity we have developed by means of a Space Act Agreement among CSRM, the Johns Hopkins University, and NASA. The agreement allows us use of the Discover computer at the Goddard Space Flight Center, with 15,000 processors, and the computation facilities at Ames Space Flight Center, with 200,000 processors. Access to a great number of processors allows us to conduct parallel computing, which reduces the time required for production of probability maps from days to a few hours.

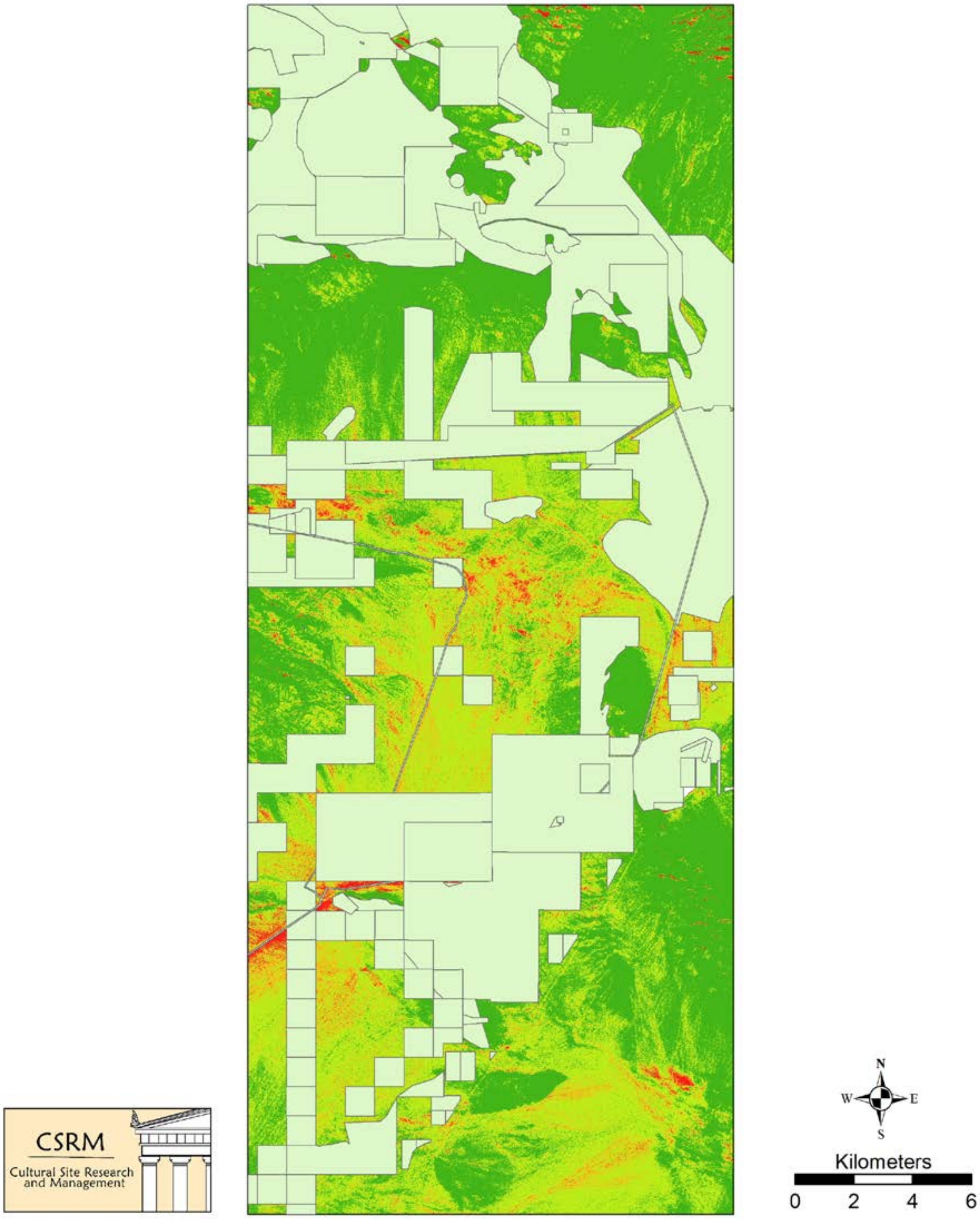


Figure 19: Fort Irwin Eastern Swath with previously surveyed areas overlaid.

2) Utilize the DDM to Strategically Plan Surveys

In the test areas of China Lake and Fort Irwin, certain areas have been surveyed, and other have not (see Figure 19). This is typical for military lands everywhere. Surveyed areas and the sites that have been located in them are available as GIS shape files, the former as polygons, the latter as point or polygons. When these are superimposed over the DDM probability map, those areas that are less likely to contain archaeological sites can be seen clearly. Survey polygons that have been surveyed in areas of in low probability display fewer sites, as might be expected, than survey polygons in higher probability areas.

As range boxes and training areas are moved, low probability areas should be surveyed first. That they are low probability, and because the surveyed polygons in low probability areas bears this out, it is likely that survey protocols can be negotiated with the State Historic Preservation Office (SHPO) that provide for faster survey of these areas. Because little will be found, all of the low probability areas should be cleared for any use desired by the military quickly. The great majority of sites that might be found will probably not be significant, as described in more detail below. As time goes on, survey areas that are more likely to contain sites can be surveyed. Among the areas that are indicated by the DDM as being somewhat likely to contain archaeological sites (those areas seen in yellow, sometimes with pinkish strips) are subareas that show very clear signs of disturbance (see Fig. 19). These areas seem to be places where once intact sites have been scattered over the landscape by vehicular movement; in fact, vehicle tracks stand out clearly. This is probably because they share attributes common at archaeological sites, among them that vegetation is more sparse in these areas than in surrounding ones. These areas, too, of course, can be surveyed quickly, and so should be surveyed first. There are other locations that have generated sensor scatter that is typical of those received from archaeological sites. These include playas and roadbeds. In all of these cases, not only is vegetation more sparse, but they are like archaeological sites in that they occupy relatively level ground. Finally, the NDVI change detection feature used in the site detection model displayed little or no change in these areas, just as archaeological sites did not change in terms infrared and red band scattering, either.

3) Utilize the DDM to Interpret Survey Results and Organize the Archaeological Data Base

Site boundaries are problematic, but important, useful, and necessary in order to maintain compliance with legislation, regulation, and policy that deal with archaeological resources. Humans have crisscrossed most of the landscapes on earth many thousands, and in some cases many millions, of times. Conceptually, an archaeological site is an area with well-defined boundaries that contains a concentration of material from the past. The importance of an archaeological site depends in large part upon how much of that material remains in the condition and place it was deposited. Yet, an archaeological site can as accurately be thought of, in many if not most cases, as an island of intact material, which was once part of a much larger area that was altered by human activity in the past. Further, landscapes in which archaeological sites are located have changed many times. At China Lake and Fort Irwin, there have been perhaps more than 13,000 years of human occupation. Uses changed over time, and layers of archaeological deposits formed in different places or in the same places, burying earlier ones. What are now recognized are sites are the results not only of varying levels of human use over time, but also removal or disturbance of archaeological remains near the ground surface by numerous means, but especially modern construction and military maneuvers. What remains resembles an oil slick of archaeological material in many places: material is more densely distributed in some areas, less so in others, and in many places modern human use has dispersed or removed it. This can be seen quite clearly in Figure 20 (next page). The DDM can be used to review the delineation of site boundaries as seen on the ground, where these can be very difficult to determine. It can also be used to review estimates of how much an archaeological sites has been disturbed.

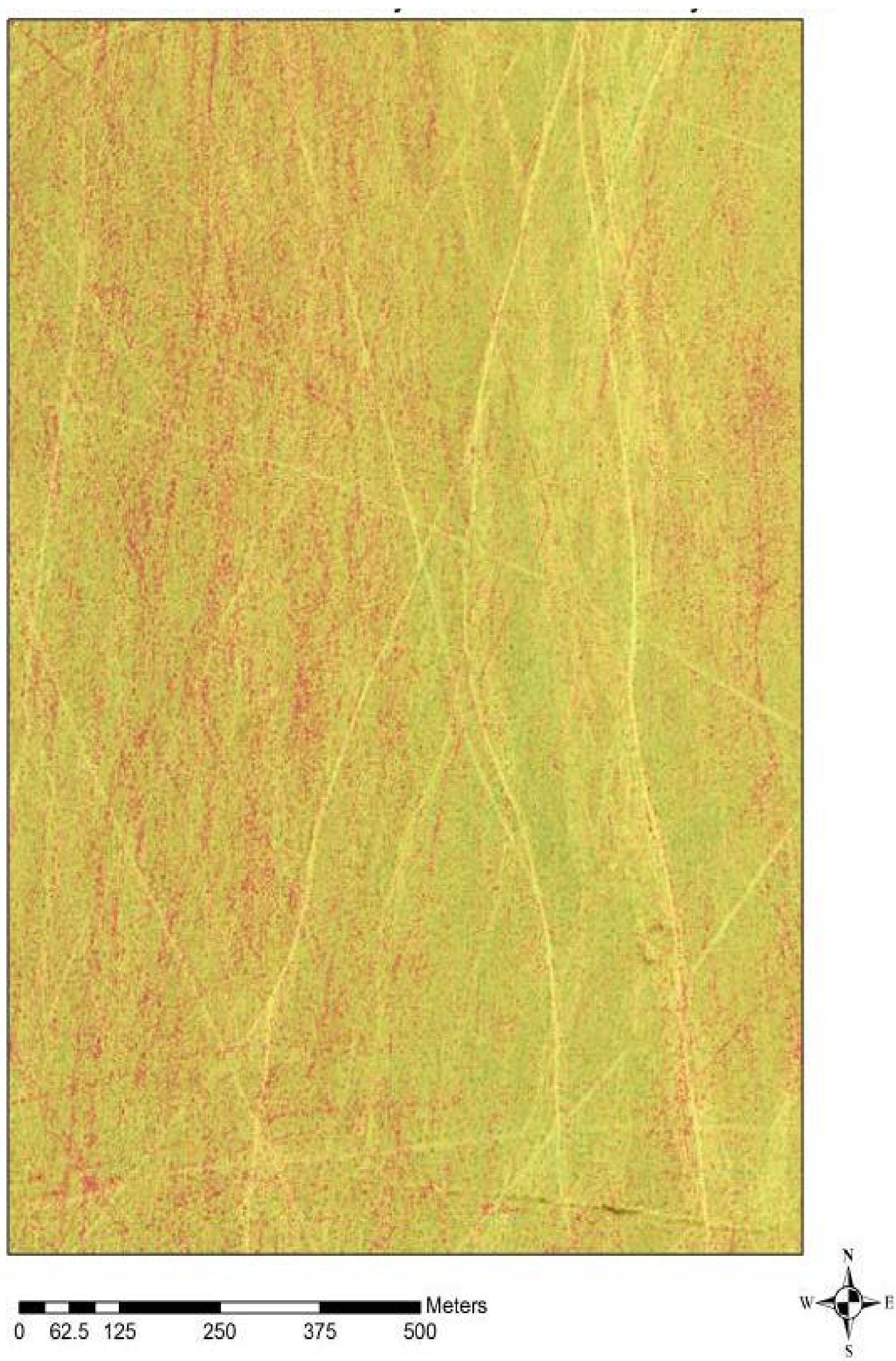


Figure 20: Fort Irwin Eastern Swath detail showing the correspondence between vehicle tracks and returns from the model.

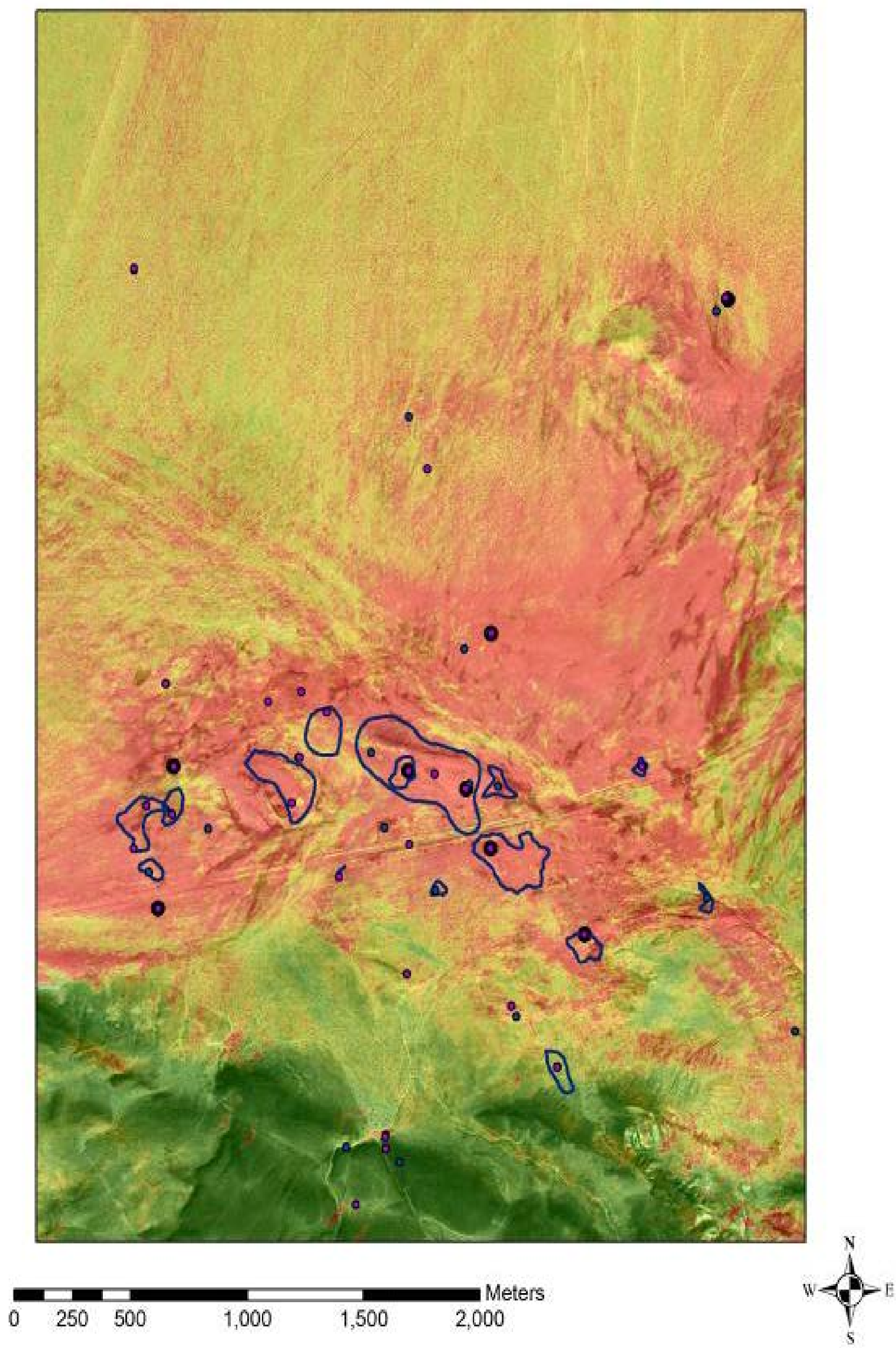


Figure 21: Fort Irwin Eastern Swath detail showing correspondence between high-return areas and site boundary polygons.

4) Utilize the DDM in Evaluation of Sites for National Register Eligibility

The DDM is not intended to replace on-ground surveys entirely, but instead to provide a decision support tool for

1) Strategic planning of surveys

2) Review and interpretation of survey results, and

3) More effective, faster, and more cost effective evaluation of archaeological sites. This is described below.

Sites that fall within DDM high probability areas and those that do not can be thought of as comprising a number of sets. Here, we will make use of Venn diagrams to consider some of those sets. Among them are:

A_1= the set of sites within areas predicted to contain high concentration of sites by the DDM. This set of sites will contain those that display relatively dense concentrations of archaeological material and that have been relatively undisturbed. They are therefore sites that are more likely to be evaluated as significant.

B= The set of sites found by on-ground archaeological survey, which will contain some sites that are significant and some that are not. Typically, most sites will not be eligible.

As illustrated in this figure, the subset formed by the intersection of these two sets ($A \cap B$) contains those sites that are most likely to be evaluated as being eligible for listing in the National Register of Historic Places. It is, after all, the dense concentration of archaeological materials in a relatively undisturbed context that is the basis for identifying high probability areas by use of direct detection.

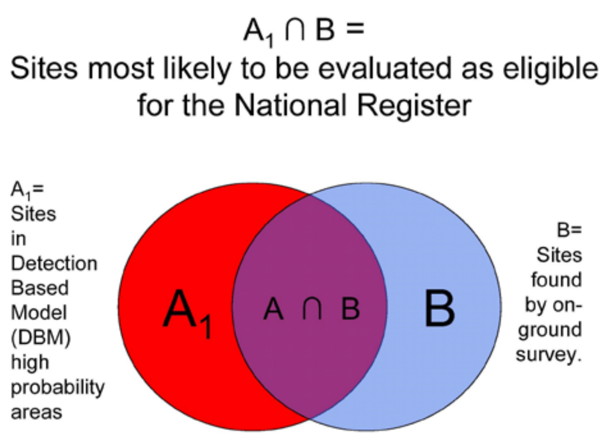


Figure 22: A Venn diagram showing the subset of sites most likely to be evaluated as eligible for listing on the National Register of Historic Places.

Eligibility for listing on the National Register would be because of the wealth of potentially important scientific and historic information that might be extracted by careful analysis of these deposits.

Areas identified as being moderately probable to contain archaeological sites are also instructive. Sites in these comprise subset A_2. Close examination of these areas, as in Figures 19 and 20, reveals that some of these have undergone a good deal of disturbance. The DDM highlights such disturbance, and so can be systematically examined to determine how disturbance might have affected eligibility for listing on the National Register.

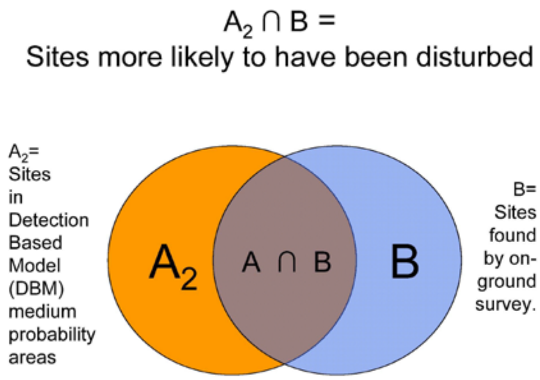


Figure 23: A Venn diagram showing the subset of sites most likely to have been disturbed, and so less likely to be evaluated as eligible for listing on the National Register of Historic Places.

In contrast to subset A1, subset A3 is of sites found by on-ground surveys in areas identified by the DDM as low probability areas. These sites, because they were not detected by the DDM, are not likely to contain high concentrations of archaeological material, and are more likely to have been disturbed. Therefore, the subset of sites formed by the intersection of subset A3 with subset B, which is of sites detected by on-ground survey, are less likely to be eligible for listing on the National Register.

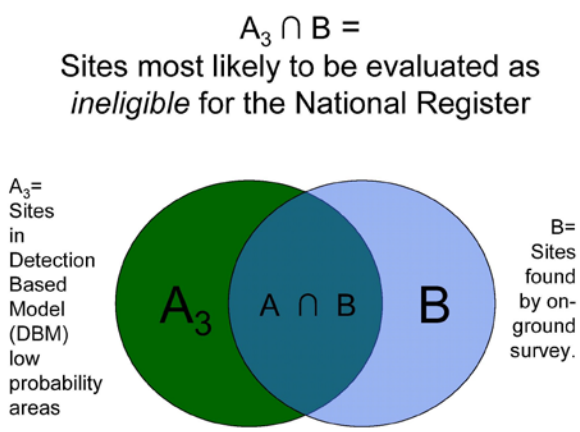


Figure 24: A Venn diagram showing the subset of sites most likely to be evaluated as *ineligible* for listing on the National Register of Historic Places.

Cost Avoidance and Applicability to all DoD Landholdings

1) General Applicability

The Direct Detection Model (DDM) that we propose to refine for use at vegetated and highly vegetated environments was developed in less complex environments; specifically, at test sites in arid regions, where archaeological materials are more easily observed on the ground surface than they are in vegetated environments. With a successful modification of the DDM model for use in vegetated environments, it can be made available for use at all lands for which the military has stewardship, within and outside of the boundaries of the United States and its protected territories.

The successful application of the DDM to Fort Benning's heavily vegetated landscape would also demonstrate the utility of this model in difficult terrain. In the past, it has proven very time consuming and therefore very costly to conduct Section 110 and Section 106 compliance activities in such areas.

2) Cost Avoidance through the use of the DDM to Create a de facto Section 110 Survey

The Direct Detection Model provides the means for a de facto Section 110 survey, one that does not require on-ground survey in addition to those that have been conducted previously at lands under the stewardship of the military.

Section 110 of the National Historic Preservation Act of 1966 (as amended) stipulates the following:

> (2) Each Federal agency shall establish (unless exempted pursuant to Section 214), in consultation with the Secretary, a preservation program for the identification, evaluation, and nomination to the National Register of Historic Places, and protection of historic properties. Such program shall ensure-
>
> > (A) that historic properties under the jurisdiction or control of the agency, are identified, evaluated, and nominated to the National Register;

Below is a comparison of the cost of complying with Section 110 by means of 1) traditional, on-ground survey and evaluation, and 2) the use of the DDM at three different areas under the stewardship of the military. The current proposal is for the application of the DDM at the third of the three. The first two areas are in arid environments, which have sparse vegetative cover. The proposed project area is the third area addressed, and is covered by thick vegetation. The DDM can be used at all three areas: the first two lie at one end of an environmental spectrum and the third at the other. It should be emphasized that virtual all lands under the stewardship of the military fall within the two ends of this spectrum.

China Lake

Section 110 Survey Cost

A recent 110 survey at the Naval Air Weapons Station at China Lake, California cost US$ 35.94 per acre. The survey was designed to find and record archaeological sites and no National Register of Historic Places (NRHP) evaluations were included. In order to perform comprehensive 110 survey activities across the entire base (1.1 million acres), it would cost approximately $39,534,000 in 2013 U.S. dollars.

Section 110 Evaluation Cost

Twenty-one percent (21%) of China Lake's total area has been subject to traditional, pedestrian survey so far (note caveats about the difficulties in ensuring complete, accurate, and precise coverage with this method). 3,837 sites have been found in the area so far surveyed; while there are a multitude of environmental and historical factors that affect the distribution of ancient sites and their modern discovery, if sites are similarly distributed in the other 79% of China Lake, there would be 18,271 total sites on the base's grounds.

Not all discovered sites require evaluation; this varies dramatically by circumstances peculiar to a survey area. Based on previous CSRM projects in the region, we will assume that 25% of discovered sites, or 4,567, require evaluation that involves some excavation. For those sites, an average cost for the evaluation is approximately $12,500, meaning that the full cost for evaluation could reach or exceed $96,621,500.

Total cost of both survey and evaluation at China Lake would be **$136,155,500**.

<u>Fort Irwin</u>

Section 110 Survey Cost

Discussion with management personnel at Fort Irwin gave an estimated cost of Section 110 survey at $27.65 per acre; following the pattern above, that would put the cost of surveying the total area of Fort Irwin (684,200 acres) at $18,918,130.

Section 110 Evaluation Cost

At Fort Irwin, 1,423 sites have been found in the 41% of the installation surveyed thus far; again assuming a similar distribution across the rest of the installation, there would be 3471 sites in total.

At Fort Irwin, many of the archaeological sites in the central corridor of the installation have been extensively disturbed by many decades of tank maneuvers over the fairly level terrain there; a disturbed or destroyed site has a much lower possibility of being eligible for the NHRP, and this disturbance can typically be seen from the surface. Evaluation is typically a matter of re-examination of surface remains, and therefore might cost as little as $1,000 per site. There are approximately 715 of such sites, so evaluation costs for this set of sites would be approximately $715,000. For the remaining 2,756 sites outside of the central corridor, however, evaluation costs might approximate those that were estimated for China Lake: $12,500. Again, we will assume that 25% of such sites require evaluation that involves some excavation. That being so, the 689 sites needing subsurface evaluation would require $8,612,500 in funding. Therefore, the total estimate for 100% Section 110 evaluation would be $9,346,418. In turn the, total cost for both survey and evaluation would be **$28,264,548**.

<u>Fort Benning</u>

Section 110 Survey Cost

Fort Benning, with 181,626 acres, is one of the very few tracts of land under military stewardship that has completed a Section 110 survey for all accessible areas of the installation. 20% of Fort Irwin contains unexploded ordinance; in the 80% surveyed, 3,975 sites were found. Section 110 survey costs at Fort Benning totaled approximately $ 7,265,040 in 2005 dollars; in 2013 dollars, this would be about $9,479,229. The cost per acre is higher at Fort Benning than at Fort Irwin or China Lake because the landscape is vegetated, obscuring visibility and hampering pedestrian transit, whereas the landscape at Irwin and China Lake is very arid and has only sparse vegetation.

Section 110 Evaluation Cost

At Fort Benning, evaluation requires about $13,659 per site. Assuming that one-third of the 3,975 sites there (or 1,324) would require this sub-surface evaluation, evaluation costs for 100% of known sites would be $18,080,076. The assumption of a higher proportion of sites needing sub-surface investigation to determine eligibility is based on a) dense vegetation making it more difficult to ascertain salient features of the site from the surface and b) a tendency for Eastern prehistoric sites to be larger and more complex than those found in the Mojave.

Total estimate for 100% Section 110 survey and evaluation: **$27,559,305**

3) Additional Cost Avoidance

As seen above, the cost avoidance that can be achieved by use of the DDM as a replacement for on-ground Section 110 surveys is very substantial. As importantly, there are additional cost avoidances to be realized that are not directly related to conducting the survey.

a) The first, additional, cost to be avoided will continue far into the future; in theory, it will burden the military forever. They are costs associated with storage of archaeological material in environmentally controlled conditions, and maintaining accession records. In the case of the latter, this will mean constantly adapting to new digital technologies for data storage and retrieval. Not doing this will result in the loss of all data associated with the archaeological material in the collection, and the loss of this often renders the collection worthless.

Fort Irwin provides a cogent example; the installation currently has 930 cubic feet of artifact collections and 199 linear feet of associated records from Section 110 and 106 compliance activities. Storage costs are $20 per cubic foot of material or $18,600 annually, with an additional $24,000 required for materials to support curation and conservation of the artifacts. To fully meet compliance with California's State Historic Preservation guidelines, a new facility with better climate stability it required – while refitting of an existing structure has not yet begun, personnel at Fort Irwin estimate that it will cost $250,000 alone to simply transfer the artifacts between buildings.

The annual costs described above only calculate expenditures for material and facilities and unskilled labor to maintain collections; wages for skilled labor to catalog, analyze, and properly store materials represent another substantial investment on the part of the installation. A particular difficulty is maintaining institutional knowledge of the collections if contract labor is routinely employed in this process. Many bases have also sought to digitize their records to streamline data entry, reduce storage costs, and have a queryable database of holdings. Database maintenance, interoperability, and longevity are persistent problems in computer and library science, and requires constant upgrades of software, hardware, and the format of records. While this is a broader issue in facility administration, the unique parameters of artifact records, the inclusion of multiple data formats (e.g. text, images, georeference points), and the specialist knowledge of archaeology required for data entry makes this especially troublesome for management. These costs will vary dramatically from facility to facility based on existing infrastructure, division of duties, and history of management, but must be considered in addition to the direct costs (listed above) for Section 110 survey and evaluation activities and are frequently substantial.

b) Secondly, we must examine the cost of noncompliance with Section 110. This varies dramatically. A Section 106 survey will be done for the area of potential effect (APE) for each individual undertaking. The cost of not knowing where sites are, or are very likely to be, located includes finding and evaluating more sites than would be necessary were the locations of sites or locations of areas highly likely to contain sites known. Finding sites during a survey increases costs because each site must be comprehensively recorded; this is time-consuming, and therefore costly. Each site that is found must be evaluated. The greater the number of sites and the more archaeological material they contain, the greater the cost of evaluation will be.

In cases where planning and design has proceeded to a point where changes would be highly problematic, mitigation of excavation of the site might be required. Mitigation costs vary dramatically, but since they involve substantial excavation and the recovery of artifacts that must be analyzed, treated for conservation if they are subject to deterioration, and then stored in environmentally controlled facilities, they are generally very high. If mitigation excavation is not done, or if a partial excavation for mitigation is conducted (which is typical), archaeological materials (such as human burials) might be encountered during construction. If this occurs, construction must be halted until consultation with recognized Native American tribes has been conducted and the remains are recorded and preserved. In some cases, construction plans must be altered to prevent further disturbance of remains. Again, it is difficult to generalize regarding the cost of non-compliance in this case, but in some instances it has added tens or even hundreds of thousands of dollars to project costs.

4) Cost of Developing DDMs

There are a variety of factors that will determine the cost of developing a DDM for any specific area. Among these are:

- The variety of archaeological sites within the given area. A model is prepared for each type of site. The greater the variety of sites, the longer developing the DDM for the entire area in question will take.
- The reliability of site descriptions. Developing a DDM for each type of site will require at least 15, and preferably as many as possible, of such sites. The location for each should be accurate within a few meters, and each site must have been characterized accurately as to type and degree of disturbance.
- The availability of aerial and satellite data sets. The model in most cases requires high resolution multispectral or hyperspectral imagery, as well as Lidar. Synthetic aperture radar data, and, in fact, data acquired by other technologies, such as (LIST), can be enormously useful and informative in certain environments.
- The availability of an archaeologist with comprehensive knowledge of the archaeological record in the area in question.

Assuming that U.S. military bases and installations have access to Lidar data and to high resolution multispectral imagery through NGA, the costs of DDM development are primarily labor costs for specialists to access and process the data, site visits to confirm accuracy and collect parameters, tailor the statistical protocols to the dimensions of parameters of the area under study, code the protocols for use with NASA Goddard Space Flight Center Discovery supercomputing cluster, and interpret and ground truth the results. The roles outlined in our development programs to date have been the following: archaeological staff, consisting of a Principal Investigator, Field Director, and field and laboratory crew, responsible for preliminary field data collection and ground-testing results; a geospatial specialist and a statistical specialist, responsible for processing the data, adapting the protocols, and interpreting the output, under the direction of the Principal Investigator; and collaborating with our development partners (the NASA Goddard Space Flight Center, the NASA Jet Propulsion Laboratory at Caltech (JPL/NASA), and the and The Johns Hopkins University Whiting School of Engineering) during this process for expertise on the statistical approach taken, the computational resources required, and the sensor instruments that provided the starting imagery.

The estimated costs for employing this process at China Lake, Fort Irwin, and Fort Benning do not include the work already done by CSRM at the former two installations in order to present a full comparison with the estimated costs for traditional survey and evaluation activities. Because of differences in terrain, types of sites that need to be modeled, and other factors such as expected disturbance of sites, presenting a comparable ‘per acre’ rate is not feasible.

Cost comparisons are as follows:

- DDM cost estimate for China Lake (1,100,000 acres): **$424,022.78**, or **less than 1/3rd of 1%** of the cost for traditional survey and evaluation.
- DDM cost estimate for Fort Benning (181,626 acres): **$281,259.26**, or approximately **1%** of the cost for traditional survey and evaluation.
- DDM cost estimate for Fort Irwin (684,200 acres): **$339, 037.50**, or approximately **1%** of the cost for traditional survey and evaluation.

Over and above the avoidance of the costs of traditional, on-the-ground archaeological surveys and evaluations are the avoidance of cost associated with non-compliance. Non-compliance costs, as described above can be enormous. Utilizing the DDM will avoid these costs, as well.

Summary and Conclusion

In this report, we have described the development of a direct detection model (DDM) for use as a decision support tool that will provide very substantial cost avoidance, facilitating full compliance with the National Historic Preservation Act of 1966 (as amended). The statistical protocols, from sampling to classification, are the culmination of a decade of research that has found support from the Department of Defense, NASA, the National Park Service, and the National Science Foundation. We have also put in place a cooperative research unit that can provide continued support for the development and application of cutting-edge statistical treatments and the capacity to process extremely large volumes of data. This research unit stands ready to deploy the decision support tool immediately.

We envision this decision support tool as being primarily employed as a cost avoidance mechanism. The use of the DDM to establish areas that are likely to contain significant (and therefore more likely to be NRHP eligible) archaeological remains does not necessarily proscribe the use of those zones for DoD training activities or other facility needs; performing de facto Section 110 survey through the application of the DDM a) avoids the considerably greater cost of performing traditional pedestrian survey over installation territory and the associated costs of analysis and curation described above and b) allows installation personnel to make informed decisions about the likelihood of invoking Section 106 compliance through choosing to utilize a particular area. This could potentially mean either expansion of training areas, through greatly reduced cost in initial surveying and by extension known reduction of Section 106 commitments in these areas, or greater selectivity in choosing training area rotations for ecological cycling.

Because the primary focus of this project is in the interests of facilitating Section 110 compliance, we do not envision the need to identify the age and cultural affiliation of materials on the surface as posing a significant problem for the application of the DDM. Since any material over 50 years of age is potentially *equally* eligible for listing on the National Register of Historic Places and every military facility in the United States has building records stretching past that point, any significant material appearing in the DDM can safely be assumed to fulfill that potential requirement and further investigation may be reserved for meeting Section 106 compliance as necessary for future construction and training activities.

In the interests of avoiding situations in which Section 106 investigation and mitigation are necessary, the DDM provides a method of delineating protected zones as described above. These zones would encompass areas of the highest signal returns (which are very frequently grouped together across the landscape) into cohesive blocks to provide facility personnel with an intuitive way of making planning decisions concerning land use. It is difficult to compare the relative impact of these protected zones with traditional site buffers for several reasons; as protected zones would be based on probabilities and not known sites, the effects that each have on base management are not directly comparable. The first consideration is raw acreage – while protected zones may or may not be larger in size than traditional site buffers, their cohesion is a powerful management tool in comparison to a multitude of small, medium, and large ‘islands’ of archaeological material situated in otherwise cleared training or building areas, which both invites site mitigation and associated costs and rules out some potential land uses, e.g., artillery and armored vehicle exercises. The second issue, related to the first, is the size of traditional site buffers. As seen from both the results of the DDM and previous CSRM work at Fort Irwin, archaeological material, especially when associated with nomadic groups over long periods of time, tends to spread across the landscape in an 'oil slick' rather than discrete spots. While the wording of current legislation requires that the 'site' is the minimal unit of designation and management, the actual size and necessary buffers for any given site will vary dramatically. This uncertainty is part of the reason that many installations and bases have not instituted site buffers at all, making direct comparisons of raw acreage impossible in our current test cases. A final consideration is the cost avoidance moving forward of incurring pedestrian Section 110 survey, Section 106 compliance, and curation and analysis costs in order to institute new site parameters and buffers – while for specific projects this may be a necessary component of facility management, employing it as a comprehensive strategy quickly leads to impossible burdens on the base or installation as we have shown above.

Next Steps

We have presented the development and use of the Direct Detection Model (DDM) to the China Lake Base Archaeologist, Michael Baskerville, and to Clarence Everly, who is Natural and Cultural Resources Program Manager at Fort Irwin. During our Legacy funded development of the DDM, we modeled on the areas at China Lake and Fort Irwin shown in Figure 2. Because of this Legacy project, we now have the capability to produce DDMs for the entirety of China Lake and Fort Irwin. The China Lake Base Archaeologist has said that he would like the team that we have organized for the Legacy project to develop a DDM for China Lake, and is attempting to identify funding for this. If he is successful, we propose to develop the China Lake DDM as an extension of the Legacy project. The Natural and Cultural Resource Manager at Fort Irwin is also very interested in the development of a DDM for all of Fort Irwin.

Our informal communication with key personnel in the California SHPO shows high levels of interest for the use of the Direct Detection Model as a decision support tool. We propose to formally present the results of the Legacy project to the California SHPO within the next few months, ideally in the company of a representative from Legacy and the DFPO at OSD. The SHPO would then be able to endorse and advise China Lake, Fort Irwin, and other DoD bases and installations on the use of the DDM as a decision support tool.

We have also presented the results of this Legacy project to the Desert Manager Group (DMG), located in Barstow, California, and staffed by personnel from the National Park Service, the Bureau of Land Management, NASA, and DoD. BLM, in particular, has also expressed great interest in developing and using a DDM for lands under their stewardship. Guidance from the California SHPO would therefore be valuable to BLM, and, in fact, all agencies with land stewardship responsibilities in the State of California.

We suggest the establishment of a Cooperative Research Unit (CRU) composed of CSRM, or the non-profit CSRM Foundation, and the organization that we have worked with on this Legacy project, including The Johns Hopkins University and NASA. At Johns Hopkins, we propose, specifically, both Whiting School of Engineering and the Applied Physics Lab (APL), which we have begun working with on several projects. APL has developed over 30 high resolution Lidar/hyperspectral sensors that have been used in Afghanistan and Iraq for several years, but are now available for other applications. From NASA, we propose including both the Goddard Spaceflight Center, like Johns Hopkins, located very near to the CSRM and CSRM Foundation offices, and also the NASA Jet Propulsion Laboratory at Caltech (JPL/NASA), with which we have worked from many years, and which would bring state-of-the-art synthetic aperture radar (SAR)technology and expertise to the CRU. Dr. Craig Dobson, Space Archaeology Program and Earth Surface and Interior focus area Lead at the NASA Washington Office, has offered to organize a meeting of interested agencies, including BLM, NPS, and DoD, to discuss how this might be done. In coordination with the Legacy Resource Management Program and Office of the DFPO at OSD, we will work with Dr. Dobson to begin making arrangements for this meeting. Funding for the production of DDMs could thereafter be provided by these agencies, providing an economy of scale in many places, such at the Mojave Desert, where lands under the stewardship of all of these agencies are adjacent to one another.

Acknowledgements
The principal author of this report would like to thank the Department of Defense Legacy Resource Management Program for its support of this project. In particular, I would like to thank Peter Boice, who oversees the Legacy Program along with serving as Deputy Director, Natural Resources, as well as Cecilia Brothers and Kelly Merrifield, who at different times during our project development were Cultural Resource Management Specialists for the Legacy Program.

John Brantley Jackson, Installation Archaeologist at Fort Irwin, provided strong support and expressed great interest throughout this project, as did Clarence Everly, who is Natural and Cultural Resources Program Manager at Fort Irwin. Their deep involvement in contract work that we conducted in the years prior to this project, the information that they shared with us then and the access to data sets that they provided were extremely valuable. Michael Baskerville was also steadfast in his support, and provided us with great insights into the archaeology of the Mojave Desert based upon not only his great experience with that archeology, but his understanding of it. Ryan Brown, the GIS specialist at China Lake, met with us several times to provide us information about site location there. Devin White, now of Oak Ridge Laboratories, contributed to this project greatly by sharing some of his vast knowledge of Lidar technology, and ultimately producing the precision digital surface model (DSM) for Fort Irwin. Fon Duke, of the Desert Research Institute, provided numerous connections to the data, people, and places that are highly pertinent to our project development. Without the participation of Prof. Carey Priebe of The Johns Hopkins University Whiting School of Engineering, we would not have become aware of how Bayesian Statistics have been used with great success in engineering and medical research. He provided guidance to his Ph.D. students, Daniel Sussman, Lee Chen, and Zachary Lubberts, who did much of the heavy statistical lifting. Bryce Davenport and Rob Wanner of CSRM did the bulk of the GIS work, and participated in numerous brainstorming sessions about the project. Dr. James C. Tilton of the NASA Goddard Space Flight Center consulted with us several times, and provided access to the Goddard Discovery supercomputer, which made it possible to generate probability maps in hours instead of days. His postdoctoral research fellow, Dr. Edoardo Pasolli, worked late into the night along with Bryce and Zach many times. Finally, I would like to thank Brian Lione, who provided many useful comments on the draft of this report, sharing as he did this his great familiarity with the cultural resource management program needs of the military.

Bibliography

Chase, Arlen F, Diane Z. Chase, and John F. Weishampel, 2013"The Use of LiDAR at the Maya Site of Caracol, Belize," in D. Comer and M. Harrower, Eds., *Mapping Archaeological Landscapes from Space: In Observance of the 40th Anniversary of the World Heritage Convention*, pp. 179-189, Springer, New York.

Chen, Li, Douglas C. Comer, Carey E. Priebe, Daniel Sussman, James C. Tilton, 2013, Refinement of a Method for Identifying Probable Archaeological Sites from Remotely Sensed Data. In *Mapping Archaeological Landscapes from Space (in Observance of the 40th Anniversary of the World Heritage Convention)*, D.C. Comer and M.J. Harrower (eds.), New York: Springer Press.

Cohen and Cohen, 1983, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (2nd ed.).* Hillsdale, N.J.: Erlbaum.

Comer, Douglas C., 2008, Wide-Area, Planning Level Archaeological Surveys Using SAR and Multispectral Images. In the *Proceedings of the 28th International Geophysical and Remote Sensing Symposium*, Boston, MA, 2008.

Comer, Douglas C., 2007, Merging Aerial Synthetic Aperture Radar (SAR) and Satellite Multispectral Data to Inventory Archaeological Sites (2007-11). Report on file, National Park Service, National Center for Preservation Technology and Training (NCPTT), Natchitoches, Louisiana. (see also http://ncptt.nps.gov/merging-aerial-synthetic-aperture-radar-sar-and-satellite-multispectral-data-to-inventory-archaeological-sites-2007-11/)

Comer, Douglas C. and Ronald G. Blom, 2007a, Detection and Identification of Archaeological Sites and Features Using Synthetic Aperture Radar (SAR) Data Collected from Airborne Platforms. In *Remote Sensing in* Archaeology, James R. Wiseman and Farouk El-Baz (eds.), New York: Springer Science + Business Media, LLC.

Comer, Douglas C. and Ronald G. Blom, 2007b, Remote Sensing and Archaeology: Tracking the Course of Human History from Space. *Earth Imaging Journal*, March/April.

Comer, Douglas C. and Ronald G. Blom, 2007c, Wide Area Inventory of Archaeological Sites Using Aerial and Satellite Data Sets: Prologue to Resource Monitoring and Preservation (with Ronald G. Blom). In *Proceedings of the 32cd International Symposium on Remote Sensing of Environment*. Ann Arbor, MI: Environmental Research Institute of Michigan.

Kvamme, Kevin, 1988. Development and Testing of Quantitative Models. In *Quantifying the Present and Predicting the Past: Theory, Method, and Application of Archaeological Predictive Modeling*, W. James Judge and Lynne Sebastian (eds.), Denver, CO: U.S. Department of the Interior, Bureau of Land Management, pp. 325-428.

Mason, S.J. and N.E. Graham, 2002, Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. *Q. J. R. Meteorological Society*, Vol. 128, pp. 2145–2166.

Metz, C.E., 1978, Basic principles of ROC analysis. *Seminars in Nuclear Medicine*, Vol. 8 (no. 4), pp. 283-298.

Moore, Elizabeth, 2000, Angkor Water Management, Radar Imaging, and the Emergence of Urban Centres in Northern Cambodia. *The Journal of Sophia Asian Studies*, 18 . pp. 39-51.

Moore, Elizabeth and Freeman, Anthony,1998, Circular sites at Angkor: a radar scattering model. *The Journal of the Siam Society*, 85 (Part 1 & 2). pp. 107-119.

Obuchowski, N.A., 2005, Estimating and comparing diagnostic tests' accuracy when the gold standard is not binary. *Academic Radiology*, Vol. 12, pp. 1198-204.

Ruiz, Marilyn O., 2003, The Development and Testing of an Archaeological Predictive Model. Report on file, U.S. Army Corps of Engineers Construction Engineering Research Laboratory: Champlain, IL.

Tilton, J. C., 2006, Parallel Implementation of the Recursive Approximation of an Unsupervised Hierarchical Segmentation Algorithm. In *High Performance Computing in Remote Sensing*, C.-I. Chang and A. J. Plaza (eds.), Boca Ranton, FL: CRC Press.

Tilton, J.C., 2012, Douglas C. Comer, Carey E. Priebe, Daniel Sussman, Li Chen, Refinement of a Method for Identifying Probable Archaeological Sites from Remotely Sensed Data. In *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XVIII*, SPIE Proceedings Vol. 8390, Sylvia S. Shen and Paul E. Lewis (eds.), Bellingham, WA: SPIE Press.

Tilton, J.C. and Douglas C. Comer, 2013, Method for Identifying Probable Archaeological Sites from Remotely Sensed Data. In *Mapping Archaeological Landscapes from Space (in Observance of the 40th Anniversary of the World Heritage Convention),* D.C. Comer and M.J. Harrower (eds.), ICAHM Series, Multidisciplinary Studies in Archaeological Heritage Management, New York: Springer Press.

Saturno, W.A., Sever, T.I, Irwin, D.E, Howell, B.F, Garrison, T.G., 2007, Putting us on the map: remote sensing investigation of the ancient Maya landscape. In: Wiseman, J.R., El-Baz, F. (Eds.), *Remote Sensing in Archaeology*, New York: Springer Press, pp. 137-1

Institutionalizing Protocols for Wide-Area Inventory and Evaluation of Archaeological Sites by the Analysis of Aerial and Satellite Imagery

Project #11-158

Background:

Within Federal preservation law, Section 106 is treated as a compliance obligation, while Section 110 requirements are often unmet because of the extraordinary time and costs associated with a comprehensive, on-ground survey that, until now, has been the only means by which to comply with Section 110. Section 110 requires each Federal agency to establish "a preservation program for the identification, evaluation, and nomination to the National Register of Historic Places, and protection of historic properties." On the ground survey and evaluation in the service of this has proven prohibitively expensive. Because there are no clear timelines for completing Section 110 requirements, no Federal agency has completed survey and evaluation. Yet in the absence of 110 survey and evaluation, important sites remain undocumented and unprotected, and there is no sound decision support prior to conducting 106 surveys for directing undertakings away from areas that are likely to contain important (significant, in terms of National Register criteria) sites.

Objective:

The objective of this effort was to develop practical and immediately available protocols for the analysis of aerial and satellite imagery to produce a direct detection model (DDM) that would serve the intended purposes of a Section 110 survey and evaluation.

Summary of Approach:

We employ Bayesian statistical analysis of direct returns from a variety airborne and satellite remote sensors. In our test areas within China Lake Naval Air Weapons Station and Fort Irwin National Training Center, we utilized high-resolution eight-band WorldView-2 satellite multispectral data and Lidar data, but our approach can also employ hyperspectral and synthetic aperture radar data. Matlab scripts are run in parallel on multiple processors. Images are treated as data matrices: the approach is not one of image enhancement. Nor is the model an archaeological predictive model (APM); it is based on direct detection of archaeological sites and materials.

Benefit:

The decision support tool that we have developed offers substantial and immediate cost avoidance to the military by minimizing activities required to comply with the National Historic Preservation Act of 1966(as amended), specifically Sections 106 and 110. The tool described in this report will provide:

- A substantial reduction in the number and intensity of required Section 106 archaeological surveys
- A practical way to find, categorize by age and materials, and evaluate sites and the context of those sites without the use of traditional survey methods
- A way to greatly increase training areas without fear of damage to archaeological resources
- The tool by which to protect the truly important archaeological sites on DoD lands
- *De facto* Section 110 surveys
- The means by which to strategically plan Section 106 Surveys by merging DoD training and facility needs with information about distribution and significance of archaeological sites
- A rich source of information by which to review, interpret, and improve upon the results of previously conducted archaeological surveys without conducting additional fieldwork, allowing for fewer mission delays.

Accomplishments:

The DDM is completely developed, the culmination of research that began in 2004. Our statistical treatments are modifications of those that have been developed for engineering and medical science applications for the past 20 years. We have also developed a research unit that can process the extremely large body of data needed to detect subtle landscape differences that demarcate most archaeological sites on military lands.

Contact Information:

Dr. Douglas C. Comer
Principal
Cultural Site Research and Management, Inc.
2113 St. Paul St.
Phone: 410 244-6320
Fax: 410 244-6324
Email: dcomer@culturalsite.com