FINAL REPORT

ESTCP Pilot Program
Classification Approaches in Munitions Response
Camp Butner, North Carolina

JUNE 2011

Anne Andrews
Herbert Nelson
ESTCP

Katherine Kaye
ESTCP Support Office, HydroGeoLogic, Inc.

ESTCP

TABLE OF CONTENTS

FIGURES

TABLES

ACRONYMS

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
HRR	Historical Record Review
IDA	Institute for Defense Analyses
IMU	Inertial Measurement Unit
IVS	Instrument Verification Strip
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

ACKNOWLEDGEMENTS

This demonstration in the ESTCP Classification Pilot Program would not have been possible without the assistance of numerous individuals. We would like to acknowledge Shelley Cazares, Mike Tuley, and Elizabeth Ayers 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, U.S. Army Corps of Engineers, Sacramento District
Harry Craig, U.S. EPA Region 10, Oregon Operations Office
Bryan Harre, Naval Facilities Engineering Service Center, Port Hueneme, CA
Robert Kirgan, Army Environmental Command
Ray Livermore, U.S. Army Corps of Engineers, Wilmington
Doug Maddox, US Environmental Protection Agency
Martha Morgan, North Carolina Department of the Environment and Natural Resources
Doug Murray, Naval Ordnance Safety and Security Activity
Andrew Schwartz, U.S. Army Corps of Engineers, Huntsville
Stephen Sterling, California Department of Toxic Substances Control
Jeff Swanson, Colorado Department of Public Health and Environment
Jon Ussery, AFCEE Technical Directorate
Ken Vogler, Colorado Department of Public Health and Environment
Dan Ward, California Department of Toxic Substances Control
Roger Young, U.S. Army Corps of Engineers, Huntsville (retired)

We would like to credit the technology demonstrators.

- CH2M HILL, Dartmouth, Geometrics, NAEVA Geophysics, Inc, Parsons, SAIC, and Sky Research for classification data analysis;
- Geometrics and Sky Research, Nova Research, and NAEVA Geophysics Inc, for data collection for MetalMapper, TEMTADS, and the EM61 cart system respectively.

For program support, we acknowledge

- Parsons for seeding and validation excavations;
- Bob Selfridge from the USACE for supplying inert rounds used for the seeding.

Finally, ESTCP thanks Ms. Louise Johanson and Mr. and Mrs. William Rouse, owners of the land involved in this demonstration, for their patience and cooperation.

EXECUTIVE SUMMARY

Munitions response is a high-priority problem for the Department of Defense (DoD). Approximately 3,800 sites, comprising tens of millions of acres, are suspected of contamination with military munitions, which include unexploded ordnance (UXO) and discarded military munitions. The Military Munitions Response Program (MMRP) is charged with characterizing and, where necessary, remediating munitions-contaminated sites.

When a site is remediated, 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 99% of objects excavated during the course of a munitions response are found to be nonhazardous items. 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 this expense could be avoided and the available funding applied to more sites.

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 are related to the physical attributes of the object that resulted in the signal, such as its physical size, aspect ratio, wall thickness, and material properties. 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.

The Environmental Security Technology Certification Program (ESTCP) is charged with demonstrating and validating innovative, cost-effective environmental technologies. ESTCP recently initiated a Classification Pilot Program, consisting of demonstrations at a number of sites, to validate the application of a number of recently developed technologies in a comprehensive approach to munitions response.

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.

The first two demonstrations in this series, at former Camp Sibert, AL, and former Camp San Luis Obispo, CA, showed good classification ability from all demonstrators. Camp Sibert was deliberately chosen as an easy site but Camp San Luis Obispo had four known targets of interest prior to the study including 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets and more difficult terrain. During the San Luis Obispo demonstration, three unexpected munitions were excavated.

ESTCP sponsored a third study in 2010 on a range at the former Camp Butner, NC, expected to contain 37-mm projectiles. Many MMRP sites contain this munition and it has proven to be difficult to classify using commercial sensors and traditional analysis methods. The range chosen is also potentially contaminated with much larger munitions items, 105-mm and 155-mm projectiles, making this a stringent test of the classification process.

Both survey and cued data were collected at Camp Butner. The primary survey instrument was the EM61-MK2, the most commonly used sensor on munitions response projects. The anomalies detected from these data were used as the primary anomaly list for the demonstration. Cued data were collected over these anomalies using two of the advanced EMI sensors, TEMTADS and MetalMapper.

Analysts from a number of firms used these data to classify each anomaly. In all cases, the process involved extracting parameters from analysis of a data chip corresponding to each anomaly and using these parameters to label the item as either a munition, harmless clutter, or unable to decide. For some of the analyses of the EM61 survey data, these parameters were data-based parameters such as the decay rate of the measured signal. For other analyses of the survey data and all the analyses of the cued data, target-based parameters that relate to the physical size of the item, material properties, and wall thickness were derived from model fits to the data and used for classification.

Each analyst prepared a ranked anomaly list with the anomalies that were classified as high-confidence clutter at the top, followed by those anomalies for which the analyst was unable to make a decision, then the anomalies classified as high-confidence munitions. In some cases the analysis failed for a small number of anomalies due to data problems; these anomalies must be dug and are placed at the bottom of the list.

Analyses were scored based on the demonstrator's ability to eliminate nonhazardous items while retaining all detected targets of interest. The results are presented as receiver operating characteristic (ROC) curves, examples of which are shown in Figures ES–1 and ES-2. 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 any training digs provided to the 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.

Analysis of the EM61-MK2 data, Figure ES-1, was not particularly successful at this site; all demonstrators missed a number of munitions after their threshold and only correctly identifying about 10% of the clutter once they achieved 100% identification of the munitions present. There were several differences from the previous demonstrations that led to this result. The small size of many of the targets at Butner resulted in low signal-to-noise anomalies in the EM61 data which, coupled with the high density of anomalies at this site, made it difficult to extract reliable parameters from many of the anomalies. In addition, many of the clutter items at Butner consisted of fragments from larger projectiles which were roughly similar in overall size and wall thickness to the 37-mm projectiles. Thus, neither of the parameters available from the EM61-MK2 data was useful as a discriminant at Camp Butner.

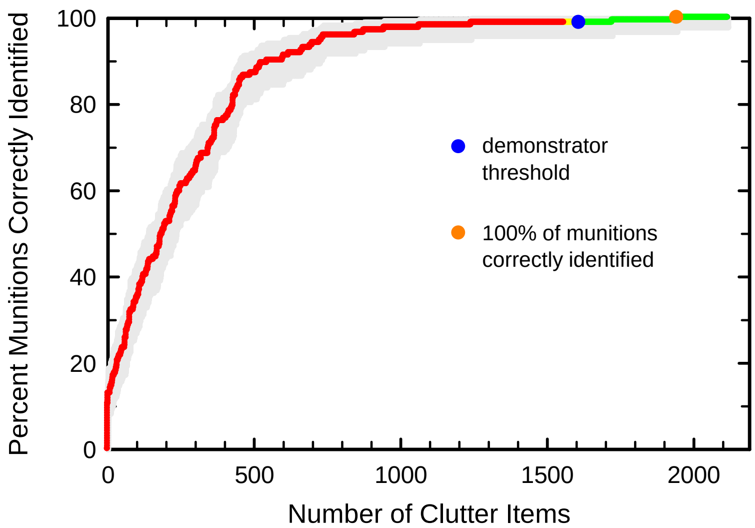

Figure ES-1. Example ROC curve resulting from analysis of the EM61-MK2 survey data. About 90% of the clutter must be dug to identify all of the munitions on the site.

At other sites in this series the EM61-MK2 has been able to successfully eliminate as many as one half of the clutter at the site. This site is more typical of a "hard" classification site and the results here indicate the limitations of the commonly-used sensors for this use.

Dramatically better results were obtained using the cued data from the advanced sensors. An example using the TEMTADS data is shown in Figure ES–2. These analysts were able to correctly identify almost 95% of the clutter while retaining 100% of the munitions.

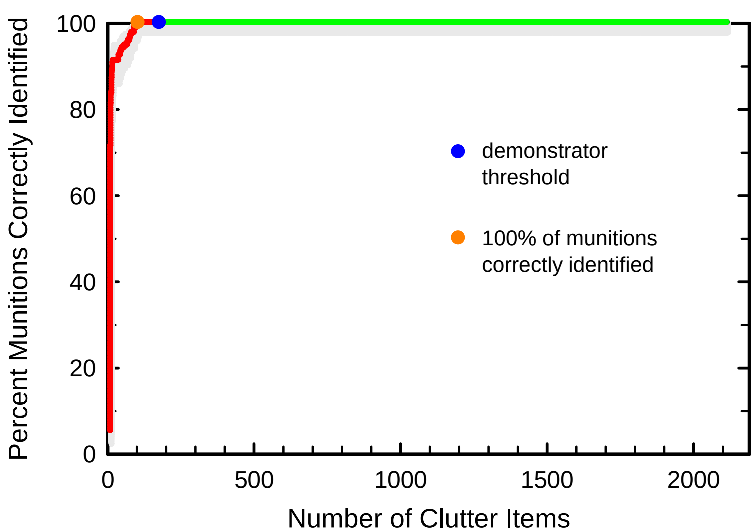

Figure ES-2. Example ROC curve resulting from the analysis of the TEMTADS cued data. Almost 95% of the clutter was correctly identified while retaining all the munitions on the site.

Not all analysts and methods were able to achieve these impressive results using the advanced sensor data although all but a handful were able to correctly identify more than 50% of the clutter while missing no targets of interest. A primary objective of the remaining demonstrations in this series will be to identify ways for all analysts to perform up to the potential demonstrated by the best performers.

The motivation for applying classification in munitions response is to more effectively use the available resources: if the digging of non-munitions targets is minimized, then the limited resources of the munitions response program can be applied to clean up more land more quickly. We developed a simple cost model with realistic assumptions for production costs of various model elements in the report describing the San Luis Obispo demonstration. Using that same model here, we have shown how the savings from the use of classification can be expected to increase the productivity of the MMRP program. If 70% of the clutter can be confidently identified, the area remediated for a fixed budget will increase by at least a factor of 1.75. If the classification efficiency can be increased to 90%, the area remediated on a fixed budget will increase by a factor of 2.4.

1 INTRODUCTION

1.1 BACKGROUND

Munitions response is a high-priority problem for the Department of Defense (DoD). Approximately 3,800 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 99% 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 cleanup 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 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. Electromagnetic Induction data are typically fit to a three-axis polarizability model that can yield parameters that 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 methodology 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, consisting of demonstrations at a number of sites, to validate the application of a number of recently developed technologies in a comprehensive approach to munitions response. This report summarizes the results of the third of these demonstrations at the former Camp Butner, NC.

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 prepared 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 THE FIRST TWO DEMONSTRATIONS

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 EMI 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. The Berkeley UXO Discriminator (BUD) is a next generation sensor designed to maximize classification information. It achieved nearly perfect results at Camp Sibert. More information on the first phase of the program approach and results is available in the ESTCP Program Office Final Report. (Ref. 3)

A hillside range at the former Camp San Luis Obispo, CA was selected for the second of these demonstrations. 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. The site is open, with good sky view, but the terrain is more challenging than that at Camp Sibert.

As in the first demonstration, the San Luis Obispo 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.

The demonstration was scored based on the demonstrator's ability to eliminate nonhazardous items while retaining all detected targets of interest (TOI) defined as UXO and related items that the site team decided must be removed from the site. The results are presented as receiver operating characteristic (ROC) curves, an example of which is shown in Figure 1-1. 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.

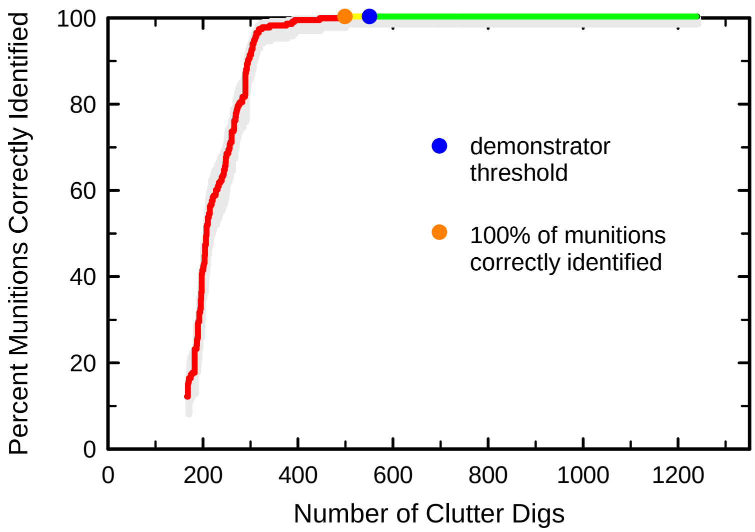

Figure 1-1. Receiver operating characteristic curve resulting from analysis of the EM61-MK2 CART data collected at former Camp San Luis Obispo.

The ROC curve in Figure 1-1 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 commonly-used EM61-MK2 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, she 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 1-2 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.

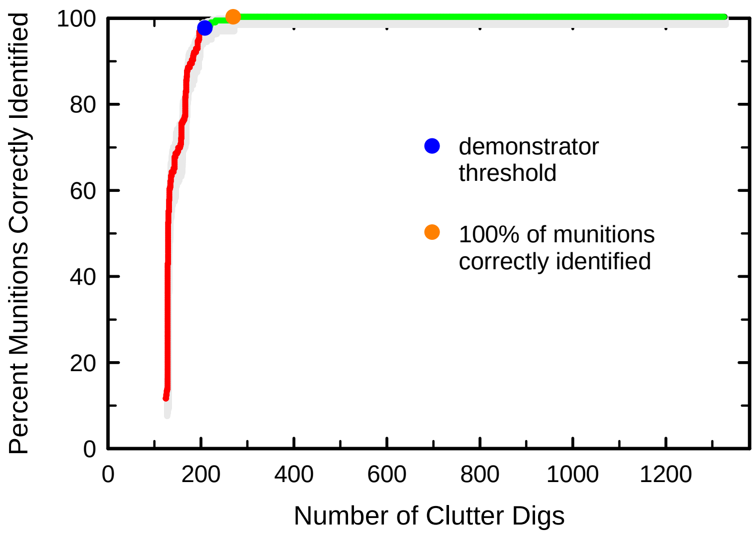

Figure 1-2. ROC curve resulting from analysis of the MetalMapper data collected at former Camp San Luis Obispo.

The Program Office Final Report describing this demonstration in detail is available as Reference 4.

1.5 ABOUT THIS REPORT

ESTCP sponsored a third study in 2010 on a range at the former Camp Butner, NC, expected to contain 37-mm projectiles. Many MMRP sites contain this munition and it has proven to be difficult to classify using commercial sensors and traditional analysis methods. The range chosen is also potentially contaminated with much larger munitions items, 105-mm and 155-mm projectiles, making this a stringent test of the classification process.

This report is intended to provide an overview of the key results from the third 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. More information about the entire demonstration and these topics in particular may be found in the individual demonstrator reports (Refs. 6-16) and an independent performance assessment by the Institute for Defense Analyses. (Ref. 19)

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 BUTNER

A range at the former Camp Butner 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 and item "size" was an effective discriminant. At former Camp San Luis Obispo, there were four targets of interest expected from historical records: 60-mm, 81-mm, and 4.2-in mortars and 2.36-in rockets. Three additional munitions types were discovered during the course of the demonstration. This site is expected to be contaminated with 37-mm projectiles as well as larger items which introduces another layer of complexity into the process.

2.1 SITE HISTORY AND CHARACTERISTICS

The site description material reproduced is here is taken from the recent EE/CA report (Ref. 5). More details can be obtained in that report. The former Camp Butner Site is a 40,384 acre site located approximately 15 miles north of Durham, partly in Durham, Granville, and Person Counties, North Carolina. The demonstration was conducted in the northern part of area defined as "Area A" in Reference 6. An aerial photo of the initial demonstration area is shown in Figure 2-1.

On February 12, 1942, the War Department issued an order for the acquisition of land near the Durham, North Carolina area to be used as a training and cantonment facility during World War II. At the time, the land use was primarily low density residential in nature. The original authorization was for 60,000 acres of real property; however, the actual amount of land acquired was approximately 40,000 acres. Although the Camp was considered active until 1946, its use for training exercises lasted only for approximately 18 months from early 1942 to June 1943.

The construction of Camp Butner began February 25, 1942 and proceeded at a high rate until its completion in August of the same year. The camp was primarily established for the training of infantry divisions (including 78th, 89th, and 4th) and miscellaneous artillery and engineering units. Camp Butner was designed to house up to 40,000 troops. In addition to infantry training, the site was the location of the one of the Army's largest general and convalescent hospitals and the War Department's Army Redeployment Center.

The primary mission of Camp Butner was to train combat troops for deployment and redeployment overseas. There were approximately 15 live-fire ammunition-training ranges encompassing a combined approximately 23,000 acres. Other training ranges included a grenade range, a 1000-inch range, a gas chamber, and a flame-thrower training pad. There was also an ammunition storage area. In September of 1943, the first Prisoners of War (POWs) arrived at the camp.

On January 31, 1947, the War Department declared Camp Butner excess. At that time, the Federal government was negotiating with the State of North Carolina for a lease on the hospital. The State was interested in using the hospital as a State mental hospital. The State was also negotiating the purchase of 10,000 acres to be used to support the hospital. On November 3, 1947, the State purchased the hospital, later named the John Umstead Hospital, and 1,600 acres of the cantonment area to be used for various projects and agricultural development. The North Carolina National Guard was conveyed 4,750 acres of the former Camp Butner for training purposes.

After Camp Butner was declared surplus, dedudding operations were initially conducted in 1947 and continued through 1950. The Recapitulation Dedudding Report presented in the ASR stated that

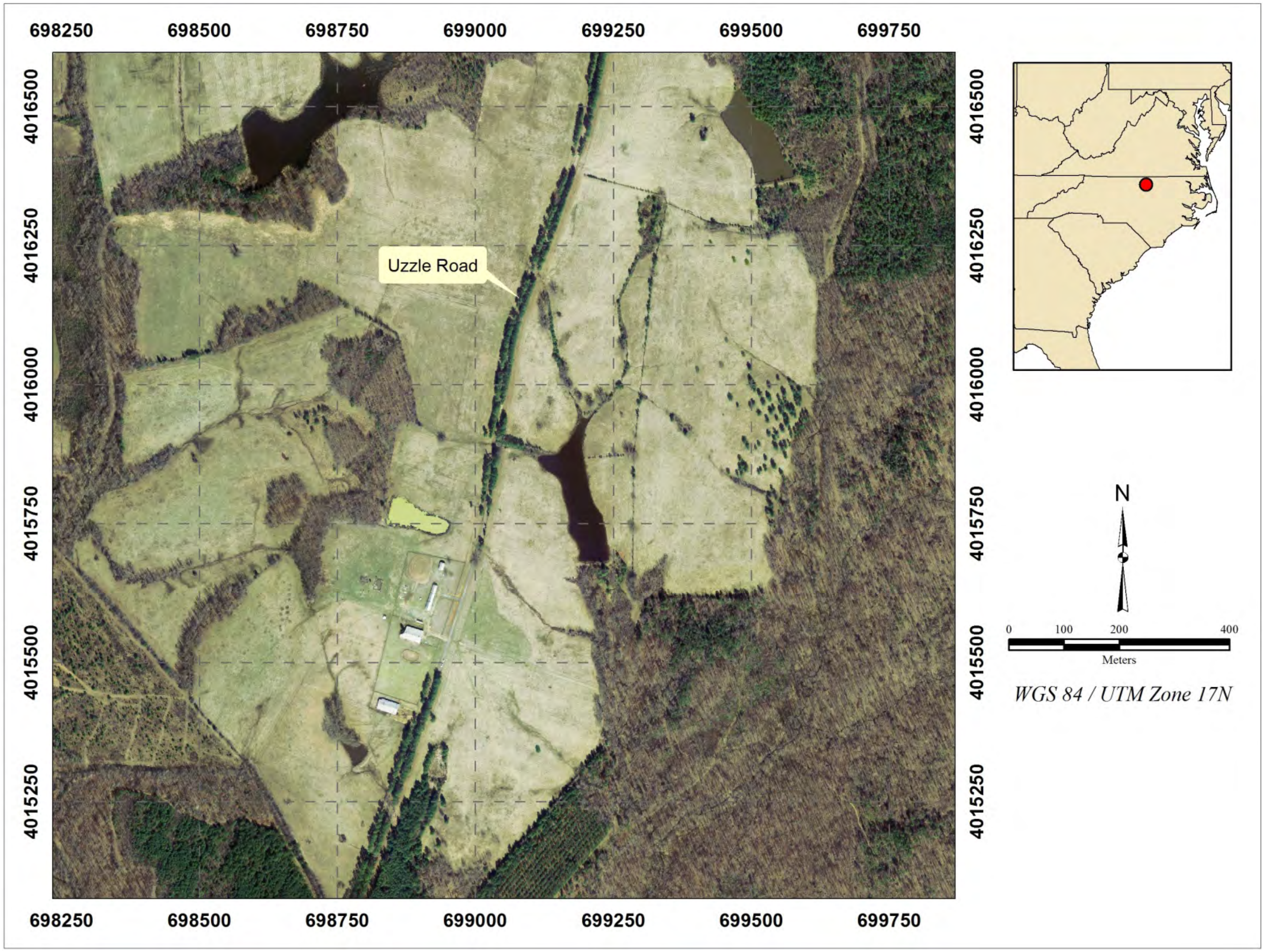

Figure 2-1. Aerial photo of a portion of former Camp Butner showing the access road and the approximate location of the site in North Carolina.

1366 UXO/OE items had been discovered and destroyed by the completion of dedudding operations. Six areas were identified during dedudding inspections as warranting land restrictions to 'surface use only' due to the number of HE duds found. Among these six was the one termed "Area A," an artillery impact area, which contains the site of this demonstration. Much of the property was sold back to the original owners, with provisions outlined in the property deed restricting land use to 'surface use only'.

Periodic inspections of the six areas with land restrictions were conducted between 1958 and 1969. During the inspections and removal of munitions from the restricted areas other property owners identified munitions for disposal that had been found in unrestricted areas. Munitions including rifle grenades, 2.36-inch rockets, 37-mm, 40-mm, 81-mm mortar, 105-mm, 155-mm, and 240-mm projectiles have been found in Area A during these period inspections. In the immediate vicinity of the area used for this demonstration, 37-mm, 105-mm, and 155-mm projectiles have been found.

The area chosen for this demonstration is open with good sky view, Figure 2-2. The ground is level with few interfering trees. It abuts Uzzle Rd. so site access is good.

Figure 2-2. Photograph of the site for this demonstration.

2.2 DEMONSTRATION PREPARATION

Several activities occurred prior to data collection to ensure the resulting data would support a successful demonstration. These activities included EM61 transects to define the initial area of interest and guide selection of site characterization grids; intrusive investigation of a 100-ft' x 100-ft grid to provide site-specific information to guide the selection of targets of interest for the site, establish the depth distributions required for the seed items, and be available for use by the demonstrators; surface clearance of the site; EMI survey of approximately 30 acres to guide selection of the 10-acre demonstration site; and emplacement of seeds.

2.2.1 EM61 Transects Surveys

Prior to selection of the location for this demonstration, initial EM61-MK2 transects were collected on several parcels (Figure 2-3) in October 2009. In addition, total coverage surveys were conducted over three 100-ft x 100-ft grids. The transect data were used to calculate rough anomaly densities for each parcel to be used in the selection of the final demonstration area. An anomaly selection threshold of 20 mV on the sum channel, roughly corresponding to the minimum signal expected from a 37-mm projectile at 30 cm depth, resulted in the anomaly densities listed in Table 2-1.

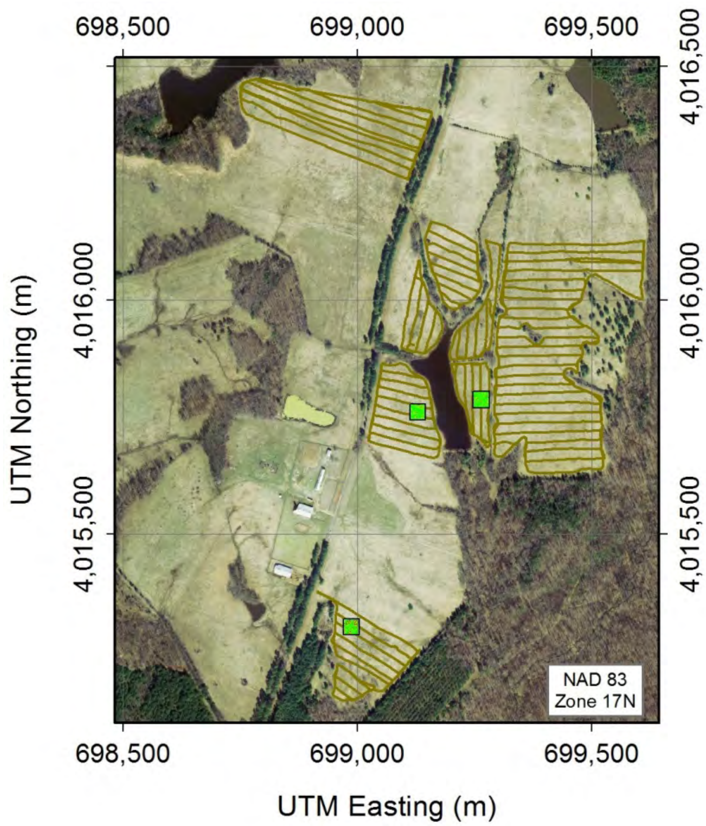

Figure 2-3. Location of initial transect survey lines and three potential characterization grids.

Table 2-1. Estimated anomaly densities for the three parcels mapped with transects

Area	Targets	Acres Mapped	Anomalies per Acre
Northern Parcel	179	1.20	149
Middle Parcel	1177	4.28	275
Southern Parcel	363	0.62	585

2.2.2 Site Characterization Grid

One of the 100-ft x 100-ft grids discussed above was excavated in December 2009 to provide information about the types and depths of munitions and clutter on the site. The grid was selected in the southern parcel which had the highest anomaly density to maximize the information obtained. A map of the EM61-MK2 survey data of this grid is shown in Figure 2-4.

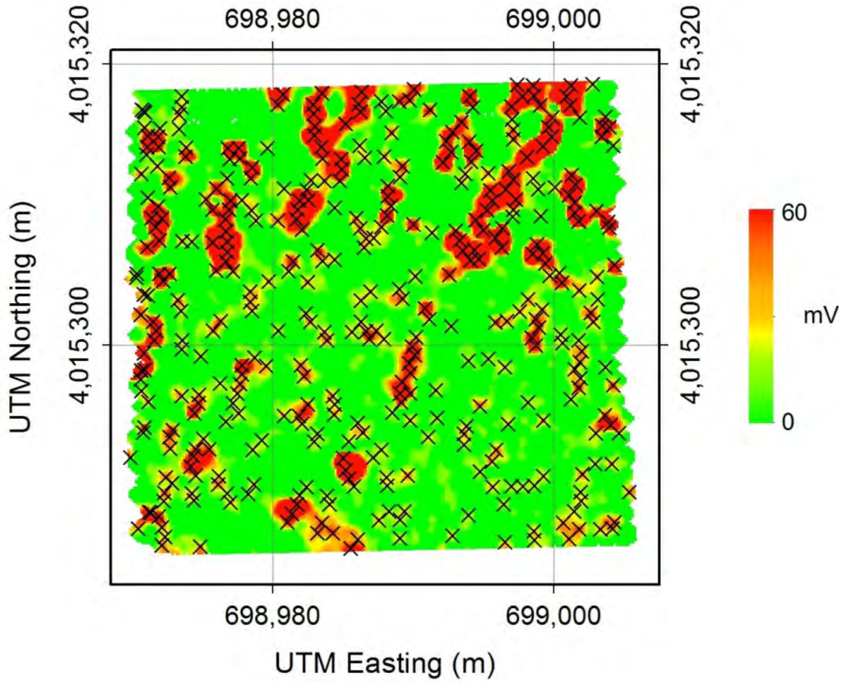

Figure 2-4. EM61-MK2 survey data from the southern grid at the Camp Butner site. Anomalies identified in the intial survey are indicated with x's.

A total of 404 anomalies were identified in the initial EMI data in this grid. The intrusive team was instructed to investigate these initial contacts and then remap the grid and investigate any remaining anomalies. Weather conditions prevented the team from finishing the intrusive investigation in the time allotted; only 65% of the grid was completed. The items were separated into classes as shown in Table 2-2, and examples of the excavated items are shown in Figure 2-5.

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

	Number of Items Recovered			Depth Range (cm)	
Class	**Initial Anomalies**	**Remapped Anomalies**	**Total**	**Initial Anomalies**	**Remapped Anomalies**
Intact Munitions	0	0	0		
Munitions Debris	295	143	438	5 - 45	0 – 50
Cultural Debris	91	64	155	5 - 35	0 – 50
Hot Soil/No Contact	7	8	15		
Total	393	215	608		

Figure 2-5. Examples of items recovered during the excavation of the site characterization grid.

The UXO technicians on the intrusive team reported that "the bulk of the fragmentation and fuze components appeared to be from 105mm and 155mm projectiles, although one piece of 37 mm projectile was recovered from the grid." This observation confirmed the historical information on this site. The identities and depths of all items recovered from the site characterization grids were provided to the demonstrators as background information about the site.

Based on the target density in the characterization grid, a transect survey was conducted on the 14-acre parcel directly north of the investigated grid using the same procedures and thresholds as the October 2009 mapping. 373 targets above 20 mV on the sum channel were identified on 1.5 linear miles of transect data.

2.2.3 Define the Demonstration Site

Four areas totaling approximately 30 acres were chosen for surface clearance and initial EM61 mapping, Figure 2-6. The results from this mapping were used to select the final 10-acre demonstration site and guide the emplacement of inert seed items.

The final 10-acre demonstration area is shown in Figure 2-7 subdivided into 44 30-m x 30-m grids established by the EM61 contractor, NAEVA Geophysics, Inc. (NAEVA). The two survey instruments, EM61-MK2 cart and MetalMapper, covered the entire 10-acre area. Because of the high anomaly density across this site, program resources limited the intrusive validation efforts to a subset of the site containing approximately 2500 anomalies. This sub-area is denoted as the "Cued Area" in Figure 2-7; the cued sensors were only deployed to anomalies within this sub-area and only targets in this area were dug and scored.

2.2.4 Seeding the Cued 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 4.5-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

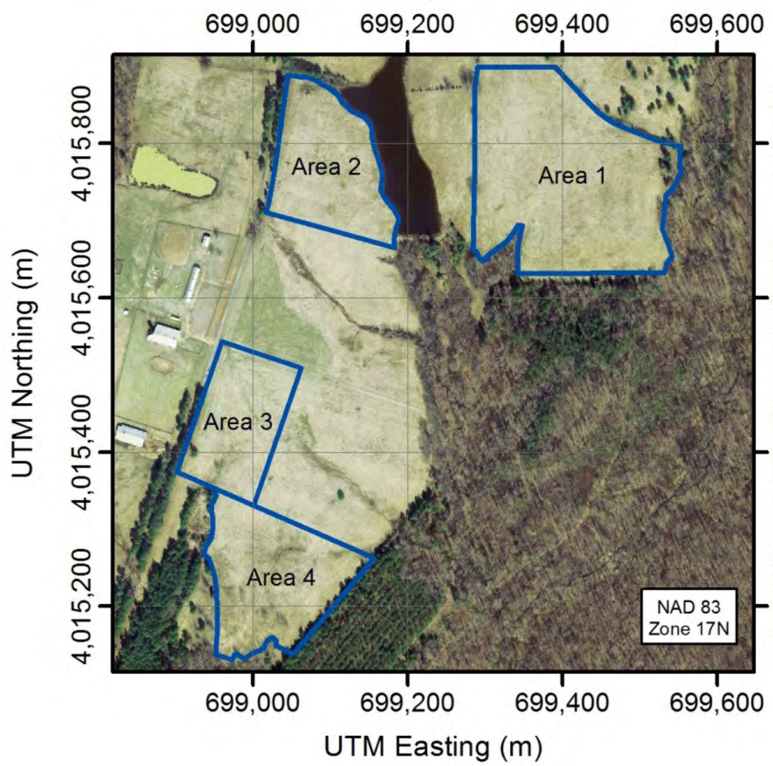

Figure 2-6. Former Camp Butner areas chosen for surface clearance and initial EM61 mapping.

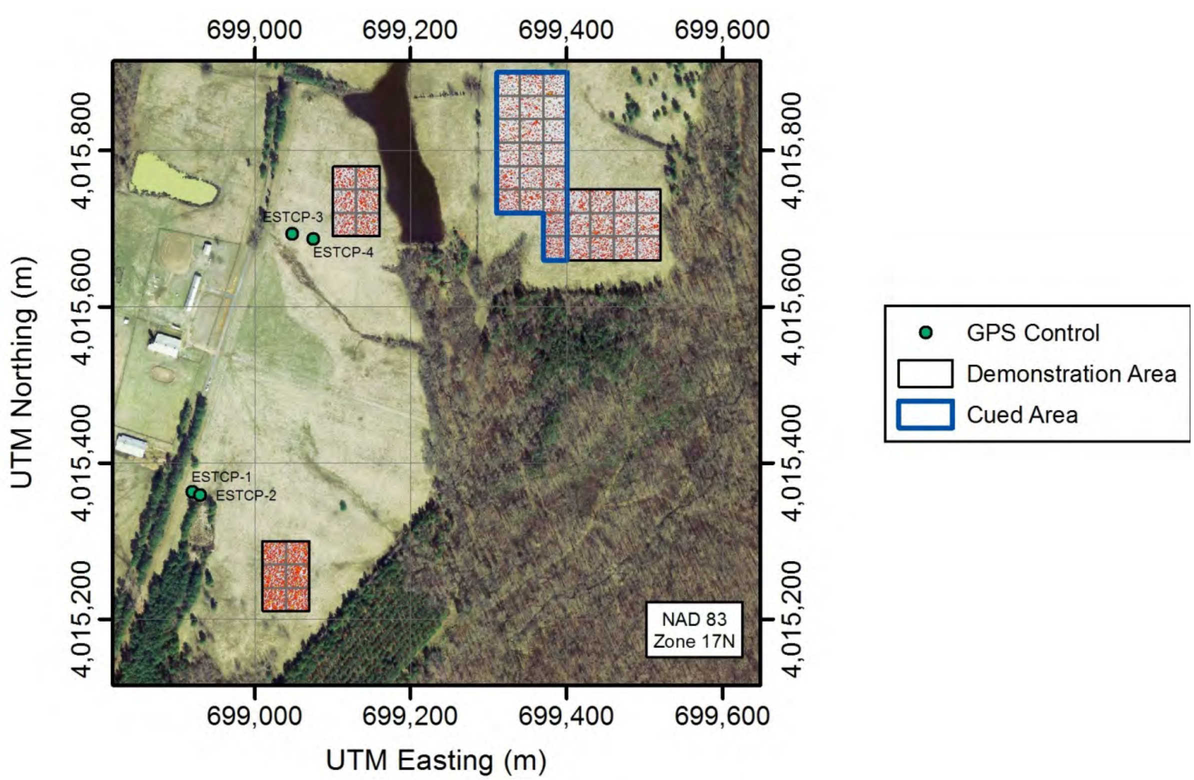

Figure 2-7. Final Camp Butner demonstration area showing the GPS control points established, the EM61-MK2 cart data, and the portion of the site chosen for cued investigation.

extent possible, items recovered from other live ranges were used as seeds; this was not possible for all items however.

A total of 160 inert items were seeded in the cued area. The seeds comprised items expected at the site, 37-mm projectiles and 105-mm projectiles, as well as M48 fuze simulants which were identified by the Advisory Group as hazardous items that must be removed if present. The identity and depth distribution of the seeds is detailed in Table 2-3. Seed locations were determined by examining the initial EM61 survey to identify locations where the seed anomaly would not overlap any above-threshold anomaly on the site. This is an artificial constraint on seed location that will not be repeated in subsequent demonstrations. The exact (*x, y*) location, depth to the center of the target, and orientation were recorded for each emplaced item and the item was photographed before burial. These details were unknown to the demonstrators. Only *in situ* clutter was used in this study, and no additional cultural clutter, munitions-related scrap, or geology was seeded.

Table 2-3. Details of inert seed items in the cued area.

Item Description	Number Emplaced	Depth Range (cm)
37-mm projectile	110	10 – 30
M48 fuze stimulant	23	10 – 30
105-mm projectile	13	20 – 60
105-mm HEAT	13	20 – 60
fuze from inert 105-mm projectile	1	15

2.2.5 Instrument Verification Strip and Training

A quiet area on the west side of the cued area was located to establish an instrument verification strip (IVS) to be used for daily verification of proper sensor operation and a training pit to be used to collect sensor data for algorithm training. Details of the contents of the IVS are given in Table 2-4.

Table 2-4. Details of the Instrument Verification Strip.

Item ID	Description	Depth (m)	Inclination	Azimuth (° cw from N)
1001	shotput	0.45	N/A	N/A
1002	37-mm projectile	0.15	Horizontal	0
1003	small ISO	0.30	Horizontal	0
1004	small ISO	0.15	Horizontal	0
1005	small ISO	0.30	Horizontal	0
1006	shotput	0.45	N/A	N/A

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 three 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 for the two survey data sets was performed in a standard fashion as dictated by the ESTCP Program Office. 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 available to be used as training data or test data. Demonstrators could choose to perform their classification based on no site specific training data, a standard set of training data collected by digging all the targets in one grid, or a demonstrator-requested training data set. If requested, all 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. At Camp Butner, the project team determined that targets of interest that must be removed would include:

- seeded munitions,
- intact munitions recovered at the site, both live and inert, and
- fuzes from the large projectiles with booster tubes attached.

One hundred sixty items were seeded and all are TOI. Seven 37-mm projectiles were recovered that were classified by the UXO specialists as UXO and four more were found that were classified as munitions debris by the intrusive team because they were empty. These latter four projectiles were intact (Figure 3-1) so they were deemed TOI for this study.

Figure 3-1. Two "empty" 37-mm projectiles recovered in this demonstration.

3.3 DATA COLLECTION

The classification pilot study consisted of several combinations of data-collection platforms and analysis approaches, ranging from careful application of a commercial EM61 survey instrument to two prototype systems specially designed to maximize 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. 6-8).

- SURVEY MODE: The cart-mounted EM61-MK2 was deployed to collect data on 100% of the site, called SURVEY mode.
- CUED MODE: Two sensors, TEMTADS and MetalMapper, were deployed to collect data at the locations of individual anomalies detected by the EM61 Cart.
- SELF-CUED MODE: The MetalMapper (MM) is intended to operate in both survey and cued mode. MM performed a detection survey and, in addition to the anomalies cued by the EM61 Cart, collected cued data over all the distinct anomalies it detected.

Table 3-1. Summary of Data Collection at Camp Butner.

Survey	Cued from EM61 Cart Data	Self-Cued
EM61 Cart	TEMTADS	MetalMapper
MetalMapper	MetalMapper	

3.3.1 Survey Mode

In survey mode, the cart-mounted EM61-MK2 covered 100% of the site. Data were acquired by running the sensor in closely spaced lines, similar to the pattern of a lawnmower cutting grass. The site was divided into 30-m x 30-m grids, Figure 2-7, and data collected one grid at a time.

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.

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 to the bottom of the coil housing. Figure 3-2 shows this system collecting data at Camp Butner. (Ref. 6) Data were collected by NAEVA.

Figure 3-2. EM61-MK2 cart deployed at Camp Butner.

3.3.2 Cued Data

Two sensors were used to collect cued data at the locations of anomalies detected by the EM61 cart. These purpose-built EMI systems were designed to collect sufficient data to fully characterize the EMI signature from a single measurement location. Approximately 2,300 anomalies in the EM61 cart survey data met the anomaly selection criteria; the TEMTADS and MetalMapper systems collected data over all of them.

TEMTADS. The TEMTADS, shown in Figure 3-3, 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 and it is towed by the same vehicle used for all the MTADS systems. The sensor height above ground is variable depending on the targets of interest and site conditions; at this demonstration it was 17 cm above the ground surface. Three cm-level GPS units are used for navigation, geolocation and orientation. Data were collected by Nova Research. (Ref. 7)

MetalMapper. The MetalMapper (MM), shown in Figure 3-4, 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 and was used in a sled configuration either mounted to a front loader tractor or pulled by an ATV. Centimeter-level GPS is used for navigation and geolocation and an IMU is used to measure platform orientation. 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 and Sky Research (Sky) (Ref. 8) with slightly different sensors; details of the two sensor configurations are attached to the report. For both cued and dynamic surveys, the Geometrics sensors was 17 cm above the ground while the Sky version was 7 cm above the ground.

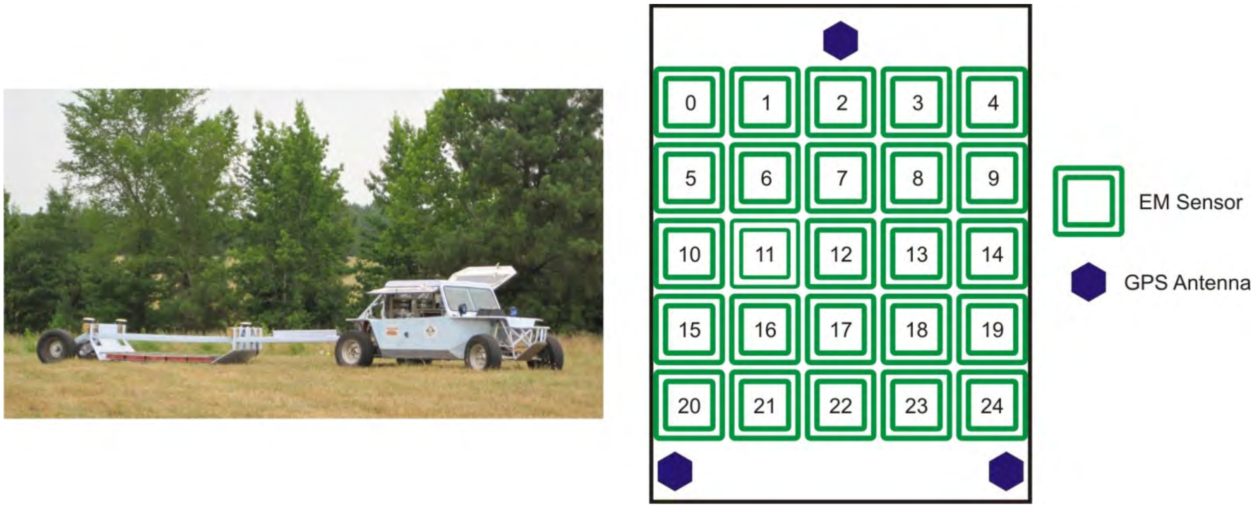

Figure 3-3. Photo (left) and schematic (right) of TEMTADS. Total dimension of the array is 2 m square.

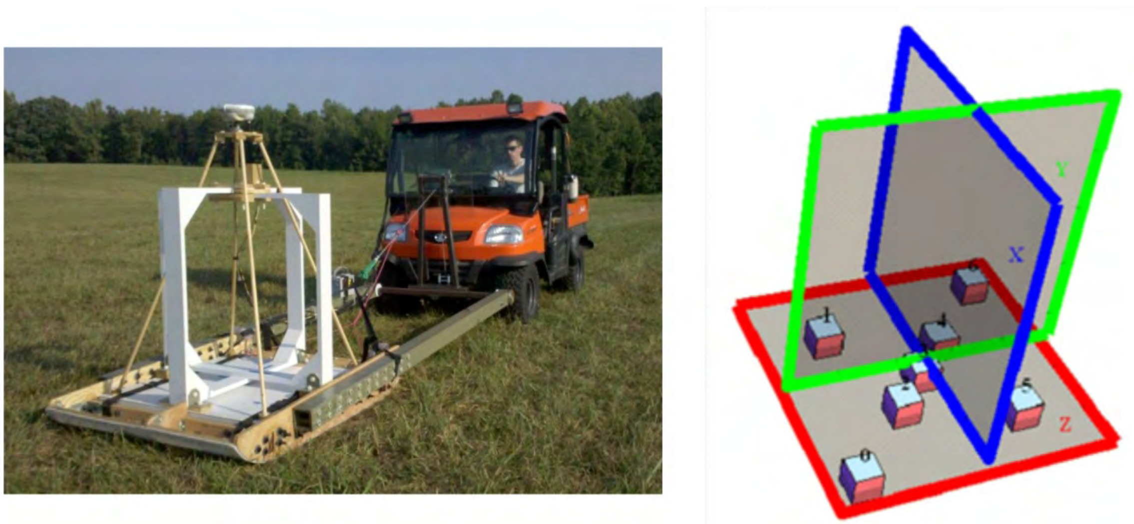

Figure 3-4. Photo (left) and schematic (right) of MetalMapper. The transmitters form a 1-m cube.

3.3.3 Self Cued

The MetalMapper is designed to be a stand-alone survey and cued detection system, and collected survey data at Camp Butner as well as the cued data discussed above. 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 data was only used by one analysis team as part of this demonstration. Several other teams are analyzing these data after completion of the demonstration.

Figure 3-5. MetalMapper configured to collect survey data at Camp Butner.

3.4 CLASSIFICATION APPROACHES

3.4.1 Processing Flow

The basic flow of the classification approaches is summarized in the flow chart in Figure 3-6. A geophysical survey of the area was performed and anomalies identified as described in Section 4. Cued data were collected at the locations of these anomalies by both TEMTADS and MetalMapper. Classification demonstrators could analyze the survey data, the cued data, or a mix of the two; for some anomalies in some analysis schemes decisions could be made from the survey data so there was no need to bear the cost of a cued measurement. The data corresponding to each anomaly were 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.

Most classification algorithms require some training to select the parameters or features that are most useful for classification and set thresholds in the decision process. For this demonstration, the analysts had the choice of using training data previously collected at other sites only, supplementing those data with data from the IVS and training pit, or adding training data obtained from excavation of a limited number of anomalies from the site. The on-site data could consist of a standard training set made up of all the anomalies in one grid or a custom set specified by the demonstrator. For those demonstrators that chose to use on-site training data, the anomaly list was divided into training and blind testing sets; for those who did not choose on-site training, the test set consisted of all selected anomalies. After training, the decision process for each algorithm was finalized and documented, and the demonstrators provided ranked dig lists for the blind test set.

3.4.2 Parameters Based on Geophysical Models

Multiple groups demonstrated processing approaches based on geophysical models. The basic classification method 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 but one of the geophysical models are based on a dipole approximation. The Dartmouth analyst used a model that more accurately captures the 3-D properties of the targets. Results from this analysis will be discussed below.

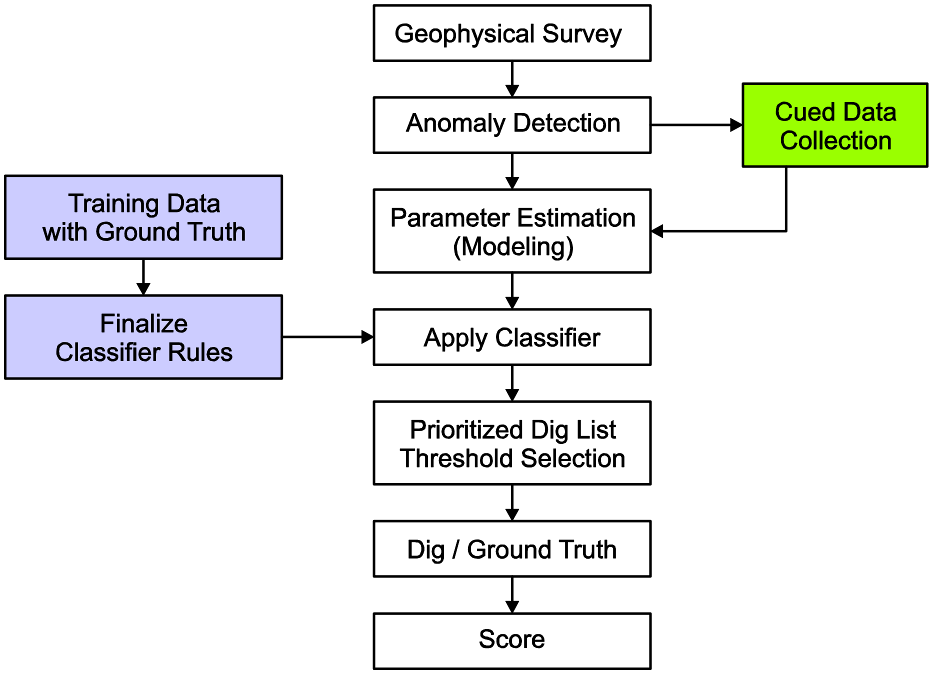

Figure 3-6. Work flow of Camp Butner classification demonstration.

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 3-7. 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.

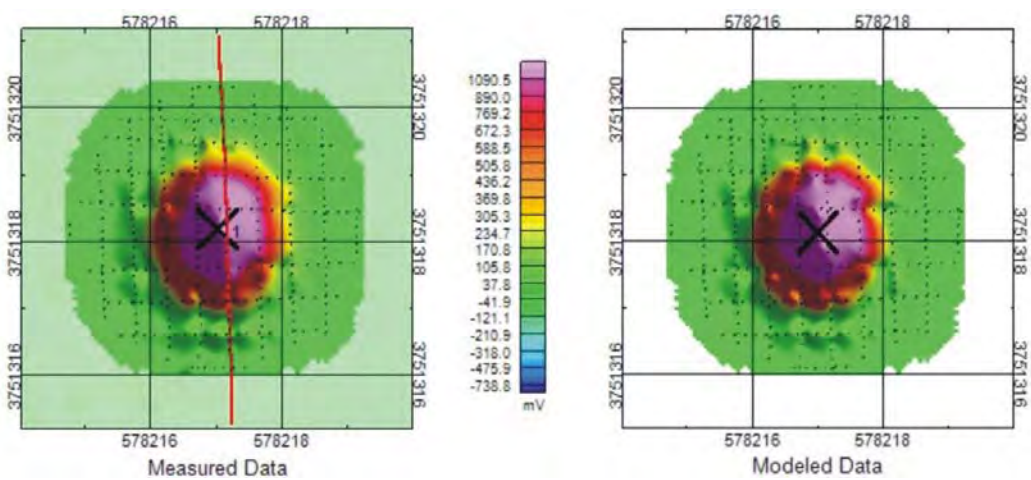

Figure 3-7. Example of a measured EM61-MK2 data chip of an anomaly (left) and the corresponding model result (right). The axes labels refer to distance in meters.

The cued sensors, TEMTADS and MetalMapper, collect sufficient data at a single spatial location to support model-based parameter estimation using either the standard analysis packages mentioned above or custom software developed by the system developers.

Some of the parameters that were considered included:

- the electromagnetic polarizabilities, which relate to the object's physical size and aspect ratio, and
- the electromagnetic decay constants, which relate to the object's material properties and wall thickness.

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 a very few 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

Several of the groups analyzing the EM61-MK2 survey data found that signal decay rates taken directly from the measured data provided some classification value. These "decay rates" were calculated as the difference in signal amplitude between various pairs of the EM61-MK2 sampling gates either for the highest amplitude sounding in the anomaly or averaged over some high-signal subset of the anomaly.

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.

3.5 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		
3	69	…		
…	…	…	High confidence NOT TOI	
…	…	…		
…	…	…		Threshold
…	…	…	Can't make a decision	
…	…	…		
…	…	…		
…	…	.03	High confidence TOI	
…	…	.02		
	…		Can't analyze	

- 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 3-8. The colored regions on the plot in Figure 3-8 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 correctly classifying the detected munitions items, versus the number of non-TOI. A perfect classifier would correctly identify 100% of the munitions and no clutter. We have modified the traditional ROC curve slightly to reflect both the TOI and non-TOI dug for training. This is done to account for the fact that different methods used different amounts of training data.

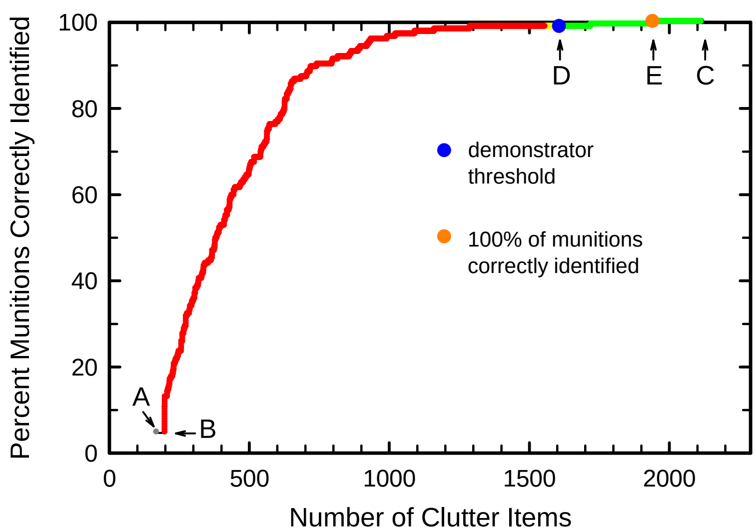

Figure 3-8. Example receiver operating characteristic curve.

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

- **A**: Targets to the left and below this point were dug for training data. Site specific training data were used in many of the processing approaches and these digs would be required. Different approaches required differing amounts of training data; the ROC curves for those that used no site-specific training data start at the origin.
- **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 30 of the can't analyze targets were false positives, reflected in the position of the point on the horizontal axis. No TOI were included in the can't analyze list.
- **C**: In the absence of any classification, this sensor detected all the TOI and had more than 2100 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 missed one TOI at her threshold.
- **E**: This demonstrator's best threshold chosen retrospectively. If the threshold had been chosen perfectly, only 200 targets could have been left in the ground.

4 ANOMALY SELECTION AND INVESTIGATION RESULTS

4.1 ANOMALY SELECTION

After 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 this sensor detection list was the basis for all subsequent analyses, a rigorous process was used to set this threshold.

4.1.1 Anomaly Selection Threshold

The known targets of interest in this demonstration were 105-mm, 155-mm, and 37-mm projectiles. Of these, the 37 mm is, obviously, the most difficult to detect. Prior to the demonstration, the site team determined that detection of 37-mm projectiles to 1 foot (30 cm) depth was a reasonable objective for this demonstration. Accordingly, the anomaly selection threshold for this demonstration was set as the smallest signal expected from a 37-mm projectile at 30 cm depth.

An example of this process is shown in Figure 4-1 for the EM61-MK2. The predicted signal from the EM61-MK2 for 37-mm projectiles (Ref. 17) in their least favorable orientation is plotted in the figure along with a vertical line marking the 30 cm depth of interest. The gate 2 anomaly selection threshold for this sensor system was set at 5.2 mV based on this curve. Also plotted on Figure 4-1 is the observed noise in the cued area. 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.

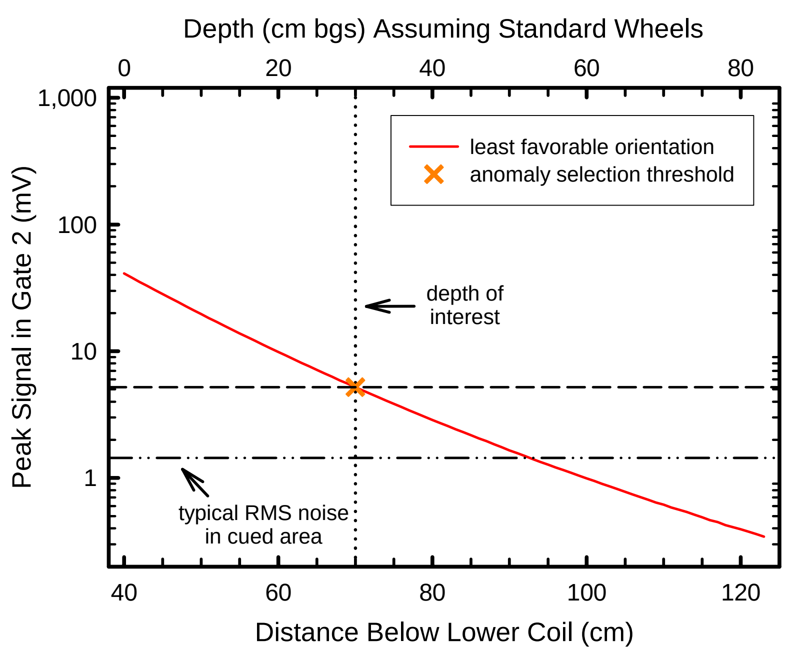

Figure 4-1. Predicted EM61-MK2 anomaly amplitude in gate 2 for a 37-mm projectile in its least favorable orientation. Also shown are the RMS noise measured at the site, the 30 cm depth used to set the threshold and the anomaly selection threshold used in this demonstration.

The goal of this process is to create a list of all positions that should be interrogated by the cued sensors. Many targets, especially those that with high length to diameter aspect ratios, result in multiple, closely-spaced exceedances. To avoid having redundant locations on the final anomaly list, 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. This consolidation resulted in 2304 identified anomalies in the cued area. A similar process was used to set the threshold for the MetalMapper survey.

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 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 anomalies are necessarily low signal-to-noise anomalies, which are often difficult to extract reliable parameters for and end up in the "unable to analyze, must dig" category.

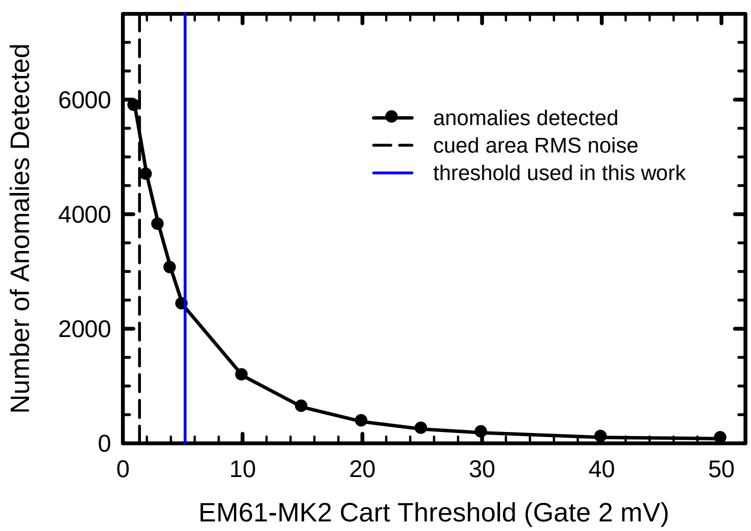

Figure 4-2. Number of EM61-MK2 threshold exceedances in the cued area as a function of the selection threshold applied. Also plotted are the system noise floor and the threshold used for this demonstration.

4.1.2 Detection of Seed Items

Using the anomaly selection thresholds described above and a detection halo of 0.6 m, the EM61-MK2 detected all seeds. Figure 4-3 shows a histogram of this detection performance. The mean miss distance was 0.20 m with a standard deviation of 0.10 m. Somewhat better location performance can be expected from the positions estimated from inversion of the measured data. The results obtained from inversion of the EM61-MK2 survey data are shown in the left hand panel of Figure 4-4 and those from inversion of the TEMTADS cued data are shown in the right panel. The inverted EM61 positions are slightly worse than those obtained from peak exceedances while from inversion of the TEMTADS data, the mean miss distance is 0.05 m with a standard deviation of 0.05 m

Although the MetalMapper dynamic survey failed to detect two seeds, both 37-mm projectiles, cued data was collected over these targets because the EM61 selections were the basis of the cued list. Based on test stand and IVS data, the MetalMapper should have easily detected these two targets so the missed detections are presumably due to the operation of the sensor rather than an inherent failure of the sensor. A retrospective analysis is in process in SERDP project MR-1772 (Ref.18).

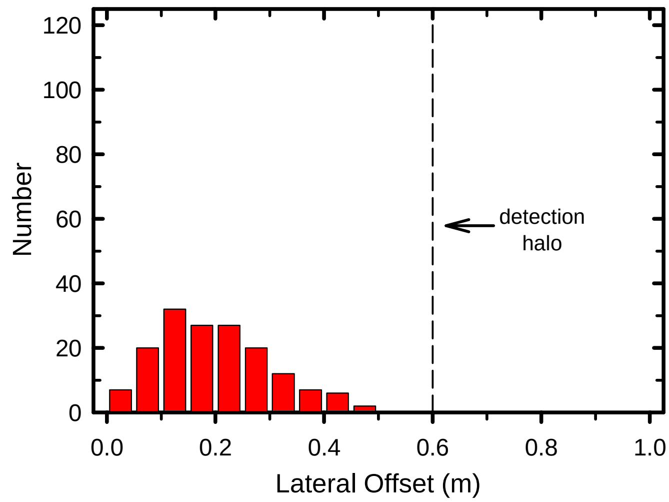

Figure 4-3. Histogram of the offsets between the actual location of the seeded items and their closest EM61-MK2 threshold exceedance.

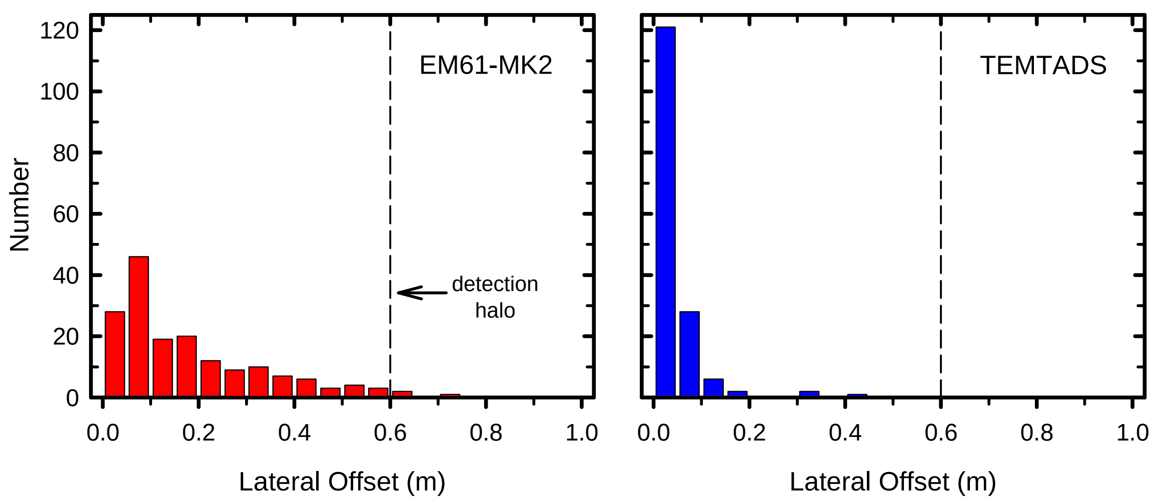

Figure 4-4. Histogram of lateral offsets between the actual location of the seeded items and their location as determined from inversion of the EM61-MK2 survey data (left panel) and the locations determined from inversion of the TEMTADS cued data (right panel).

4.2 DIG LIST

A list of locations for intrusive investigation, the dig list, was prepared starting with the EM61-MK2 anomaly list. Based on the performance on the seed items shown in Figure 4-3, the x,y positions resulting from inversion of the TEMTADS cued data were used for the target list. If the EM61-MK2 threshold exceedance's location was more than 0.6-m from either the TEMTADS or

MetalMapper location, both locations were added to the list to ensure that all metal objects associated with each exceedance were found. This resulted in 120 extra entries on the dig list.

4.3 INTRUSIVE INVESTIGATION

The distribution of recovered items by class is shown in Figure 4-5. The vast majority of items recovered at this site were classified by the UXO crew as munitions debris. Although it is difficult to see in Figure 4-5, there were 7 items recovered that were classified as UXO and an additional 4 items that were not UXO but, as discussed in section 2, were intact but empty 37-mm projectiles so they were declared as targets of interest.

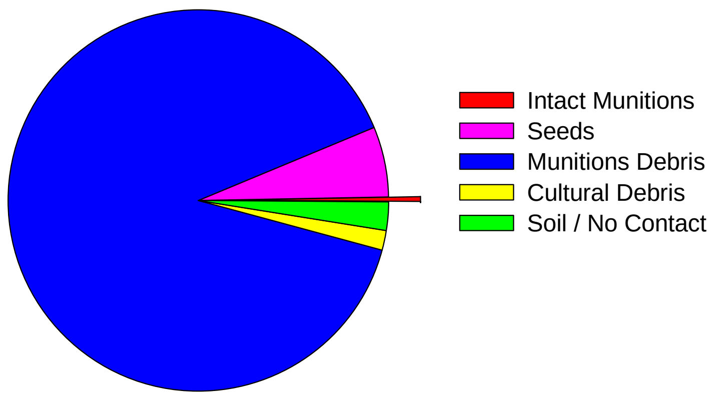

Figure 4-5. Distribution of recovered items by class.

The measured depths of the recovered items are plotted in Figure 4-6. As expected, most recovered items were quite shallow; the bin corresponding to recovered depths of 5 to 10 cm is, by far, the largest with nearly half the total recoveries in this bin. In fact, 95% of all recoveries corresponded to less than 22 cm to the center of the target.

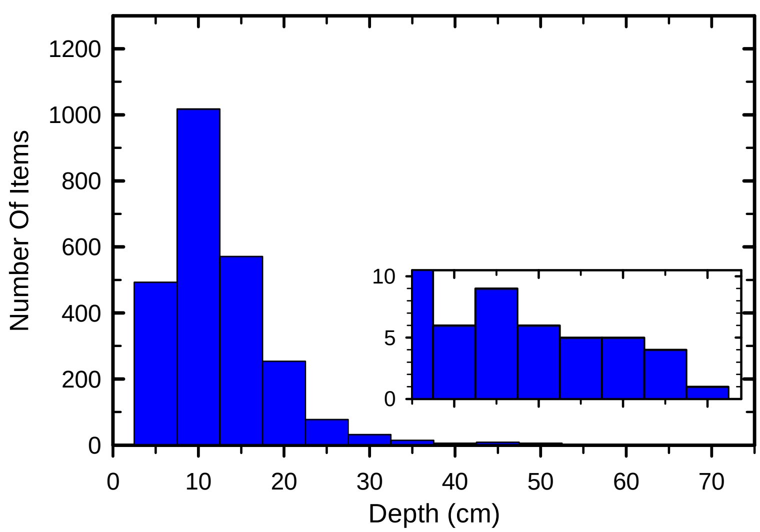

Figure 4-6. Measured depth distribution of all items recovered during the Camp Butner demonstration. The inset enlarges the scale to make the handful of targets recovered deeper than 40 cm visible. Depths tabulated are measured from the ground surface to the center of the recovered target.

Of the 52 items recovered at depths greater than 30 cm, 22 were seeded 105-mm projectiles; the rest were labeled "frag" by the UXO specialists. The seven live 37-mm projectiles recovered were all less than 16 cm deep and the four empty 37-mm projectiles that were called TOI were recovered less than 18 cm deep. These results confirm our expectation that detection of 37-mm projectiles at 30-cm (1-foot) depth was a reasonable goal for this site.

5 CLASSIFICATION RESULTS

A number of the demonstrators investigated multiple methods for training, parameter estimation, and classification during this demonstration. As a result, total of 30 dig lists were scored in the blind phase of the demonstration, representing the various combinations of sensor data collection systems and processing approaches used, with several more submitted later in the process. All of the results may be found in the report by IDA (Ref. 19). In the following sections, 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. Following these examples we present an overview of all classification results from Butner.

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 (△).

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 geophysicists at the contractor that collected the data at Camp Butner, NAEVA. NAEVA used a rules-based classifier based on decay constants calculated from different combinations of the four EM61 gates. They used the standard training data to set the rule for classification. The ROC curves showing their results based on gates 1, 2, and 3 are shown in Figure 5-1.

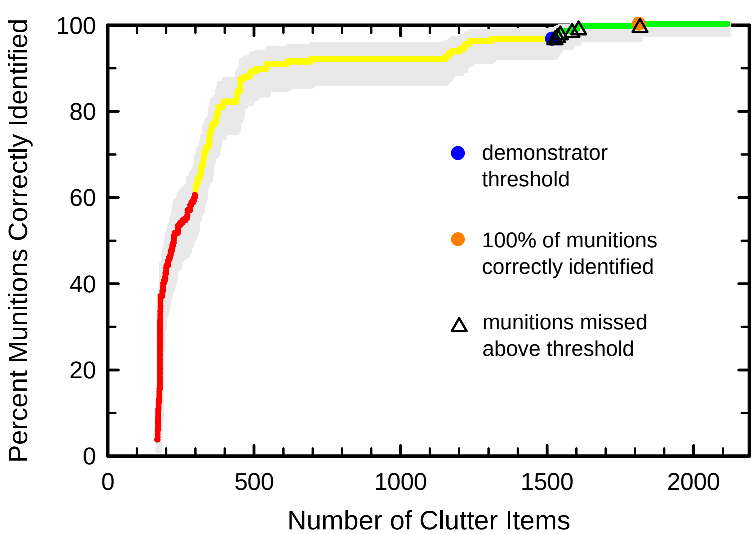

Figure 5-1. NAEVA analysis of the EM61-MK2 cart data.

The EM61 data were much less valuable for classification at this site than at the first two sites in this series (compare Figure 1-1). At the demonstrator-specified threshold, nearly 600 of the 2100 clutter items were left, but so were six 37-mm projectiles. The best possible performance for this sensor and analysis combination, denoted by the orange dot (●), would only correspond to ~300 avoided clutter digs.

The EM61-MK2 cart data were also analyzed by one of the algorithm developers, Sky. Sky used a combination of the total polarizability and the polarization decay as inputs to a statistical classifier. This team used no site-specific training data beyond what was collected in the test pit. These results of this analysis are shown in Figure 5-2. As in the case of the NAEVA analysis, these demonstrators were too aggressive in setting their threshold; a 37-mm projectile and a M48 fuze stimulant were missed. The best possible performance in this case would only leave ~200 clutter items in place.

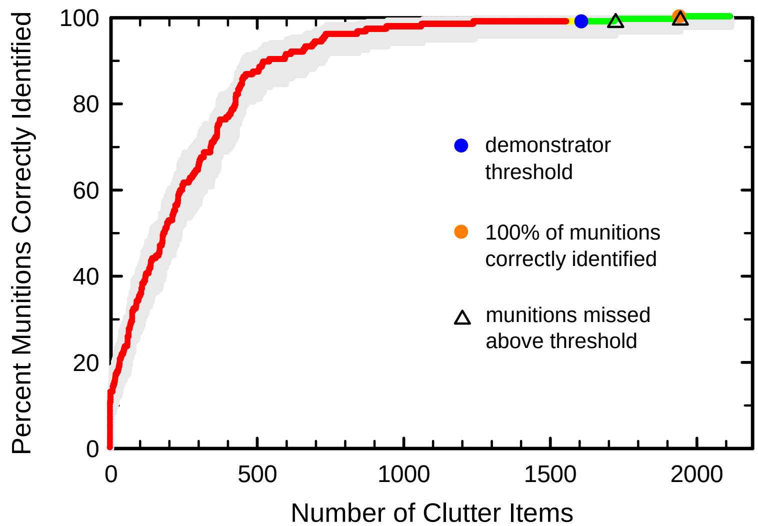

Figure 5-2. ROC curve resulting from Sky's statistical classifier applied to the EM61-MK2 cart data.

5.2 TEMTADS

Results of analyses of the TEMTADS cued data by Sky and Dartmouth are shown in Figure 5-3 and Figure 5-4 respectively. The Sky team submitted a number of analyses of these data; the one shown involves using features resulting from geophysical inversion of the measured data using a model that handles multiple sources as input to a support vector machine classifier. For most of the targets, this analysis used the time-dependant response coefficients (or polarizabilities), $\beta_1(t)$, $\beta_2(t)$, and $\beta_3(t)$, as inputs to the classifier. For the few targets nearest the boundary between the TOI and non-TOI classes the analysis switches to using the sum of the three polarizabilities as the classifier input. Note that in the original submission of this analysis, the one shown in Figure 5-3, only 22 training labels were requested. These analysts were able to correctly label ~1900 targets as clutter with no missed munitions.

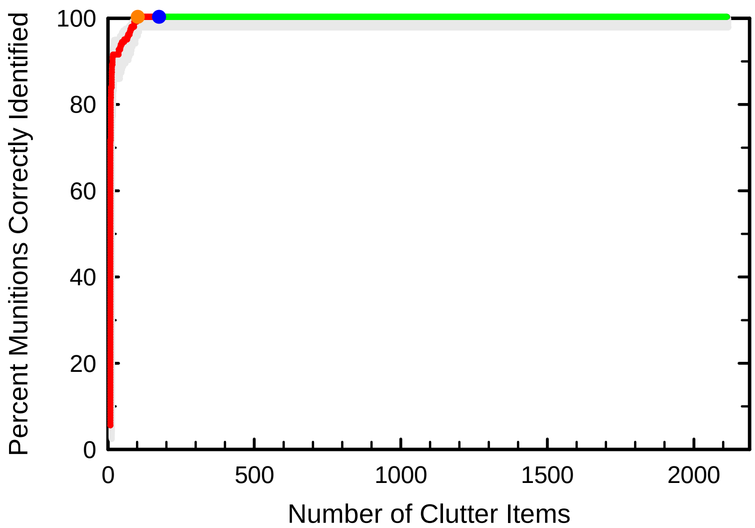

Figure 5-3. Sky analysis of the TEMTADS cued data.

Even better results were obtained from the Dartmouth analysis of the TEMTADS data, Figure 5-4. This analysis involved model fits to a multi-source, non-dipole model with a classifier that was a mix of rule-based and statistical. This analyst requested training labels on 75 targets; sixty-five of these were to identify "difficult" targets and ten were to confirm the identity of the four clusters of targets, Figure 5-5, found from the classifier. As was the case in most analyses of the advanced sensor data, these clusters corresponded to 1) 105-mm projectiles, 2) M48 fuze simulants, 3) 37-mm projectiles with rotating bands, and 4) 37-mm projectiles without rotating bands. As can be seen in Figure 5-4, only 41 false positives were required to identify 100% of the munitions, leaving over 2000 correctly labeled clutter items.

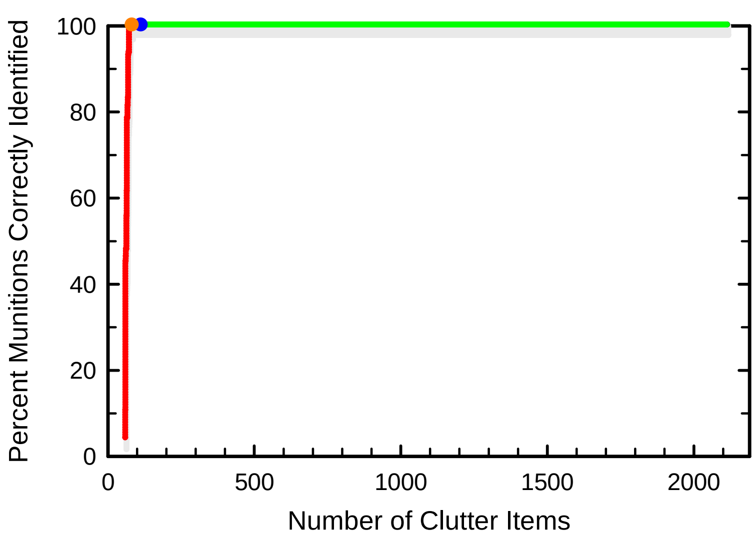

Figure 5-4. Dartmouth analysis of the TEMTADS cued data.

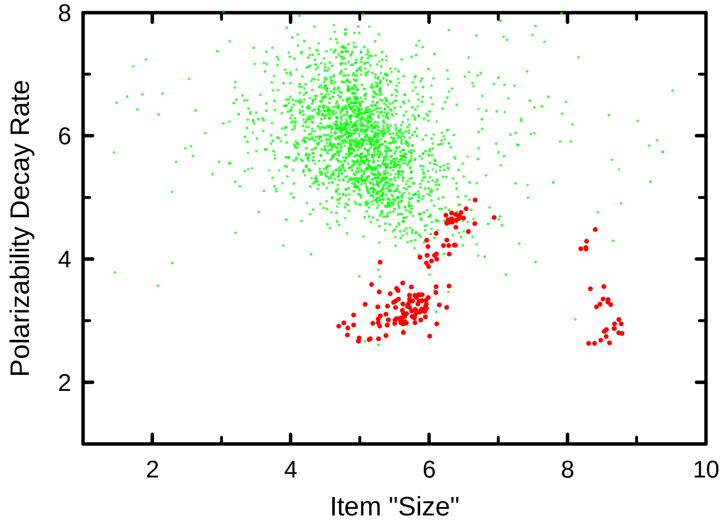

Figure 5-5. Features resulting from the Dartmouth analysis of the TEMTADS cued data illustrating the feature space "clusters" corresponding to the four munitions types on the site.

Figures 5–3 and 5–4 illustrate what is achievable using the advanced sensors. Not all analysts' results were this good. The results of the analysis of the TEMTADS data using the UX-Analyze module of Oasis montaj is shown in Figure 5-6. The results look very good for the first 98% of the munitions but the last four items proved difficult to identify. This analysis was carried out using a developmental version of the UX-Analyze software that only considered a single metallic object in the field of view of the sensor. The anomaly density at this site is high enough that this is a poor assumption in many cases as will be discussed in the next section. The current version of UX-Analyze incorporates a multi-source solver which has been shown in a retrospective analysis to eliminate these problems.

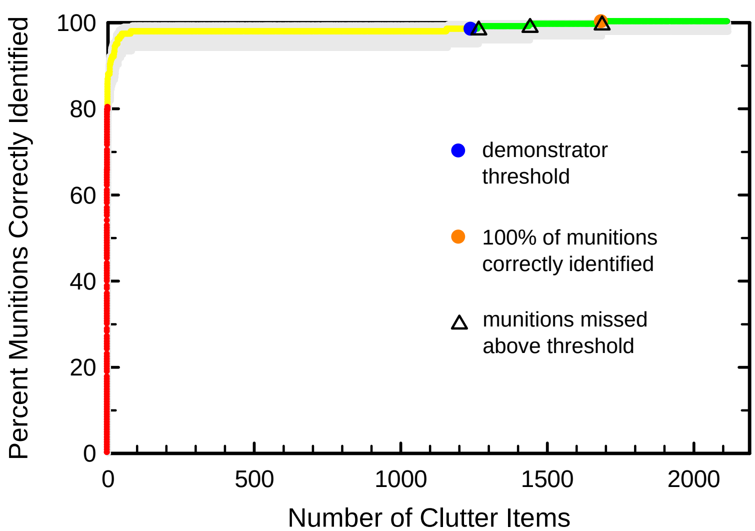

Figure 5-6. SAIC analysis of the cued TEMTADS data using the UX-Analyze module of Oasis montaj.

5.3 METAL MAPPER

Results from the MetalMapper data as analyzed by Sky and Dartmouth are shown in Figure 5-7 and Figure 5-8, respectively. The methods employed to produce the two curves shown are the same as discussed above for the two TEMTADS analyses so they will not be repeated here. Overall, the results are very good. It is clear from the steep rise in the red portion of the ROC curves that most TOI are readily recognized and classified as such with high confidence. It is equally clear from the distance that the green lines extend along the top axis that most non TOI are also readily recognized and classified as such with high confidence.

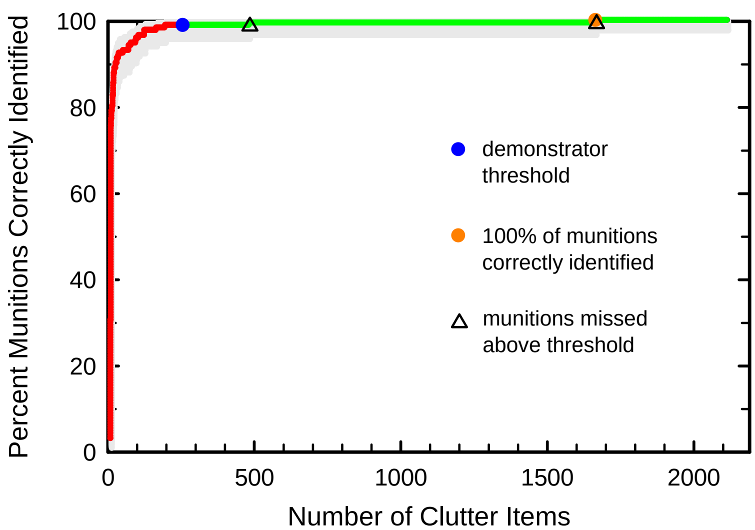

Figure 5-7. Sky analysis of the MetalMapper cued data.

There are more targets in this data set that caused difficulties for the analysts. In the case of Sky, Figure 5-7, this results in two munitions missed past the demonstrator threshold; one of them is quite far to the right of the ROC curve. In the case of the Dartmouth analysis, Figure 5-8, this results in more training data requests required to clarify the "difficult" targets and thus slightly more clutter digs to identify all the munitions. There are still 1950 targets correctly identified as clutter in this analysis.

As for the case with TEMTADS, the results from the method developers show the potential available from the MetalMapper data. These data were also analyzed by geophysicists from two production companies. An example of their results is shown in Figure 5-9. The production geophysicist was less skilled with the parameter extraction step in the process; they were unable to extract reliable parameters from nearly 350 anomalies. After that though, their results are quite good. They only missed one TOI at their operating point and were able to correctly eliminate half of the clutter while detecting 100% of the UXO. This was the first experience for the production geophysicists with advanced sensor data; as they gain familiarity with the classification concept and UX-Analyze, we can expect their results to improve.

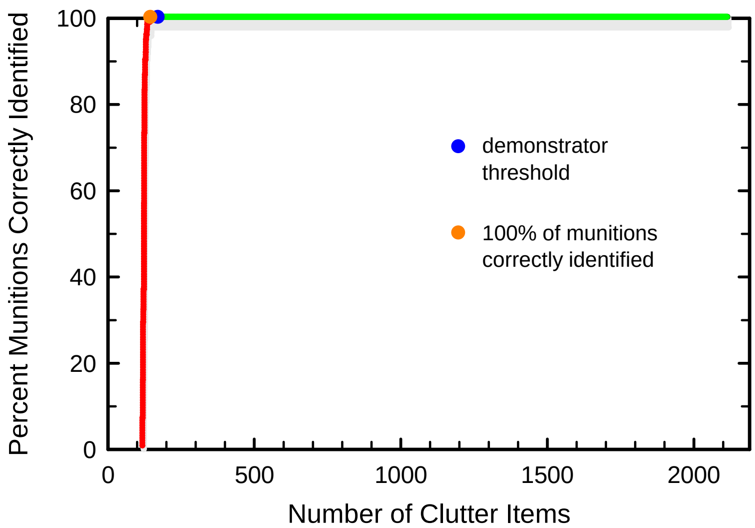

Figure 5-8. Dartmouth analysis of the MetalMapper cued data.

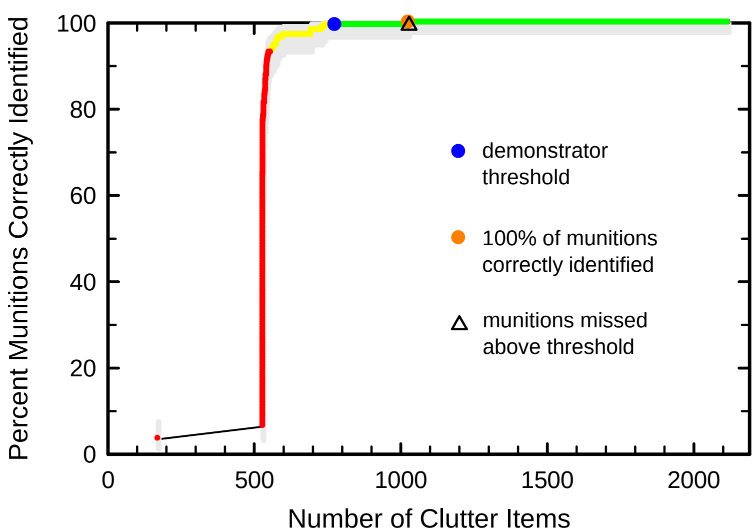

Figure 5-9. NAEVA analysis of cued MetalMapper data using the UX-Analyze module of Oasis montaj.

5.4 RESULTS OVERVIEW

An overview of the results from all 54 ranked anomaly lists scored in this demonstration is shown in Figure 5-10. In the left panel, the number of munitions correctly identified at the demonstrator's operating point is plotted versus the number of clutter items at that same point. The goal, of course, is to be as close to the upper left corner of the plot as possible. The right panel of the figure shows the best possible performance for each analysis; it is the number of clutter items that must be dug from each list to identify all the munitions. This point is only known after the fact; it requires digging all the anomalies.

Several general trends can be seen in Figure 5-10. It was difficult for all analysts to achieve good classification performance using the EM61-MK2 data. The demonstrator operating points for

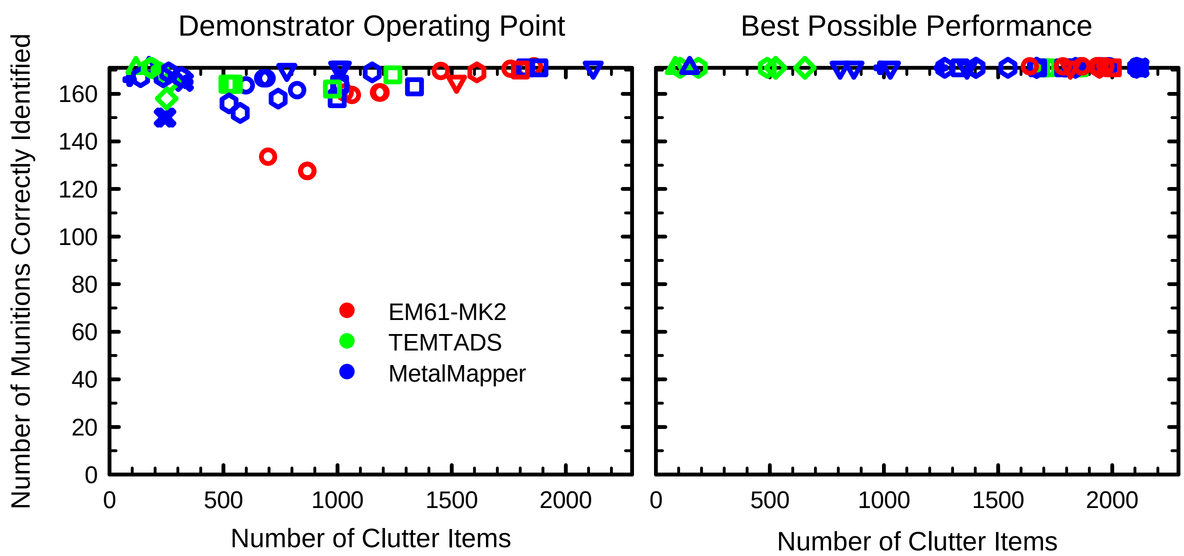

Figure 5-10. Overview of the performance of all analysis demonstrators at Butner. Performance at each demonstrator's operating point is plotted in the left panel and the best possible performance for each analysis is plotted in the right panel. The points are color coded for the sensor data set used.

EM61 data in the left panel are either well below 100% identification of the munitions or require a large number of clutter digs. Even when the best possible operating point is established after the fact for the EM61 analyses (right panel), the performance points are near the upper right corner of the plot. This corresponds to minimal classification ability; nearly all the clutter must be dug to correctly identify all the munitions.

The best performers (those nearest the upper left corner) at the operating point involved either TEMTADS or MetalMapper data. This trend continues in the right panel of Figure 5-10, the best possible operating points. Although the performance with the advanced sensors varied widely depending on the analyst, the EM61 points all cluster on the right side of the plot.

There appear to have been a few anomalies for which the MetalMapper data were incomplete or in error (e.g. anomalies 1344, 1346, and 2504) because all but one analysis of the MetalMapper data results in a point far from the upper left corner in this plot. Most analysts placed these anomalies well into the clutter portion of their ranked anomaly lists.

5.5 DISCUSSION

5.5.1 Features

In the first two demonstration in this series, analysis of the EM61-MK2 data provided considerable classification using the signal decay parameter. This was not the case at this demonstration. This is illustrated in Figure 5-11 which shows the decay parameters (τ) calculated by NAEVA for two pairs of EM61-MK2 gates. There is very little separation between the munitions and clutter evident in this plot except for the 105-mm projectiles. Much of the clutter at this site consists of fragments of large projectiles which have similar sizes and thicknesses as the 37-mm projectiles. The EM61-MK2 will not have much classification ability at a site like this.

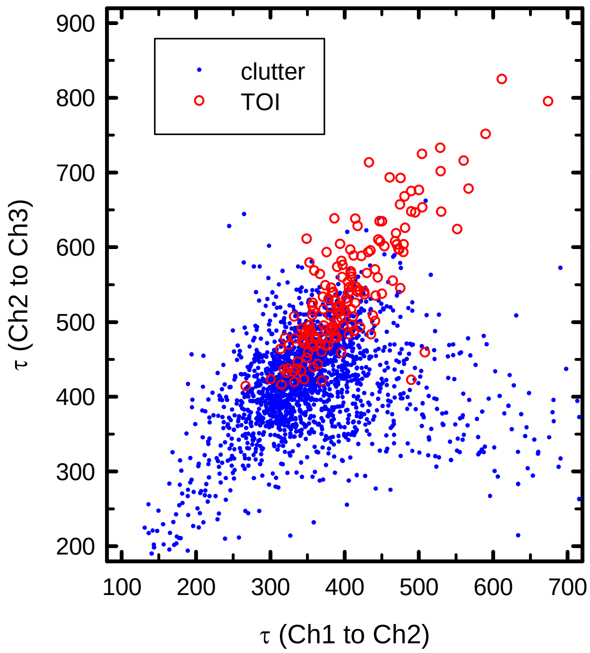

Figure 5-11. EM61-MK2 decay parameters as analyzed by NAEVA.

Because they illuminate the target completely and produce much higher signal-to-noise ratios, the advanced sensors provide much more accurate estimation of the target polarizabilities, which are the basis of the key features used in the successful classification approaches. Not only does this present the analyst with a more reliable and reproducible estimate of polarizability decay (as opposed to simply decay of the observed signal), it allows the use of polarizability amplitudes and patterns in classification. Figure 5-12 compares the polarizabilities as a function of decay time estimated using TEMTADS cued data for four 37-mm projectiles. The blue curves result from a reference projectile measured in air. The other three curves result from analysis of data collected over buried targets. All four sets of curves are virtually indistinguishable. 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.

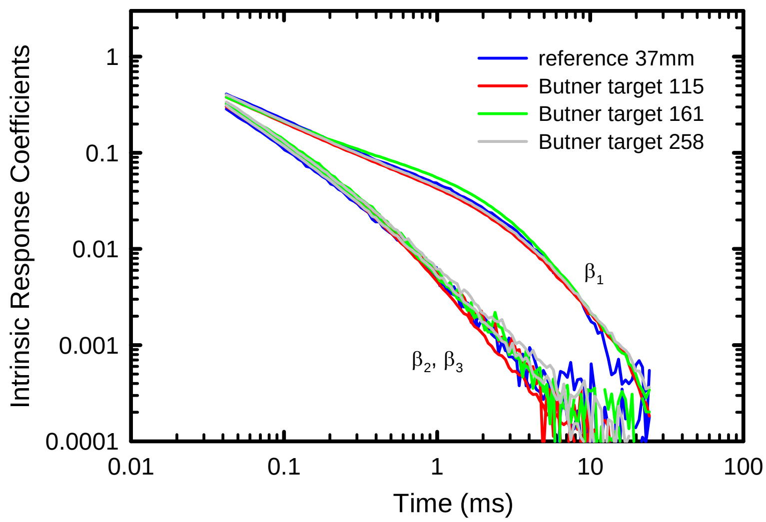

Figure 5-12. Superposition of the polarization as a function of decay time for TEMTADS for a reference 37-mm projectile measured in air and three buried 37-mm targets from Camp Butner.

An illustration of these extra features for classification is shown in Figure 5-13. The left hand panel shows the polarizabilities as a function of time estimated by inversion of the TEMTADS cued data collected over target 13. We can learn several things about this target by inspection of these curves. First, the magnitudes of the recovered polarizabilities are much larger than those for the 37-mm projectiles shown in Figure 5-12 indicating this target is substantially larger than a 37mm. Second, the decays observed in the left panel of Figure 5-13 are similar to those from Figure 5-12 indicating that the wall thicknesses of the two targets are roughly similar. Finally, the polarizabilities indicate one large response and two smaller, and roughly equal, responses, a pattern characteristic of a cylindrical object such as a projectile. These three factors led the analysts to declare this item a high-confidence munition.

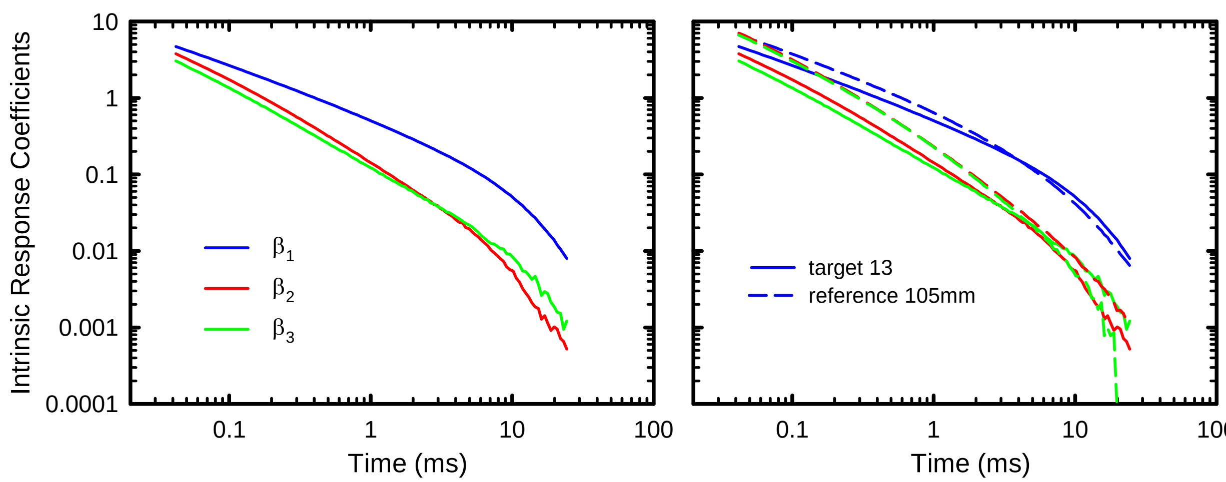

Figure 5-13. The estimated polarizabilites of Butner Target 13 as a function of time (left panel). These polarizabilities are compared to a reference 105-mm projectile in the right panel.

The right hand panel of Figure 5-13 compares the polarizabilities of target 13 to those of a reference 105-mm projectile. Although the curves do not match exactly, they are close enough that item 13 has to be declared a munition in cases like this where false negatives must be avoided. Figure 5-14 is a photograph of the two objects recovered in the intrusive investigation of target 13. The two pieces of frag were oriented parallel to each other in the hole producing a roughly symmetric object.

Figure 5-14. Photograph of the two objects responsible for anomaly 13 at Camp Butner.

A final illustration of the power of the advanced sensors is given in Figure 5-15 which compares the polarizations as a function of time of the reference 37-mm projectile from Figure 5-12 with those estimated for three different targets recovered during this demonstration. Although the three targets shown exhibit similar polarizabilities to the reference 37mm, they are distinctly, and reproducibly, different. Photographs of these three items are shown in Figure 5-16; they are all 37-mm projectiles without rotating bands.

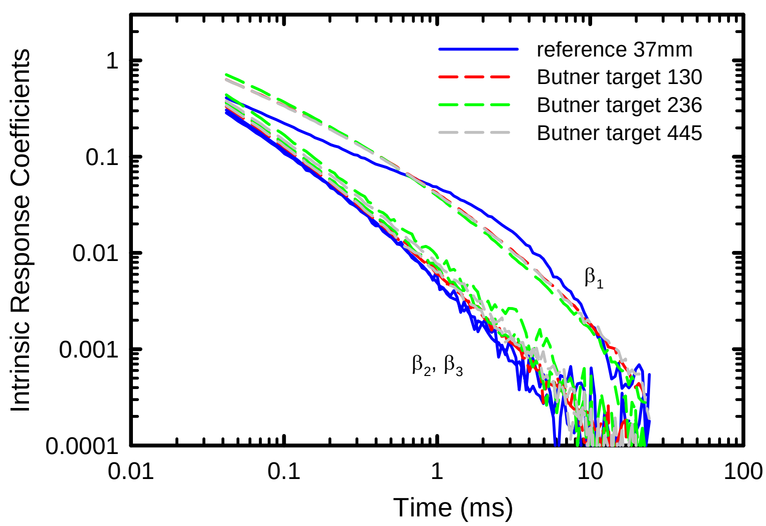

Figure 5-15. Comparison of the polarizabilities of the reference 37-mm projectile from Figure 5-12 with three different 37-mm projectiles recovered during this demonstration.

Figure 5-16. Photographs of three 37-mm projectiles without rotating bands seeded at Camp Butner.

5.5.2 Importance of Multi-object Solvers

The anomaly density at Camp Butner was much higher than at the previous two demonstrations in this series; over 400 anomalies per acre above threshold with many more low-amplitude anomalies (see Figure 4-2). Because of this, the majority of cued measurements involved two or more metal objects in the field-of-view of the sensor. Analysis algorithms that assumed the measured signal was due to only one object will naturally return erroneous results in these cases. This is illustrated in Figure 5-17 which plots the polarizabilities estimated from MetalMapper cued data over target 1346 (a seeded inert 37-mm projectile). The left hand panel plots the results from a solver that assumes only one source with a reference 37mm for comparison. Obviously, the estimated responses look nothing like the reference munition. In this case, any classifier would report this target as clutter. The right hand panel plots the two sets of responses returned from a solver that assumes two sources are present. While not perfect, the responses for object 1 are an acceptable match to the reference 37mm with object 2 being a smaller, non-symmetric item.

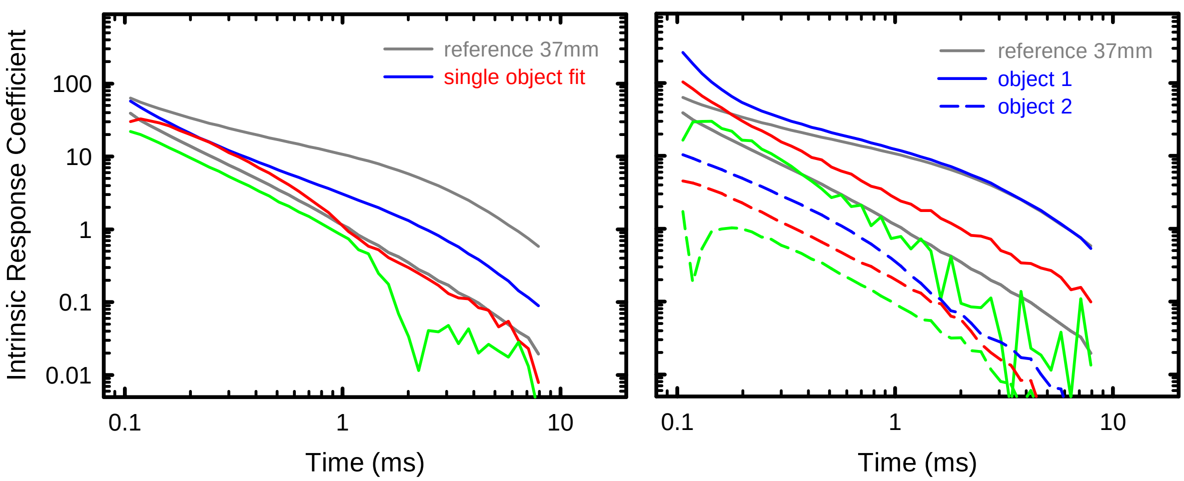

Figure 5-17. Results of the analysis of the cued MetalMapper data from target 1346 assuming one source (left panel) and two sources (right panel). In both cases, the response of a reference 37-mm projectile is plotted in gray.

Theses multi-source solvers are just emerging from the research program. Their refinement and universal implementation will be important as these classification demonstrations are extended to more and more difficult sites.

6 COST CONSIDERATIONS

The motivation for applying classification in munitions response is to more effectively use the available resources: If the digging of non-munitions targets is minimized, then the limited resources of the munitions response program can be applied to clean up more land more quickly. The actual costs of a demonstration in this series 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 developed a simple cost model with realistic assumptions for production costs of various model elements in the report describing the San Luis Obispo demonstration (Ref. 4). We use that same model here to examine the cost implications on the FUDS MMRP budget.

6.1 FUDS BUDGET IMPLICATIONS

The Defense Science Board Task Force on Unexploded Ordnance reported in 2003 (Ref. 2) that over 75% of the budget for a typical munitions response was spent on removal of items that turned out to be non-hazardous. If we apply the percentages in that report to a nominal $200M annual FUDS budget when the MMRP program reaches the remediation stage, we get the breakdown in site activities shown in Figure 6-1 for cleanups conducted using current practice. Also plotted on the right side of the figure is a bar representing the number of acres per year that can be remediated for this budget assuming remediation costs of $25K per acre.

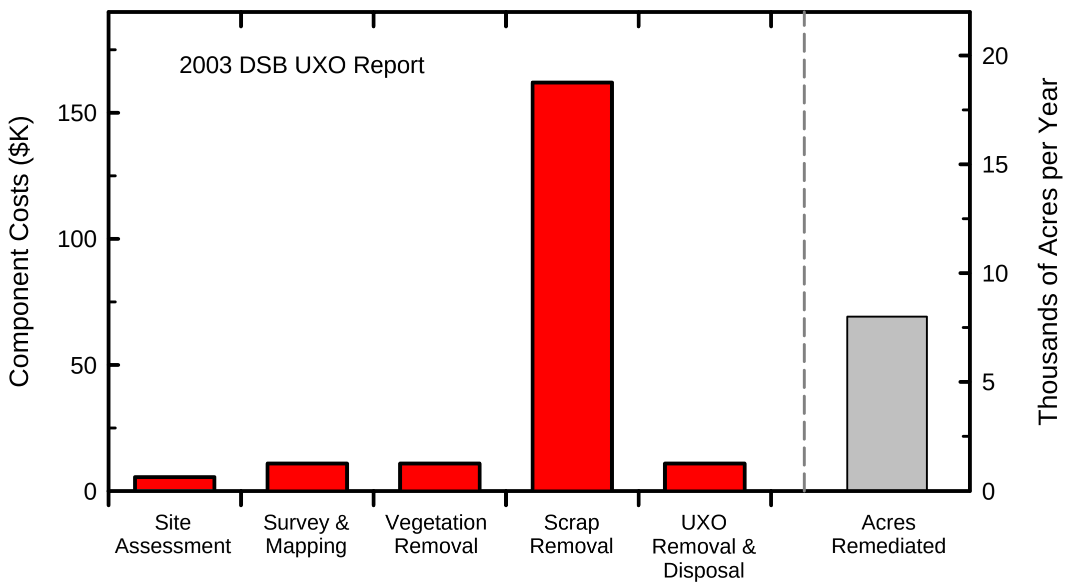

Figure 6-1. Distribution of activity funding in a nominal $200M FUDS MMRP using the breakdown discussed in the 2003 Defense Science Board UXO Task Force report.

The factor driving the overwhelming portion of the budget that must be devoted to scrap removal is that UXO make up 1% or less (and often much less) of the metal items on a munitions site. Thus, a reduction in the number of clutter items that must be treated as potential UXO has a large budgetary impact. Figure 6-2 plots the funding distribution possible if a 70% reduction in false alarms can be achieved. Based on the results at this demonstration, this reduction should be possible at most sites.

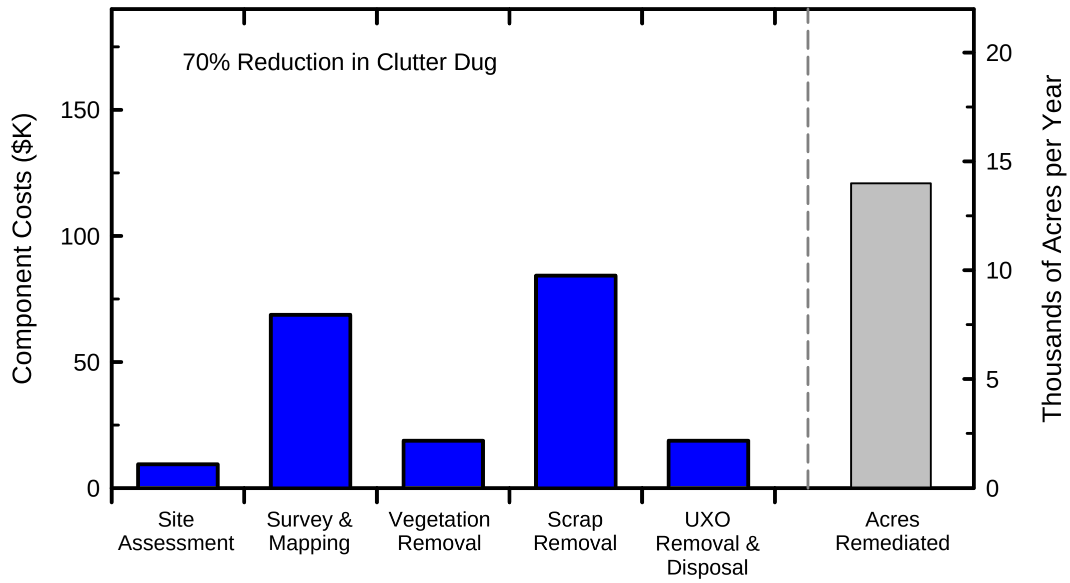

Figure 6-2. Data from Figure 6-1 assuming a 70% reduction in false alarms.

Notice in Figure 6-2 that the relative funding devoted to Survey and Mapping is increased. As we have demonstrated, the data collection and analysis required to implement classification requires more resources than a typical detection survey. These added up-front costs are more than repaid by the savings in the intrusive phase; the number of acres that can be remediated under this scenario increases by 75%.

The DSB Task Force called for a research effort devoted to reducing the false positive rate by 90%. This standard was achieved by the better demonstrators at Camp Butner and should be possible at many sites. The distribution of funding assuming a 90% reduction in false alarms is plotted in Figure 6-3. In this case, the relative area cleared increases by a factor of 2.4 compared to the base case.

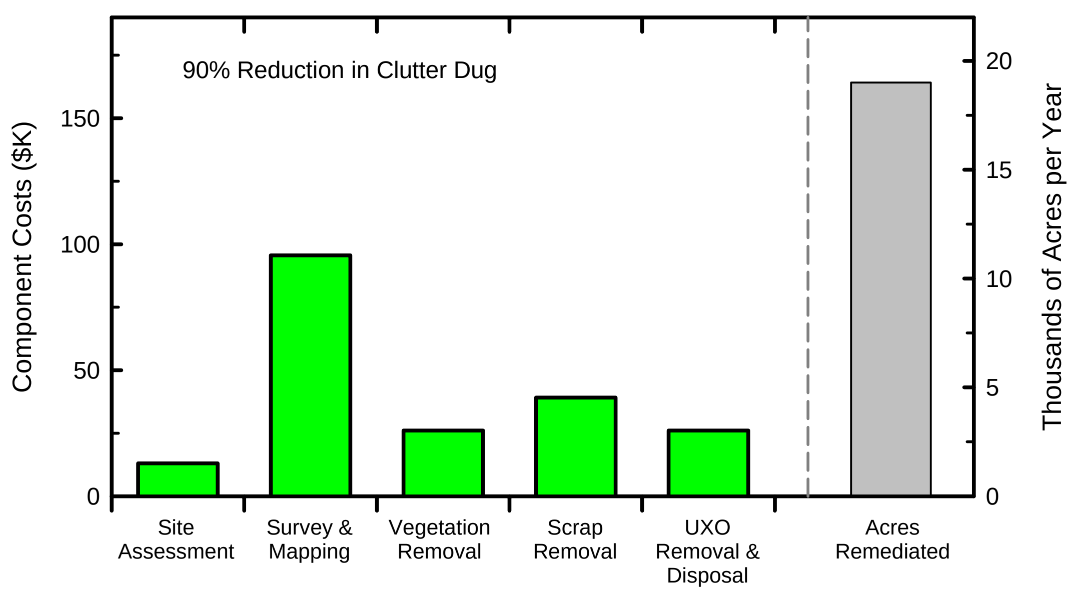

Figure 6-3. Data from Figure 6-1 assuming a 90% reduction in false alarms.

7 CONCLUSIONS

Continuing the performance established at the San Luis Obispo demonstration (Ref. 4), this demonstration at the former Camp Butner showed outstanding classification potential using advanced EMI sensors on a site that contained the challenges of higher anomaly density and the presence of 37-mm projectiles with larger munitions. Because of these complications, carefully collected survey data from commercial EM61-MK2 sensors were of much less value than in previous demonstrations. The best analysts using the EM61-MK2 data were only able to correctly classify about 10% of the non-hazardous clutter at the point where they identified all the munitions.

7.1 DEFINING SUCCESS

In munitions response, success should be judged from the perspective of risk reduction. With existing technology, no cleanup can guarantee that 100% of munitions are detected and removed from a site. Sensors have known limitations with regard to the types of targets that can be detected and to what depth. Even with the most careful QC, uncertainty remains that all munitions were detected and removed. At the end of a cleanup, even one where the objective is to "remove all detected 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 best performers at Camp Butner working with the best data sets correctly identified all the munitions on the anomaly list with very few false positives. Others had one or two munitions quite far into the high confidence non-TOI part of their prioritized lists. This indicates 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 the potential for real risk reduction, the Butner study was successful. For the purposes of the study, a target of interest was defined as an intact munition or a projectile fuze with booster attached. Importantly, the classification process allows for iterative expansion of the targets of interest. 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 targets of interest can be revised at any time.

7.2 DETAILED OBSERVATIONS

7.2.1 EM61-MK2

The EM61-MK2 cart data were collected by NAEVA and analyzed by NAEVA and Sky. NAEVA based their analysis primarily on signal decay while Sky used both signal decay and total polarizability, which serves as a marker for the physical size of the item. Neither of these was particularly successful at this site with both demonstrators missing a number of munitions after their threshold and only correctly identifying about 10% of the clutter once they achieved 100% identification of the munitions present.

There were several differences from the previous demonstrations that led to this result. The small size of many of the targets at Butner resulted in low signal-to-noise anomalies in the EM61 data which, coupled with the high density of anomalies at this site, made it difficult to extract reliable parameters from many of the anomalies. In addition, many of the clutter items at Butner consisted of fragments from larger projectiles which were roughly similar in overall size and wall thickness to the 37-mm projectiles. Thus, neither of the parameters available from the EM61-MK2 data was useful as a discriminant at Camp Butner.

At other sites in this series the EM61-MK2 has been able to successfully eliminate as many as one half of the clutter at the site. This site is more typical of a "hard" classification site and the results here indicate the limitations of the commonly-used sensors for this use.

7.2.2 Detection of Seeds

Detection of all emplaced seeds continues to be an important element in building site team confidence in the geophysical data quality, data analysis methods, and overall demonstration design. As has been the practice in the demonstration series, the anomaly selection threshold at Camp Butner was set based on the expected signals from the targets of interest rather than referenced to the observed survey noise. This is an important aspect of the classification method; it eliminates low signal-to-noise targets that are too small to be targets of interest but are be difficult to extract parameters for successfully from the analysis. Using this approach, analysis of the EM61 data resulted in the detection of all emplaced seeds.

The dynamic MetalMapper data did miss two seeds. Based on test stand and IVS data, the MetalMapper should have easily detected these two targets so the missed detections are presumably due to the operation of the sensor rather than an inherent failure of the sensor. A retrospective analysis is being conducted to understand the cause of this failure. Obviously, an actual remediation would not proceed until the results of this failure analysis were known.

7.2.3 Advanced Sensors

Two recently-developed EMI sensors, optimized for UXO classification, were demonstrated at Camp Butner. The NRL TEMTADS was operated in cued mode working against the detection list from the EM61 cart. The Geometrics MetalMapper system first surveyed the field in detection mode and then revisited all locations on both its own detection list and the distinct EM61 detections (so that comparisons could be easily drawn) in cued mode.

The best analysts were able to achieve remarkable results when working with the data collected by either of these sensors. The Dartmouth analysis of the TEMTADS data required only 75 ground truth labels for training and then was able to identify all the remaining munitions with only 41 false positives. This resulted in over 2000 correctly identified clutter items out of 2219 at the site (96%).

Results using the MetalMapper data were nearly as good although there were more targets in this data set that caused difficulties for the analysts. Upon retrospective analysis, this handful of "difficult" targets could largely be attributed to insufficient SNR. This underscores the role of careful field QC when deploying these sensors. This point will be discussed in more detail in a later section.

Not all analysts and methods were able to achieve these impressive results using the advanced sensor data although all but a handful were able to correctly identify more than 50% of the clutter while missing no targets of interest. A primary objective of the remaining demonstrations in this series will be to identify ways for all analysts to perform up to the potential demonstrated by the best performers.

7.3 REMAINING CHALLENGES

Although the analysts in this demonstration made substantial progress in resolving the limitations identified from previous demonstrations in this series, continuing work is required:

Partial, corroded, or bent rounds – All of the items identified by the UXO specialists at Butner as UXO were intact 37-mm projectiles. There were also four "empty" 37-mm projectiles recovered that were classified as targets of interest for this demonstration. Unlike previous demonstrations, nothing was recovered that could be classified as a partial round; thus we were not able to test the ability of the demonstrators to correctly identify partial, corroded, or bent rounds. As we move forward, it is important to recognize that there is a continuum of partial or damaged rounds, clearly define what constitutes a TOI, and set classification criteria appropriately.

Smaller Munitions – The smallest munition of interest to date has been a 37-mm projectile. The applicability of these techniques on sites containing smaller munitions and submunitions remains unknown.

Unexpected Munitions – Three unexpected munitions types were ultimately found and successfully classified in the San Luis Obispo demonstration (Ref. 4) but none were encountered here. An important aspect in building stakeholder confidence in the classification process will be continued demonstration of the ability to successfully identify unexpected munitions.

Overlapping targets – In earlier demonstrations, care was taken to only include anomalies as part of the demonstration that were well separated. Even then, a number of items of interest were missed because they were located close to another object. This demonstration was conducted at a site with higher anomaly densities (over 400 per acre above the selection threshold) and all anomalies in the demonstration area were included in the scoring. This is one reason that classification using the EM61 data was less successful at this site.

Many of the model developers working with the advanced sensor data used a model that can handle multiple items in the field-of-view of the sensor which led to very successful classification. For the most part, the production geophysicists did not use these advanced models. The identification and handling of multiple objects is an active area of research and further demonstrations should show even better success as the advanced models are disseminated more widely .

Thresholds – While there continue to be examples 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 remain cases, however, where a handful of TOI are ranked very near their 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.

Difficult Site Conditions – These techniques have not been demonstrated on sites with difficult vegetation and incomplete sky view that will require man-portable sensors and non-GPS sensor location. Two of the next four demonstrations will address this challenge. Similarly, these methods have not been demonstrated on sites with extreme geologic interference.

Variability in Performance – As seen above, not all analysts are able to achieve the impressive results demonstrated by the best performers. In general, the method developers were able to correctly classify many more anomalies as clutter than analysts from the production firms. This is presumably due to the large difference in experience with the software and methods between the two groups. A challenge going forward will be to bring the achievement of the production geophysicists more in line with that of the experts.

7.4 LESSONS LEARNED

Several of the implementation issues that have arisen in prior demonstrations were confirmed at Camp Butner and several new issues emerged.

Importance of Careful QC/QA Procedures – Well defined quality control procedures are a good predictor of success in all munitions response actions. This is particularly true in classification using advanced sensor data where one demonstrator remarked "if you collect high-quality data with these sensors, the decision makes itself."

The importance of a careful quality plan is important in both the data collection and analysis segments of the process. As discussed above, a number of MetalMapper data points were difficult for even the best analysts primarily due to signal-to-noise limitations. These issues could have been easily corrected before the data collection team left the field. Less obvious is the QC failure that led to the misclassification of the target of interest on the far right in Figure 5-7. This was anomaly 1346 that was discussed in Figure 5-17. The incorrect submission of the single-object solver results to the classifier resulted in this failure. A careful QC procedure could have detected this error before it propagated into the ranked anomaly list.

Seeds are Critical for Confidence – All targets are investigated in these demonstrations permitting a careful retrospective analysis of the results. In a real-life implementation of classification only those targets below the dig threshold would have ground truth available. Even if the site managers chose to selectively sample those targets classified as non-hazardous, the identity of most of them would remain hidden. The seed targets then become critical in providing performance confirmation to the site team. If the analysts successfully classify all the seeded targets and place them well before the dig threshold, the site team will be more likely to leave many items buried with confidence.

Seed items can be chosen to be representative of the munitions expected on the site or they can include a class of targets not expected by the demonstrators to provide extra assurance to the site team. They should not however be emplaced outside the bounds set for the remedial action. Seeds smaller than would be expected from site conditions or buried deeper than would be reasonable in order to "check on the performance of the system" set the geophysical team up for failures that provide no instructive value for the site team.

Assumptions Should be Re-examined after the Initial Excavation – All members of the Advisory Group agree that the standard operating procedure for classification should include a careful examination of the results and assumptions once the targets from the initial dig list are excavated. Anything uncovered that invalidates assumptions made in designing the response action such as unexpected munitions or items deeper than expected is cause for a detailed discussion by the site team. If the anomaly selection thresholds need to be revised or the classification procedures modified, this is the time to make these changes. It is also the time for all stakeholders to plan any selective sampling of the targets classified as clutter (and thus left unexcavated) that will give the site team the confidence to proceed.

Take Advantage of All Features Available from the Advanced Sensors – Analysis of data from the advanced sensors not only gives information that can lead to successful classification, it provides highly accurate extrinsic features (location, depth, and orientation) parameters as well. As was seen in Figure 4-3, the locations are often within 15 to 20 cm and the depths are even better. If the depth and orientation are included on the dig list, the remediation crew quickly gains confidence in the estimates and the remediation proceeds more efficiently and quickly and ensures that the intended target is excavated.

7.5 COST AND PRODUCTION

We have shown how the savings from the use of classification can be expected to increase the productivity of the MMRP program for two assumptions about the number of false positives that can be confidently eliminate. If 70% of the clutter can be confidently identified, the area remediated for a fixed budget will increase by at least 1.75. If the classification efficiency can be increased to 90%, the area remediated on a fixed budget will increase by a factor of 2.4.

REFERENCES

1. *Fiscal Year 2009 Defense Environmental Programs Annual Report to Congress*, http://www.denix.osd.mil/arc/ARCFY2009.cfm

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, http://www.acq.osd.mil/dsb/reports/ADA419970.pdf

3. Nelson, Kaye, and Andrews, *ESTCP Pilot Program: Classification Approaches' in Munitions Response,* 2008, http://serdp-estcp.org/content/download/7423/94821/file/ESTCP-Classification-CampSibert-FinalReport%2012-17-09.pdf

4. Nelson, Andrews, and Kaye, *ESTCP Pilot Program: Classification Approaches in Munitions Response San Luis Obispo, California,* 2010, http://serdp-estcp.org/content/download/7424/94825/file/ESTCP-Classification-CampSLO-Final-Report-26May2010.pdf

5. Parsons, Inc., *Final Engineering Evaluation/Cost Analysis, Former Camp Butner, Camp Butner, NC*, July 2004.

6. ESTCP Project MR-200910, Analysis of Multi-Axis, Multi-Coil Electromagnetic Induction Sensor Data for UXO Discrimination, NAEVA Data Collection Report, in preparation.

7. ESTCP Project MR-201034, UXO Classification Using EMI Sensors, TEMTADS Data Collection Report, in preparation.

8. ESTCP Project MR-200603, MetalMapper: A Multi-Sensor TEM System for UXO Detection and Classification, MetalMapper Data Collection Report with cfg files appended, in preparation.

9. ESTCP Project MR-201004, Practical Strategies for UXO Discrimination, Sky Analysis Report, in preparation.

10. ESTCP Project MR-200910, Analysis of Multi-Axis, Multi-Coil Electromagnetic Induction Sensor Data for UXO Discrimination, SAIC Analysis Report, in preparation.

11. SERDP Project MR-1572, A Complex Approach to UXO Discrimination: Combining Advanced EMI Forward and Statistical Signal Processing, Dartmouth Analysis Report, in preparation.

12. ESTCP Project MR-200603, MetalMapper: A Multi-Sensor TEM System for UXO Detection and Classification, Geometrics Analysis Report, in preparation.

13. ESTCP Project MR-200910, Analysis of Multi-Axis, Multi-Coil Electromagnetic Induction Sensor Data for UXO Discrimination, NAEVA Analysis Report, in preparation.

14. ESTCP Project MR-200910, Analysis of Multi-Axis, Multi-Coil Electromagnetic Induction Sensor Data for UXO Discrimination, Parsons Analysis Report, in preparation.

15. SERDP Project MR-1708, Optimal Sensor Management and Signal Processing for New EMI Systems, SIG Analysis Report, in preparation.

16. ESTCP Project MR-201004, Practical Strategies for UXO Discrimination, CH2M HILL Analysis Report, in preparation.

17. Nelson, Bell, Kingdon, Khadr, and Steinhurst, *EM61-MK2 Response of Standard Munitions Items*, NRL/MR/6110--08-9155, October 2008, http://serdp-estcp.org/Tools-and-Training/Munitions-Response/Geophysical-System-Verification/GSV-Response-Calculator

18. SERDP project MR-1772 Fact Sheet, *Expanded Processing Techniques for EMI Systems*, http://serdp-estcp.org/Program-Areas/Munitions-Response/Land/Modeling-and-Signal-Processing/MR-1772

19. Cazares, Tuley, and Ayers, Institute for Defense Analyses, The UXO Classification Demonstration at the former Camp Butner, NC, September 2011, IDA document number D-4379.