FINAL REPORT

# ESTCP Pilot Program
# Classification Approaches in Munitions Response
# San Luis Obispo, California

May 2010


Herb Nelson,
ESTCP

Anne Andrews,
ESTCP

Katherine Kaye,
ESTCP Support Office, HydroGeoLogic, Inc.


Environmental Security Technology
Certification Program

# Report Documentation Page

*Form Approved*
*OMB No. 0704-0188*



| | | |
|---|---|---|
| 1. REPORT DATE<br>**MAY 2010** | 2. REPORT TYPE | 3. DATES COVERED<br>**00-00-2010 to 00-00-2010** |
| 4. TITLE AND SUBTITLE<br>**ESTCP Pilot Program Classification Approaches in Munitions Response San Luis Obispo, California** | | 5a. CONTRACT NUMBER<br>5b. GRANT NUMBER<br>5c. PROGRAM ELEMENT NUMBER |
| 6. AUTHOR(S) | | 5d. PROJECT NUMBER<br>5e. TASK NUMBER<br>5f. WORK UNIT NUMBER |
| 7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)<br>**Strategic Environmental Research and Development Program (SERDP),Environmental Security Technology Certification Program (ESTCP),4800 Mark Center Drive, Suite 17D08,Alexandria,VA,22350-3605** | | 8. PERFORMING ORGANIZATION REPORT NUMBER |
| 9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) | | 10. SPONSOR/MONITOR'S ACRONYM(S)<br>11. SPONSOR/MONITOR'S REPORT NUMBER(S) |

12. DISTRIBUTION/AVAILABILITY STATEMENT
**Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

| 16. SECURITY CLASSIFICATION OF: | | | 17. LIMITATION OF ABSTRACT | 18. NUMBER OF PAGES | 19a. NAME OF RESPONSIBLE PERSON |
|---|---|---|---|---|---|
| a. REPORT<br>**unclassified** | b. ABSTRACT<br>**unclassified** | c. THIS PAGE<br>**unclassified** | **Same as Report (SAR)** | **75** | |

# TABLE OF CONTENTS

# FIGURES

# TABLES

# ACRONYMS

| | |
|---|---|
| BUD | Berkeley UXO Discriminator |
| CNG | California National Guard |
| DoD | Department of Defense |
| DSB | Defense Science Board |
| EM(I) | Electromagnetic (Induction) |
| ESTCP | Environmental Security Technology Certification Program |
| FUDS | Formerly Used Defense Site |
| GPS | Global Positioning System |
| GSA | General Services Administration |
| HRR | Historical Record Review |
| IDA | Institute for Defense Analyses |
| IMU | Inertial Measurement Unit |
| LBNL | Lawrence Berkeley National Laboratory |
| MM | MetalMapper |
| MMRP | Military Munitions Response Program |
| MR | Munitions Response |
| MRS | Munitions Response Site |
| MSEMS | Man-portable Simultaneous Electromagnetic-Magnetic Sensor |
| MTADS | Multi-sensor Towed Array Detection System |
| NRL | Naval Research Laboratory |
| QC | Quality Control |
| ROC | Receiver Operating Characteristic |
| SIG | Signal Innovations Group |
| SLO | San Luis Obispo |
| SNR | Signal to Noise Ratio |
| TEMTADS | Time Domain Electromagnetic MTADS |
| TOI | Target of Interest |
| UXO | Unexploded Ordnance |

# ACKNOWLEDGMENTS

The ESTCP Classification Pilot Program would not have been possible without the assistance of numerous individuals. We would like to acknowledge Shelley Cazares, Michael Tuley, and Jennifer Soult from the Institute for Defense Analyses who were instrumental in the overall program design and analysis of the demonstrator results. We would also like to thank the ESTCP Classification Study Advisory Group listed below. They were involved in site selection, program design, data review, and the development of conclusions and methods as a whole.

Jim Austreng, California Department of Toxic Substances Control
Harry Craig, U.S. EPA Region 10, Oregon Operations Office
Jon Haliscak, AFCEE Technical Directorate
Doug Maddox, US Environmental Protection Agency
Andrew Schwartz, U.S. Army Corps of Engineers, Huntsville
Jeff Swanson, Colorado Department of Public Health and Environment
Ken Vogler, Colorado Department of Public Health and Environment
Roger Young, U.S. Army Corps of Engineers, Huntsville

We would like to credit the technology demonstrators.

- SAIC, Signal Innovations Group, Sky Research Inc., and RML Technologies, Inc., Parsons Inc., and NAEVA Geophysics, Inc. for classification data analysis;
- Lawrence Berkeley National Laboratory, Geometrics, and NAEVA Geophysics Inc, for data collection and supporting analysis for the Berkeley UXO Discriminator (BUD), MetalMapper, and EM61 CART system respectively;
- SAIC and NAEVA Geophysics Inc. for the MSEMS data collection; and
- Nova Research, Inc. for the EM-ARRAY, Magnetometer ARRAY and TEMTADS data collections.

For program support, we acknowledge

- Explosive Ordnance Technologies Inc. (EOTI) for excavation of the characterization grids;
- Parsons for seeding and validation excavations;
- Nagi Khadr from SAIC, Inc., who served as a representative for the ESTCP Program Office, for data analysis;
- BJ Allen and Lloyd Godard from the USACE for their assistance in supporting the Former Camp San Luis Obispo demonstration;
- Doug Murray and Shawn Jorgensen from the Naval Ordnance Safety and Security Activity for their support with the Explosive Safety Submission; and
- Clif Youmans from the Montana Army National Guard, Cheryl Edwards from the U.S. Army Aberdeen Test Center, and Bob Selfridge and Amy Walker from the USACE for supplying inert rounds used for the seeding.

Finally, ESTCP thanks California Polytechnic State University, owners of the land involved in this demonstration, for their patience and cooperation. We would particularly like to thank David Ragsdale for the time he dedicated to supporting this study.

# EXECUTIVE SUMMARY

The Military Munitions Response Program (MMRP) is charged with characterizing and, where necessary, remediating munitions-contaminated sites. When a site is cleaned up, it is typically mapped with a geophysical system, based on either a magnetometer or electromagnetic induction (EMI) sensor, and the locations of all detectable signals are excavated. Many of these detections do not correspond to munitions, but rather to other harmless metallic objects or geology: field experience indicates that often in excess of 90% of objects excavated during the course of a munitions response are found to be nonhazardous items. Current technology, as it is traditionally implemented, does not provide a physics-based, quantitative, validated means to discriminate between hazardous munitions and nonhazardous items.

The MMRP is severely constrained by available resources. Remediation of the entire inventory using current practices is cost prohibitive, within current and anticipated funding levels. With current planning, estimated completion dates for munitions response on many sites are decades out. If the savings possible from classifying objects as either munitions or other harmless objects were realized, the limited resources of the MMRP could be used to accelerate the clean up of munitions response sites that are currently forecast to be untouched for decades.

The Environmental Security Technology Certification Program (ESTCP) has initiated a Classification Pilot Program to validate the application of a number of recently developed technologies in a comprehensive approach to munitions response. The pilot program envisions a series of such demonstrations at live sites of increasing difficulty. A hillside range at the former Camp San Luis Obispo, CA was selected for the second of these demonstrations. At this site, there were four known targets of interest prior to the study including 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets.

Prior to the demonstration, an EMI survey was conducted over approximately 30 acres of the hillside. These data guided the selection of the final 11.6-acre demonstration area and the location of two 50-ft x 50-ft grids to be used for site characterization. Excavation of thee two grids yielded 369 items including 60-mm, 81-mm, and 4.2-in mortar fragments. Although no evidence for 2.36-in rockets was found in these grids, rockets debris were observed during earlier site reconnaissance.

No item was found deeper than 30 cm and the dig team reported that the soil density was such that it was unlikely that anything would be found deeper than that. Based on these results and, in consultation with the Program Advisory Group, a 50% safety margin was applied to this observation and the depth of interest was established as 45 cm.

Using the EMI data and this depth of interest as a guide, a seed plan was developed for the site and an instrument verification strip (IVS) established. The IVS comprised two examples each of the four munitions expected and two shotputs for calibration. Similarly, the four known munitions were used for the 200 blind seeds.

The demonstration consisted of several combinations of data-collection platforms and analysis approaches, ranging from careful application of commercial survey instruments to three prototype systems specially designed to maximize detection and classification of munitions. The systems demonstrated fell into three broad classes:

- SURVEY MODE: The commercial survey systems were deployed to collect data on 100% of the site, called SURVEY mode.
- CUED MODE: Two sensors, the Time Domain Electromagnetic Multi-sensor Towed Array Detection System (TEMTADS) and the Berkeley UXO Discriminator (BUD), were deployed to collect data at the locations of individual anomalies detected by the EM61 ARRAY.
- SELF-CUED MODE: The MetalMapper system (MM) is intended to operate in both survey and cued mode. MM performed a detection survey and collected cued data over all the anomalies it detected.

After each of the survey systems completed data acquisition, anomalies were selected from the data using a procedure designed by the program office. A detection list was generated by recording all locations for which the sensor signal exceeded a system-specific threshold. Since these individual sensor detection lists were the basis for all subsequent analyses, a rigorous process was used to set this threshold.

The targets of interest in this demonstration were 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets. Based on the results of the exploratory intrusive investigation conducted at the beginning of the demonstration, no targets are expected deeper than 30 cm below the surface. Since the signal from each of these targets can be predicted accurately as a function of depth, the threshold for each sensor can be chosen to ensure detection of all items of interest while leaving smaller targets off the list. For this demonstration, the anomaly selection threshold for each sensor was set as the smallest signal expected from any of the items at 45 cm below the surface. This threshold corresponds to the 30 cm depth of interest established from the digs in the characterization grids with a 50% safety margin.

Using this threshold, the EM61-MK2 ARRAY and magnetometer array detected all seeds in the area they were able to survey. One of the seeds was not selected in the EM61-MK2 CART survey, even though it was covered during the survey. The item missed was a 60-mm mortar body without fins or nose. The threshold of 11.3 mV was developed from the response of a complete 60-mm mortar. The partial round missed produced a response of 10.7 mV, just under the threshold. This reinforces the point that target-based thresholds need to be chosen with care and a complete understanding of the items expected to be encountered at the site.

A list of locations for intrusive investigation, the dig list, was prepared by combining the anomaly lists from each sensor in a stepwise fashion. The starting point of the list was the 1464 detections by the EM61-MK2 ARRAY which were used to cue the TEMTADS and BUD systems. To these were added unique detections by the EM61 CART and the MSEMS EM61 sensor. The MetalMapper system performs self-cueing so the detection list for this system was compiled separately. There were 387 unique MetalMapper detections added to the final dig list for a total of 2588 dig locations.

At each dig location, the unexploded ordnance specialists carefully excavated all metal objects, photographed them, and measured their exact locations. These dig data make up the ground truth for the demonstration. In addition to the 200 seed items, 26 UXO were found as well as 25 items that, while not presenting an explosive hazard, were so close in character to potential UXO that they were declared as targets of interest (TOI).

Multiple groups demonstrated processing approaches. The basic classification method for all but one demonstrator involved using a geophysical model to estimate target parameters that may be useful in making a classification decision. Although the processing approaches differ in their manner of implementation, all the geophysical models are based on a dipole approximation.

For the mapping sensors, this process involves using data from multiple spatially diverse locations that together fully characterize the signature. The cued sensors, BUD, TEMTADS and MetalMapper, collect sufficient data at a single spatial location to support model-based parameter estimation using custom software developed by the system developers.

The anomaly lists were divided into training and blind testing sets. Some of the truth data was provided to each demonstrator for algorithm training, so that the parameters or features that were most useful for classification could be determined and thresholds in the classification process set. After training, the decision process for each algorithm was finalized and documented, and the demonstrators submitted ranked dig lists for the blind test set arranged such that the items which they were most confident were not hazardous were at the top of the list, followed by those items for which they could not make a decision. The anomalies which they were confident were targets of interest were placed at the bottom of the list followed by targets for no analysis could be performed. For most demonstrators, this "can't analyze" class contained only a handful of items.

The demonstration was scored based on the demonstrator's ability to eliminate nonhazardous items while retaining all detected TOI. The results are presented as receiver operating characteristic (ROC) curves, an example of which is shown in Figure ES-1. This curve plots the percentage of the targets of interest recovered as a function of the number of non-TOI that had to be dug. The points are color-coded according to how they were classified by the analyst with red corresponding to high-confidence TOI, yellow to can't decide, and green to high-confidence not TOI. The first point plotted is offset from the origin to reflect the 200 training digs provided to this analyst. Two additional points are plotted on the figure. The orange dot indicates the point where 100% of the TOI have been found. The blue dot indicates the demonstrator's dig threshold.

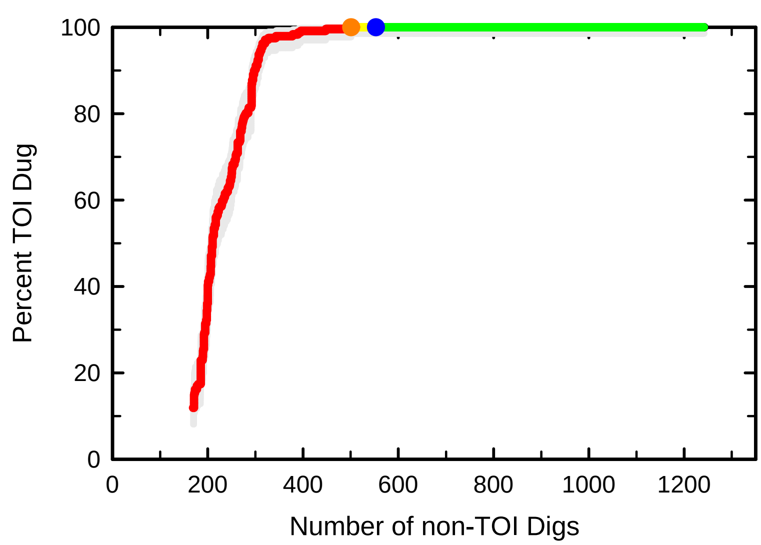


**Figure ES-1. Receiver operating characteristic curve resulting from analysis of the EM61-MK2 CART data.**

The receiver operating characteristic (ROC) curve in Figure ES-1 results from analysis of the EM61-MK2 CART data. A feature based on decay of the induced current was the primary discriminant used in this analysis. Using the data from the commercial sensor, this analyst was able to analyze all targets and was able to correctly classify more than 600 of the 1250 non-TOI. In addition, they set their threshold appropriately, slightly beyond the point where all TOI were identified.

Even better results were obtained using the data collected by the advanced EMI sensors. Figure ES-2 plots the ROC curve resulting from analysis of the data collected by the MetalMapper system in cued mode. Notice that the red portion of the curve is much more vertical indicating that the analyst was able to efficiently identify targets of interest with few false positives. Even more impressively, this analyst was able to correctly classify nearly 1000 items as nonhazardous. The dig threshold from this analysis is slightly too aggressive, resulting in a few missed TOI at the demonstrator threshold.

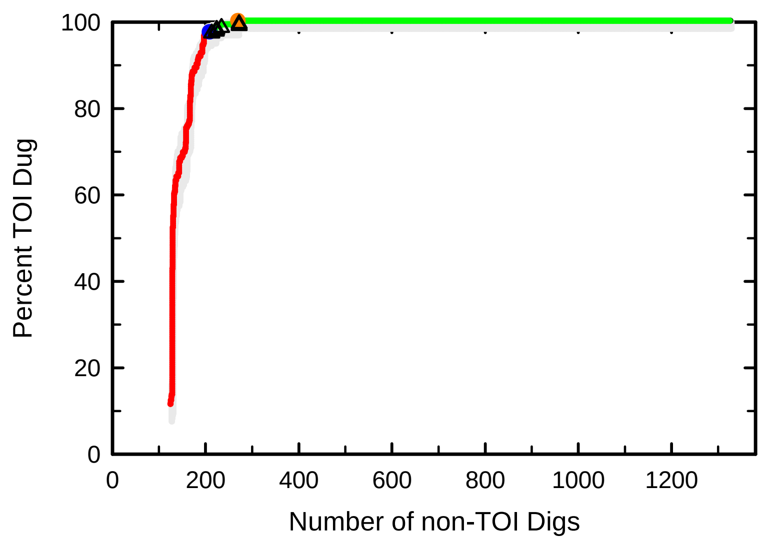


**Figure ES-2. ROC curve resulting from analysis of the MetalMapper data.**

Using these MetalMapper results as an indication of what is possible from the advanced sensors, we have constructed a notional cost model for munitions response projects. The scenario imagines a site that is 100 acres, with 15,000 total detections above the target selection threshold and 200 TOI recovered. The digging costs are assumed to be $90 per target. We assume 200 training targets were dug and provided and the fraction of items correctly identified as non-TOI at SLO by MetalMapper was applied to the 15,000 total detections. Figure ES-3 shows the results of this model. The $1.4M cost of the detection only process could have been reduced to approximately $400K using the MetalMapper classification results to guide the digging, for a savings of about $1M on this 100-acre site.

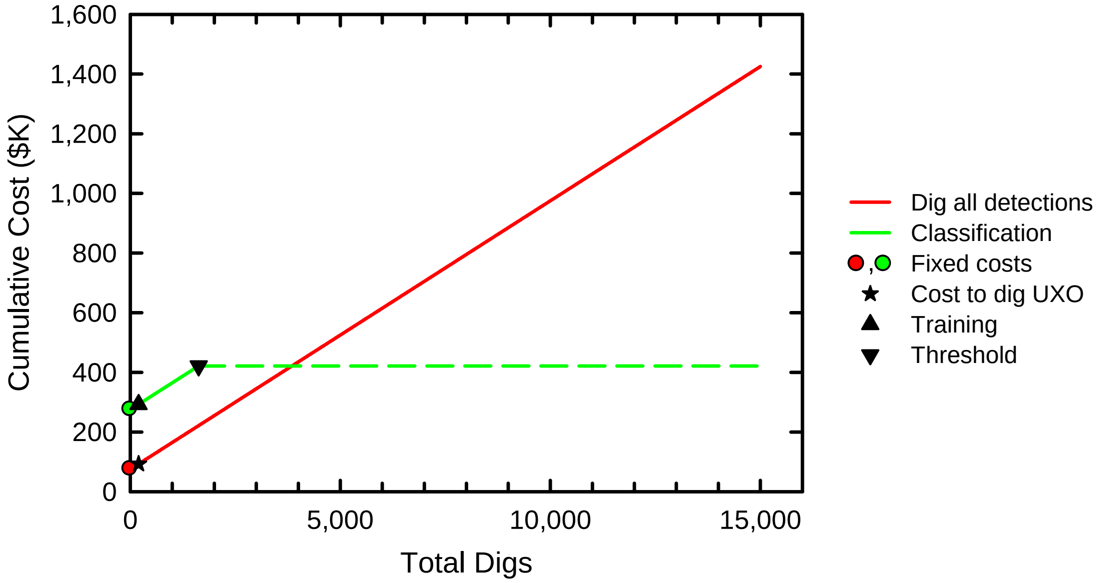


**Figure ES-3. Cost model for the use of MetalMapper on a 100 acre site.**

In summary, the demonstration at the former Camp San Luis Obispo showed substantial classification ability on a site that was more challenging than the Camp Sibert demonstration site with a wider mix of targets of interest (TOI). In the application of classification analyses to carefully collected survey data from commercial EM sensors, the best performers correctly classified all or nearly all of the targets of interest, while achieving reductions of up to 50% in the number of non-TOI. Recently-developed EMI sensors, optimized for UXO classification, demonstrated at San Luis Obispo showed outstanding results. A simple cost model based on these results shows the potential for large savings as the classification approach is adopted on munitions response sites.

# 1 INTRODUCTION

## 1.1 BACKGROUND

Munitions response is a high-priority problem for the Department of Defense (DoD). Approximately 3600 sites, comprising tens of millions of acres, are suspected of contamination with military munitions, which include unexploded ordnance (UXO) and discarded military munitions. (Ref. 1) Many of these are formerly used defense sites (FUDS), which are no longer under DoD control, and are used for a variety of purposes, including residential development, recreation, grazing, and parkland, often without restriction.

The Military Munitions Response Program (MMRP) is charged with characterizing and, where necessary, remediating munitions-contaminated sites. When a site is cleaned up, it is typically mapped with a geophysical system, based on either a magnetometer or electromagnetic induction (EMI) sensor, and the locations of all detectable signals are excavated. Many of these detections do not correspond to munitions, but rather to other harmless metallic objects or geology: field experience indicates that often in excess of 90% of objects excavated during the course of a munitions response are found to be nonhazardous items. Current technology, as it is traditionally implemented, does not provide a physics-based, quantitative, validated means to discriminate between hazardous munitions and nonhazardous items.

With no information to suggest the origin of the signals, all anomalies are currently treated as though they are intact munitions when they are dug. They are carefully excavated by certified UXO technicians using a process that often requires expensive safety measures, such as barriers or exclusion zones. As a result, most of the costs to remediate a munitions-contaminated site are currently spent on excavating targets that pose no threat. If these items could be determined with high confidence to be nonhazardous, some of these expensive measures could be eliminated or the items could be left unexcavated entirely.

The MMRP is severely constrained by available resources. Remediation of the entire inventory using current practices is cost prohibitive, within current and anticipated funding levels. With current planning, estimated completion dates for munitions response on many sites are decades out. The Defense Science Board (DSB) observed in its 2003 report that significant cost savings could be realized if successful classification between munitions and other sources of anomalies could be implemented. (Ref. 2) If these savings were realized, the limited resources of the MMRP could be used to accelerate the clean up of munitions response sites that are currently forecast to be untouched for decades.

## 1.2 CLASSIFICATION CONCEPT

Classification is a process used to make a decision about the likely origin of a signal. In the case of munitions response, high-quality geophysical data can be interpreted with physics-based models to estimate parameters that may be useful for classification. The parameters in these models are related to the physical attributes of the object that resulted in the signal, such as its physical size and aspect ratio. The values of these parameters may then be used to estimate the likelihood that the signal arose from an item of interest, that is, a munition. Other approaches attempt classification using data-based features, such as the amplitude, footprint, shape and statistical moments of the anomaly directly.

Magnetometer data are typically fit using a simple model of a single dipole moment, which is related to the physical size of the object. EM data are fit to a more complex polarizability model that can yield a larger set of parameters that more completely describe the source of the signal. The EM parameters relate to the physical size of the object, its aspect ratio, the wall thickness, and the material properties.

Munitions are typically long, narrow cylindrical shapes that are made of heavy-walled steel. Common clutter objects can derive from military uses and include exploded parts of targets, such as vehicles, as well as munitions fragments, fins, base plates, nose cones and other munitions parts. Other common clutter objects are man-made nonmilitary items. While the types of objects that can possibly be encountered are nearly limitless, common items include barbed wire, horseshoes, nails, hand tools, and rebar. These objects and geology give rise to signals that will differ from munitions in the parameter values that are estimated from geophysical sensor data.

Once the parameters are estimated, a means must be found to sort the signals to identify items of interest, in this case munitions, from the clutter. This is termed classification. In a simple situation, one can imagine sorting items based on a single parameter, such as object size. A rule could be made that all objects with an estimated size larger than some value will be treated as potentially munitions items of interest, such as large bombs, and those smaller could not possibly correspond to intact munitions.

In reality, many classification problems cannot be handled successfully based on a single parameter. Because the parameter-estimation process is imperfect and the physical sizes of the objects of interest may overlap with the sizes of the clutter objects, it is rare to get perfect separation based on one parameter. For complex problems, sophisticated statistical classifiers can combine the information from multiple parameters to make a quantitative estimate of the likelihood that a signal corresponds to an item of interest.

## 1.3 ESTCP PILOT PROGRAM

The Environmental Security Technology Certification Program (ESTCP) is charged with demonstrating and validating innovative, cost-effective environmental technologies. In response to the DSB Task Force report (Ref. 2) and Congressional interest, ESTCP initiated a Classification Pilot Program to validate the application of a number of recently developed technologies in a comprehensive approach to munitions response. The pilot program envisions a series of such demonstrations at live sites of increasing difficulty. This report summarizes the results of the second of these demonstrations at the former Camp San Luis Obispo.

Some form of classification is used on all munitions response projects, most often implicitly. In the case of traditional "mag and flag," the operator adjusts the sensitivity audio control and makes a decision as to whether each signal is significant. Since no data are recorded, these decisions can never be reviewed. In the case of digital geophysical mapping, a threshold is selected for determining targets of interest, and often a geophysicist uses professional judgment to decide based on a visual inspection of shape and amplitude whether anomalies are likely to arise from geology or compact metallic objects. In both cases, the sources of signals deemed insignificant are not further investigated and remain in the ground.

Significant progress has been made in explicit classification technology. To date, emerging technologies have primarily been tested at constructed test sites, with only limited application at live sites. The routine implementation of classification technologies requires demonstrations at real munitions response sites under real-world conditions. Any attempt to declare detected anomalies to be harmless will require demonstration to regulators, safety personnel, and project managers of not only individual technologies, but an entire decision-making process.

The goal of the pilot program is to demonstrate that classification decisions can be made explicitly, based on principled physics-based analysis that is transparent and reproducible. As such, the objectives of the pilot program are to:

- test and validate detection and classification capabilities of currently available and emerging technologies on a real site under operational conditions, and
- investigate how classification technologies can be implemented in cleanup operations in cooperation with regulators and program managers.

To address the second of those objectives, a Program Advisory Group composed of representatives of the Services and State and National regulators was established at the beginning of the program. This Advisory Group is involved with site selection, program design, data review, and the development of conclusions. The Advisory Group has been heavily involved in drafting this report.

## 1.4 RESULTS FROM CAMP SIBERT

The Former Camp Sibert in Alabama was selected as the first pilot site with success in mind. This site presented a single munitions type (the 4.2-inch mortar) and benign conditions where high-quality data could be collected. The motivation of this selection was to demonstrate a process under conditions where the technologies were expected to perform well, so that the advisory group could have a meaningful discussion regarding the application of classification.

The pilot program demonstrated successful classification on this simple site. With carefully collected survey data from either magnetometers or EM sensors and transitioning physics-based analysis techniques, well over half the detected clutter items were routinely eliminated with high confidence, while all or nearly all the munitions were correctly classified. Table 1-1 shows example results from the Camp Sibert demonstration for classification algorithms applied to data collected with four systems. The EM61-MK2 CART and EM61-MK2 ARRAY data were processed by algorithm developers as well as the Munitions Response contractor on site. In all cases, the classification processing correctly identified all or nearly all the munitions and a significant fraction of the clutter was successfully identified as such with high confidence. Classification processing applied to data from the commercial instruments eliminated 45%–70% of the clutter in these examples. Figure 1-1 provides an example of these commercial instrument results, highlighting the number of avoided digs using EM-61 ARRAY and CART data processed with the UX-Analyze classification approach. The Berkeley UXO Discriminator (BUD) is a next generation sensor designed to maximize classification information. It achieved nearly perfect results. More information on the first phase of the program approach and results is available in the ESTCP Program Office Final Report. (Ref. 3)

Table 1-1. Example classification results from the Camp Sibert demonstration.

| Sensor & Performer | # Munitions in Survey Area | % Munitions Correctly Identified | # Non-munitions Detected | % Non-munitions Correctly Identified |
|---|---|---|---|---|
| Mag Array Sky Research | 118 | 100 | 615 | 44 |
| EM61-MK2 CART Parsons | 118 | 99 | 428 | 44 |
| EM61-MK2 ARRAY SAIC, Inc. | 119 | 99 | 615 | 72 |
| BUD Lawrence Berkeley National Laboratory | 56 | 100 | 209 | 97 |

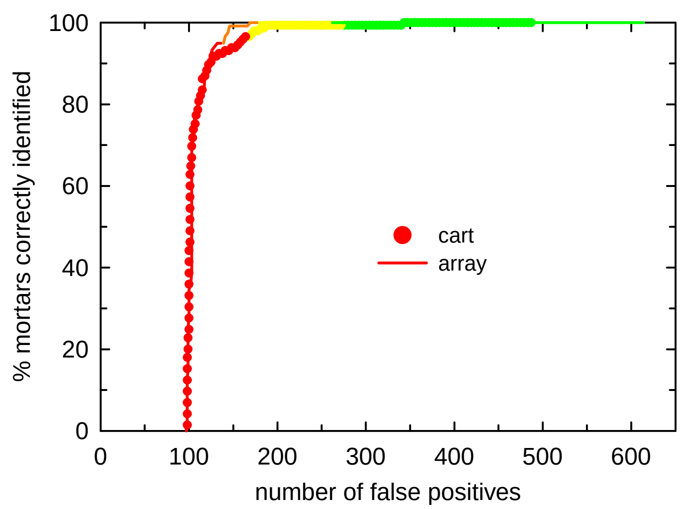


Figure 1-1. Receiver operating characteristic curve for EM-61-MK2 ARRAY and CART data from the former Camp Sibert processed with the UX-Analyze classification approach. The number of avoided digs is highlighted in green.

## 1.5 ABOUT THIS REPORT

ESTCP sponsored a second study in 2009 at a site with more challenging topography and a wider mix of targets of interest. A hillside range at the former Camp San Luis Obispo, CA was selected for this demonstration. Camp Sibert had only one target-of-interest so the physical "size" of the item was an effective discriminant. At Camp San Luis Obispo, there were at least four known targets of interest prior to the study including 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets. This site was chosen as the next in a progression of increasingly more complex sites for demonstration of the classification process.

This report is intended to provide an overview of the key results from the second phase of the pilot program for project managers, regulators and contractors. The focus of this report is on commercial instruments with available processing and emerging purpose-built munitions classification sensors. However, the material covered in this report represents only a small part of a much larger study, and notably absent here is any discussion of advanced and innovative signal processing techniques. More information about the entire demonstration and these topics in particular may be found in the individual demonstrator reports (Refs. 4-14) and an independent performance assessment by the Institute for Defense Analyses. (Ref. 15)

The report begins with a description of the site and an overview of the program approach. We then describe the detection and classification performance. This is followed by a discussion of costs and a summary of the program conclusions.

# 2 FORMER CAMP SAN LUIS OBISPO

The former Camp San Luis Obispo, California was selected for this study. The site spans a hillside that is a historical mortar target and has limited vegetation and geologic interference, and a variety of munitions types. This site was chosen as the next in a progression of increasingly more complex sites for demonstration of the classification process. The first site in the series, Camp Sibert, had only one target of interest, the 4.2-in. mortar, and the physical "size" of the item was an effective discriminant. At this site, there were four targets of interest known prior to the study: 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets.

## 2.1 SITE HISTORY AND CHARACTERISTICS

The site description material reproduced in this section is taken from the recent Site Inspection report (Ref 16). More details can be obtained in the report. The former Camp San Luis Obispo is approximately 2,101 acres situated along Highway 1, approximately five miles northwest of San Luis Obispo, California. The majority of the area consists of mountains and canyons. The demonstration area is a mortar target on a hilltop in Munitions Response Site (MRS) 05 within former Rifle Range #12 as seen in Figure 2-1.

Camp San Luis Obispo was established in 1928 by California as a National Guard Camp. Identified at that time as Camp Merriam, it originally consisted of 5,800 acres. Additional lands were added in the early 1940s until the acreage totaled 14,959. During World War II, Camp San Luis Obispo was used by the U.S. Army for infantry division training including included artillery, small arms ranges, mortar, rocket, and grenade ranges. According to the Preliminary Historical Records Review (HRR), there was a total of 27 ranges and thirteen training areas located on Camp San Luis Obispo during World War II. Construction at the camp included typical dwellings, garages, latrines, target houses, repair shops, and miscellaneous range structures. Following the end of World War II, a small portion of the former camp land was returned to its former private owners.

The U.S. Army was making arrangements to relinquish the rest of Camp San Luis Obispo to the State of California and other government agencies when the conflict in Korea started in 1950. The camp was reactivated at that time. The U.S. Army used the former camp during the Korean War from 1951 through 1953 where the Southwest Signal Center was established for the purpose of signal corps training. The HRR identified eighteen ranges and sixteen training areas present at Camp San Luis Obispo during the Korean War. A limited number of these ranges and training areas were used previously during World War II.

Following the Korean War, the camp was maintained in inactive status until it was relinquished by the Army in the 1960s and 1970s. Approximately 4,685 acres was relinquished to the General Services Administration (GSA) in 1965. GSA then transferred the property to other agencies and individuals beginning in the late-1960s through the 1980s. Much of the transfer was for educational purposes to California Polytechnic University, San Luis Obispo and Cuesta College. A large portion of Camp San Luis Obispo (the original 5,880 acres) has been retained by the California National Guard (CNG) and is not part of the FUDS program.

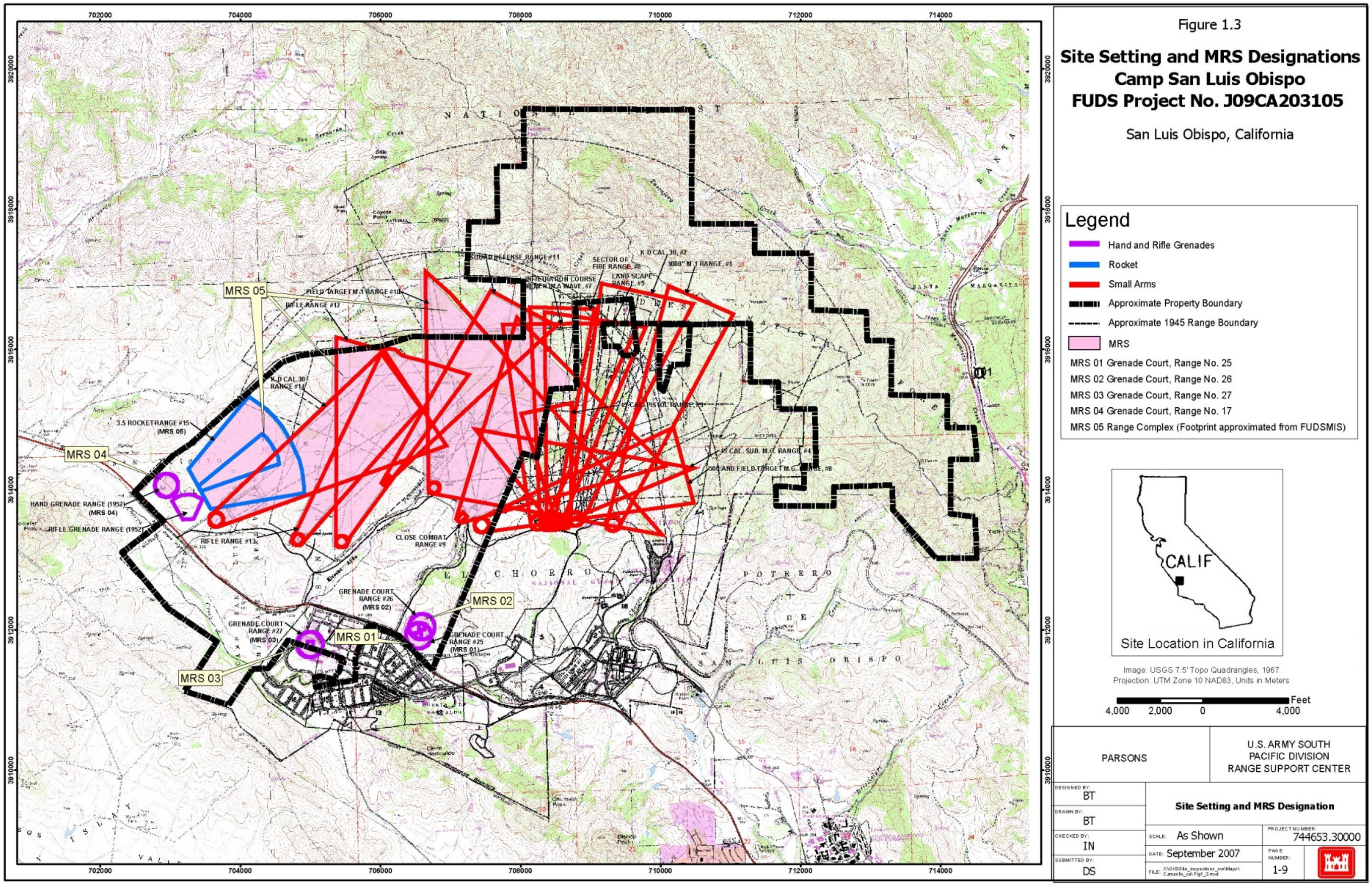


Figure 2-1. Map of the Camp San Luis Obispo FUDS showing the location of MRS 05 taken from Ref 16.

The Classification Study site is shown in Figure 2-2. Historic evidence and recent field observations document intact munitions and debris associated with 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets in this area. Other items can be considered likely as the historical record indicates extensive and diverse uses of overlapping ranges. The demonstration area was selected a sufficient distance from the target center to ensure the average anomaly density was 100-200 anomalies per acre, low enough to allow for classification of individual items.

Figure 2-2. Former Camp San Luis Obispo Classification Study area.

## 2.2 DEMONSTRATION PREPARATION

Several activities occurred prior to data collection to ensure the resulting data would support a successful demonstration. These activities included a screening magnetometer survey, an initial EMI survey, excavation of two 50-ft × 50-ft site characterization grids, construction of an instrument verification strip and training pit, and emplacement of seed targets.

### 2.2.1 Initial Magnetometer and Electromagnetic Induction Surveys

Prior to selection of the location for the second demonstration, initial magnetometer transect surveys were conducted at two candidate sites. At San Luis Obispo, these transects sampled about 12% of the 37 acre hillside and were used to estimate the anomaly density and geologic background.

Once San Luis Obispo was selected for the demonstration, the anomaly densities estimated from the initial magnetometer transects were used to guide the selection of approximately 30 acres to conduct a 100% coverage EMI survey. The EMI survey results guided selection of the final 11.6-acre demonstration site, Figure 2-3, to ensure an anomaly density between 100 and 200 anomalies per acre. This density provided a sufficient number of targets to support statistical analysis of the results, while still allowing for successful classification of isolated targets. One contiguous area was chosen to make up the demonstration site. It was designed to avoid the large rock outcroppings on the top of the hillside and the high density target center to the southwest.

The EMI survey also was used to guide the location of the instrument verification strip, the characterization grids, and the emplacement of seed targets.

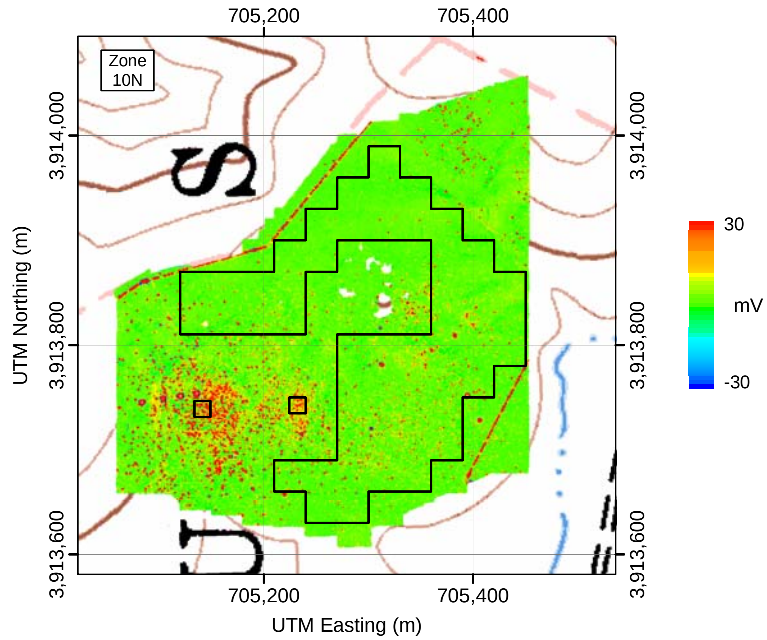


**Figure 2-3. Initial EM61-MK2 CART survey of the former Camp San Luis Obispo study area showing the final demonstration area outline and the two site characterization grids.**

### 2.2.2 Site Characterization Grid

Two 50-ft x 50-ft grids were excavated to provide information about the types and depths of munitions and clutter on the site. The grids were selected in the high density target center identified in the initial EMI survey to provide the maximum information about the items on the site. A total of 156 anomalies identified in the initial EMI data were initially dug and then the dig team re-surveyed the site and excavated any remaining anomalies above background. In total, the team excavated 369 items within the two grids including 60-mm, 81-mm, and 4.2-in mortar fragments, unknown fragments, and fuzes. No evidence 2.36-in rockets was identified.

The items were separated into classes as shown in Table 2-1, and examples of the excavated items are shown in Figure 2-4. No item was found deeper than 30 cm and the dig team reported that the soil density was such that is was unlikely that anything would be found deeper than that. In consultation with the Advisory Group, a 50% safety margin was applied to this observation and the depth of interest at the site established as 45 cm.

**Table 2-1. Class of items excavated from the site characterization grid.**

| Class | Number |
|---|---|
| Intact Munitions | 0 |
| Munitions Debris | 366 |
| Cultural Debris | 3 |
| Hot Soil/No Contact | 0 |

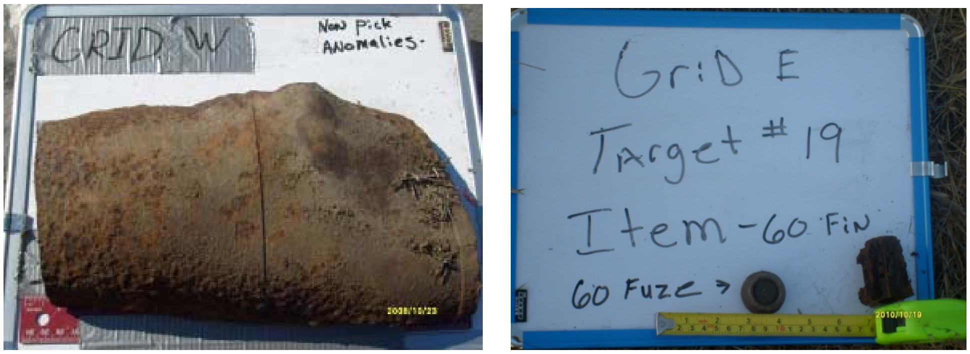


Figure 2-4. Items recovered during the excavation of the site characterization grids. A fragment is on the left and a fin and fuze are on the right.

The identities and depths of all items recovered from the site characterization grids were provided to the demonstrators as background information about the site.

### 2.2.3 Instrument Verification Strip

A quiet area was located near the entrance to the demonstration site to establish an instrument verification strip to be used for daily verification of proper sensor operation (Ref. 17). Each data collection day, each demonstrator surveyed the strip morning and evening, before and after survey activities. By comparing the twice-daily responses to the expected responses from these targets (Ref. 18), the demonstrators verified that their equipment was operating properly. Table 2-2 provides details on the buried targets. All targets were buried to a nominal depth of 30 cm.

Table 2-2. Instrument Verification Strip targets.

| Munition type | Quantity | Orientations |
|---|---|---|
| Shotput | 2 | NA |
| 60-mm mortar rounds | 2 | Vertical and horizontal at 120° |
| 81-mm mortar rounds | 2 | Vertical and horizontal at 120° |
| 4.2-in mortar rounds | 2 | Vertical and horizontal at 120° |
| 2.36-in rockets | 2 | Vertical and horizontal at 120° |

### 2.2.4 Training Pit Measurements

A pit was dug near the site to support collection of sensor data for algorithm training. Based on the input from the data processors, signature data was collected from the each of the targets of interest at a variety of depths and orientations. These data were used to confirm the detection thresholds for each sensor and as training data for the analysis demonstrators.

### 2.2.5 Seeding the Survey Area

At a live site such as this, the ratio of clutter to targets of interest is such that only a small number of targets of interest may be found in a 10-acre area; not nearly enough are expected to determine any demonstrator's classification performance with acceptable confidence bounds.  To avoid this problem, the site was seeded with enough targets of interest to ensure reasonable statistics  To the extent possible, items recovered from other live ranges were used as seeds; this was not possible for all items however.

The demonstration area was seeded with 200 inert munitions.  The locations and depths of these targets were unknown to the demonstrators.  The exact (*x, y*) location, depth to the center of the target, and orientation were recorded for each emplaced item.  Most objects were emplaced at a depth of 30 cm and none were deeper than 45 cm, based on the depth distribution of items recovered from the site characterization grids and observations regarding the soil conditions from the excavation crew.  The emplacement distribution of these inert rounds is summarized in Table 2-3.  Only *in situ* clutter was used in this study, and no additional cultural clutter, munitions-related scrap, or geology was seeded.

**Table 2-3. Blind seeded targets.**

| Munition type | Quantity | Depth Range | Orientations |
|---|---|---|---|
| 60-mm mortar | 76 | Majority at 30 cm, few at 45 cm | Mix of random horizontal orientations and vertical targets |
| 81-mm mortar | 56 | Majority at 30 cm, few at 45 cm | Random horizontal orientations, one vertical target |
| 4.2-in mortar | 54 | Majority at 30 cm, few at 45 cm | Random horizontal orientations, one vertical target |
| 2.36-in rocket | 14 | Majority at 30 cm, few at 45 cm | Random horizontal orientations, one vertical target |

# 3 PROGRAM DESIGN

## 3.1 OVERALL APPROACH

The objective of the study was to evaluate classification, as opposed to detection. Multiple classification approaches were applied to data collected using seven different sensor platforms. For comparisons of different classification approaches to be straightforward, a common set of detections for each data set was required. The detection stage was conducted by a program office team that was separate from the classification demonstrators. The approach to detection is described below. For each data set, a common list was passed to all of the classification demonstrators to attempt classification.

All the targets on the detection lists were dug and assigned ground-truth labels designating whether or not each was a target of interest (TOI). These labeled data, including the seeded targets, were segregated into training and testing data. All the truth information for the training data was provided to the processors and used to train their algorithms. The truth labels for the remaining data were sequestered, and these were used for blind testing. The processors were required to provide their assessment of the TOI/not-TOI labels for each item in the test data part of the detection list. The labels were compared to truth by an independent third party to score performance.

## 3.2 TARGETS OF INTEREST

The main goal of classification in the pilot program is to identify with high confidence items that can be safely left behind. We will refer to items that must be removed as Targets of Interest (TOI) and those that may remain as non-TOI. At SLO, the project team determined that targets of interest would include:

- seeded munitions,
- intact munitions recovered at the site, both live and inert, and
- sizeable pieces of munitions, which would be sufficiently munitions-like to alarm the public

The intact munitions are straightforward. Two hundred items were seeded and all are TOI. Four munitions types were recovered as expected from the geophysics detections: 60-mm mortars, 81-mm mortars, 4.2-in mortars and 2.36-in rockets. In addition, three other munitions types were discovered in the digging, including a 5-in rocket warhead, a 3-in stokes mortar and a 37-mm projectile. These items were also added to the TOI list.

The demonstrators were not informed about the specifics of the other munitions items found. Since the purpose of the classification is ultimately to leave items unexcavated, information about the presence of these other items would be available in the routine application of classification only if the items were correctly classified as TOI. However, it was made clear throughout the project that the possibility existed that items in addition to those known to be present could be encountered and that all munitions recovered would be considered TOI.

Deciding which partial rounds to include as TOI was less straightforward. Figure 3-1 shows four recovered 2.36-in rockets and rocket parts. There is essentially a continuum between complete rounds and fragments that are clearly not hazardous. Most of the recovered items were easily classified as either TOI or non-TOI. However, there were a few dozen items like the images in Figure 3-1 that could arguably have gone either way. In these cases, the decision was made to be conservative in defining TOIs most inclusively, in terms of both their munitions likeness and their axial symmetry. The 2.36-in rockets that had a symmetric body and substantial recognizable fins were included as TOI. In addition, 60-mm mortar bodies like those shown in Figure 3-2 that were broken open and empty were also TOI.


Figure 3-1. Recovered 2.36-in rocket and rocket parts.


Figure 3-2. 60-mm mortar parts labeled TOI.

## 3.3 DATA COLLECTION

The classification pilot study consisted of several combinations of data-collection platforms and analysis approaches, ranging from careful application of commercial survey instruments to three prototype systems specially designed to maximize detection and classification of munitions. Data-collection plans were generated by all data collectors and shared with the data processors prior to deployment. The data collection assets are listed in Table 3-1 and briefly described below. Details may be found in the reports provided by the performers. (Refs. 4-9)

- SURVEY MODE: The commercial survey systems were deployed to collect data on 100% of the site, called SURVEY mode.
- CUED MODE: Two sensors, TEMTADS and BUD, were deployed to collect data at the locations of individual anomalies detected by the EM61 ARRAY.

- SELF-CUED MODE: The MetalMapper (MM) is intended to operate in both survey and cued mode. MM performed a detection survey and collected cued data over all the anomalies it detected.

**Table 3-1. Summary of Data Collection at SLO.**

| SURVEY | CUED from EM61 ARRAY Data | Self-Cued |
|---|---|---|
| EM61 CART | TEMTADS | MetalMapper |
| EM61 ARRAY | BUD | |
| MSEMS | | |
| Magnetometer Array | | |

### 3.3.1 Survey Mode

In survey mode, the sensors covered 100% of the site. Data were acquired by running a sensor in closely spaced lines, similar to the pattern of a lawnmower cutting grass. For the cart-based systems, the site was divided into 30-m x 30-m grids and data collected one grid at a time, Figure 3-3. The vehicular systems did not make use of the grids, but surveyed the site in long lines that extended from one site boundary to the other, as they normally would do. This resulted in several additional grids being surveyed by the vehicle systems. These data have been archived, but were not used in the demonstration.

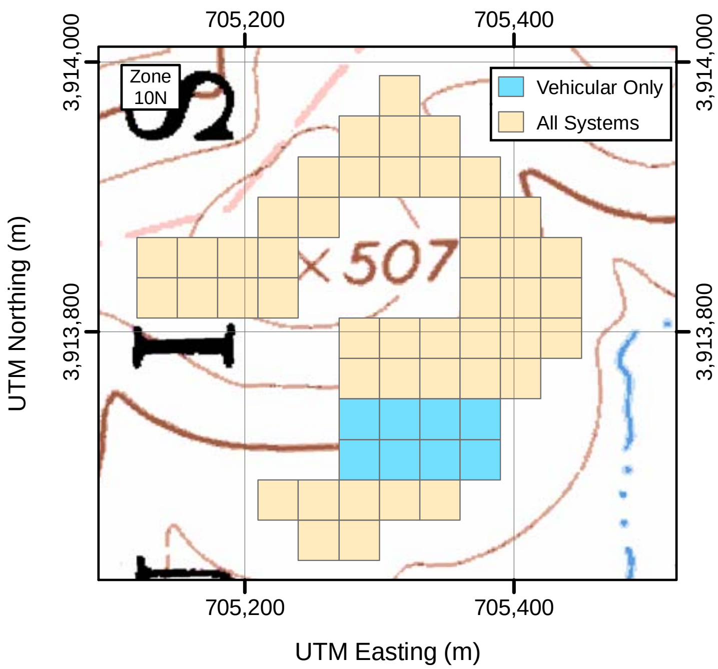


**Figure 3-3. Layout of the demonstration site showing the grids surveyed by the cart systems (10 acres) and the additional 8 grids (1.6 acres) surveyed by the vehicular systems.**

The survey mode data are intended to be representative of what can be achieved with careful data collection using standard equipment and field techniques. As such, care was taken when designing the data-collection protocols to ensure that data of a sufficient quality to support advanced analyses would result. For the most part, this involved controlling data density and system noise. However, no extraordinary measures, such as adding Inertial Navigation devices to cart platforms that do not otherwise employ them, were taken.

**EM61-MK2 CART:** Data were collected with a standard cart platform EM61-MK2 system. Typical industry-standard centimeter-level accuracy Global Positioning System (GPS) equipment was used for geolocation. The survey lane spacing was specified as 0.5 m and was marked on the ground using measuring tapes and rope. The sensor height above ground was the standard 40 cm. Figure 3-4 shows this system, which will be referred to throughout the report as EM61-MK2 CART, collecting data at SLO. (Ref. 4) Data were collected by NAEVA Geophysics.

Figure 3-4. EM61-MK2 CART collecting data at SLO.

**Multi-sensor Towed Array Detection System (MTADS) EM61-MK2 ARRAY.** The MTADS EM system uses an array of three overlapping 1 $m^2$ EM61-MK2 sensors that have been modified to increase the transmit current and adjust the receiver time gates from the standard sensor. The three sensors are pulsed and sampled simultaneously. The sensors are positioned to provide 0.5 m spacing between centers within the array and data are collected with approximately 0.25 m overlap in adjacent passes of the array. The array rides 34 cm above the ground surface. Two orthogonal survey passes are used to completely illuminate all targets. Figure 3-5 shows this system, which is referred to throughout the report at EM61-MK2 ARRAY. Data were taken in the MTADS standard configuration using an array of three centimeter-level accuracy GPS and an Inertial Measurement Unit (IMU) for geolocation and platform orientation. (Ref. 5) Data were collected by Nova Research.

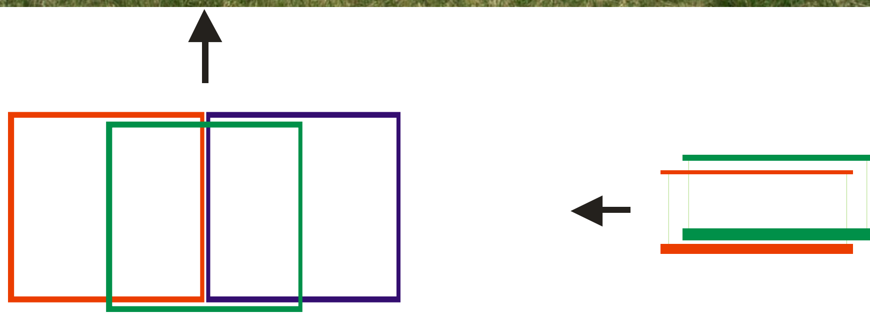

Figure 3-5. Photograph and schematic of the MTADS EM61 ARRAY.

**MTADS Magnetometer array**: The MTADS magnetometer (MAG ARRAY) platform houses a 1.75-m wide array of eight G-822A Cs-vapor magnetometers, spaced 0.25 m apart and 25 cm off the ground. Data were taken in the MTADS standard configuration using two centimeter-level accuracy GPS antennae for geolocation and platform orientation. Data were collected by Nova Research. (Ref. 5)

**MSEMS:** The Man-Portable Simultaneous Electromagnetic-Magnetic Sensor, shown in Figure 3-6, collects both magnetometer and EM61 data in a single pass. It incorporates a G-822 magnetometer with an EM61-MK2 on its factory cart platform. The data collection for the two sensors is interleaved to avoid interference. The line spacing of 0.5 m was marked by tapes and ropes and the platform height was 40 cm above ground level. Geolocation was provided by cm-level GPS. Data were collected by NAEVA Geophysics and SAIC. (Ref. 6)

Figure 3-6. MSEMS collecting data at SLO.

### 3.3.2 Self Cued

**MetalMapper.** The MetalMapper (MM), shown in Figure 3-7 and Figure 3-8, is designed to be a stand-alone survey and cued detection system, and was operated independently. The system is composed of three orthogonal 1-m x 1-m transmitters for target illumination and 7 three-axis receivers for recording the response. For this demonstration, it measured the decay curve up to 8 ms after the transmitters were turned off. It was used in a sled or a wheeled configuration mounted to a front loader tractor. Centimeter-level GPS is used for navigation and geolocation and an IMU is used to measure platform orientation.

In survey mode, MetalMapper covered the entire site with 0.75-m line spacing, with down track point spacing of approximately 5 cm and the base of the sensor 21 cm above the ground. For the survey mode, only the vertical field transmitter is used and the receive data recording is truncated at 0.9 ms after the turn off of the transmitter. The MetalMapper survey resulted in about 1700 anomalies.

In cued mode, MetalMapper is positioned over each anomaly on its target list and collects the full suite of data while stationary. Data were collected by Geometrics. (Ref. 9)

Figure 3-7. MetalMapper collecting data at SLO.

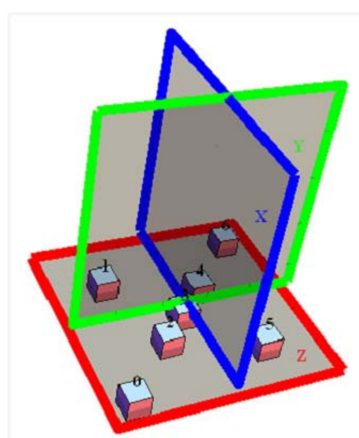


Figure 3-8. Schematic of MetalMapper.

### 3.3.3 Cued Data

Two sensors were used to collect cued data at the locations of anomalies detected by the EM61 ARRAY. These developmental EM systems were designed to collect sufficient data to fully characterize the EM signature from a single measurement location. Approximately 1500 anomalies in the EM61 ARRAY survey data met the target selection criteria. The TEMTADS collected data at the locations of all 1500 anomalies. The BUD is a prototype that is less rugged and was not designed for extensive field work. BUD collected data on approximately 500 targets. The subset of targets on which BUD was deployed consisted of all anomalies in 15 of 45 total grids on the site.

**TEMTADS.** The TEMTADS, shown in Figure 3-9, is positioned over each anomaly on its target list and collects data in a stationary mode. The system is a 5 x 5 array of elements oriented parallel to the ground. Each array element is 0.35 m on a side and contains both transmit and receive coils. The 25 transmit elements are pulsed in sequence and data are collected from all receivers for each transmit pulse. The receive coils collect data until 25 ms after the transmit current has been turned off. The total array dimension is 2-m x 2-m, it collects data at a height of 30 cm above the ground surface, and it is towed by the same vehicle used for all the MTADS systems. Three cm-level GPS units are used for navigation, geolocation and orientation. Data were collected by Nova Research. (Ref. 7)

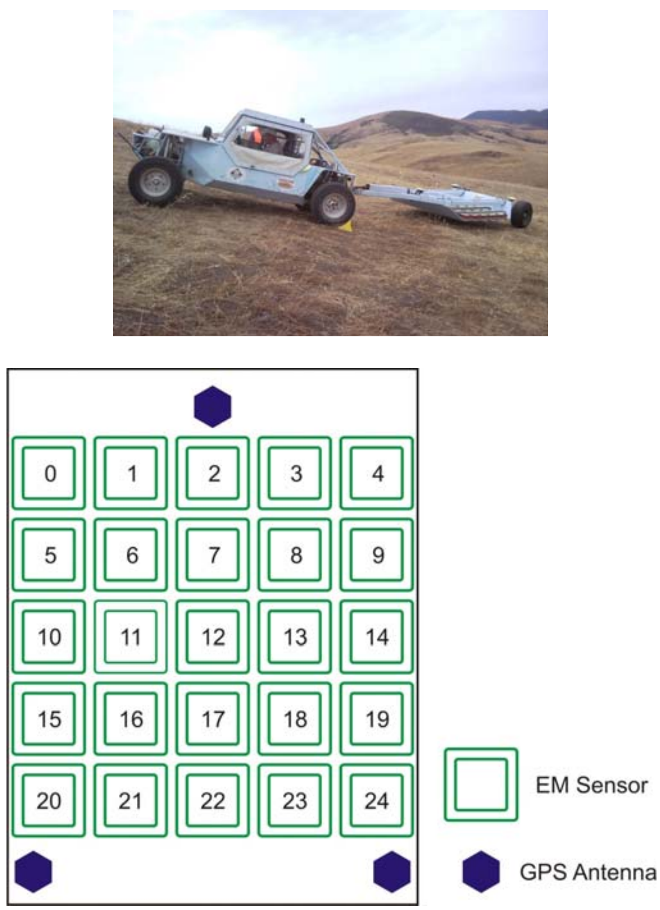


**Figure 3-9. Photo (upper) and schematic (lower) of TEMTADS. Total dimension of the array is 2m square.**

**BUD:** The Berkeley UXO Discriminator is composed of three orthogonal transmitters for target illumination, and eight pairs of differenced receivers for recording the response. It measures the entire decay curve up to 1.2 ms after the transmitters are turned off. It collected data on all anomalies in 15 grids of the site, for a total of approximately 500 cued targets. A photo and schematic of BUD are shown in Figure 3-10. (Ref. 8) Although BUD is intended to collect all the data needed to estimate target parameters from a single location, at SLO data were acquired at multiple points for each target as illustrated in Figure 3-11. Data were collected by Lawrence Berkeley National Laboratory.

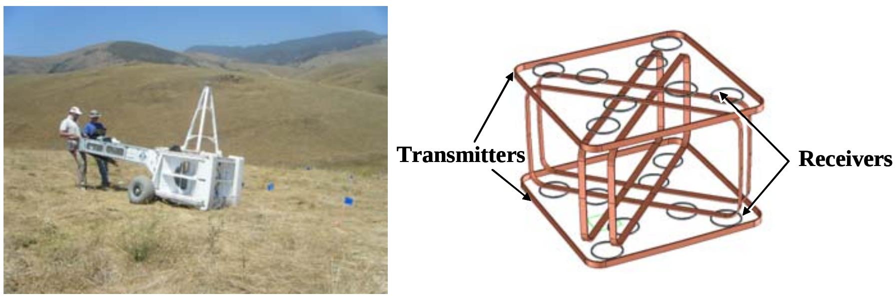


Figure 3-10. Photo and schematic of BUD.

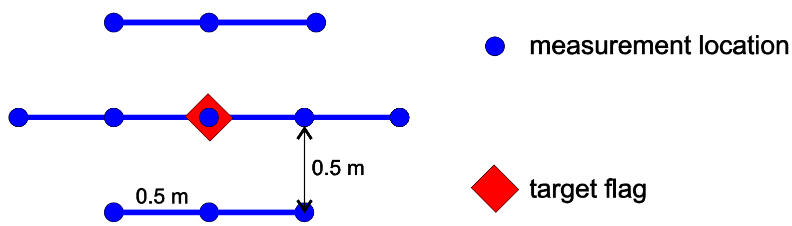


Figure 3-11. BUD data collection protocol at SLO.

## 3.4 CLASSIFICATION APPROACHES

### 3.4.1 Processing Flow

The basic flow of the classification approaches was the same for all demonstrators and is summarized in the flow chart in Figure 3-12. The classification demonstrators began with target lists provided by the program office. These lists contained all the anomalies detected by each sensor, as described in Section 4. Each anomaly was analyzed by the processing teams to extract parameters by fitting the data to a model or by selecting features of the data upon which to perform classification.

The anomaly lists were divided into training and blind testing sets. Some of the truth data were provided to each demonstrator for algorithm training, so that the parameters or features that were most useful for classification could be determined and thresholds in the classification process set. It is expected that some parameters and features will be more useful than others. After training, the decision process for each algorithm was finalized and documented, and the demonstrators provided ranked dig lists for the blind test set.

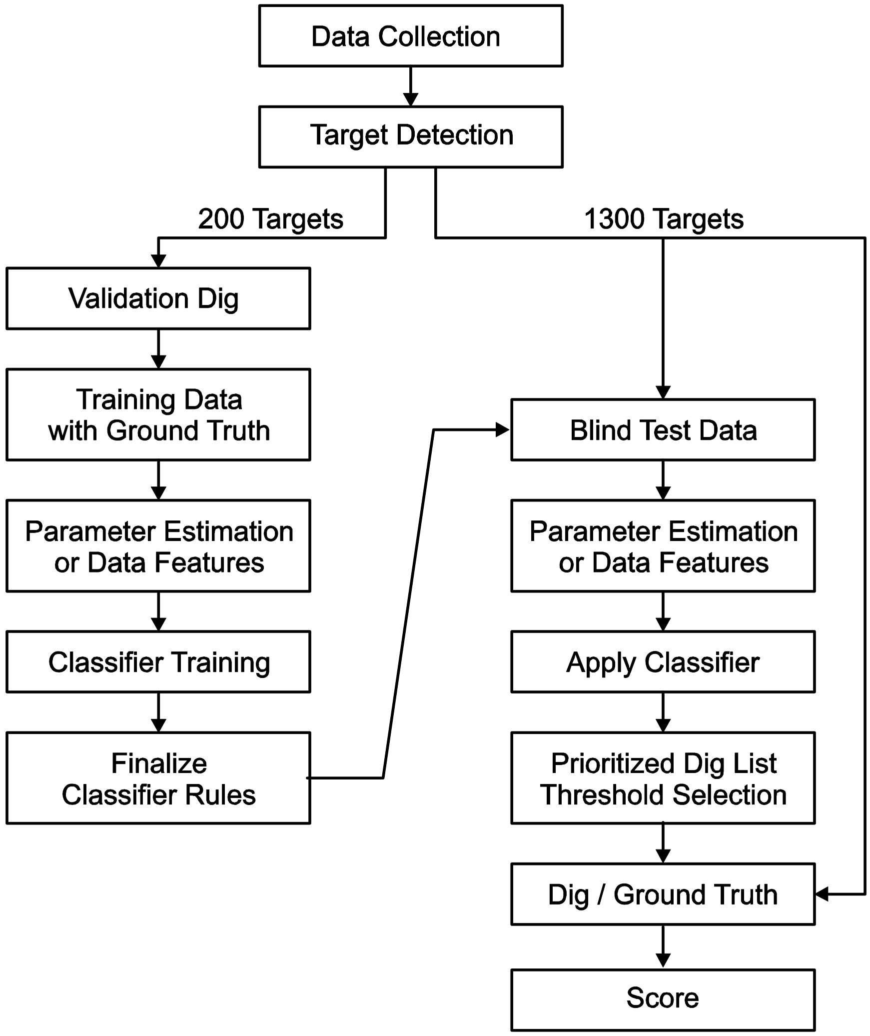


Figure 3-12. Work flow of classification demonstrations.

### 3.4.2 Parameters based on Geophysical Models

Multiple groups demonstrated processing approaches. The basic classification method for all but one demonstrator involved using a geophysical model to estimate target parameters that may be useful in making a classification decision. Although the processing approaches differ in their manner of implementation, all the geophysical models are based on a dipole approximation.

For the mapping sensors, this process involves using data from multiple spatially diverse locations that together fully characterize the signature. An example of a small section of field data encompassing an anomaly, called a data chip, is shown on the left panel of Figure 3-13. During the processing, the field data are used to extract the values of the model parameters. The right panel shows the modeled chip, which depicts the anomaly as it is predicted using the best fitted parameter values. When meaningful parameter values are arrived at, the two should look substantially similar. Quantitative measures of their similarity are used to determine whether the fit is reliable. This procedure was implemented in the UX-Analyze package by SAIC and UXO-Lab by Sky Research.

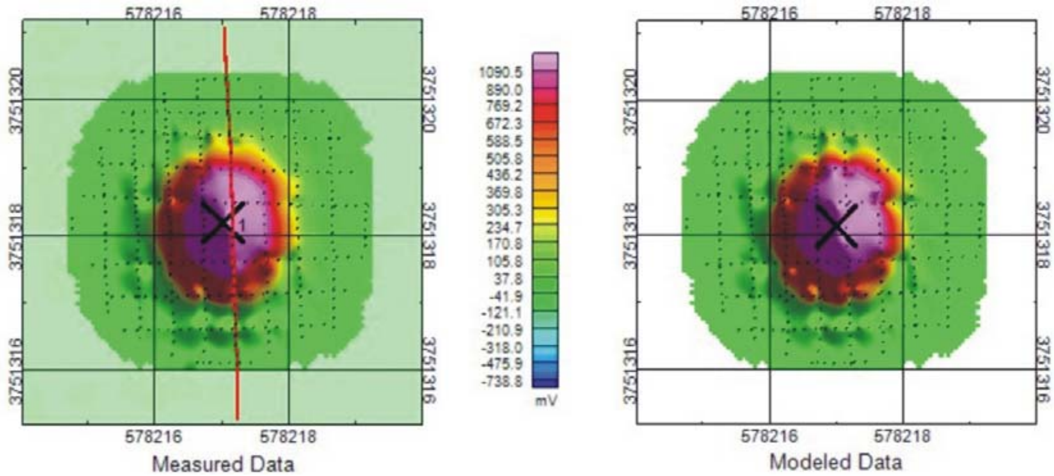


**Figure 3-13. Example of a measured EM61-MK2 data chip of an anomaly (left) and the corresponding model result (right). The axes labels refer to distance in meters.**

The cued sensors, BUD, TEMTADS and MetalMapper, collect sufficient data at a single spatial location to support model-based parameter estimation using custom software developed by the system developers.

Some of the parameters that were considered included:

- the electromagnetic polarizability, which relates to the object's physical size and aspect ratio,
- the electromagnetic decay constant, which relates to the object's material properties and wall thickness, and
- the magnetic dipole moment, which is related to the physical size of the object.

Here, the estimated size of the object should not be confused with the spatial size or footprint of the anomaly. While it is true that large, deep objects will give rise to anomalies with a greater spatial dimension than small, shallow objects that may have comparable amplitudes, anomaly size is not a rigorous, direct substitute for object size.

Inadequacies in the model, noise in the data, or difficulty in the mathematical process used to fit multiple parameters to the measured data will result in variation in these parameter estimates. Sometimes noisy data or a model insufficiency will yield a result that is nonsensical or will cause the estimation process to fail to converge on an answer at all. Although the demonstrators were requested to provide estimated parameters for each target analyzed, in some cases where meaningful fits could be not be obtained, items were identified as "Can't Analyze." Since no classification decision can be made, all items in this category must be treated as potential munitions.

### 3.4.3 Parameters Based on Data Features

One of the processing teams, SAIC/RML, had a fundamentally different approach and based its classification on features of the data, rather than model fit parameters. Examples include peak amplitude, anomaly statistical moments, anomaly footprint size and shape, and decay rate from a single sounding. These features are automatically extracted directly from the data for each anomaly and the classifier is used to determine which combinations of data features are good indicators of munitions based on the training data.

### 3.4.4 Classifiers

Once the parameters are estimated, a mechanism is needed to decide whether the corresponding object is a target of interest or not. Several types of classification processing schemes were evaluated in the classification study. These included both

> **Statistical classification:** Computer algorithms evaluate the contributions of each parameter to defining munitions likeness based on "training" on a subset of the data for which the identities of the objects are known. Then the unknown objects are prioritized based on whether their parameters are statistically similar to known objects in the training data.
>
> **Rule-based classification:** A data analyst inspects the training data and the associated parameters to make a "rule" about how unknown objects will be sorted. For example, a rule may be defined so that all objects are sorted based on their "size" and decay constant, which relate to intrinsic physical target parameters, such as wall thickness and material.

The final step in classification is delineating the targets of interest from those that are not. For example, in the case of a statistical classifier, all the anomalies are ordered by the likelihood that they do not belong to the class of the targets of interest. These likelihood values do not represent a yes/no answer, but rather a continuum within which a dividing line or threshold must be specified. Depending on the application, this threshold may be set to try to avoid false positives, which may come at the expense of missing some items of interest, or it may be set to try to avoid false negatives, which will come at the expense of a greater number of non-TOI. In this program, where missing an item of interest represented the most serious failure, demonstrators selected thresholds to try to retain all the detected munitions.

### 3.4.5 Summary of SLO Classification Approaches

The focus of this report is on (1) commercial instruments and available processing that were successfully demonstrated and could be implemented today and (2) the potential of specialized emerging sensors. However, the material covered in this report represents only a small part of a much larger study, and notably absent here is any discussion of advanced and innovative processing techniques. Details of each approach and all of the results can be found in the individual demonstrator reports. (Refs. 8-14)

## 3.5 CLASSIFICATION PRODUCT

Demonstrators were asked to produce a ranked dig list for each sensor and processing combination. These lists were constructed as shown in Table 3-2.

Table 3-2. Model of Ranked Dig List

| Rank | Anomaly ID | $P_{clutter}$ | Comment |
|---|---|---|---|
| 1 | 247 | .97 | |
| 2 | 1114 | .96 | High confidence NOT TOI |
| 3 | 69 | … | |
| … | … | … | |
| … | … | … | |
| … | … | … | Can't make a decision |
| … | … | … | |
| … | … | … | |
| … | … | … | |
| … | … | … | High confidence TOI |
| … | … | .03 | |
| … | … | .02 | |
| | … | | |
| | … | | Can't analyze |
| | … | | |

**Threshold**

- GREEN: The top item in the list was that which the demonstrator was most certain does NOT correspond to a TOI.
- YELLOW: A band was specified indicating the targets where the data can be fit in a meaningful way, but the derived parameters do not permit a high confidence determination of TOI or not-TOI.
- RED: The bottom items were those that the demonstrator was most certain are TOI.
- GRAY: Targets where the signal-to-noise ratio (SNR), data quality, or other factors prevent any meaningful analysis were deemed "can't analyze" and appended to the bottom of the list.
- THRESHOLD: A threshold was set at the point beyond which the demonstrator would recommend all anomalies be treated as TOI, either because they are determined to be so with high confidence or because a high-confidence determination that they are not TOI cannot be made. This is indicated by the heavy black dashed line.

## 3.6 SCORING METHODS

The demonstration was scored based on the demonstrator's ability to eliminate nonhazardous items while retaining all detected TOI. A common way to evaluate performance of detection and classification is the receiver operating characteristic (ROC) curve. An example is shown in Figure 3-14. The colored regions on the plot in Figure 3-14 correspond to the colors of the various sections of the ranked dig list in Table 3-2. The ROC curve is a plot of the percent of the TOI dug, that is it reflects the probability of detecting and correctly classifying the munitions items, versus the number of non-TOI. A perfect detector and classifier would detect 100% of the munitions and no clutter. We have modified the traditional ROC curve slightly to reflect the total of non-TOI that must be dug on the x-axis, including those for training. This is done to account for the fact that different methods used different amounts of training data.

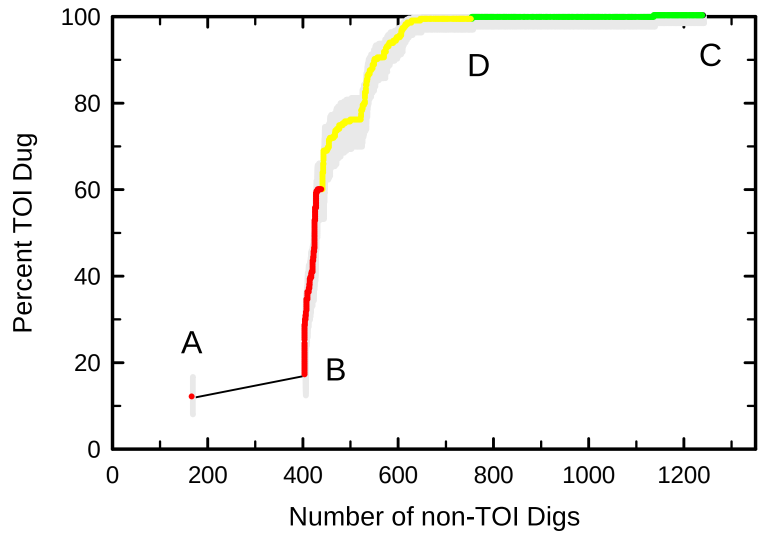


**Figure 3-14. Example receiver operating characteristic curve.**

The key regions to interpret the ROC curves used in this program are:

- **A**: Targets to the left of this point were dug for training data. Training data were used in all of the processing and these digs would be required. Different approaches required differing amounts of training data.
- **B**: Targets from point A to this point were categorized as can't analyze and would need to be treated as potential TOI because no meaningful classification could be done. In this example, about 200 of the can't analyze targets were false positives, reflected in the position of the point on the horizontal axis. Three TOI were also included in the can't analyze list, reflected in the point's position on the vertical axis.
- **C**: In the absence of any classification, this sensor detected all the TOI and had more than 1200 non-TOI items in the detection list.
- **D**: Based on classification, this is the demonstrator's threshold for the dividing point between TOI and not-TOI. This demonstrator correctly identified all of the TOI and about 800 non-TOI items remained on the dig list after classification, that is 400 non-TOI did not have to be dug.

# 4 ANOMALY SELECTION AND INVESTIGATION RESULTS

## 4.1 ANOMALY SELECTION

After each of the survey systems completed data acquisition, anomalies were selected from the data using a procedure designed by the program office. A detection list was generated by recording all locations for which the sensor signal exceeded a system-specific threshold. Since these individual sensor detection lists were the basis for all subsequent analyses, a rigorous process was used to set this threshold.

The targets of interest in this demonstration were 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets. Based on the results of the exploratory intrusive investigation conducted at the beginning of the demonstration, no targets were expected deeper than 30 cm below the surface. Since the signal from each of these targets can be predicted accurately (Ref. 18) as a function of depth, the threshold for each sensor can be chosen to ensure detection of all items of interest while leaving smaller targets off the list. For this demonstration, the anomaly selection threshold for each sensor was set as the smallest signal expected from any of the items at 45 cm below the surface. This threshold corresponds to the 30 cm depth of interest established from the digs in the characterization grids with a 50% safety margin.

An example of this process is shown in Figure 4-1 for the EM61-MK2 ARRAY. For this system, the item of interest with the lowest predicted signal at a particular depth is the 60-mm mortar. The predicted signal from the EM61-MK2 ARRAY for this target in its least favorable orientation is plotted in the figure along with a vertical line marking the 45 cm depth of interest. The anomaly selection threshold for this sensor system was set at 29.0 mV based on this curve. Also plotted on Figure 4-1 is the observed noise at the site. As can be seen from the figure, the anomaly selection threshold is well above the measured noise so the anomaly selection process should be relatively unambiguous for this sensor system. The response curve plotted was confirmed using measurements in the pit.

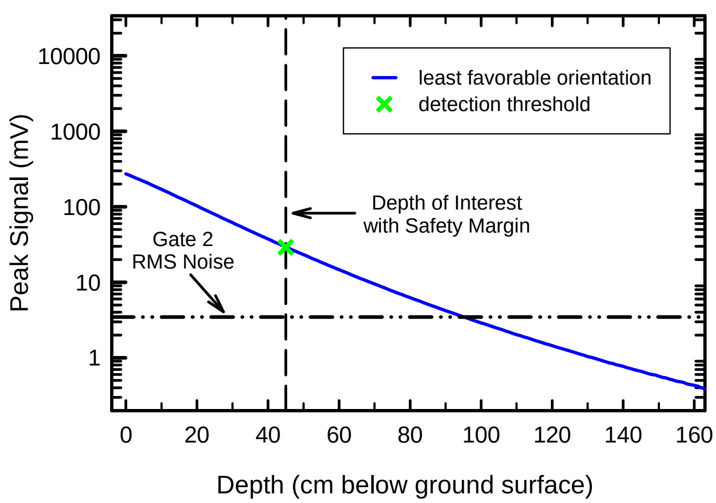


**Figure 4-1. Predicted EM61 ARRAY anomaly amplitude in gate 2 for a 60-mm mortar in its least favorable orientation. Also shown are the RMS noise measured at the site, the 45 cm depth used to set the threshold and the anomaly selection threshold used in this demonstration.**

These raw threshold exceedances were further processed to yield the sensor-specific detection list. For the EM61-MK2 ARRAY data, each threshold exceedance was used as the starting location of a geophysical inversion routine. All inversion results that converged to a common target position were grouped into one detection. Then, all exceedances within the distance of 0.6 m were grouped into a single detection. Finally, all pairs of exceedances between 0.6 m and 1.0 m apart were examined by a trained analyst who made a judgment whether they corresponded to a single source or not.

A similar process was used to set the threshold for the other survey sensors, although inversions were performed only on the EM61-MK2 ARRAY data. The number of detections for each survey instrument is listed in Table 4-1. Although the initial magnetometer transects of the site did not reveal this, the magnetometer data presented two problems that prevented complete analysis in this program. First, there are a number of large geologic anomalies that obscure significant portions of the site. Second, there are a large number of small discrete anomalies near the threshold. The result is that there were three times as many anomalies above threshold in the magnetometer data than the EM61-MK2 ARRAY with many of these corresponding to large, diffuse geological features. Since the program budget could not support investigation of this large number of anomalies, the magnetometer detection list was discarded and only those locations corresponding to detections by the EM61-MK2 ARRAY were used for magnetometer analysis.

**Table 4-1. Sensor-specific anomaly detection and scoring details.**

| Sensor | Anomalies Detected | Anomalies Scored | Comment |
|---|---|---|---|
| EM61-MK2 ARRAY | 1464 | 1464 | • all grouped threshold exceedances<br>• used to cue TEMTADS and BUD |
| EM61-MK2 CART | 1506 | 1506 | • 1151 common targets with EM61-MK2 ARRAY<br>• 355 additional exceedances |
| MSEMS EM61 | 1375 | 1375 | • 1139 common targets with EM61-MK2 ARRAY<br>• 236 additional exceedances |
| MAG ARRAY | 5511 | 1464 | • only EM61-MK2 ARRAY target locations scored |
| MetalMapper | 1745 | 1745 | • self cued<br>• 1358 common with EM61-MK2 ARRAY, EM61-MK2 CART, or MSEMS EM61 used for initial scoring<br>• All dug and scored in a second round of analysis |
| TEMTADS | n/a | 1464 | • cued by EM61-MK2 ARRAY |
| BUD | n/a | 500 | • cued by EM61-MK2 ARRAY |

The target-based selection threshold employed in this demonstration is an important component of the classification process. The number of threshold exceedances in the EM61-MK2 ARRAY data as a function of threshold chosen is shown in Figure 4-2. As the selection threshold approaches the measured site noise, the number of exceedances increases dramatically. These extra targets are necessarily low signal-to-noise targets, which are often difficult to extract reliable parameters for and end up in the "unable to analyze, must dig" category. The accuracy of the threshold chosen for this demonstration was tested by digging targets below the threshold in a few grids; these results will be discussed in the section on intrusive results.

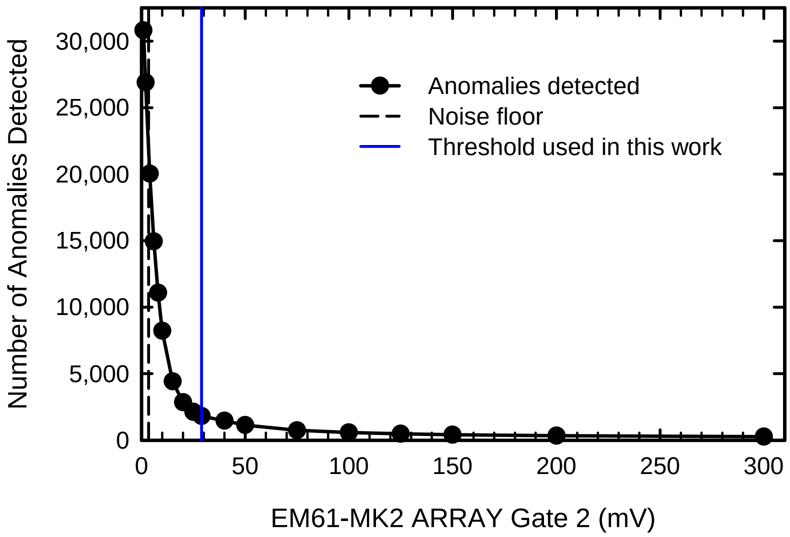


**Figure 4-2. Number of EM61-MK2 ARRAY threshold exceedances as a function of the anomaly selection threshold applied. Also plotted are the system noise floor and the threshold used for this demonstration.**

### 4.1.1 Detection of Seed Items

Using the anomaly selection thresholds described above, the EM61-MK2 ARRAY and MAG ARRAY detected all seeds in the area they were able to survey. One seed, Seed ID 170, was emplaced between two large rocks in an area too small for the arrays to survey. This seed was detected by the EM61 cart-based systems however. The EM61 CART survey did not select one of the emplaced seeds even though it was covered during the survey. The item missed was a 60-mm mortar body without fins or nose. The threshold of 11.3 mV was developed from the response of a complete 60-mm mortar. The partial round missed produced a response of 10.7 mV, just under the threshold. This reinforces the point that target-based thresholds need to be chosen with care and a complete understanding of the items expected to be encountered at the site.

## 4.2 DIG LIST

A list of locations for intrusive investigation, the dig list, was prepared by combining the anomaly lists from each sensor in a stepwise fashion (Ref. 15), Table 4-2. The starting point of the list was the 1464 detections by the EM61-MK2 ARRAY which were used to cue the TEMTADS and BUD systems. To these were added unique detections by the EM61 CART and the MSEMS EM61 sensor. This resulted in an initial dig list with 1819 entries with which the intrusive crew began work.

The MetalMapper system performs self-cueing so the detection list for this system was compiled separately. Of the 1819 locations on the initial dig list, 1358 corresponded to MetalMapper detections. There were an additional 387 unique MetalMapper detections that were added to the final dig list.

There were two additional sets of locations added to the final dig list. Forty five "large, deep" mag anomalies as judged by the MSEMS magnetometer data analysts were investigated to test for the presence of items too deep to be detected by the EM sensors. In eight out of the 45 grids on the site, targets from the EM61-MK2 ARRAY list below the threshold of 29 mV but larger than 14.5 mV were dug to check the appropriateness of the threshold chosen. This accounted for 337 additional locations on the final dig list for a total of 2588 locations for digging.

Table 4-2. Origin of the locations on the final dig list.

| **Used in Scoring** | |
|---|---|
| EM61-MK2 ARRAY | 1464 |
| Added EM61-MK2 CART and MSEMS EM61 | 355 |
| MetalMapper | 387 |
| **Other Digs** | |
| Below Threshold from 8 grids | 337 |
| "Large, Deep" Mag Targets | 45 |
| **Total Digs** | **2588** |

## 4.3 INTRUSIVE INVESTIGATION

The distribution of recovered items by class is shown in Figure 4-3. The left hand chart shows the results of digging items on all the EM sensors detection lists (2210 items, larger than the number of entries on the dig list since some locations yielded a number of items in close proximity). The vast majority of items recovered at this site were classified by the UXO crew as munitions debris. Note that there were 26 items recovered that were classified as UXO and an additional 25 items that were not UXO but, as discussed in section 2, were so close in character to potential UXO that they were declared as targets of interest.

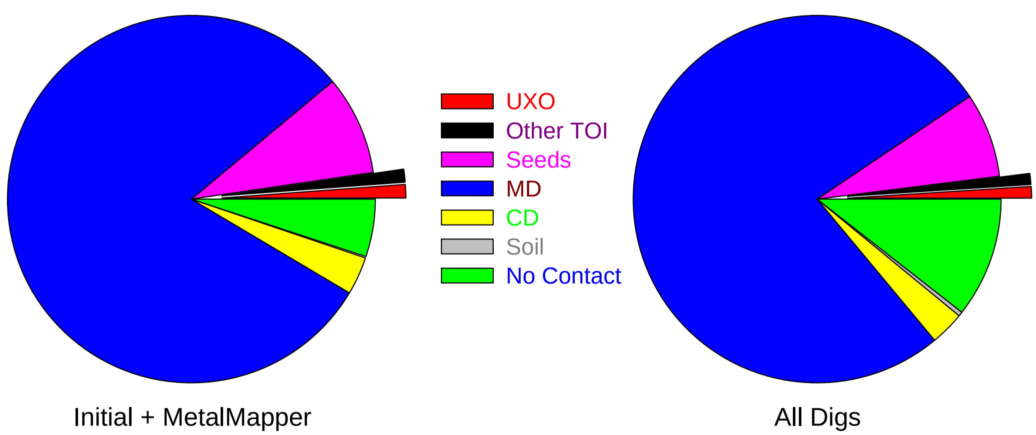


Figure 4-3. Distribution of recovered items by class. The chart on the left shows the items recovered from the EM survey lists. The chart on the right includes those targets plus the "large, deep mag" targets and the below threshold digs from a portion of the site.

The right hand chart in Figure 4-3 shows the distribution of all recovered items at the site. Adding the "large, deep mag" targets and the below threshold locations to the list did not result in the identification of any more UXO, seeds, or other targets of interest. Adding these targets did result in substantially more digs classified as soil or no contact. This confirms the importance of choosing the anomaly selection threshold carefully so as to include all targets of interest while excluding as much of the sensor noise as possible.

The measured depths of the recovered items are plotted in Figure 4-4. The bin corresponding to recovered depths of 0 to 5 cm is, by far, the largest with half the total recoveries in this bin. In fact, 95% of all recoveries corresponded to less than 30 cm to the center of the target.

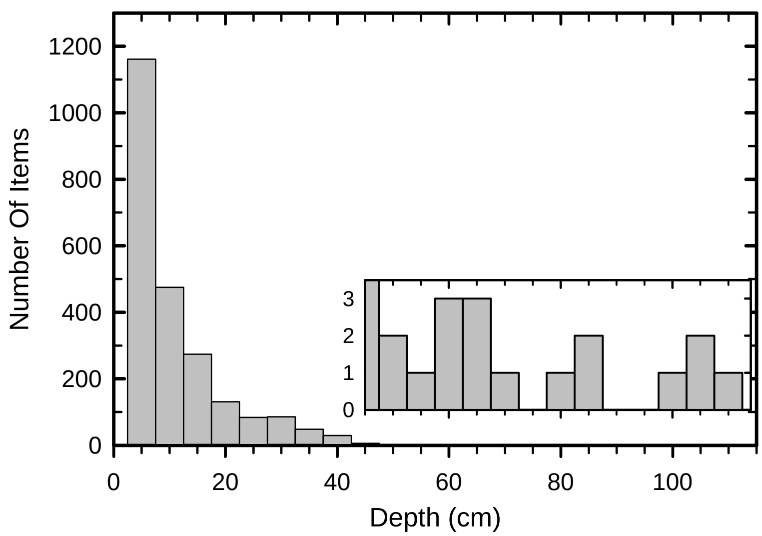


**Figure 4-4. Measured depth distribution of all items recovered at San Luis Obispo. The inset enlarges the scale to make the handful of targets recovered deeper than 50 cm visible. Depths tabulated are measured from the ground surface to the center of the recovered target.**

Except for three items, the UXO and other items identified as targets of interest were consistent with the historical records and the results of the exploratory excavations. The unexpected items were a 37-mm projectile, a 3-in stokes mortar, and a 5-in warhead for a 3.5-in Naval rocket. These three items were, of course, not considered when the anomaly selection threshold was chosen. Figure 4-5 plots the predicted response of the EM61-MK2 CART system to these items as a function of depth along with the anomaly selection threshold used for this sensor. Based on these data, 37-mm projectiles would have been detected with 100% certainty down to 0.2 m below the surface, 3-in stokes mortars to approximately 0.7 m, and the 5-in warhead to 1.10 m. The discovery of unexpected items on a production application would present the site team with a decision. Depending on what is found, it may be necessary to revisit the anomaly selection criteria and perform another analysis of the survey data for detection.

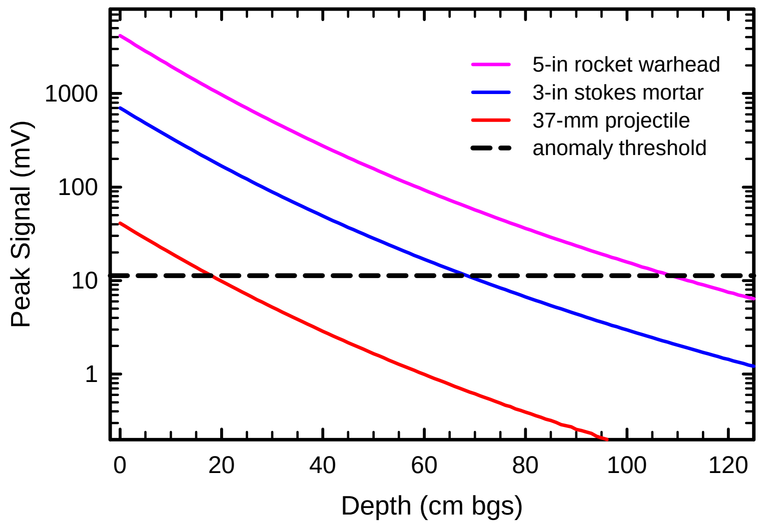


**Figure 4-5. Predicted EM61-MK2 CART minimum response in gate 2 of the unexpected munitions found at SLO as a function of depth. The horizontal dashed line indicates the anomaly selection threshold used for this sensor.**

The measured depths of the 26 UXO recovered are shown in Figure 4-6. All but two are within the 45 cm depth of interest for this demonstration, an 81-mm mortar recovered at 0.52 m and the 5-in warhead recovered at 1.06 m.

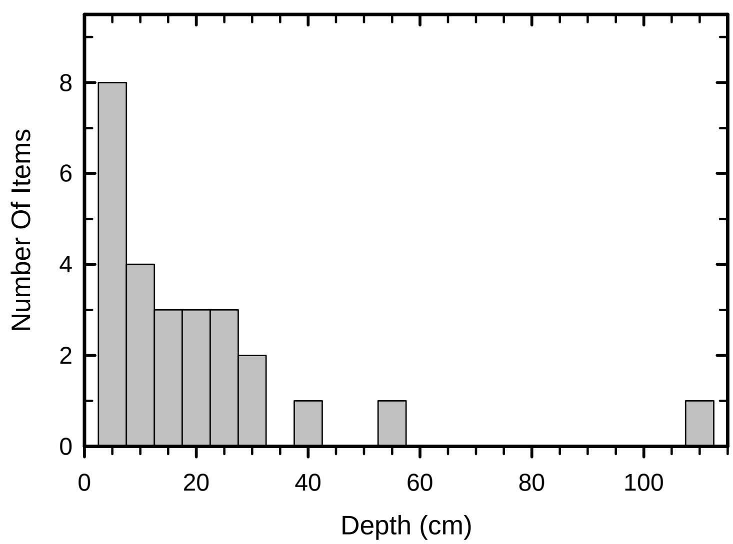


**Figure 4-6. Measured depth distribution of the 26 UXO recovered in this demonstration. Depths are measured from the ground surface to the center of the item.**

# 5 CLASSIFICATION RESULTS

A total of 54 dig lists were scored, representing the various combinations of sensor data collection systems and processing approaches used at SLO. All of the results may be found in the report by IDA (Ref. 15). In this section, we present selected results that illustrate important conclusions of the demonstration, focusing on what can be achieved with currently available technologies and the value added of emerging advanced sensors and processing. Results are organized by data collection system.

The results in this section are presented as ROC curves, which plot the percent of correctly classified munitions versus the number of false positives (i.e., unnecessary digs). Their interpretation is described in Section 3.6. The colored segments of the ROC curve correspond to the categories specified on the dig list and two threshold values are shown on the ROCs. The dark blue dot (•) indicates the demonstrator's threshold beyond which all targets are considered high confidence non-TOI. The orange dot (•) indicates the best that the demonstrator could have done had the threshold been set in the optimal place, the point at which the first TOI would be incorrectly classified as non-TOI, which would produce the first false negative. Missed TOIs on the ROC curve are indicated by open black triangles (△) and a picture of the item(s) missed is inset. In some cases missed TOIs were seeds or other items that clearly belonged in the TOI category and must be correctly classified. However, in other cases the missed TOIs corresponded to the ambiguous items discussed in Section 3.2. These items were substantial pieces of munitions items that were not hazardous, but that were sufficiently munitions-like that they were classified as TOI. In the application of classification, it will be a matter to be decided by a site team as to whether missing such items should really be considered a failure.

## 5.1 EM61-MK2 CART

The EM61-MK2 is the most common geophysical sensor in use for Munitions Response (MR) projects today. This sensor on a cart platform is our benchmark for what could be accomplished with carefully collected production geophysics data. These data were analyzed by two geophysicists at contractors that perform production work on MR projects, NAEVA Geophysics and Parsons. NAEVA used UX-Analyze to estimate the target's polarization parameters, made an estimate of the decay constant for each anomaly, and constructed a rules-based classifier. Parsons calculated two decay constants (Ch1-Ch2 and Ch3-Ch4) and prioritization was based on the weighted sum of the two decay constants. The ROC curves showing their results are shown in Figure 5-1 and 5-2.

Both NAEVA and Parsons demonstrated substantial classification ability. NAEVA labeled about 200 targets as "can't analyze." At their threshold, about 500 of the 1200 non-TOI were correctly identified, but two TOI, both seeded 4.2-in mortars, were also missed. Parsons labeled only a few targets as "can't analyze." At their threshold, only about 100 of the 1200 non-TOI were correctly identified, but no TOI were misclassified. However, the orange dot (•), indicates that their threshold could have been set much more aggressively, and about 550 non-TOI correctly identified before the first non-TOI was misclassified.

The EM61-MK2 CART data were also analyzed by two of the algorithm developers, SAIC and Sky Research. SAIC used UX-Analyze to extract the three principal axis polarizations, which were summed to estimate a size parameter. The ratio of the first to fourth time gate was used as the decay constant. A statistical classifier was used to rank targets based on size and decay constant.

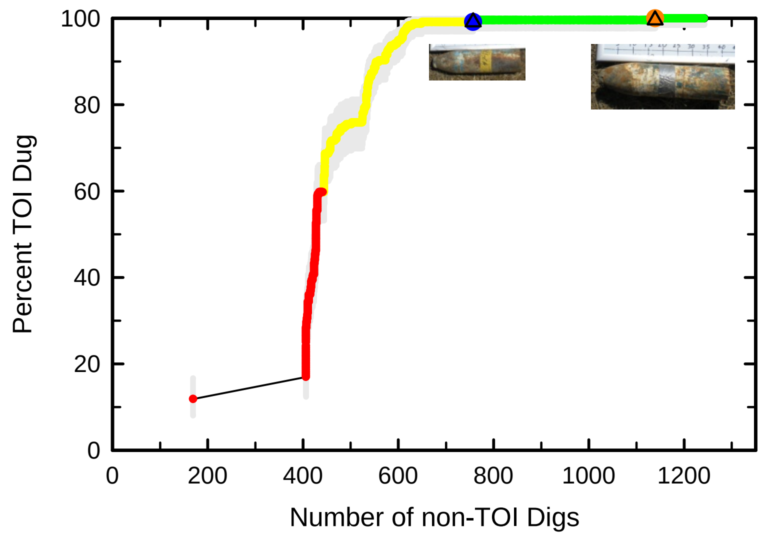


Figure 5-1. NAEVA Analysis of the EM61-MK2 CART Data.

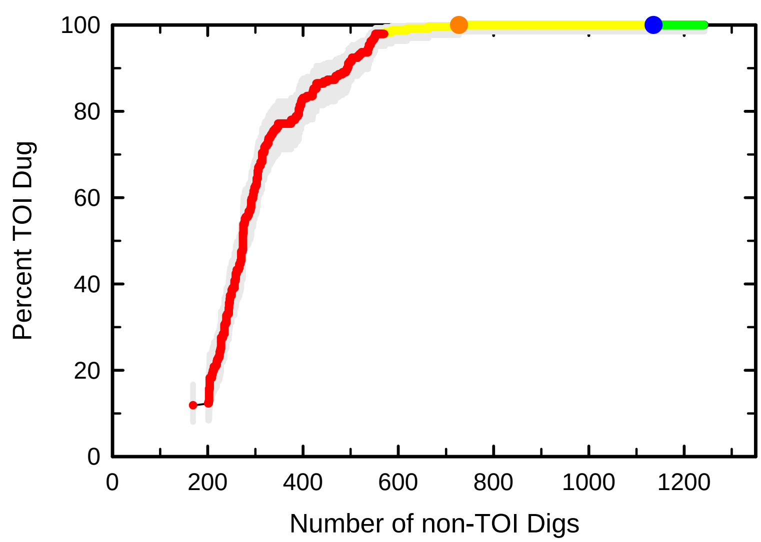


Figure 5-2. Parsons Analysis of the EM61-MK2 CART Data.

Sky used only the polarization decay parameter for the EM61 CART data. Since only a single parameter was used, the decision boundaries were set in a single dimension and no statistical classifier was needed. These results are shown in Figure 5-3 and Figure 5-4. The results for the two are fairly similar.

Both SAIC and Sky labeled few or no targets as "can't analyze." The close proximity of the blue and orange dots, indicating their threshold and the point of their first missed TOI indicates that both set their thresholds well, although SAIC misses one 60-mm mortar very near the threshold. These 60-mm teardrop shapes were among the most commonly missed items at SLO. At their threshold, SAIC is able to correctly classify about 600 of the non-TOI and Sky about 650. The main difference is that SAIC labels a number of targets as "can't decide" where Sky made definitive TOI or non-TOI calls on most items.

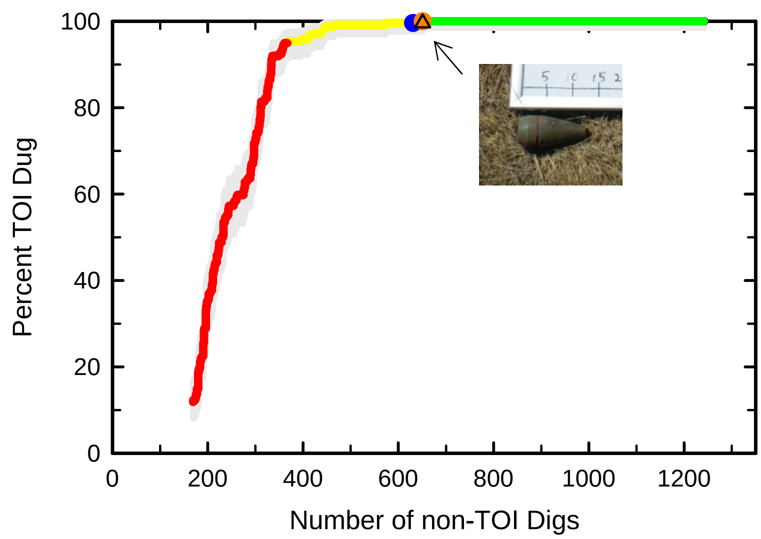


**Figure 5-3. SAIC Analysis of the EM61-MK2 CART Data.**

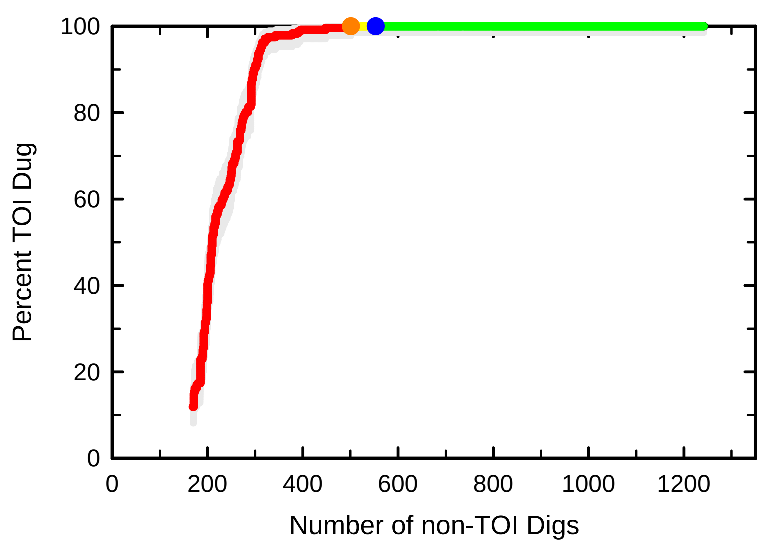


**Figure 5-4. Sky Research Analysis of the EM61-MK2 CART Data.**

It is important to note that the CART results shown here were obtained using a commercial sensor and analysis by NAEVA, Parsons and SAIC was performed using a freely available module of Oasis montaj. The cart data were collected using tighter lane spacing than would be standard for a detection-only survey (0.5-m spacing in this case), but otherwise normal commercial data collection and operator procedures were used. Analysis of the EM61-MK2 CART data by either the two contractors yield substantial classification ability, showing the ease of use and power of this module.

## 5.2 EM61-MK2 ARRAY

The EM61-MK2 ARRAY results were similar to the EM61 CART results. Figure 5-5 compares the CART and ARRAY results obtained by SAIC using the same parameter extraction and classification approaches in UX-Analyze. The two curves match quite closely. The main difference is in the placement of the threshold, which is set somewhat more conservatively for the ARRAY, resulting in no missed TOI, but fewer correctly identified non-TOI, about 350 for the cart compared to 600 for the array.

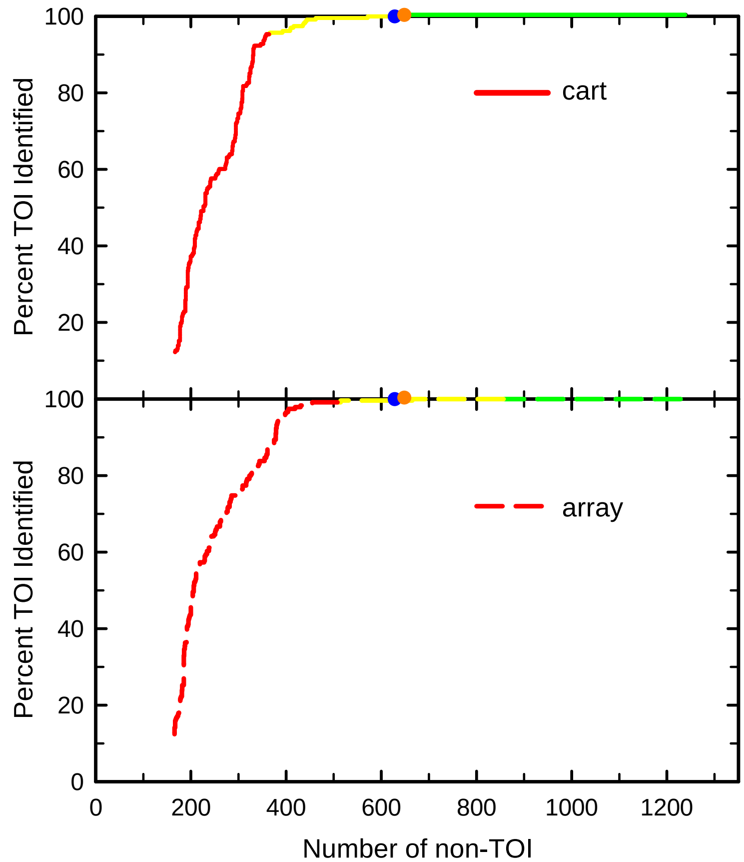


**Figure 5-5. SAIC Analysis of the EM61-MK2 CART and ARRAY Data.**

The SAIC/RML team analyzed the EM61-MK2 ARRAY data using data features and a classifier based on linear genetic programming. Two iterations were performed on the classifier and their results are shown in Figure 5-6. The panel on the left shows the first iteration in which 182 labeled anomalies were provided for training. Another 226 items in the test set were labeled as "can't analyze." At the threshold, about 300 non-TOI are correctly classified, but three TOI are missed. The missed TOI included three 60-mm mortars. It is interesting that there is little overlap in the TOI missed using this method and those missed by the model-based methods, which tended to share commonality within data sets.

As implemented by SAIC/RML, this approach allows for multiple iterations with requests for additional ground truth labels for training until it is determined that the algorithm has been optimized. In a second iteration, the vendor requested an additional 256 ground truth labels. This total of 438 training points for the second iteration is reflected in starting point of the plot in the right panel in Figure 5-6. About 231 items were labeled as "can't analyze." After the second iteration, about 400 non-TOI are correctly classified, but again three TOI are missed.

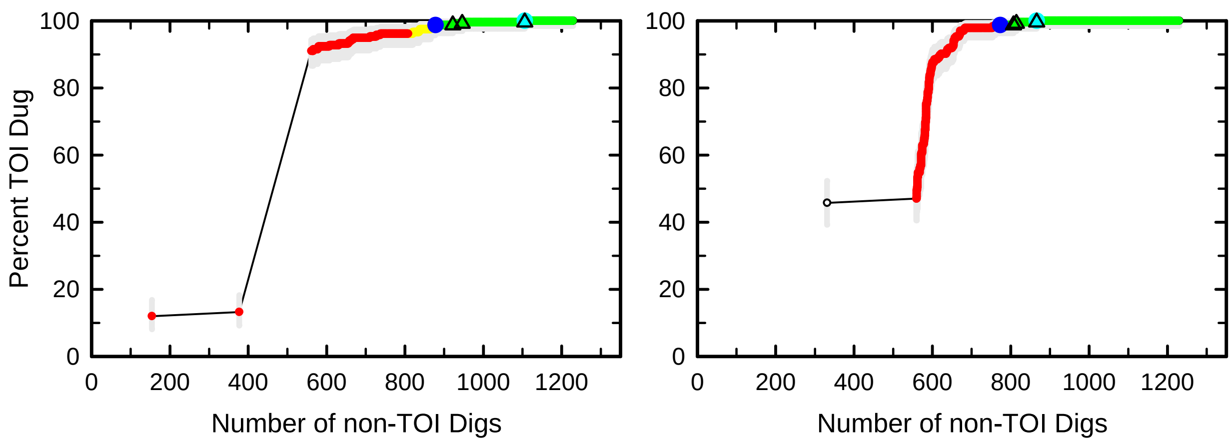


**Figure 5-6. SAIC/RML Analysis of the EM61-MK2 ARRAY Data Using Data Features on the LGP Classifier. The two panels show two iterations of the classifier.**

## 5.3 METAL MAPPER

Results of the MetalMapper data as analyzed by Geometrics and Signal Innovations Group (SIG) are shown in Figure 5-7 and Figure 5-8, respectively. Geometrics is the developer of the system and has developed analysis routines that invert for the full suite of polarization parameters and decay rates. They used a three-stage analysis process: First, the anomalies were screened based on SNR, fit quality and the estimated physical size of the object. Second, an artificial neural network is used to rank targets. Finally, curve matching to the known TOI is applied. The SIG results are shown for analysis of the features extracted by the Geometrics team using SIG's semi-supervised Probabilistic Neighborhood Based statistical classifier. Here semi-supervised means that in addition to using the training data for which truth is known in the classifier, the unknown test data is also considered in setting the decision boundaries.

Overall, the results are outstanding. It is clear from the steep rise in the red portion of the ROC curve that most TOI are readily recognized and classified as such with high confidence. It is equally clear from the distance that the green line extends along the top axis that most non TOI are also readily recognized and classified as such with high confidence. There are approximately 1300 non-TOI in the MM data set. Geometrics is able to correctly classify about 1000 of them, but sets their threshold too aggressively and misses three TOI, and one of them is deep in the high-confidence non-TOI list. The inset figures reveal that these are, in the order that they are missed, a 60-mm mortar that was collocated with several small munitions parts, the recovered 37-mm projectile, and a collection of 2.36-in rocket parts. SIG can similarly correctly classify about 1000 of the non-TOI, but misses six TOI at their threshold, including the same three missed by Geometrics. The SIG misses, however, all fall very near the threshold.

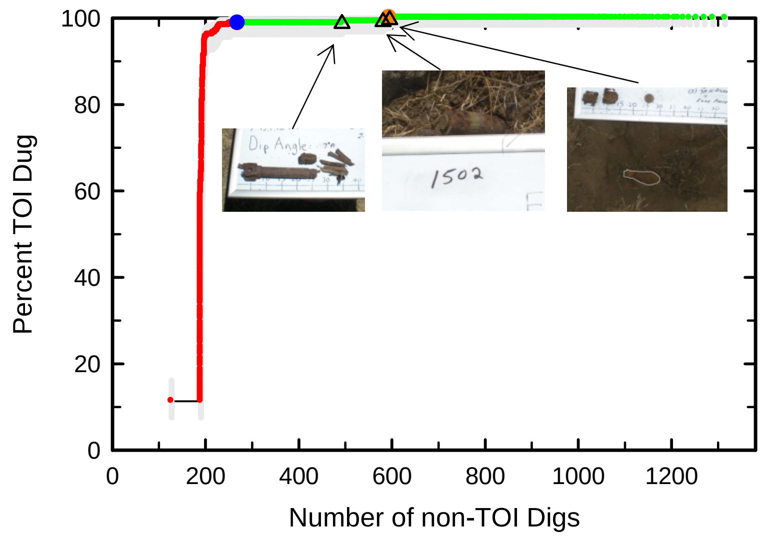


Figure 5-7. Geometrics analysis of MetalMapper data.

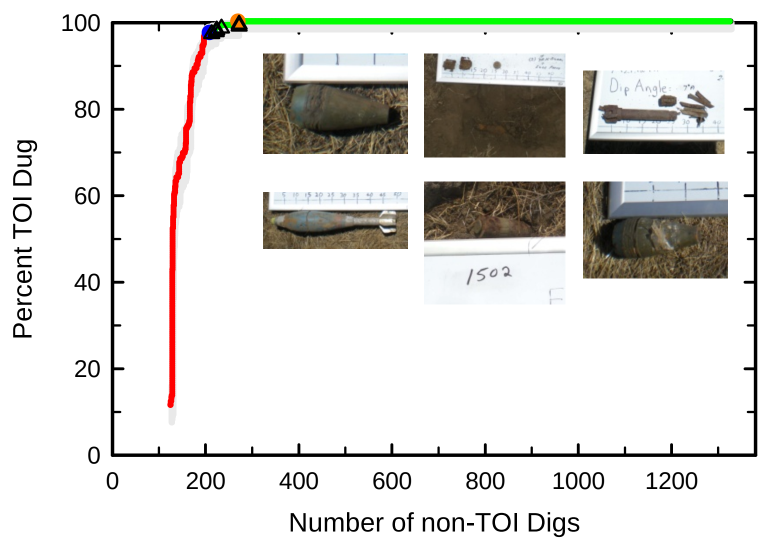


Figure 5-8. SIG analysis of MetalMapper data.

## 5.4 TEMTADS

The results of the SAIC analysis of the TEMTADS data are shown in Figure 5-9. SAIC is on the development team for TEMTADS and they have developed analysis routines that solve for the three principal polarizations for each time gate in the decay. The results shown in Figure 5-9 are for a classifier that is based on a library match of the largest polarization value $\beta 1$, which gives information about the targets size, and the ratio of $\beta 1/\beta 2$, which gives information about its shape.

The shape of the SAIC TEMTADS ROC curve is similar to the MetalMapper and it shows a steep rise in the red area indicating TOI are readily identified with high confidence. There are two important differences however. The first is that there are a large number of targets in the yellow area, indicating the “can’t decide” region. The second is that although the green area remains close to the 100% correct classification of TOI for a considerable length, there are three missed TOI and they are scattered throughout the ranked list, with one being ranked very near the most confident non-TOI. As seen in the inset, the missed targets include two collections of 2.36-in rocket parts and a 60-mm mortar. The rocket parts again fall into the TOI that are not hazardous, but were included in the conservative definition of TOI. However, the 60-mm mortar is a mostly intact round that undoubtedly would need to be correctly classified in any implementation.

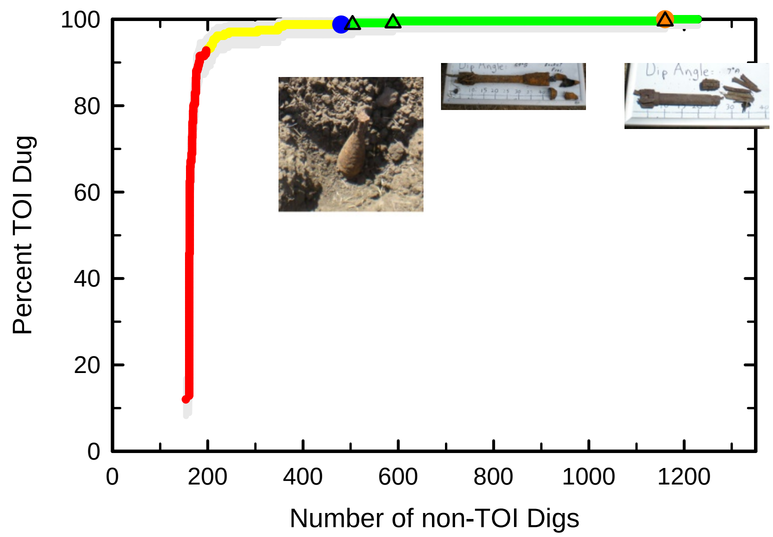


**Figure 5-9. SAIC analysis of TEMTADS data.**

## 5.5 MAGNETOMETER DATA

Minimal analysis was done on the magnetometer data from SLO. Figure 5-10 shows a false color map of the magnetometer data over the whole site. Clearly, there are a number of strong large anomalies from geology that make processing these data challenging. Figure 5-11 shows a magnified chip of the magnetometer data beside the EM61-MK2 ARRAY data for the same area. The crosses indicate threshold exceedances at the threshold set as described in Section 4.1. It is clear that in addition to the large scale geology, there are far more discrete anomalies in the magnetometer data that would need to be classified. Many of these are weak signals that pose analysis challenges. Because there were not sufficient funds to dig all of the 5500 magnetometer targets, the classification was performed on only those anomalies that corresponded to EM61-MK2 ARRAY detections.

The results of the analysis of the magnetometer data performed by Sky Research are shown in Figure 5-12. Several conclusions about this sensor/analysis combination can be seen in the figure. First, there were more than 100 “can’t analyze” targets. Second, the green region indicating high confidence non-TOI is very small compared to other analyses, so few targets could be confidently left behind. Third, the rise in the red portion of the curve is much less steep than in the EM results, indicating that the analysis was not as successful at identifying the TOI with high confidence: this list contains far more non-TOI in the red region than those for other sensors. Finally, although it does not appear in the figure, analysis of magnetometer data alone would require classifying all of the nearly 4000 additional threshold exceedances that were eliminated from this process. Most of these likely would be low SNR targets or obscured by the large geologic anomalies and likely end up as either "can't analyze" or "can't decide."

In short, the magnetometer is not a good choice for attempting classification at this site. One implication of this is that the added value of the concurrent collection of magnetometer and EM data offered by the MSEMS system could not be assessed.

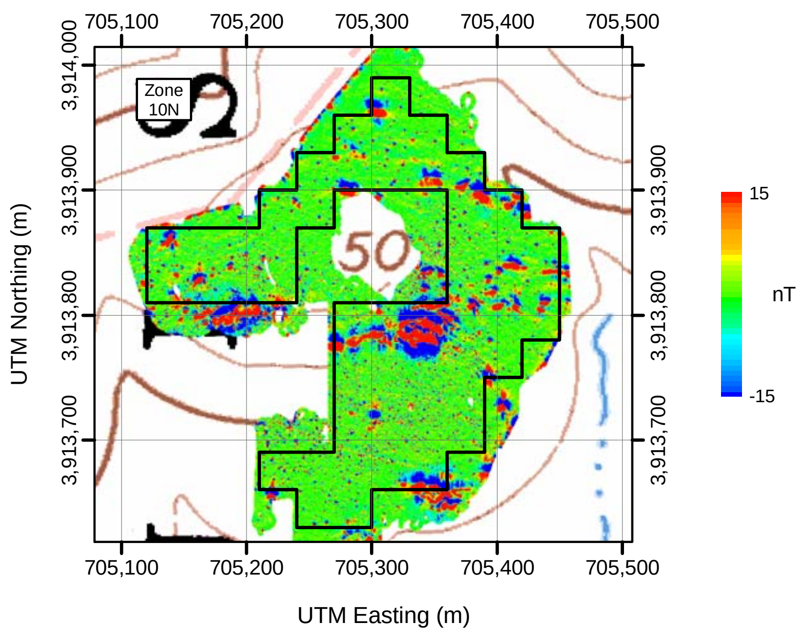


Figure 5-10. Magnetometer data at SLO.

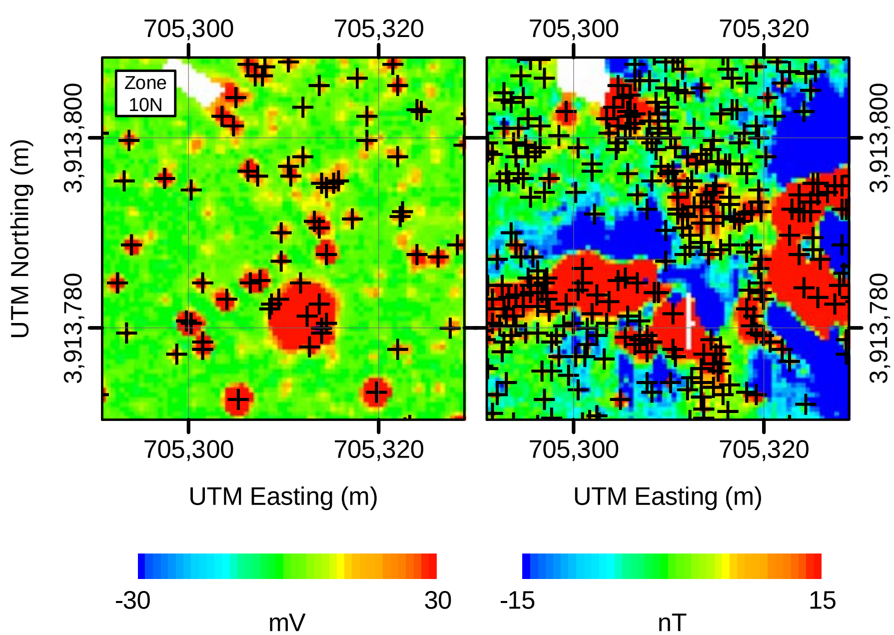


Figure 5-11. Data chip of EM61-MK2 ARRAY data (left) and magnetometer data (right) at SLO. The crosses indicate threshold exceedances that were selected as detections requiring classification analysis.

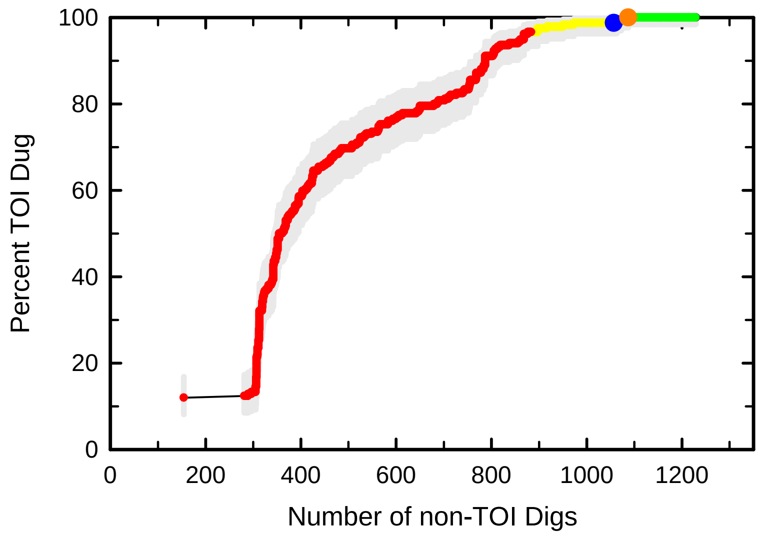


**Figure 5-12. ROC curve resulting from the analysis of magnetometer features by Sky Research.**

## 5.6 DISCUSSION

### 5.6.1 Features

In the first demonstration of this series at Camp Sibert, there was only one munition of interest, the 4.2 in mortar. Since this is a large object, it was expected and borne out by the demonstration results that estimates of the physical size of the object would be a valuable parameter in classification. Here, size refers to the estimated physical size of the object, as opposed to the size of the anomaly footprint. At Camp SLO, with multiple munitions types expected, ranging in size from the 60-mm mortar to the 4.2-in mortar, it was expected that classifiers would need to include other parameters in addition to size estimates.

All demonstrators that analyzed the EM61 data, from both the CART and the ARRAY, observed that the decay parameter was of primary importance. Figure 5-13 shows plots of asymmetry vs size and decay constant versus size for the EM61 CART data. Visual inspection reveals that the decay constant provides much better separation between the TOI and non TOI.

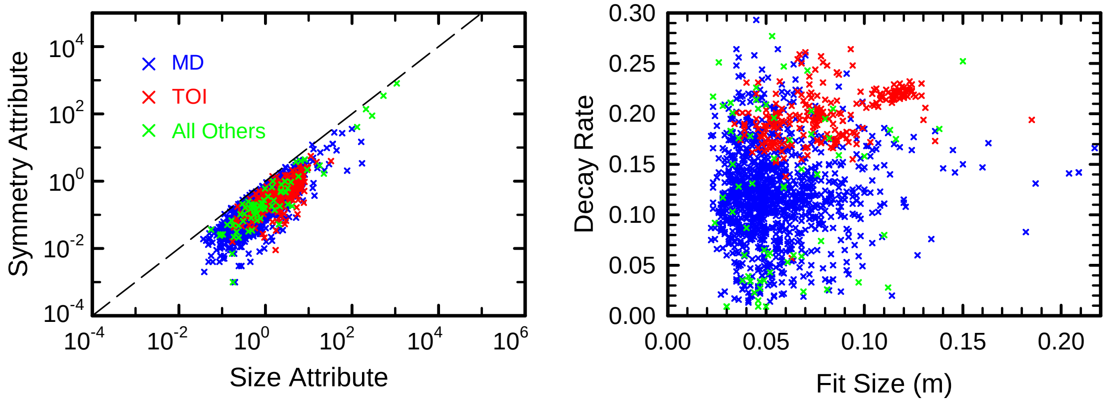


**Figure 5-13. EM61-MK2 CART parameters as analyzed by SAIC. The left panel shows asymmetry versus size and the right panel decay rate versus size.**

The advanced sensors provide much more accurate estimation of the polarizabilities, which are the basis of the key features used in all of the classification approaches. Figure 5-14 shows the extracted polarizations as a function of decay time for TEMTADS. It is clear that both are well reproduced for this series of 60-mm mortar targets measured in the test pit. Such results present many more possibilities for parameters to be used in statistical classification, including size and time decay useful for the EM61 systems, but also adding options for shape and asymmetry.

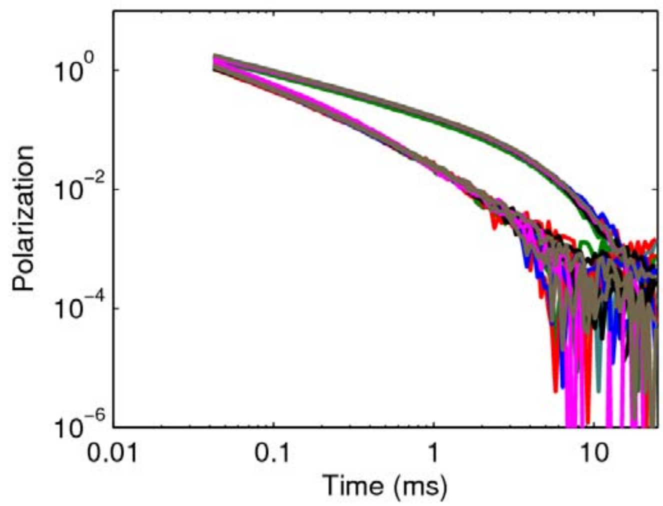


**Figure 5-14. Superposition of the polarization as a function of decay time for TEMTADS for six 60-mm mortars measured in the test pit.**

## 5.6.2 Unexpected TOI

Three munitions types were uncovered in the digging that were not expected on the site based on the history of munitions found, as well as the digging of the two grids prior to the experiment. These included a 37-mm projectile, a 5-in rocket warhead, and a 3-in Stokes mortar. The demonstrators were expected to correctly classify all targets of interest, including any unanticipated munitions types, so incorrectly classifying these items is scored as a false negative failure.

**37-mm projectile:** The single 37-mm projectile that was recovered appeared only in the EM61-MK2 CART, MSEMS EM61, and MetalMapper data set. It was located in a rocky area that was inaccessible to the EM61 ARRAY platform. Since the cued target list was generated using the EM61-MK2 ARRAY data, it did not appear in any of the cued lists. The 37-mm projectile was classified as "high confidence TOI" in three of the EM61-MK2 CART dig lists presented above and as "can't decide" in the fourth. It was, however, misclassified in the MetalMapper analysis by all processing demonstrators.

**5-in rocket warhead:** The 5-inch rocket warhead appeared in all the data sets analyzed. In all cases discussed in this report and almost all cases overall, it was correctly classified as a TOI, likely because it is similar in size to the 4.2-in mortar, which is a known TOI.

**3-in Stokes mortar:** The 3-in stokes mortar was detected by all the survey systems. It was not in the subset of the site covered by BUD but was correctly classified in all the other dig lists.

### 5.6.3 Issues with Missed Targets of Interest

There are a few targets or classes of targets that seemed to give all analysis teams problems. These included the 2.36-in rocket and the 60-mm mortar bodies (teardrops). Examples are shown in Figure 5-15 and Figure 5-16. Specific examples of 2.36-mm partial rounds showed up repeatedly across many data sets. In contrast, the missed 60-mm mortars all looked similar, but the actual items seem randomly distributed.


Figure 5-15. Examples of commonly missed 2.36-in rockets.

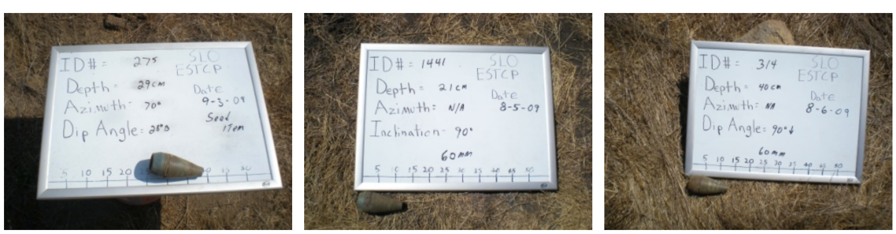


Figure 5-16. Examples of missed 60-mm mortars.

Other items that commonly caused classification problems were TOI collocated with other objects. Examples are shown in Figure 5-17. Extraction of accurate parameters for multiple objects is known to be a continuing difficulty and remains the subject of research.


**Figure 5-17. Two commonly missed target are a 2.36-in rocket warhead collocated with a piece of barbed wire (left) and a 60-mm mortar with other munitions debris (right).**

The polarizabilities for all complete 60-mm mortars and 60-mm mortar bodies encountered in this demonstration are shown plotted in Figure 5-18. The symbols correspond to the average polarizability and the vertical lines represent the standard deviation. There is very little spread in the derived values within a class but the mortar bodies have about a 40% lower response. This becomes important when these items are encountered in the field.

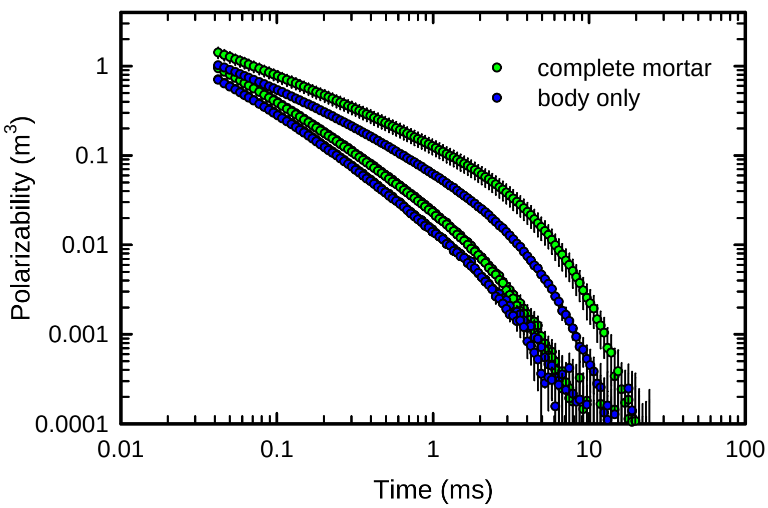


**Figure 5-18. Calculated polarizabilities derived from test stand measurements of all complete 60-mm mortars and 60-mm mortar bodies encountered in this program. The symbols correspond to the average polarizability and the vertical line represents the standard deviation.**

Figure 5-19 shows the signals measured in the field from the center sensor in the TEMTADS array when the sensor was positioned over two 60-mm mortar bodies. Anomaly 498 corresponds to the strongest signal measured from a 60-mm body while anomaly 16 corresponds to the weakest. Although the weaker signal shown can be analyzed, it decays to the noise floor much sooner than the strong signal, so the late-time behavior cannot be modeled accurately. It is signals such as this that caused the demonstrators trouble. In many cases, this difficulty could have been avoided by better field QC procedures to ensure that adequate SNR signals are collected over every anomaly.

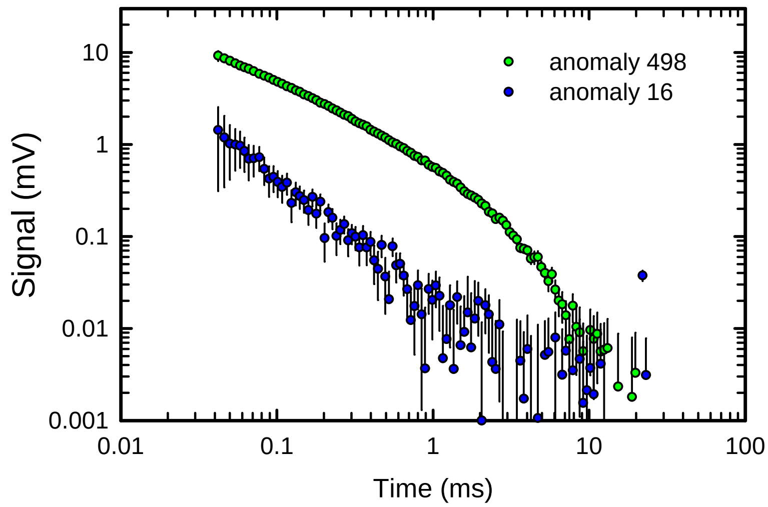


**Figure 5-19. Monostatic response of the center sensor in the TEMTADS array when the sensor was positioned over the 60-mm mortar body with the largest signal and the one with the smallest signal.**

Some TOI were missed only by one sensor system or another, suggesting an isolated data problem with the data collected by that sensor at that target. For example, Figure 5-20 shows the attempt by Sky Research to fit TEMTADS data from target 103. This was a 60 mm mortar that was misclassified as a non-TOI. The figure shows the predicted and fit values for the polarizabilities calculated for each sensor element in the array. When compared to the noise floor, indicated by the dotted line, it is evident that the SNR on this target was not sufficient to support extraction of reliable parameters. This target was classified as either "high confidence TOI" or "can't decide" by the other demonstrators who submitted a TEMTADS dig list.

The RML method missed 3 targets. Not all of them are common to misses by the other classification methods. Absent a physical interpretation of the signature, it is difficult to determine physical cause.

Two main conclusions can be drawn from this study of the missed TOIs at SLO. First, the targets of interest need to carefully specified and appropriate examples provided in the training data, and the data processors must carefully consider the effects of TOI variability in assessing the training data. The tailboom missing from the 60-mm mortar teardrop shapes made a substantial difference in signature from the intact round and the 2.36-in rocket parts spanned a continuum in which it was difficult to apply a consistent, bright-line rule about what was a TOI. Both factors caused missed TOIs at SLO. Second, for the advanced sensors, where the sufficiency of data collected at each location is determined by the operator's decision that adequate SNR has been achieved, it is important that careful QC verify these decisions while still in the field.

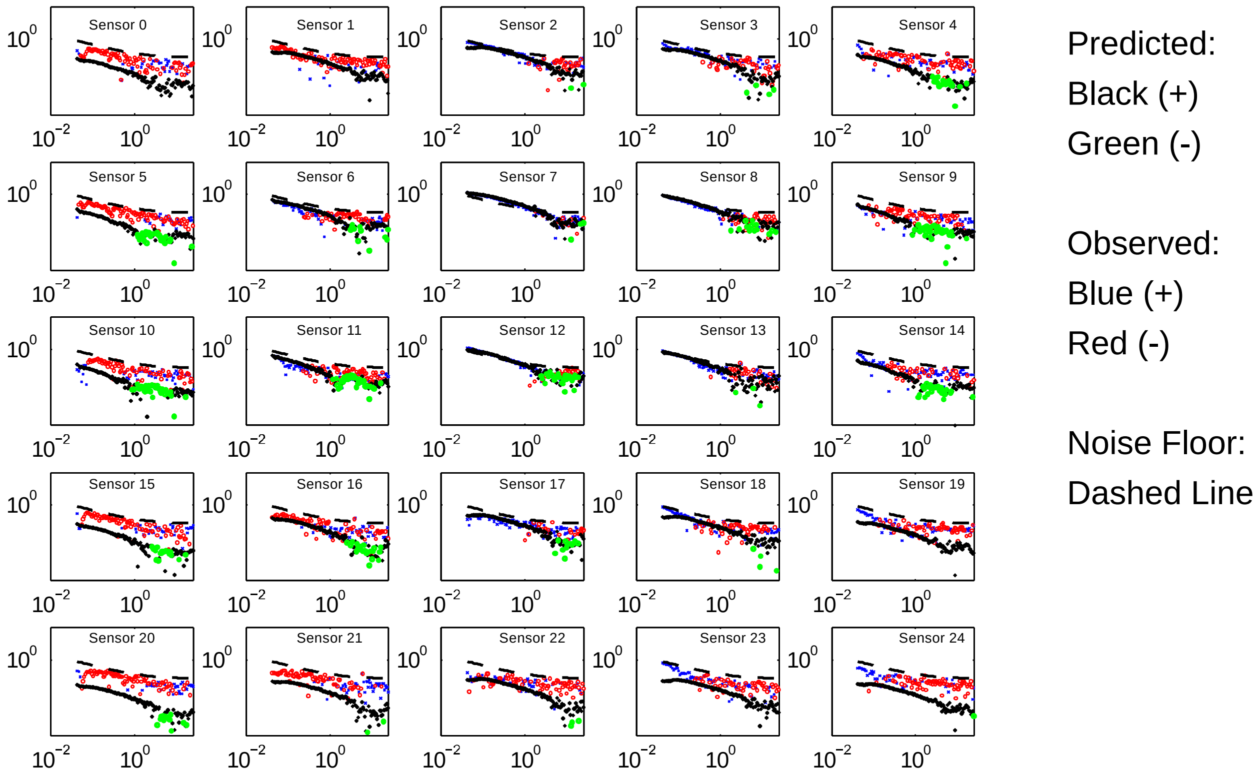


**Figure 5-20. Predicted and observed data for the TEMTADS target 103, a 60-mm mortar, as analyzed by Sky research.**

# 6 COST CONSIDERATIONS

The motivation for applying classification in munitions response is cost savings: if the digging of non-munitions targets is minimized, then the limited resources of the munitions response program can be more effectively applied to clean up more land more quickly. The actual costs of a demonstration include extensive planning, reporting, and coordination, as well as redundant data collection and processing by developers that has not yet been standardized for field use. These costs are not representative of what would be expected for production application. We use a simple cost model and realistic assumptions for production costs of various model elements to estimate costs for two scenarios.

## 6.1 COST MODEL

The cost model shown notionally in Figure 6-1 compares the current approach of digging all detections above a pre-set threshold to using classification to eliminate digs of items determined with high confidence to be non-TOIs. The elements included in the model are:

"Dig Everything" shown on the red plot

- Data collection and processing: the value on the y-axis includes the costs of data collection and processing
- Digging: It is assumed that all of the detected targets are dug at a constant cost

"Classification-Guided Digging" shown on the green plot

- Data collection and processing: the value on the y-axis includes the costs of data collection and processing. It expected that these costs will be higher for classification than for detection only.
- Digging: the per dig costs are assumed to be same as for detection only. However, the digs are broken out as follows:
    - Training: The cumulative cost of the digs needed to collect the training data are represented by the ▲
    - TOIs: the cost for digging all of the targets that are classified as TOI, can't analyze and can't decide are represented by the ▼
    - Non-TOIs: Items below threshold indicating high confidence non-TOI are not dug so no further costs are incurred.

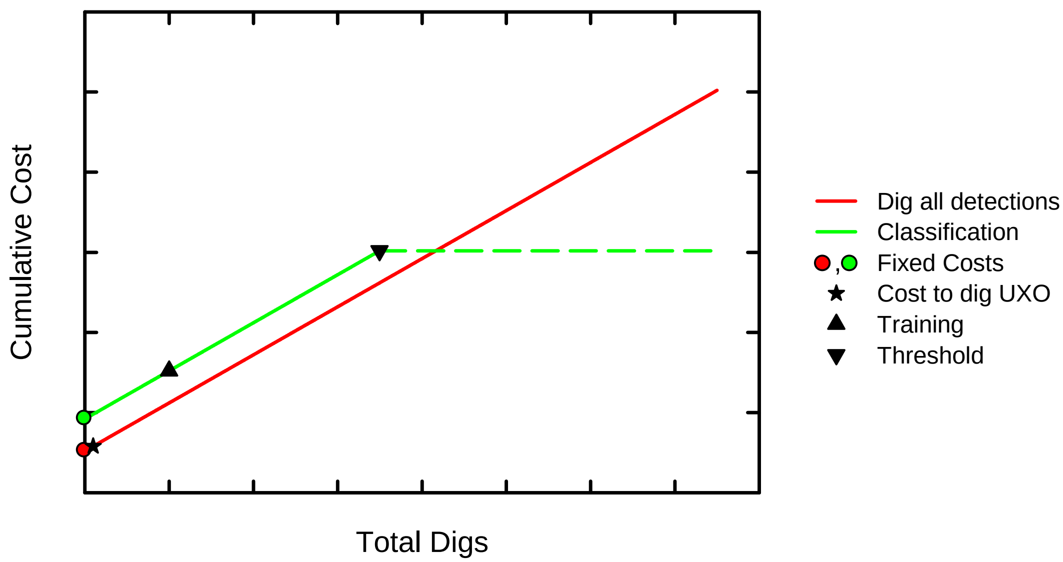


**Figure 6-1. Notional Cost Model.**

There are several cost variables that will determine the real-world trade off in total project costs. The first is the cost of data collection and processing. It is expected that these costs will be higher for classification than for detection only surveys to support digging all targets. The second is efficiency of the classification. The more targets that can be identified as non-TOI with high confidence, the greater the cost savings. The final variable is the cost per dig. We look at two scenarios for cost comparison using the commercial EM61 sensor as well as the advanced MetalMapper.

## 6.2 ESTIMATED COSTS FOR A SMALL SITE

The scenario for a small site is similar in size to the demonstration at SLO. The site is 10 acres in size, with 1500 detections above the target selection threshold, and a total of 20 TOI were recovered. The cost for each dig is $150.

For the small site, the EM61 is operated in cart mode. The survey and analysis costs for the detection only survey are $2600 per acre and for the classification survey with higher data density and additional processing are $4600 per acre. The classification requires 200 training items and, using the classification achieved at SLO, results in another 499 total digs. This comes to a total of 699 digs required in the classification approach out of the 1500 total detections. Compared to the $250K cost of digging every object, a savings of about $100K can be realized on this site with classification.

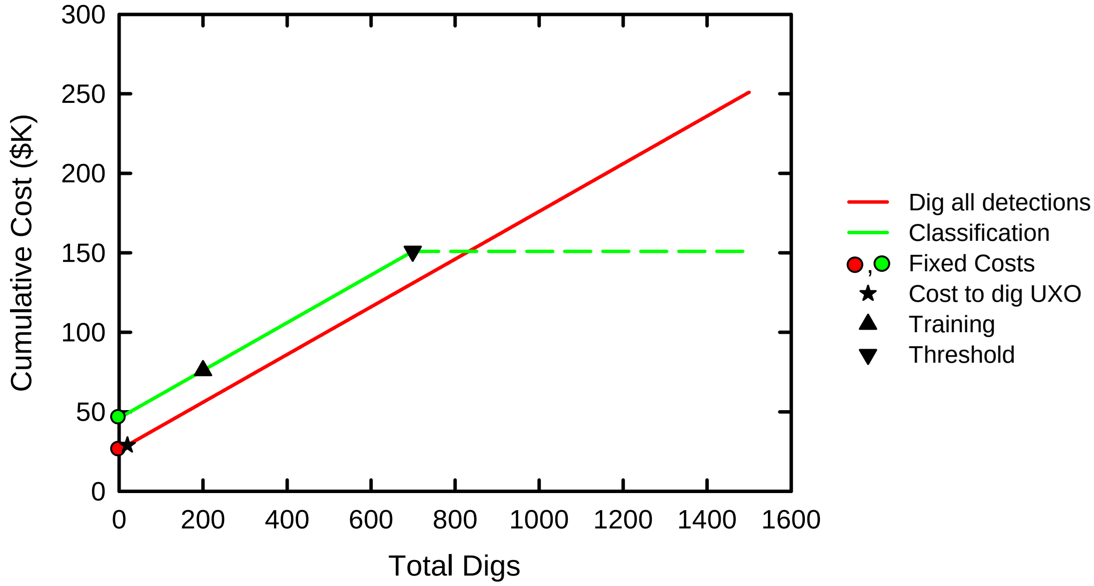


**Figure 6-2. Cost Model for an EM61 CART on a 10-acre site.**

A similar analysis for the MetalMapper on the small site is seen in Figure 6-3. The same assumptions are made about the costs per dig. The survey costs for detection only are $2600 per acre and the additional costs for the cued data collection and classification processing bring the price to $4200 per acre. The same 200 targets were provided for training data. However, the classification efficiency was far greater: only an additional 140 TOI required digging, for a total of 340 total digs out of the 1500 total detections. Compared to the $250K cost of digging everything, the MetalMapper can achieve a savings of about $150K.

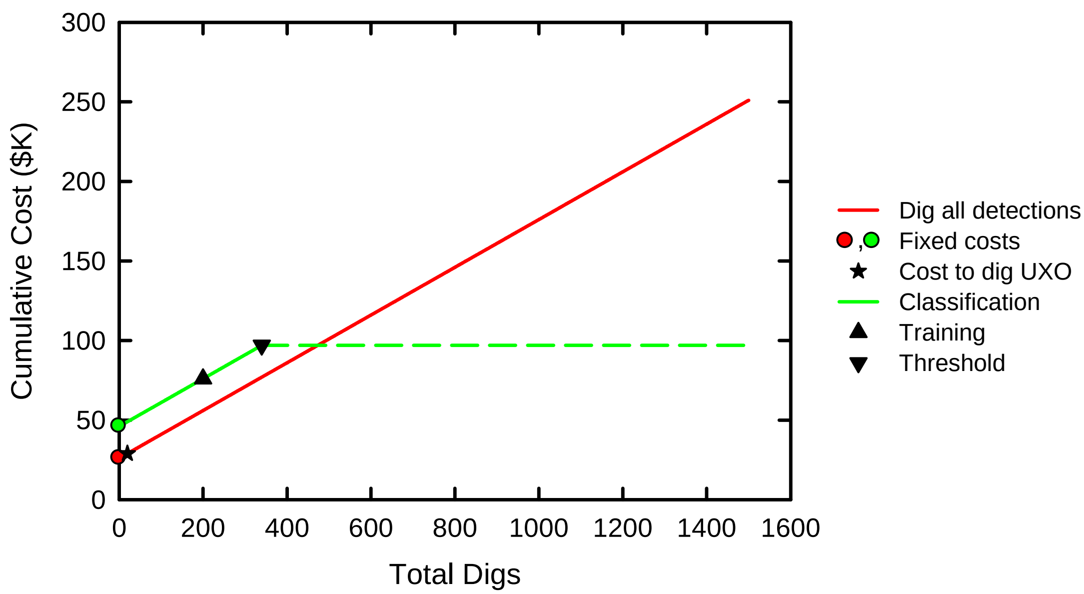


**Figure 6-3. Cost Model for MetalMapper on a 10-acre site.**

## 6.3 ESTIMATED COSTS FOR A LARGER SITE

Many sites are likely to be far larger than the 10-acre demonstration at SLO. To illustrate the economics for a larger site, the second scenario imagines a site that is 100 acres, with 15,000 total detections above the target selection threshold and 200 TOI recovered. The digging costs are assumed to be lower at $90 per target. As with the small site, 200 training targets were dug and provided and the fraction of items correctly identified as non-TOI at SLO was applied to the 15,000 total detections.

For the larger site, the EM61 is deployed as a towed array. Detection only and digging everything on this site would cost $1.4M. As shown in Figure 6-4, digging only the targets that were not high confidence non-TOI in the classification processing of EM61 data could have saved about $650K.

Figure 6-5 shows a comparable analysis of MetalMapper on the larger site, using the same cost assumptions and 200 training targets. The $1.4M cost of the detection only process could have been reduced to approximately $400K using the MetalMapper classification results to guide the digging, for a savings of about $1M on a 100-acre site.

Table 6-1 summarizes the results of the cost estimates.

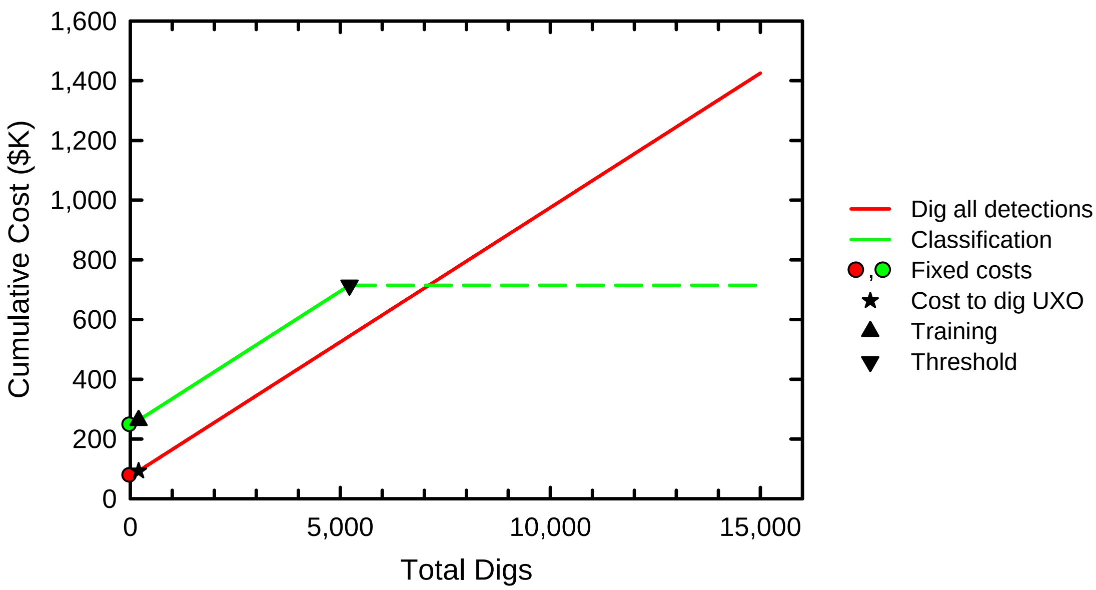


Figure 6-4. Cost Model for an EM61 on Large Site.

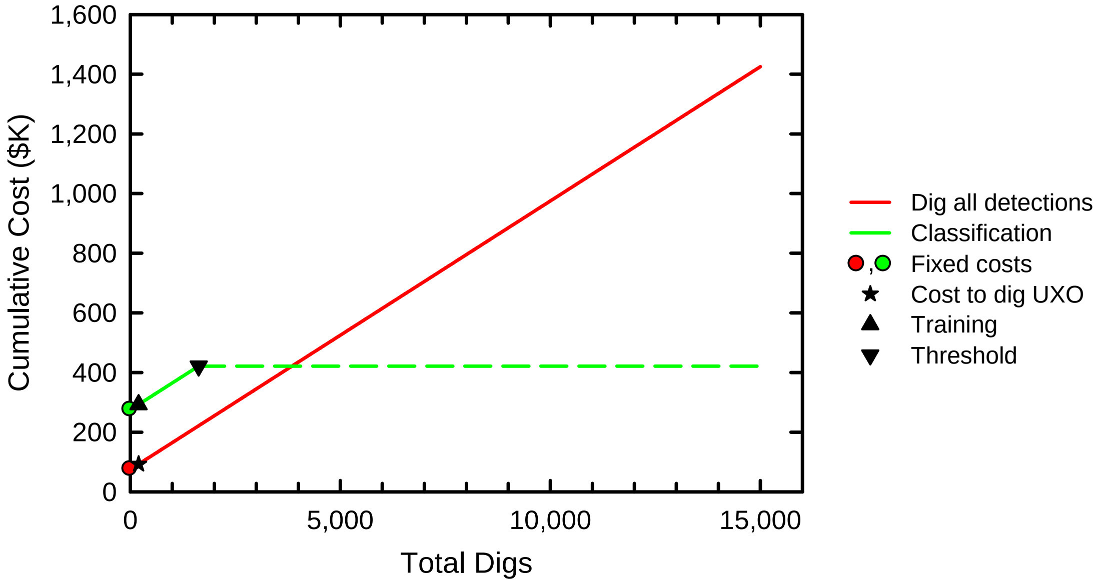


**Figure 6-5. Cost Model for MetalMapper on Large Site.**

**Table 6-1. Summary of Cost Model in $K.**

| | 10-acre Site | | 100-acre Site | |
|---|---|---|---|---|
| | **EM61** | **MetalMapper** | **EM61** | **MetalMapper** |
| **Detection Only** | 250 | 250 | 1,400 | 1,400 |
| **Classification** | 150 | 100 | 750 | 400 |
| **Savings** | 100 | 150 | 650 | 1,000 |

# 7 CONCLUSIONS

The demonstration at the former Camp San Luis Obispo showed substantial classification ability on a site that was more challenging than the Camp Sibert demonstration site with a wider mix of targets of interest. In the application of classification analyses to carefully collected survey data from commercial EM sensors, the best performers correctly classified all or nearly all of the targets of interest, while achieving reductions of up to 50% in the number of non-TOI. Recently-developed EMI sensors, optimized for UXO classification, showed outstanding results.

## 7.1 DEFINING SUCCESS

In munitions response, success should be judged from the perspective of risk reduction. All cleanups are imperfect – there are limits on the detection capabilities of current sensors and, even with careful quality control, human errors occur in the tedious and repetitive work. At the end of a cleanup, even one where the standard is to "remove all metal from the site," there remains residual risk and uncertainty that is unknown and unquantifiable.

In this context, how are the merits of classification judged? In the reality of residual uncertainty at the conclusion of a traditional response action, there will always need to be some risk management plan. The results from SLO indicate that the application of classification may add modest additional uncertainty. However, it is unlikely that it would change how residual risk is managed. In fact, the classification process is extremely well documented, transparent, and auditable, and its residual risks are quantifiable.

From the perspective of demonstrating real risk reduction, the SLO study was hugely successful. For the purposes of the study, a target of interest was defined by explosive hazard and the potential to alarm the public. Although this assured that nothing on our non-TOI list that would be left behind would cause concern, this resulted in the inclusion of many partial rounds in the TOI that were not in fact hazardous. Reasonable people can disagree on definition of a TOI and, ultimately, it will be the task of the project team to make this decision.

At SLO, a detailed examination of the TOI that were correctly identified indicates that successful demonstrators reliably identified the things that posed an immediate threat. Where TOI were mis-classified, errors were in the partial rounds. These item may contain residual HE , and were added to the list of TOI, but did not present an explosive hazard. At this site, there is predominantly one type of partial round (60-mm mortar body that was shaped like a teardrop). As the analysis of the demonstration progressed, many of the demonstrators learned to correctly identify these targets.

Finally, the classification process allows for iterative expansion of the TOI. If, during the course of digging targets, evidence of unexpected items is uncovered, the decisions throughout the process can be revisited. Since all of the data and decision criteria are archived, the anomaly selection criteria and the decision criteria for selecting TOI can be revised at any time.

## 7.2 LIMITATIONS

This demonstration revealed three primary challenges:

**Partial rounds** – Our definition of a TOI included partial rounds, which accounted for the vast majority of mis-identified TOI. Some partial rounds were included in the training data, but no set of training data will ever provide every possible permutation of partial rounds. It is important to recognize that there is a continuum of partial rounds, clearly define what constitutes a TOI, and set classification criteria appropriately.

**Overlapping targets** – Some misidentified TOI were collocated with other metallic items. In these cases, it can be difficult to determine first that there are multiple items present. If the data are analyzed assuming single source, unreliable parameters are extracted and an incorrect decision can be made. The identification and handling of multiple objects is an active area of research.

**Thresholds** – While there are examples from this study where a TOI was ranked well into the non-TOI list, that is the demonstrator was very confident it was not a TOI, in general the rankings were accurate. There are a number of cases, however, where a handful of TOI just to the left or right of the threshold, that is they are the last correctly identified TOI or the first mistake. These are the subject of retrospective analyses to help improve threshold selection.

## 7.3 NOT YET TESTED

ESTCP has now conducted two large-scale-live site demonstrations of classification. These have tested many aspects of classification, but many more remain to be explored. Notably,

- many types of partial rounds, including fuzes, bursters and other hazardous components,
- submunitions, and
- smaller munitions types, including 37-mm and 20-mm projectiles.

Future demonstrations are being planned to address these issues.

## 7.4 DETAILED OBSERVATIONS

### 7.4.1 EM61

The EM61-MK2 CART data were collected by NAEVA Geophysics and analyzed by NAEVA and Parsons, each using the UX-Analyze software module. Both of the analysts based their analyses on signal decay and, while they were both able to correctly label a significant number of the non-TOI as nonhazardous, there were a few missed targets of interest.

The algorithm developers achieved better results with the EM61-MK2 CART data, SAIC using UX-Analyze and Sky using UXO Lab. Decay-based features were also the primary discriminant used by these analysts. Both of these analysts labeled few or no targets as "can't analyze" and both were able to correctly classify more than 600 of the more than 1200 non-TOI. They set their thresholds in slightly different points, resulting in SAIC missing one TOI near threshold.

A conventional EM61-MK2 ARRAY system also collected data during this demonstration. Even though this system provides better inter-sensor location accuracy and better measurement of platform orientation, the classification results achieved with the array data were virtually identical to those using the cart data when analyzed by the same demonstrator. Since data from both platforms can support similar classification results, at this site, the choice between single-sensor cart deployment and array deployment is best made on data collection efficiency grounds.

### 7.4.2 Advanced Sensors

Three recently-developed EMI sensors, optimized for UXO classification, were demonstrated at San Luis Obispo. The NRL TEMTADS and LBNL BUD system were operated in cued mode working against the detection list from the EM61 array. The Geometrics MetalMapper system was self-cued; it first surveyed the field in detection mode and then revisited all locations on its detection list in cued mode.

The MetalMapper system achieved outstanding classification performance. Most TOI were readily recognized and classified as such with high confidence, there were very few "can't decide" targets, and the analysts working with the MetalMapper data were able to correctly classify about 800 of the more than 1200 possible non-TOI correctly, although there were a handful of missed TOI at the demonstrator thresholds. The TEMTADS analysts achieved similar results although there were significantly more targets classified as "can't decide" and two partial 2.36-in rockets appeared well into the "high confidence non-TOI" portion of the anomaly list.

Upon retrospective analysis, the misses in the emerging sensor data sets could largely be attributed to insufficient SNR on a handful of targets. This underscores the role of careful field QC when deploying these sensors. As one demonstrator remarked, "if you collect high-quality data with these sensors, the decision makes itself." As we gain more experience with these systems, data collectors are likely to develop field procedures that will minimize the errors due to insufficient data quality.

### 7.4.3 Defining TOI

Once again, this demonstration employed a target-based approach to anomaly selection in which the detection threshold is set based on the minimum signal expected from the targets of interest rather than referenced to the site noise. For this approach to be successful, the definition of TOI must be carefully considered; if partial munitions are targets of interest, the target selection criteria should reflect the minimum portion of a munition that is of concern. In this demonstration, the selection threshold for one sensor was based on the predicted signal from intact 60-mm mortars causing one 60-mm mortar body to be missed in the EM61-MK2 CART data set when its signal was just below the threshold.

Careful definition of TOI is also important for the classification portion of the demonstration. To maximize classifier performance, training data must contain examples consistent with the definition of TOI. Some ambiguity will be hard to avoid on a live site but the better the initial definition of TOI, the better the classification results are likely to be.

We were very conservative in defining TOI for this demonstration. As discussed above, a number of items that were scored as TOI were, in fact, not hazardous. In a real-life implementation of these techniques, the site team needs to carefully consider the definition of TOI, focusing their attention on what misses would truly present a concern.

No matter how well this TOI definition process is accomplished, there is always the possibility that the historical records are incomplete and that unanticipated munitions types will be present on the site.  In this demonstration, three unanticipated munitions were encountered: a 37-mm projectile, a 3-in stokes mortar, and a 5-in rocket warhead.  The larger two items were easily detected and classified as either "high-confidence TOI" or "can't decide" on all anomaly lists.  The 37-mm projectile was detected by all sensors that surveyed its location but was not classified as TOI on some of the MetalMapper lists where it was rejected as being too small to be a TOI.

In all cases, and particularly for the advanced sensors, commonly missed TOI were mostly partial rounds and munitions recovered collocated with other items.  The first of these is a data QC and classifier training issue that can be expected to improve as more experience is gained in the operation of these sensors.  The second is an active area of research and will be considered by the demonstrators in their follow-on retrospective analysis programs.

### 7.4.4 Cost and Production

We developed a simple cost model for the implementation that considers the added costs to collect classification-quality data and conduct the required analysis as well as the digs avoided by successful classification.  Examples are provided for a site the size of this demonstration and a 100 acre site with 15,000 detected anomalies.  For both example sites, moderate savings can be achieved using conventional EM61-MK2 surveys and substantial savings are possible using the MetalMapper system.

# 8 FREQUENTLY ASKED QUESTIONS ABOUT CLASSIFICATION

**What is classification?**
Classification is a process that differentiates munitions from nonhazardous buried items by applying mature physics-based analysis methods to sensor data. Nonhazardous items could include munitions-related scrap, geology, and cultural clutter. The analysis methods are used to estimate parameters of buried objects, such as size, aspect ratio, and electromagnetic decay rates, that can be useful in distinguishing munitions from other sources. Advanced classification algorithms then use this information to determine whether a signal is likely to arise from a munitions item or another source.

The results from the classification algorithms are used to develop a dig list ranked from highest confidence the items are nonhazardous to highest confidence the items are munitions. A point on the dig list, termed the threshold, is selected to separate items that need not be treated as potential munitions.

**What about other anomaly prioritization methods?**
Physics-based approaches demonstrated in the pilot program, are principled, transparent, processes consisting of component methods that are well understood and have undergone several years of development and review by DoD and other interested parties. Alternative anomaly-prioritization methods have been attempted. ESTCP does not have a detailed understanding of these methods, and many differ considerably in approach. Only those methods referenced here were demonstrated as part of the pilot program. The same steps could be implemented by other analysts with the same result. The successes seen in the pilot program should not be expected to transfer to alternative and undemonstrated methods.

**How do you evaluate whether a proposed classification approach is real? How do you filter bold claims?**
Proposed classification schemes should be physics-based, principled, and transparent, with a set process for quantitative decision making. They should have a development history that provides well-documented and predictable results in controlled experiments, such as test stands and test sites, followed by comprehensive, well-documented demonstration of success at a real field site. So-called black-boxes in which the methods and decision-making criteria are not transparent should be viewed with skepticism.

**There is always some concern about items not being detected. What affect does this have on implementing classification?**
Detection and classification should be thought of as separate sequential steps. If an item is not detected, there is no opportunity to classify it. The classification step is performed only on those signals selected as detections, and applying classification can neither improve nor hinder the detection step. However, because classification generally requires better data than detection alone, the steps taken to collect such data (tighter line spacing, more data stacking, careful geolocation) may actually result in improved detection.

**Where could classification fit into the regulatory/munitions response process?**
There are several applications for the information gained from classification on a site. In the remedial phase, site managers can use the ranked dig list to select an appropriate point to stop

digging and leave the remainder of the detected items in the ground. The results of classification could also be applied to a site where all items will be dug. It is expensive to deploy UXO technicians and in some cases expensive shielding equipment and exclusion zones to support digging of all detected items at a site. The ranked dig list produced from classification could be used to manage dig teams by deploying these measures only where they are necessary.

In the Site Inspection or Remedial Investigation phase, where the objective is to determine the nature and extent of the contamination, classification can be used to guide investigative digging. Sampling items with a variety of physical parameters can lead to a more complete understanding of the site.

**What type of sites is classification applicable to?**
The first demonstration in the pilot project validated classification technology on a simple site with a single munition type, benign to moderate geology, limited vegetation, and flat terrain. The San Luis Obispo demonstration was at a site with more rocky and hilly terrain with a range of munitions and still showed good results. The influence of site characteristics such as smaller items of interest (such as 37-mm projectiles) and denser vegetation will be tested in a continuing ESTCP effort that will span several years. In principle, classification can be applied at any site where digital geophysical mapping data can be collected. The objective of this series of demonstrations is to further define types of sites where classification would be appropriate and expected to be successful.

**Is the classification process mature enough to employ on my site?**
Commercially available technologies were tested as part of this program and showed substantial classification performance. This success required careful data collection and analysis using advanced, but freely available, tools. To replicate this success on a site, care will have to be taken to follow this example. The process is not yet automated; knowledgeable and experienced geophysicists are required for success. As the advanced sensors become commercially available and the analysis methods become more widely known, this will change.

**How do I perform classification on my site?**
The classification process as demonstrated here consists of two essential steps: collection of precisely mapped, digital geophysical data and careful analysis with advanced tools.

The first step is to collect 100% coverage digital geophysical data over the survey area. For example, EM61-MK2 data can be collected with a cart using tighter lane spacing than would be currently used for a detection-only survey (0.5-m spacing in this case), but the process uses normal commercial data collection and operator procedures. A detection list is generated from the sensor data set and an appropriate threshold is chosen. For this demonstration, the threshold was set based on the signal from the smallest munitions target of interest at a depth of 45 cm. This depth was based on the results of an exploratory dig at the site with a 50% safety margin applied.

The detected anomalies can be analyzed using the commercially available UX-Analyze module of the Geosoft software Oasis montaj. Each anomaly can be analyzed to extract features such as size, depth, aspect ratio, and electromagnetic decay rate. Classification algorithms then use these features to assign a probability that the item is a munition or nonhazardous. This information is used to create a ranked dig list that orders all anomalies from highest confidence an item is nonhazardous to highest confidence an item is a munition. A point on the dig list can be selected by a project team to identify which items must be treated as potential munitions based on the site-remediation objectives. Additional sensor and data analysis technologies were successful in this project.

**What technologies are needed for classification?**
Digital geophysical data are required for classification. Both magnetometer and electromagnetic induction data can be used for analysis although, at this site, the geologic interference resulted in the magnetometer data being of little use. Each of these sensors have strengths and weaknesses, and the decision of which sensor to use will depend upon site conditions and objectives. Magnetometers can locate relatively deeper items, but can only detect ferrous materials, and, as was the case at San Luis Obispo, their effectiveness is reduced by magnetic geology. Electromagnetic induction sensors detect ferrous and nonferrous metallic objects and can be effective in geology that challenges magnetometers. EM has a limited depth of investigation due to faster signal fall-off over distance than a magnetometer. Care should be taken with data collection to ensure precise sensor location and 100% coverage, which directly impact classification performance.

Data-analysis requires feature extraction (i.e., size, depth, aspect ratio) from each of the detected anomalies and classification of these anomalies by assigning a confidence they are likely munitions or nonhazardous items. This process can be conducted in the commercially available Geosoft software package Oasis montaj as part of the UX-Analyze module.

**Why doesn't current practice classify munitions and clutter?**
Current practice is motivated by detection, which sets the data requirements and the decision process. There is no reason in principle that the decision process could not be modified to add a classification step. Current commercial instruments as they are deployed by the contractor community can collect data that may be used for classification. In most cases, doing so would require revisiting the data requirements. Classification is beginning to be adopted at munitions response sites and will become more common in the future.

**Can classification approaches identify individual items within a cluster of buried items?**
The current sensors and classification algorithms are effective only on individual isolated anomalies and do not allow reliable parameter extraction and classification of overlapping clustered anomalies. For the purposes of this study, anomalies were considered "isolated" if they were 0.5 m from the closest adjacent anomaly. For targets of interest the size of mortars, this has meant densities 100-200 anomalies per acre. For smaller, shallower targets with a smaller anomaly footprint, current methods may be applicable to higher anomaly densities on some sites.

The advanced sensors demonstrated at San Luis Obispo achieved some success with clustered targets but this remains an active area of research. High-density target centers would not be an appropriate location to apply classification. Current classification approaches can be applied to areas surrounding high-density target centers or where the high density layer had been removed where isolated anomalies are present

**How can I be confident in the results?**
The techniques used in this study to confirm technology performance can be utilized in operational classification applications. In this demonstration, examples of all known targets of interest were seeded in the survey area itself to confirm the detection and classification technologies used. Seeding is recommended in an operational classification application.

All items above the detection threshold were dug in this study to verify the performance of the sensors and the analysis methods. In an operational application, the ultimate objective is to dig only the targets of interest and leave the non-TOI. Random digging beyond the threshold is unlikely to be useful. The number of intact munitions present on most sites is very small, and randomly finding a misclassified TOI, if it is present, among overwhelming clutter is very unlikely. This will likely be effective in confirming cultural clutter and munitions-related scrap assignments rather than finding incorrectly classified munitions.

A small portion of the site could be dug completely to confirm the performance of the technologies and refine the classification algorithms. Alternatively, selected anomalies on the dig list could be sampled. The selection of these anomalies could be based on sampling various estimated parameters from the inversion step to confirm the results are physically reasonable.

**What if data collected for an anomaly cannot support reliable parameter extraction to identify its features?**
For all but one or two dig lists submitted, there were fewer than a handful of targets that were classified as "can't analyze." Those targets whose signal did not support reliable parameter estimation were added to the dig list. Demonstrators were generally successful at determining when reliable model fits were not achieved.

**How do I determine where to set the threshold once a dig list is created?**
There are several factors that can influence this decision. It is recommended the project team weigh the site objectives. Factors such as future land use should be considered. This decision process could be iterative and the point could be moved if information is gained from any validation digging that occurs.

**Is classification applicable to marine sites?**
There is no reason in principle that classification could not be performed on magnetometer or EM data collected at a marine site. The parameter-extraction process for magnetic data would be identical, and the physics that must be accommodated to account for differences in the EM signal is well understood and could be readily implemented. However, no system that can collect data to support classification has yet been demonstrated.

**Can you do classification with helicopter data?**
Classification from a helicopter platform is expected to be very limited. There are existing magnetometer systems that can take high-density, well-located data at low altitude from a helicopter platform. It is possible to analyze these data to obtain estimates of the target size and depth. The success will depend on the size of the targets of interest and the altitude that can be maintained. For EM systems, where the signal falloff is faster than for magnetometers, no system that collects data appropriate to classification has been demonstrated to date. This is expected to be more limited both because the signal falls off faster with the separation of the target and sensor and because the data requirements are more stringent for parameter extraction from EM data.

**Can you use this in the center of an impact area?**
Classification to date has been demonstrated only on isolated or mildly overlapping targets. In areas where the target density is very high and many signals will overlap, such performance has not been demonstrated. This is currently the subject of research, and progress is expected in classification of overlaps of two or three signals and in the analysis of a single strong target shielded by small surface

clutter. It is not expected that the current processes for classification will evolve to provide acceptable performance in the centers of impact areas, where individual signals are not separable.

**How long does it take to collect data and conduct analysis?**
Data collection can range from 0.5 to 10 or more acres/day depending upon if the system is man-portable or towed. If the data density required for classification is twice that required for detection only, the field deployment can be expected to approximately double. Data-analysis times will vary depending upon the number of anomalies detected and the presence of complicating factors such as geology. For the roughly 1500 anomalies analyzed in this demonstration, the data analysis required less than 1 person-week for the survey data and about 2 weeks for the advanced EM. Much of this time was devoted to quality checks and data handling. As the classification procedures become better defined, this is likely to decrease.

**How specialized is this? What contractor qualifications are required?**
Some of the algorithms used in this demonstration are available in a package called UX-Analyze, which runs in Oasis montaj. The two commercial contractors that analyzed the cart data in this program downloaded this package and used it to perform successful classification. No special qualifications are needed to run this program beyond those usually held by the geophysicist or data analyst typically involved in a project. It is necessary that the individual understand and be able to evaluate data quality, as well as assess whether reasonable answers are obtained.

**What emerging classification capabilities should I watch?**
Vastly improved classification performance was demonstrated in this study using a multi-axis data collection system called the MetalMapper. This sensor is one of several new generation EM sensors that gather substantially more information in the signals because the systems make measurements at a variety of angles over buried objects. While these sensor systems and associated data-analysis approaches are still in development, they show promise in significantly improving classification performance.

# REFERENCES


1. *Fiscal Year 2008 Defense Environmental Programs Annual Report to Congress*, https://www.denix.osd.mil/portal/page/portal/ARC/ARCFY2008

2. Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, *Report of the Defense Science Board Task Force on Unexploded Ordnance*, Washington, DC 2003.

3. Nelson, Kaye, and Andrews, *ESTCP Pilot Program,Classification Apporaches in Munitions Response,* 2008, http://www.estcp.org/Technology/upload/ESTCP-Class-Study-12-17-09.pdf.

4. NAEVA EM61-MK2 CART Final Report

5. Nova Research Inc., Demonstration Data Report, 2009 ESTCP UXO Classification Study San Luis Obispo, CA ESTCP MM-0744, Former Camp San Luis Obispo, Magnetometer and EM61 MkII Surveys, 2009.

6. SAIC, Man-Portable Simultaneous EMI and Magnetometer System San Luis Obispo Data Report. 2009.

7. SAIC, Data Collection Report, EMI Array for Cued UXO Discrimination, ESTCP MM-0601, Former Camp San Luis Obispo, 2009.

8. LBNL BUD Final Report

9. Geometrics MetalMapper Final Report

10. SAIC Analysis Report

11. Sky Research, Inc., Final Report, Feature Extraction and Classification of Magnetic and EMI Data, San Luis Obispo, CA, ESTCP MM-0504 Practical Discrimination Strategies for Application to Live Sites, in preparation.

12. Carin, Kennedy, Zhu, and Dasgupta, Final Report, 2009 ESTCP UXO Discrimination Study, San Luis Obispo, ESTCP Project # MM-200501 CA, in preparation.

13. SAIC and RML, Final Report LGP Discrimination and Residual Risk Analysis, MM-0811 2008-9 ESTCP UXO Classification Study, Camp Sibert and Camp San Luis Obispo, in preparation.

14. Parsons Analysis Final Report

15. IDA SLO Report

16. *Final Site Inspection Report, Former Camp San Luis Obispo, San Luis Obispo, CA*, Parsons, Inc., September 2007.

17. Nelson, Kaye, and Andrews, *Geophysical System Verification (GSV):A Physics-Based Alternative to Geophysical Prove-Outs for Munitions Response,* http://www.estcp.org/Technology/upload/GeoSysVerif-July-09-FINAL.pdf

18. Nelson, Bell, Kingdon, Khadr, and Steinhurst, *EM61-MK2 Response of Standard Munitions Items*, NRL/MR/6110—08-9155, 2008.