FINAL REPORT

Implementing Advanced Classification on Munitions Response Sites: A Guide to Informed Decision Making For Project Managers, Regulators, and Contractors

December 2011

Anne Andrews
Herb Nelson
ESTCP

ESTCP

Table of Contents

Executive Summary

The Munitions Response Program (MRP) is charged with characterizing and, where necessary, remediating munitions-contaminated sites. During a cleanup, a site is typically mapped with a geophysical system, based on either a magnetometer or electromagnetic induction (EMI) sensor, and the locations of all signals above some detection criterion 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 95% of objects excavated during the course of a munitions response are found not to be munitions. Current technology, as it is traditionally implemented, does not provide a physics-based, quantitative, validated means to discriminate between 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 Unexploded Ordnance (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 items that pose no threat. If these items could be determined with high confidence beforehand to be nonhazardous, some of these expensive measures could be eliminated or they could be left unexcavated entirely.

The MRP 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. If these savings were realized, the limited resources of the MRP could be redirected to accelerate the cleanup of munitions response sites, reducing real risk more quickly.

Classification is a process used to make a decision about the likely origin of a signal. Following a decade of research and development, classification technology has now been successfully demonstrated on several live sites under the Environmental Security Technology Certification Program (ESTCP). The Classification Pilot Program is validating the application of a number of recently developed classification technologies in a comprehensive approach to munitions response. The Pilot Program has thus far conducted three such demonstrations and envisions a series of continuing demonstrations at live sites representing a wide range of site conditions.

Classification has matured to the point that it is ready for use on some production projects and it is likely that it will be proposed for use in the near future, even as additional demonstrations are on-going and understanding of the capabilities and limitations is evolving. This document is intended to provide interim guidance that will allow project managers, regulators, and contractor personnel to evaluate whether classification is appropriate to a site, establish realistic expectations based on what has been demonstrated to date, and to assess whether it is being implemented correctly.

Three essential components are needed to do classification: a geophysical sensor system, a model to estimate intrinsic parameters of a target, and a classifier to make decisions about whether a signal likely came from an item of interest. Most successful applications of classification have used Electromagnetic Induction sensors.

- Successful classification has been consistently demonstrated with advanced sensors collecting stationary data over anomalies identified from a geophysical survey. These include the commercial MetalMapper, the Naval Research Laboratory's TEMTADS, and Lawrence Berkeley National Laboratory's BUD. On the same sites, analyses of these data have shown that it is possible to correctly identify 75-90+% of clutter while retaining all of the munitions on the dig list. This was achieved routinely on the Sibert and San Luis Obispo sites with medium and large munitions and by some analysts on the former Camp Butner site that also contained 37-mm projectiles.

- Some classification is possible with the commonly-used EM61 sensor data if the data collection procedures are appropriate to support classification and not just detection and the target mix is appropriate. In the demonstrations at former Camp Sibert and former Camp San Luis Obispo, all the targets of interest were either large munitions (4.2-in mortars) or a mix of medium and large munitions (2.36-in rockets and 60-mm, 81-mm and 4.2-in mortars). Classification based on EM61 data achieved correct identification of about 50% of the clutter while correctly identifying 100% of the detected munitions. However, on the Butner site which also contained 37-mm projectiles and clutter items that were of the same size, no worthwhile classification was achieved by any of the analysts using EM61 data.

The classification process relies on using the measured sensor signals to estimate parameters of geophysical models that relate to physical features of the objects giving rise to the observed data. There are important reasons that we use the model parameters instead of the data directly. The data, as measured, reflects a complex interaction of the sensor and the target. Direct data features, such as the amplitude and the shape of the anomaly, are a result of not only the intrinsic target features, but also the sensor characteristics and the relative orientation of the sensor and target. The same target measured from a different distance or orientation will exhibit different signal amplitude and decay. This clouds the interpretation of direct data features.

Finally, classifiers are used to determine the likelihood that a signal arises from a target of interest or TOI. Parameters that are meaningful in distinguishing TOI from non-TOI such as size, aspect ratio, polarizability decay, etc. are identified. In general, which parameters are meaningful will depend on the munitions of interest, the site conditions, the data quality, and other factors. Some classifiers evaluate the parameter values directly and establish mathematical relationships to determine which combinations of values make an object look like a TOI. Other classifiers rely on how well the derived parameters match to a library of signatures.

No matter what classifier is used, the final product a ranked anomaly list that orders all anomalies from highest confidence an item is a TOI to highest confidence an item is not a TOI. Since the classifier provides a likelihood and not a "yes/no" determination, a point on the anomaly list, often termed a threshold, is selected to identify which items must be treated as potential munitions.

As a cleanup project works through the classification process, there are a number of QC checks and products, many centered on QC seeds, that should be expected. These include documentation regarding selection of anomalies for inclusion on the anomaly list, correct classification of the seed items, correct classification of others items marked to be dug, and confirmation that the correct item was removed from the hole. Successful completion of all these steps will bolster confidence in the end-to-end process.

1 OVERVIEW

1.1 Motivation – Why Classification?

The Munitions Response Program (MRP) is charged with characterizing and, where necessary, remediating munitions-contaminated sites. During a cleanup, a site is typically mapped with a geophysical system, based on either a magnetometer or electromagnetic induction (EMI) sensor, and the locations of all signals above some detection criterion 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 95% of objects excavated during the course of a munitions response are found not to be munitions. Current technology, as it is traditionally implemented, does not provide a physics-based, quantitative, validated means to discriminate between 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 Unexploded Ordnance (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 items that pose no threat. If these items could be determined with high confidence beforehand to be nonhazardous, some of these expensive measures could be eliminated or they could be left unexcavated entirely.

The MRP 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. (Ref. 1) 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 MRP could be redirected to accelerate the cleanup of munitions response sites, reducing real risk more quickly.

Classification is a process used to make a decision about the likely origin of a signal. Following a decade of research and development, classification technology has now been successfully demonstrated on several live sites under the Environmental Security Technology Certification Program (ESTCP). The Classification Pilot Program is validating the application of a number of recently developed classification technologies in a comprehensive approach to munitions response. The Pilot Program has thus far conducted three such demonstrations (Ref. 3, 4, 5) and envisions a series of continuing demonstrations at live sites representing a wide range of site conditions.

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. In the evolution of geophysics approaches to munitions response, the classification process envisioned will represent another major step in providing an auditable decision record for each geophysical anomaly encountered on a site. Historically, the Mag and Flag process did not provide an auditable decision record; decisions were made in the field based on the operator's judgment which cannot be archived.

It has been documented in many demonstrations that Mag and Flag is not as reliable as once accepted. (Ref. 6,7,8) The adoption of Digital Geophysical Mapping represented a major shift to archiving and documentation and permitted fundamentally different quality control opportunities that have come to be the norm. (Ref. 9,10) Even so, anomaly selection criteria sometimes are based on amplitude or footprint shape that involve professional judgment and are not always rigorously tied to targets of interest. The process presented here will take the next step in moving to a auditable and transparent decision process applied to every anomaly detected.

1.2 About this Document

Classification has matured to the point that it is ready for use on some production projects and it is likely that it will be proposed for use in the near future, even as additional demonstrations are on-going and understanding of the capabilities and limitations is evolving. This document is intended to provide interim guidance that will allow project managers, regulators, and contractor personnel to evaluate whether classification is appropriate to a site, establish realistic expectations based on what has been demonstrated to date, and to assess whether it is being implemented correctly. This document is focused on the use of classification as a tool to avoid digging all anomalies during removal or remedial actions. It is also likely to be of value during site inspections and remedial investigations, but that is not considered here.

The document is structured in three additional sections:

- Section 2 summarizes the performance expectations based on what has been demonstrated to date and where known challenges remain.
- Section 3 contains a summary of the process by which classification could be implemented and the essential elements needed: sensors, models, and classifiers. It is assumed that the reader has a basic understanding of the MRP process and the fundamentals of geophysical sensors. References 10 and 11 contain additional background, if needed.
- Section 4 provides a description of the products that should be provided during and at the conclusion of the work and a brief description of what these products should address. Opportunities for Quality Control (QC) and Quality Assurance (QA) are discussed for each product.

It is important to note that this document deals with classification as it has been demonstrated in the Pilot Program. Most of the success has been demonstrated with EMI-based sensor systems paired with analysis methods that are well understood and have undergone several years of development and review by DoD and other interested parties. In short, these are principled, transparent processes that could be implemented by other analysts with the same result.

The label "classification" will no doubt be applied to many alternative approaches that may vary in minor details or that may be fundamentally different. The successes seen in the Pilot Program should not be expected to transfer to the latter. Careful scrutiny should be given to contractor-proposed methods that have not been independently demonstrated in blind tests. Methods that look good in retrospect will not necessarily be reproducible at other sites.

Finally, the ESTCP Classification Pilot Program will be ongoing over a period of several years. As demonstrations are completed under a variety of site conditions and as the technology continues to evolve and the remaining research problems are investigated, our expectations will change. This document will be updated periodically to reflect these changes.

2 EXPECTATIONS

The Pilot Program demonstrations have shown that successful classification is possible under the conditions tested, but have also indicated limitations to some approaches that are unlikely to be overcome and where challenges remain for others. A mix of factors will influence whether classification can be successfully applied on any given munitions response project. A general understanding of which sensors are expected to perform well against which munitions under which site conditions is emerging based on the demonstrations completed to date.

It is important to evaluate opportunities to employ classification in the context of a realistic understanding of what can be expected at the conclusion of a clean up using current practice. No clean up 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. Some form of implicit classification is already being used, either in setting the initial anomaly selection criteria or in removing anomalies from further consideration based on data characteristics such as the size or shape of the anomaly. Even with the most careful QC, some uncertainty remains that all munitions were detected and removed.

Demonstrations have been performed at the former Camp Sibert in AL, the former Camp San Luis Obispo (SLO) in CA, and the former Camp Butner in NC, with the following general characteristics:

- Sibert: Flat terrain, benign geology, low anomaly density (100-200 per acre), single munition of interest was the 4.2-in mortar; no other munitions discovered.

- SLO: Hilly terrain, moderate geology, low anomaly density (100-200 per acre), four known munitions of interest prior to the study were the 2.36-in rocket and 60-mm, 81-mm and 4.2-in mortar; others discovered during the demonstration were the 3-in stokes mortar, 5-in naval rocket, and 37-mm projectile.

- Butner: Flat terrain, benign geology, high anomaly density (up to 800 per acre), known munitions prior to the study were the 105-mm projectile, 37-mm projectile, and M48 fuze, with many other types possible; no other munitions were discovered during the demonstration.

The sections below provide rules of thumb based on the current state of the technology and performance in these demonstrations. The sections that follow are not meant to be mutually exclusive.

2.1 Sensors

Most successful applications of classification have used EM sensors.

- Successful classification has been consistently demonstrated with advanced sensors collecting stationary data over anomalies identified from a geophysical survey. These include MetalMapper, TEMTADS, and BUD, which are described in detail in Section 3. On the same sites, analyses of these data have shown that it is possible to correctly identify 75-90+% of clutter while retaining all of the munitions on the dig list. This was achieved routinely on the Sibert and SLO sites with medium and large munitions and by some analysts on the Butner site that also contained 37-mm projectiles.

- Some classification is possible with EM61 data if the data collection procedures are appropriate to support classification and not just detection and the target mix is appropriate. In the demonstrations at Sibert and SLO, all the targets of interest were either large munitions (4.2-in mortars) or a mix of medium and large munitions (2.36-in rockets and 60-mm, 81-mm and 4.2-in mortars). Classification based on EM61 data achieved correct identification of about 50% of the clutter while correctly identifying 100% of the detected munitions. However, on the Butner site which also contained 37-mm projectiles and clutter items that were of the same size, no worthwhile classification was achieved by any of the analysts using EM61 data.

2.2 Munitions Types

Classification on the Sibert site with only large munitions was demonstrated to be achievable by both commercial and advanced sensors. Using EM61 data, both developers and production geophysicists were able to routinely eliminate up to 50% of the clutter while correctly identifying all munitions. With the advanced sensor data, results were nearly perfect. On very simple sites like Sibert, it should be expected that these results would be replicated routinely.

On the SLO site, with a limited mix of medium and large size munitions, classification was also possible using both commercial and advanced EMI sensor data.

The demonstration on the Butner site which contained small munitions – 37-mm projectiles – showed that classification of these targets is possible, but it is not yet routine. Success should not be expected with commercial EM61 sensors. Some, but not all analysts, were able to achieve near-perfect results with advanced sensor data. The most successful classification employed advanced models in the processing as well.

The possible presence of unknown munitions types on a site is always a concern. In some cases, matches to libraries of signatures are used in the classification process and this method can fail if an unknown item is present or if multiple versions of a particular munition type are present and not properly taken into account. For example, the Butner site contained 37-mm projectiles with and without rotating bands, which looked very different to the advanced sensors. However, robust classifiers are more accommodating and also look at the target parameters, such as size and symmetry, regardless of a mismatch to a library. The SLO demonstration showed that unexpected munitions can be correctly classified, but it is important to understand the classifier rules. A demonstration has not yet been attempted on a site with a large mix of munitions types.

2.3 Site Conditions

The three demonstrations to date have encountered a limited range of site conditions. All of the sites have been chosen with the criteria that high quality digital data can be obtained. This series of demonstrations is planned to span several years, with the objective of further defining the types of sites where classification would be appropriate and expected to be successful.

The sensors discussed in this document are mounted on platforms that perform well in limited vegetation and terrain that can accommodate a large cart or vehicle and where the sky view allows for use of GPS. No demonstrations have been attempted in sites with severe geologic interference. On the SLO site, where the magnetometer encountered moderate geologic interference, the EM was not demonstrably affected. The SLO site also presented rocky and hilly terrain and still good results were obtained.

A wider variety of site conditions is planned for upcoming demonstrations. Emerging advanced sensors are on platforms that can go in the trees and do not require detailed spatial maps that rely on GPS for point-to-point geolocation accuracy. The first of these demonstrations is underway in the summer of 2011.

In principle, classification can be applied at any site where digital geophysical data can be collected. It has not been demonstrated using data collected from underwater or airborne platforms. It is considerably more difficult to acquire data of the quality needed for classification from these platforms, so this application is not expected in the near future.

2.4 Personnel

All of the demonstrations have involved both the developers of the sensors and the processing packages, as well as production geophysics contractors. Contractor personnel collected data with the EM61 at all sites and with the MetalMapper at Butner. The production geophysicists have been successful at analyzing the commercial EM61 data at both SLO and Sibert, and the MetalMapper data at SLO and Butner. Classification analysis requires using advanced but freely available analysis tools. The processes are documented but not yet automated, so knowledgeable and experienced geophysicists are required. As ESTCP proceeds through the demonstration series, continued involvement of production contractors is anticipated which will increase the pool of experienced geophysicists.

2.5 Anomaly Density

Classification at Sibert and SLO was demonstrated only on isolated or mildly overlapping targets. Areas of high anomaly density were deliberately excluded from the demonstrations because success was not expected. Extraction of reliable parameters from signals with multiple sources is challenging. At Butner, the target density in the demonstration was high and many signals overlapped. Classification using the EM61 was not successful: this is attributed to both the anomaly density and the presence of small munitions. Advanced analysis methods were demonstrated and some achieved success with overlapping targets. Swift progress has been made in analyzing overlapping signatures, but this remains the subject of research.

2.6 Managing Residual Uncertainty

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

The uncertainty at the conclusion of a project using classification should be thought of in the same way as the uncertainty at the end of any project. It is true that uncertainties are introduced by making an explicit decision to leave detected items undug. However, the classification process outlined here will result in a more transparent, reproducible, and better quantified final product. Finally, in the event that new information about a site becomes available or project objectives change, all of the steps in the classification process can be revisited.

3 THE CLASSIFICATION PROCESS

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.

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 target items, 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, classifiers are used to sort the signals to identify items of interest, in this case munitions, from the clutter. 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. For complex problems, sophisticated statistical classifiers combine the information from multiple parameters to make a quantitative estimate of the likelihood that a signal corresponds to an item of interest. In most cases, the required multiple parameters can only be reliably extracted from the data collected by advanced sensors.

3.1 Three Essential Components

Three essential components are needed to do classification: a geophysical sensor system, a model to estimate intrinsic parameters of a target, and a classifier to make decisions about whether a signal likely came from an item of interest. Here we will discuss the aspects of these three components that are important to determining the success of classification.

3.1.1 Sensor Systems

Digital geophysical data are required for classification. Most successful classification applications have been based on EMI data. In principle, magnetometer data can be used, but magnetometers inherently provide less information about the target being interrogated and are more susceptible to geologic interference. Both of these factors limit what is achievable with magnetometer data. This document will focus on EMI, and the discussion will be in terms of time domain systems, which are most common.

EMI sensors in the time domain transmit a pulsed electromagnetic field and sense the responses of nearby objects once the field has been turned off. Figure 3-1 shows a schematic of two cycles of this process which can be repeated as many times as is required for signal fidelity. For classification,

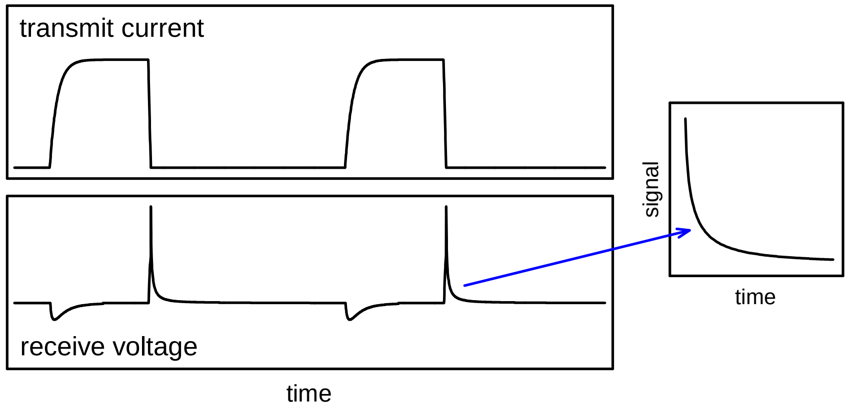


Figure 3-1 - Schematic of the time-domain EMI process. Current is pulsed through a transmit coil which results in a pulsed electromagnetic field under the sensor. While on, this field magnetizes the metal target and when rapidly turned off, excites eddy currents in the target which are sensed by a receive coil. The amplitude and decay properties of these currents are used in classification.

one important aspect of EMI is illuminating the target from multiple directions and sensing the return field in multiple directions. This allows the sensor to completely sample the target response in three dimensions. A second important parameter is when in time the signal is sampled after the transmitter is turned off. The decay of this signal is related to the wall thickness of the object and its material properties, which are important features for classification, and the longer the decay is sampled, the better this decay can be determined.

3.1.1.1 Commonly Used EMI Sensors

In this section, we discuss single-axis transmit and receive sensors that have a limited number of time gates sampling the decay of the object being sensed. The most commonly used commercial sensor is the Geonics EM-61 MK2, a schematic of which is shown in Figure 3-2.

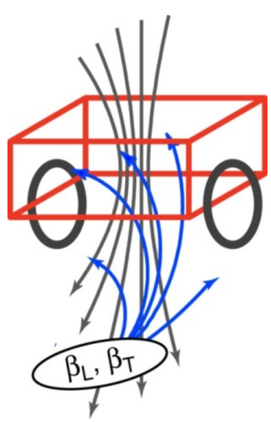


Figure 3-2. Schematic diagram of the Geonics EM61.

Most analysis approaches are based on what is known as a 3-axis target response or polarizability. This model assumes a target can be described by orthogonal dipole responses oriented along the three principal axes of the target as illustrated in Figure 3-3. This model is explained in section 3.1.2. From a data collection standpoint, the important requirement is that the object be illuminated by an EM field that stimulates the response along all three axes and that the receivers record the target response along all three axes.

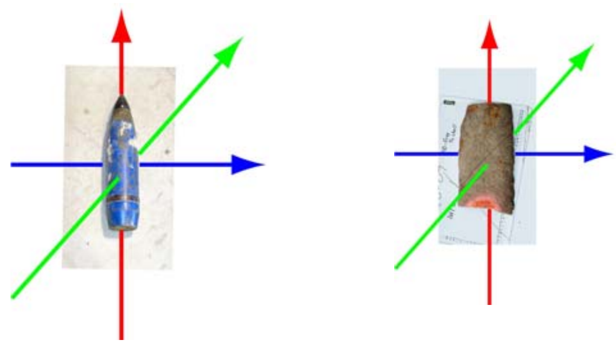


Figure 3-3. Three principal axes for a projectile (left) and a mortar fragment (right).

In a single axis system, this illumination diversity is accomplished by sampling each target at multiple locations. As shown in Figure 3-4, as the sensor passes over the item, the direction of field lines (arrows) that interact with the item changes. For example, in the second panel where the item is directly below the sensor, the field interacting with the item is almost entirely in the vertical direction. When the item is positioned far to one side of the sensor coil, as in the last panel, the field interacting with the item is primarily in the horizontal direction. Data from a number of locations over the item are combined to provide the information needed to solve for the model parameters.

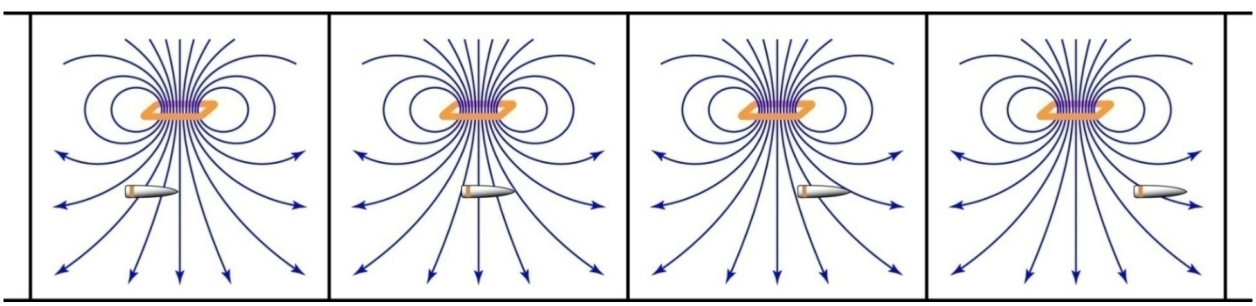


Figure 3-4. EM61 data from multiple points is combined to provide the diversity in illumination needed for classification analysis. The direction of illumination changes as the sensor is moved over an object.

The need for sampling in multiple locations introduces one of the difficulties associated with analyzing data from a single axis sensor – it is necessary to have precisely geolocated data. Positional errors of even a few centimeters can cause difficulties in doing a successful analysis and obtaining meaningful parameters. In the open, GPS can give accuracy on the order of a few cm if care is taken. Where GPS is not available, the needed accuracy is more difficult to achieve. Laser-based robotic total stations can give sub-cm accuracy anywhere there is line of sight. This is done commercially, but the equipment is more expensive and set-up is more involved.

The performance of common commercial sensors is also limited by the fraction of the signal decay that is recorded. As shown in Figure 3-5, the EM61-MK2 records four integrated readings of the target response over an interval out to 1.2 ms after transmitter turn off. Especially for large, thick walled objects, much of the time decay information that is useful for classification takes place at later time.

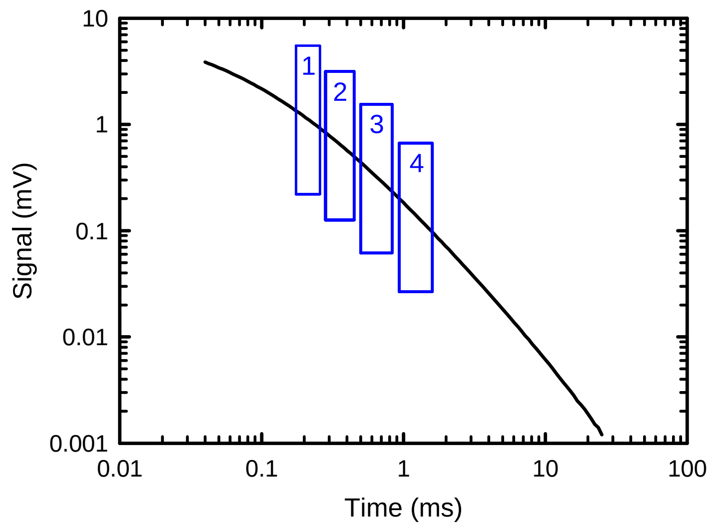


Figure 3-5. Time gates in which the EM-61 samples the decaying response of the target. The black curve is the response of a metal target.

3.1.1.2 Advanced EMI Sensors

Advanced EMI sensors that are purpose-built to support munitions classification are becoming available. These sensors have been under development in SERDP and ESTCP for nearly a decade. Detailed information can be found at www.serdp-estcp.org under Featured Initiatives\ Munitions Response\Classification.

These sensors differ from standard commercial sensors in two important aspects: they sample multiple axes at a single point in space and they are able to sample the time decay in finer steps that go out to much longer times. These features are illustrated in Figure 3-6. At present, classification-quality data are acquired by these advanced sensors in so-called “cued” mode, where the sensor is stationary over a target that has been previously detected in a survey. Work is ongoing to use dynamically acquired advanced sensor data for classification, but this has not yet been demonstrated.

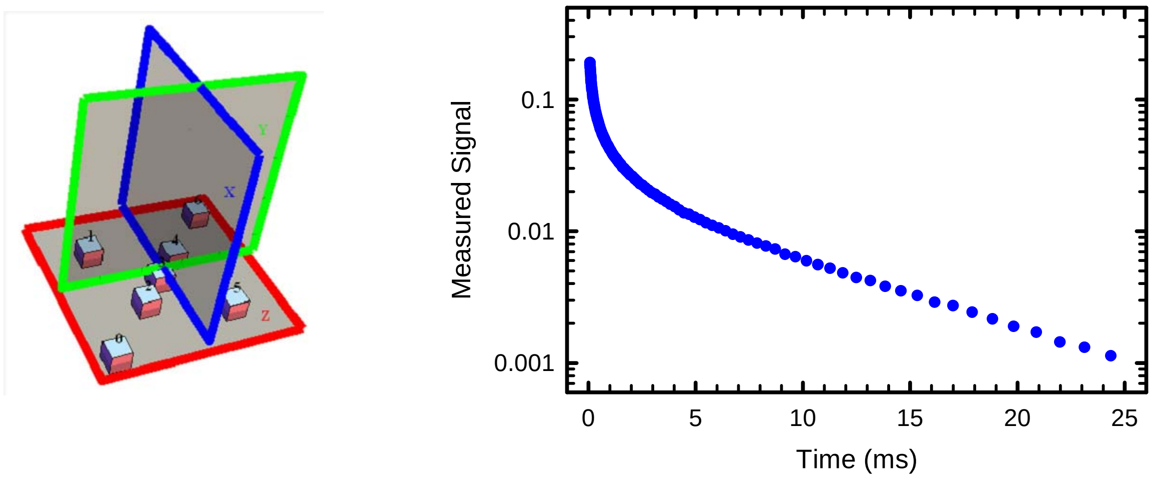


Figure 3-6. Features of advanced EM sensors. The left panel is a schematic diagram of the Geometrics MetalMapper with the three transmit coils colored red, green, and blue and the seven receive cubes shown. The right panel shows the time interval in which the TEMTADS samples the decaying response of the target.

Three advanced EMI systems, shown in Figure 3-7, have been demonstrated to date: MetalMapper, TEMTADS, and the Berkeley UXO Discriminator (BUD).

Figure 3-7. MetalMapper (top left), TEMTADS (top right), and BUD (bottom).

MetalMapper: The MetalMapper developed by Geometrics is the most mature of the advanced sensors and is available for commercial use. (Ref. 12) It is designed to be a stand-alone survey and cued detection system. The system is composed of three orthogonal 1-m x 1-m transmitters for target illumination and 7 three-axis receivers for recording the response. Its sampling is programmable, and therefore flexible. In demonstrations to date, it has measured the decay curve up to 8 ms after the transmitters were turned off. It has been used in a sled or a wheeled configuration mounted to a front loader tractor or utility vehicle. Centimeter-level GPS is used for navigation and geolocation and an inertial measurement unit (IMU) is used to measure platform orientation. (Ref. 13)

In survey mode, only the vertical field transmitter is used and the receive data recording is truncated on the order of one ms after the turn off of the transmitter. In cued mode, MetalMapper is positioned over each anomaly on its target list and collects the full suite of data while stationary.

TEMTADS: The TEMTADS, developed by the Naval Research Laboratory (NRL), is a one-of-a-kind system designed for extensive field use. The system is a 5 x 5 array of elements oriented parallel to the ground. Each array element is 0.35 m on a side and contains both transmit and receive coils. The 25 transmit elements are pulsed in sequence and data are collected from all receivers for each transmit pulse. The receive coils collect data until 25 ms after the transmit current has been turned off. The total array dimension is 2-m x 2-m, it collects data at an adjustable height between 15-35 cm above the ground surface, and it is towed by a dune buggy. Three cm-level GPS units are used for navigation, geolocation and orientation. TEMTADS is positioned over each anomaly on its target list and collects data in a stationary mode. (Ref. 14)

BUD: The Berkeley UXO Discriminator, developed by Lawrence Berkeley National Laboratory, was the first of the advanced purpose-built EMI sensors to be field tested. It is a one-of-a-kind system that is not ruggedized for extensive field use. It may be useful for projects of modest size with amenable site conditions. It is composed of three orthogonal transmitters for target illumination, and eight pairs of differenced receivers for recording the response. It measures the decay curve up to 1.2 ms after the transmitters are turned off. (Ref. 15)

OTHER EMERGING SENSORS: The three sensor systems described above are all mounted on large carts or towed platforms and can be deployed in terrain where such systems can be maneuvered. Systems on smaller platforms intended for use in more restrictive terrain and vegetation have been under development and are currently entering the field demonstration phase. These systems, shown in Figure 3-8, include Man-portable Vector Sensor (MPV), a smaller hand-held version of BUD, and smaller variants of the TEMTADS (TEMHH and TEM2X2). (Refs. 16, 17, 18, 19) They are designed with the same principles of multi-axis transmit and receive and more complete recording of the transient decay. Compromises in size, transmit moment, and other features compared to their larger counterparts will likely have some impact on their capabilities, particularly the depths to which targets can be detected and classified, but they are expected to provide advanced capability in environments not currently accessible. Initial demonstrations were conducted in the summer of 2011.

Figure 3-8. Smaller, more portable versions of advanced sensors: MPV (top left), HH-BUD (top right), TEMHH (lower left) and TEM2X2 (lower right).

Table 3-1 provides a summary of the features of all the advanced sensors above, as well as the conditions where their implementation is expected to be successful.

Table 3-1. Summary of the Advanced EMI Sensors

Sensor	Description	Effectiveness	Implementability
MetalMapper	1 meter cube 3-axis transmit 7 3-axis 10-cm receive cubes Continuous sample to 8 ms after turn off	Near-perfect discrimination demonstrated in live sites Good depth – large transmit moment	Survey and Cued Requires vehicle to maneuver Requires GPS Commercially available
TEMTADS	2 m square planar array 25 single-axis transmit 25 single-axis receive Continuous sample to 25 ms after turn off	Near-perfect discrimination demonstrated in live sites Good depth – large transmit moment	Cued only Requires vehicle to maneuver Requires GPS One of a kind Ruggedized
BUD	1 meter cube 3-axis transmit 8 single axis gradient receive pairs Continuous sample to 1.2 ms after turn off	Near-perfect discrimination demonstrated in live sites Good depth – large transmit moment	Cued only Can be maneuvered in open areas by a team of 2 Not ruggedized for extended field use
MPV	Hand carried on a wand, 12 pounds 50-cm diameter transmitter – one dimension only 5 3-axis 8-cm receiver cubes Continuous sample to 8 ms after turn off Can be manipulated in 3D to get multiple looks at the target	Good Classification on Test Site results Will have less depth capability because of smaller transmit moment	Detection and Cued modes Small and maneuverable for applications in wooded areas Does not require GPS to operate Uses locating beacon
TEMTADS HH	Hand carried on a wand, 3.5 pounds. Backpack 25 pounds. 35-cm diameter transmitter – one dimension only Continuous sample to 25 ms after turn off Can be manipulated in 3D to get multiple looks at the target	Good classification on Test Site First Live Site Demonstration Summer 2011 Will have less depth capability because of smaller transmit moment	Cued operation No GPS required Currently uses a template for data collection, IMU integration planned
TEMTADS 2X2	Transported on a small cart, 4 pounds. Overall dimension 80 cm square. Backpack 25 pounds. Four 35-cm transmitters 8-cm, 3-axis receive coils centered in each Continuous sample to 25 ms after turn off	Good classification on Test Site First Live Site Demonstration Summer 2011 Will have less depth capability because of smaller transmit moment – best to 30 cm	Detection and Cued modes Does not require GPS Fully samples target response from a single location
Handheld BUD	Mounted on small cart, 35 pounds Continuous sample to 1.2 ms after turn off 8 pairs of receivers 35-cm cube transmitter	Good classification on Test Site First Live Site Demonstration Summer 2011 Will have less depth capability because of smaller transmit moment	Cued operation Does not require GPS Fully samples target response from a single location

3.1.2 Features and Models

Forward models use the physical properties of an object to predict the signal it will produce in a sensor. For EMI, the simplest and most common model is based on the dipole response of the object along three orthogonal axes. The model uses the characteristics of the transmitted signal of the specific system being modeled to calculate the field that the object will experience and the resulting voltage measured at the receive coil. The calculated responses reflect the size, shape, and material properties of the object. The received signal is also dependent on the distance and relative orientation of the object and the sensor, as well as the properties of the receivers, which are captured in the model.

A dipole consists of two equal and opposite point charges. It is the first term in a mathematical expression commonly used to describe electromagnetic fields. At distances large in comparison to the size of the object being modeled, the electromagnetic field depends almost entirely on the dipole moment. The dipole is an approximation of the total field that is simple enough to use for efficient analysis and, in most cases, captures the important features of the target.

Figure 3-9 shows the EM61 signal predicted by a dipole model for an 81-mm mortar buried 40 cm deep at an angle parallel to the coil.

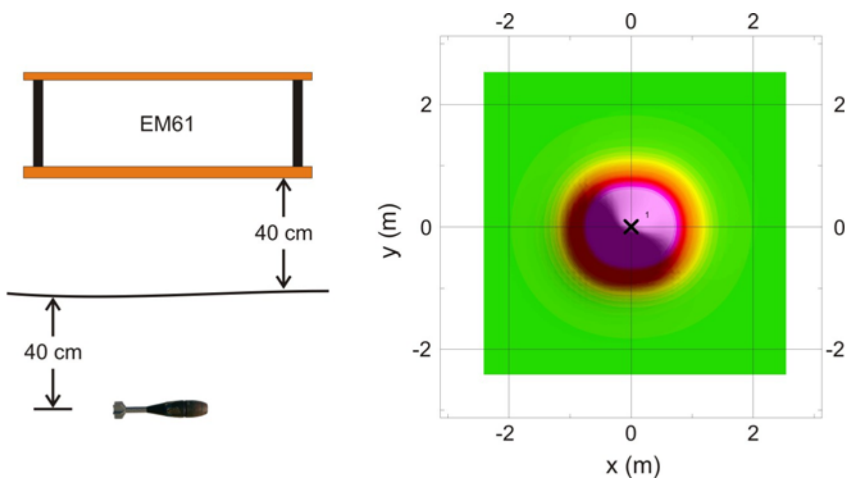


Figure 3-9. The forward model for an EM61 used to calculate the signal of a mortar at the specified depth and orientation.

The classification process relies on estimating model parameters that relate to physical features of the objects giving rise to a signal from the observed data. The forward model is at the heart of a process called inversion, which is used to interpret geophysical data in many applications. A discussion of the inversion process is available at http://www.eos.ubc.ca/ubcgif/iag/tutorials/invn-concepts/index.htm.

In the *inversion* process, the model parameter values are continuously adjusted until a solution is found that accurately reproduces the measured data.

Inversion can be used to estimate the physically meaningful parameters that appear in the forward model, related to attributes such as size and shape. A measure of how closely the measured data are reproduced gives an indication of how confident one can be that the solution is meaningful.

There are important reasons that we use the models instead of the data directly. The data, as measured, reflects a complex interaction of the sensor and the target. Direct data features, such as the amplitude and the shape of the anomaly, are a result of not only the intrinsic target features, but also the sensor characteristics and the relative orientation of the sensor and target. The same target measured from a different distance or orientation will exhibit different signal amplitude and decay. This clouds the interpretation of direct data features. Figure 3-10 shows the variability in the measured data and the consistency in the model target responses for three 37-mm projectiles.

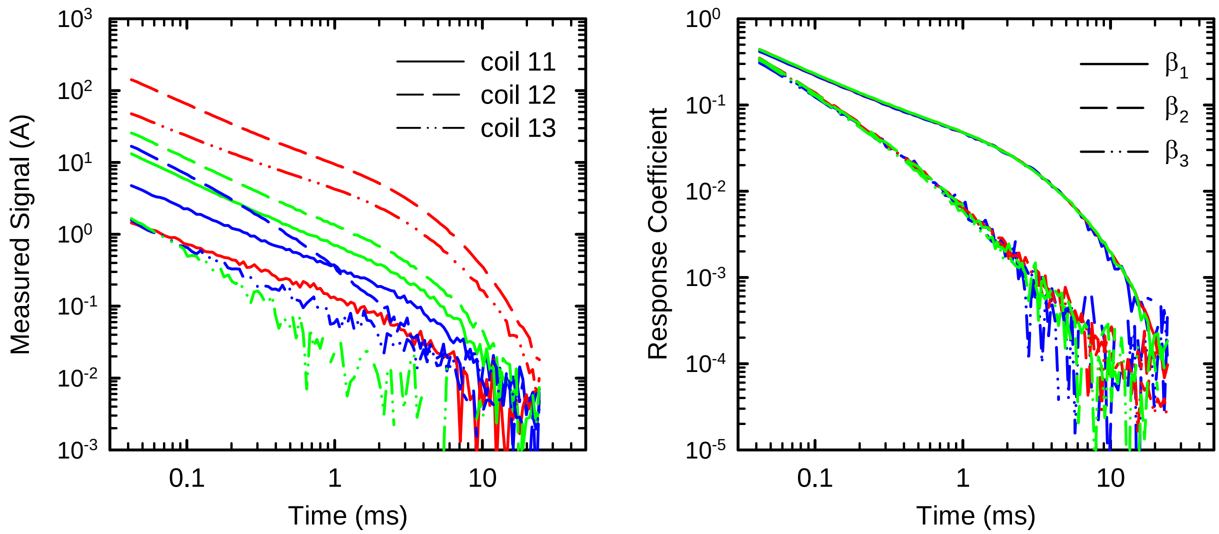


Figure 3-10. Measured data (left) from three of the middle TEMTADS sensors for three 37-mm projectiles and the resulting model parameters (right).

Commonly, the object's response coefficients, which are often referred to as polarizabilities, are represented as βs. Long cylindrical objects, such as many munitions, will have one large and two small βs, corresponding to one long axis along the body and two shorter axes perpendicular. This is the case for the 37-mm projectiles shown in Figure 3-10. In general the larger the physical item, the larger the responses or polarizabilities.

Figure 3-11 shows measured data and signal predicted from the model for a 2.75-in rocket. Note that the measured data contains noise in addition to the signal attributable to the target. Analyses are straightforward for isolated objects with good strong signals that have been sufficiently illuminated by the sensor. As the noise level increases or as the target strength decreases, for example small or deep items, the analysis can become less reliable.

Multiple objects have overlapping signatures are a known challenge. Meaningful results for objects with overlapping signatures cannot be expected with EM61 data. Data from advanced models can

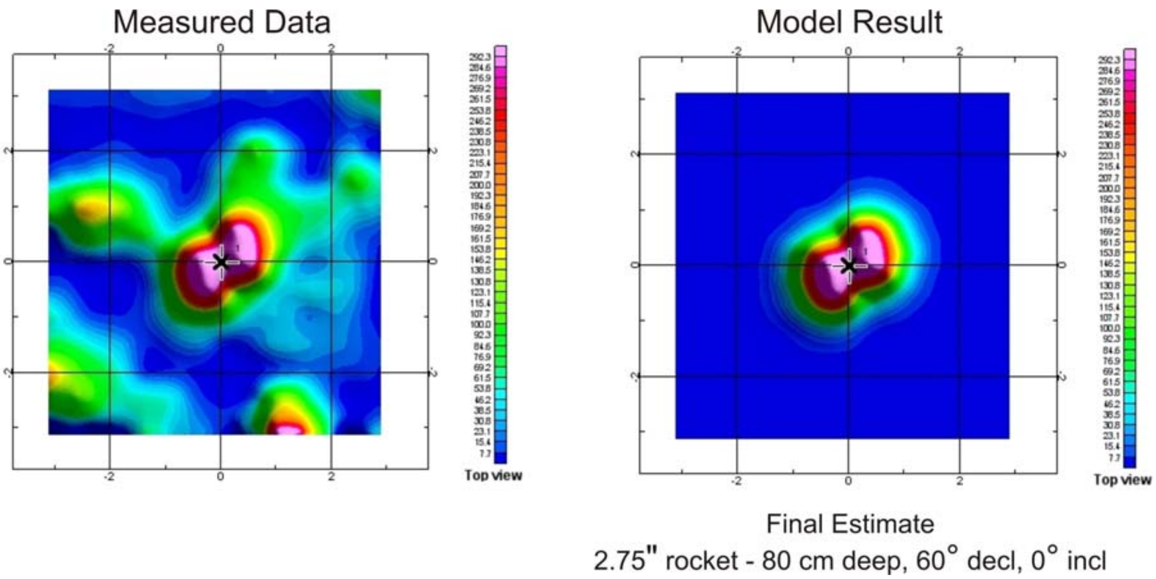


Figure 3-11. Results of inversion of geophysical data. The left panel shows the measured data. The right panel shows the model results from the best match of the type, depth and orientation of the object.

be analyzed using specialized models that account for multiple sources to obtain accurate parameter estimates that are useful for classification. However, these techniques are limited in the number of overlapping objects that can be analyzed. In some cases with severe overlap no analysis is yet possible. In this case the anomaly must be dug.

It is often not possible to do the analysis to estimate intrinsic target responses using EM61 data. The best alternative is often to revert to a simple calculation of the fall off of the measured signal using ratios of two or more of the EM61 time gates. This has the advantage that it can be calculated directly from the data and is not reliant upon a physical model. Since the EM61 has four time gates, several possible ratios exist and one or more may be useful.

The ratios of time gates derived from EM61 data directly should not be confused with the inversion-based *decay rates* that can be calculated from advanced sensor data. These analyses remove the effect of extrinsic properties such as distance and relative orientation to produce decay rates that reflect only the properties of the object. For the advanced sensors, the decay can be parameterized in a number of different ways to capture more fidelity than is possible with simple ratios.

Two methods are commonly used to describe the fall off of the EM signal. In this document, we define

Ratios calculated directly from the measured EM61 data in its four time gates

and

Decay Rates based on the inversion of the data (generally from advanced sensors) that remove the effects of distance and relative orientation.

Some common data features and model parameters are listed in Table 3-2. The model separates the intrinsic from extrinsic features. Intrinsic features are more robust for making classification decisions. Extrinsic features will be useful for improving the digging process, where good estimates of location and depth can help assure that the correct target is reacquired and dug up.

Table 3.2 Data Features and Model Parameters

Data Features	Model Parameters	
	Extrinsic	Intrinsic
Amplitude Footprint Ratio of Time Gates	Location Orientation	Polarizabilities – relate to object size and aspect ratio Decay – relates to wall thickness and material properties

Feature extraction based on a dipole model is commercially available in the Geosoft software package Oasis montaj as part of the UX-Analyze module. Models are available and documented for the EM61, MetalMapper, and TEMTADS. Decay rates can be estimated in a number of ways. The base Oasis montaj package offers this capability for the EM61.

Finally, work has been ongoing to develop more physically complete forward models. The dipole model is a greatly simplified approximation. Commercial instruments collect data of lower fidelity and cannot support an analysis more sophisticated than a dipole model. More advanced instruments take higher fidelity data that may require more physically accurate models to properly reproduce the data. These models have been the subject of research for many years and have recently been demonstrated in the field with improvements over using dipole models for the same data. (Ref. 20)

More physically complete models are not widely available in documented form yet. As these models are more widely demonstrated, they too will be incorporated into available and documented software platforms. Currently, they are best implemented through collaboration with their developers.

3.1.3 Classifiers

Classifiers are used to determine the likelihood that a signal arises from a target of interest or TOI. Features that are meaningful in distinguishing TOI from non-TOI are identified. In general, which parameters are meaningful will depend on the munitions of interest, the site conditions, the data quality, and other factors.

At the former Camp Sibert, there was only one munitions type of interest, the large 4.2-in. mortar. In this case, size alone was a good discriminant. Figure 3-12 shows the distribution of mortars and clutter and there is a clear separation, with the mortars larger than most of the clutter. Here many of the clutter items can be eliminated based on a size threshold alone, while missing none of the munitions of interest.

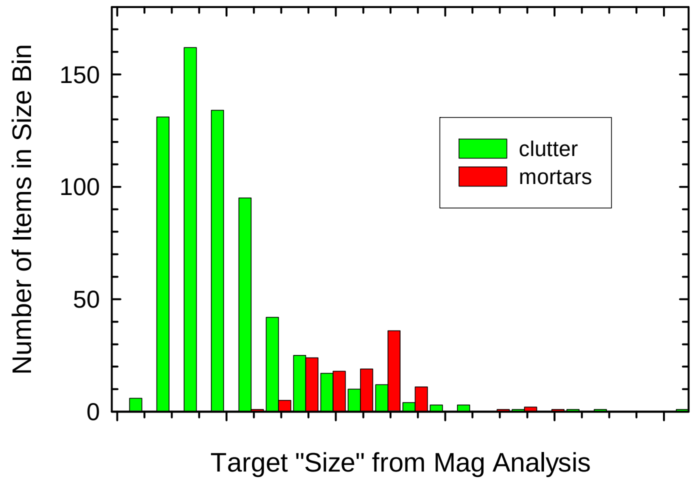


Figure 3-12. Results from Camp Sibert show a separation of mortars and clutter from only the object size estimate from magnetometer data

Figure 3-13 shows a slightly more complex example from SLO where two features, time decay and size, are used. Ideally, the TOI will cluster in one area, that is they will all look similar to one another in feature space. In this case, the rockets and mortars that make up the TOI form readily identifiable tight clusters, and all are concentrated in the area with larger sizes and longer decay times.

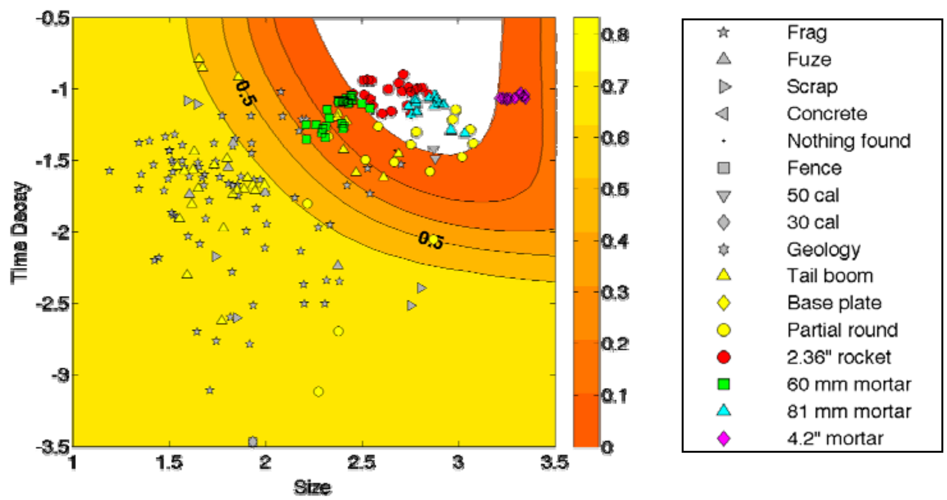


Figure 3-13. Decision boundary for classifier. The color scale refers to the probability of being clutter.

The classifier is then used to formulate a relationship to the likelihood that an item is a TOI. Some classifiers evaluate the parameter values directly and establish mathematical relationships to determine which combinations of feature values make an object look like a TOI. Other classifiers rely on how well features match to a library of signatures. In this case, if an unknown object has a set of features that are similar to the features of a known item, then the unknown object can be matched to the library. This method can often lead to very high confidence in the library matches but risks misidentifying objects that are munitions but are types not included in the library.

Classifiers can be statistical, often Bayesian, or rules-based. In either case, the ultimate product is a probability or metric that an item is a TOI. In the example in Figure 3-13, these likelihoods are represented by the false color scale that runs from 0 to 0.8, where 0.8 is the highest confidence that an item is not a TOI. These likelihoods are relative and depend on many assumptions that go into building the classifier. Generally, they do not directly translate to commonly understood probabilities, such as a coin flip. Further, the classifier does not draw a line between the TOI and non TOI. The end product is a ranked anomaly list within which a threshold must be specified.

3.2 Applying Classification to a Site - Work Flow

Figure 3-14 is a work flow diagram for how classification might be applied at a site. It is based on the classification pilot program and will serve as an outline for this section. This is not meant to be a rigid structure, with every step required exactly as described. But most classification will involve these elements.

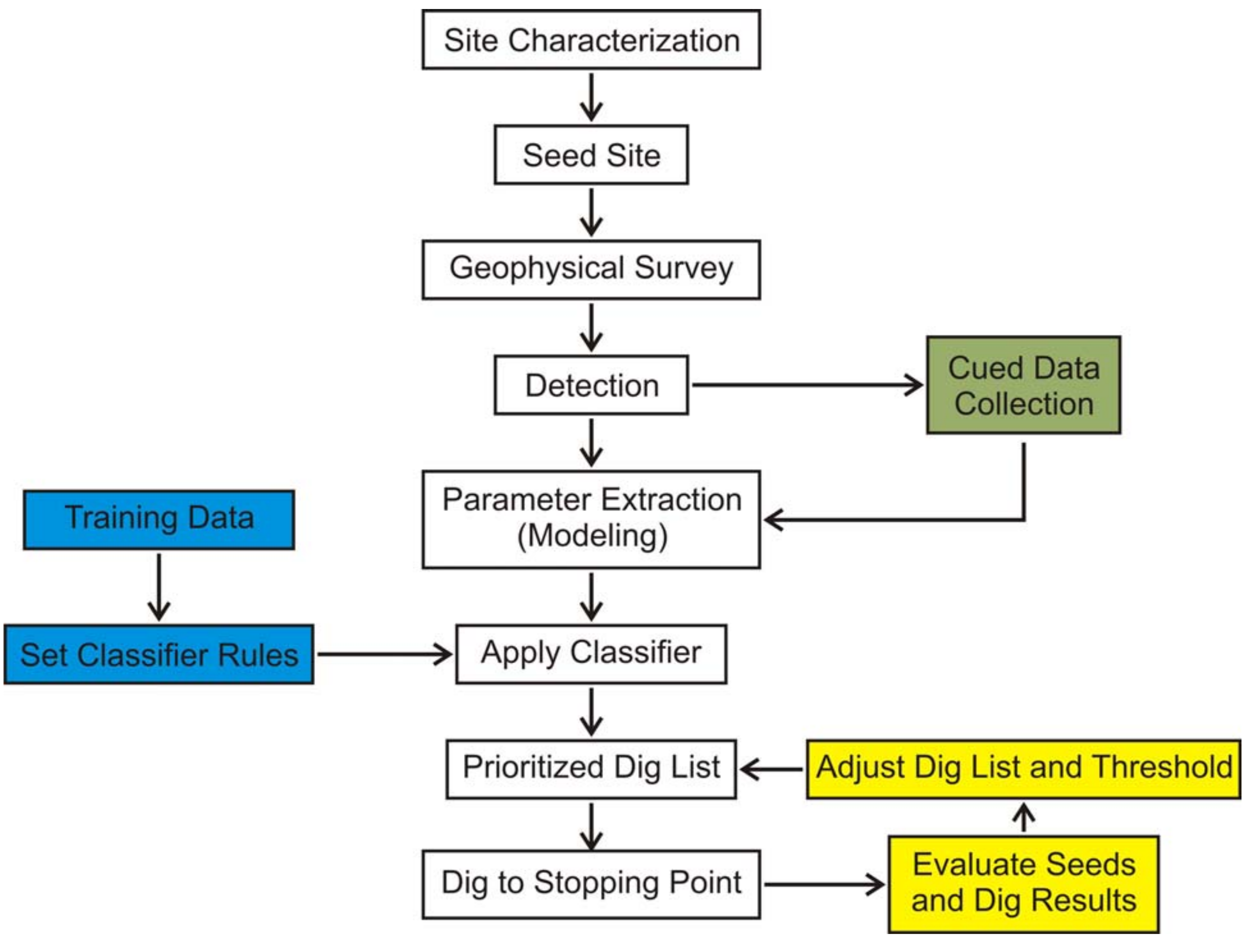


Figure 3-14. Flow chart of classification process.

3.2.1 Site Characterization

Any munitions response project requires characterization information about the site. At sites where classification will be applied, good site characterization is particularly important. The types of munitions expected, density of anomalies, clutter environment, geology and other factors will all affect whether classification is appropriate on a site and how it is best applied. In addition to usual decisions about instrument selection and deployment, this information will also define what items are included in the TOI, what is the appropriate threshold for detection, what types of items are appropriate for QC seeds, and how deeply they should be buried.

Characterization information will be available from work done during the PA, SI and RI/FS phases of the munitions response process. Additional characterization work that may be required includes transect sampling to establish anomaly density estimates and background geology and limited 100% coverage surveys of small areas with some sample digging of anomalies to gain information about the types and depths of munitions present, as well as the types of clutter.

One of the main decisions that the site team will need to make at this point is to define the targets of interest, that is the items that must be removed from the site, and those that may remain as non-TOI. Targets of interest could include:

- seeded munitions (below),
- intact munitions recovered at the site, both live and inert,
- munitions parts that are hazardous, such as certain fuzes and firing train components

The site characterization information will be useful in deciding what constitutes a target of interest, but it should be recognized that additional munitions may be discovered during the course of the project that will also be TOI.

3.2.2 Seeding

Seeding here means burying known items at recorded locations as a quality control measure and it is highly recommended. Seeds will allow a project team to confirm that targets of interest are being detected and correctly classified. Some of the considerations in a seed plan are:

- *What to seed* – The seeds should reflect the TOI. Seeds may be inert munitions items – to date we have tried to maximize use of real items for the sake of realism. If many munitions types are expected on a site, a subset may be selected as seeds. In this case, the seed items may span the types of munitions expected or may focus on items expected to be the most difficult to classify. The availability of inert versions of the actual munitions expected to use as seeds may present a challenge. In this case the use of ISOs should be considered. There is a concern that the uniformity of the ISOs may make them easier to correctly classify than munitions in situ, which are beat up and corroded. Work on how best to employ ISOs as seeds is ongoing. In either case the seeds are part of the target of interest set that must be correctly classified.

- *How many to seed* – In the pilot program we have typically seeded around 200 items for all demonstrations. This number is high enough to give reasonable statistics on detection and classification at the end of the project and, for the purpose of statistical validity, would be sufficient for a site of any size. However, seeds also serve a purpose for daily quality control. The GSV document recommends seeds be distributed such that one will be encountered on approximately a daily basis to give prompt feedback to the QC team in the event that an item is missed. (Ref. 21)

- *How deep* – The depth of the seeds will depend on the project objectives. However, the seed depths should be selected in light of the capabilities of the detection sensor – that is you can't classify what you can't detect. For example, if the most challenging TOI is a 37 mm and the expected depth to which 100% of the 37mm will be detected is one foot, there is nothing to be gained by burying classification seeds any deeper.

Industry Standard Objects (ISO) were introduced in the development of Geophysical System Verification (GSV. (Ref. 21). Three schedule 40 pipe nipples, 1-in, 2-in, and 4-in nominal pipe size, threaded on both ends, made from black welded steel, manufactured to ASTM specification serve as the ISOs. These items span munitions sizes from 37 mm to 105 mm. The items are available at most plumbing or hardware stores and produce predictable and consistent results regardless of where they are obtained. Response curves have been calculated to characterize these items, and corresponding verification measurements performed.

3.2.3 Geophysical Survey

As with any project where the objective is to detect and remove all of the hazardous munitions, a project employing classification will begin with the collection of 100% coverage digital geophysical data over the survey area. In common practice, EM61-MK2 data might be collected with a cart or an array coupled with GPS geolocation. The survey data are used for detection – to identify anomalies of interest that must be further investigated – and may be the basis for classification on some sites. If data are to be used only for detection, then industry standard practices would be adequate.

Whether standard survey data are appropriate to attempt classification will depend on site condition, the types of munitions sought, and the density of anomalies. For sites where survey data are to be used also for classification, the data specification may require tighter lane spacing than would be currently used for a detection-only survey, but otherwise normal commercial data collection and operator procedures would be used.

3.2.4 Detection - Anomaly Selection

Appropriate anomaly selection criteria are chosen and a detection list is generated from the sensor data set. The anomaly selection criteria for detection can be set in many ways. The threshold should be consistent with project objectives and consider the munitions of interest. It should also be set with awareness that the lower the amplitude threshold is set, the more small amplitude anomalies there will be on the detection list. These anomalies are hard to classify. It should also be set with awareness of the capabilities of the detection sensor.

In the pilot program, we have used anomaly selection criteria based on the amplitude of the smallest signal expected for a munitions target of interest at a specified depth, as predicted in Reference 21. The results of an exploratory dig at the site will give some information about the expected depths of items, which can be used to specify a depth. In other cases, the threshold was specified based on the known capabilities of the detection sensor against the targets of interest. For example, the threshold for EM61 survey data at Camp Butner was set at the mV level where all of the 37 mm would be expected to detected under the site noise conditions. This corresponded to the least favorable (long axis horizontal) orientation at 30 cm (1 foot) depth. Of course, some items that lie at more favorable orientations may be detected at deeper depths than that.

Anomaly Selection Criteria are used to determine which signals from a sensor will be subject to further evaluation. These signals may be referred to as **detections**.

In many cases, the single anomaly selection criterion may be an amplitude threshold. Other data-based factors such as time gate ratio or anomaly footprint may be used to further screen out potential anomalies.

We use the term **anomaly list** to refer to the anomalies that meet the selection criteria and will be passed to the classification phase.

3.2.5 Cued Data Collection

The advanced sensors currently in use were designed to work in a stationary cued mode to collect data of the quality required for classification. In the early demonstrations, cued data were collected at the locations of all detections. Currently, the demonstration program is exploring where there would be value in a hybrid survey/cued approach, in which survey data are analyzed to attempt classification. Then, cued data are collected only on those anomalies for which a high-confidence decision cannot be made based on the survey data alone. Cued data can be collected at a rate of hundreds of anomalies per system per day.

3.2.6 Feature Extraction

Each anomaly is analyzed to extract features that relate to the physical parameters of the buried object. These include model-based features, such as polarizability amplitude and decay rate, and data based features such as the time gate ratios. The analyst selects the data points for each anomaly that will used for feature estimation. In the case of model-based features, the data are submitted to an analysis routine, which returns best fit values and an indication of the quality of the match between the sensor data and the data predicted from the fit values. Criteria are established to assess whether the match is sufficiently good and, for each anomaly, an analyst decides whether the analysis has been successful and the parameter estimates can be trusted. There are some anomalies for which parameter estimation will not be successful and no classification can be attempted.

3.2.7 Apply classifier

Classification algorithms use these features to assign a probability that the item is a target of interest or not. This information is used to create a ranked anomaly list that orders all anomalies from highest confidence an item is a TOI to highest confidence an item is not a TOI. Since the classifier provides a likelihood and not a "yes/no" determination, a point on the anomaly list, often termed a threshold, is selected to identify which items must be treated as potential munitions.

Training. Classifiers typically require some amount of "training." The objective of training is to teach the classifier what the TOI on your site look like. This is typically accomplished by providing the truth data for a fraction of the excavated items. This may be 100% truth for a small portion of the site. In most cases, targeted training data may be requested to explore the origins of signals with particular features. It is also possible to derive training data from historical archived work.

3.2.8 Ranked Anomaly List and Stopping Point

The final product is a anomaly list, ranking all of the detected anomalies by the likelihood that they are targets of interest. Figure 3-15 shows a sample ranked anomaly list.

- GRAY: Anomalies where the signal-to-noise ratio (SNR), data quality, or other factors prevent any meaningful analysis were deemed "can't analyze." Nothing can be said about these anomalies so they must be dug.
- RED: Next comes those items that the analyst was most certain are TOI.
- YELLOW: A band was specified indicating the anomalies where the data can be analyzed in a meaningful way, but the derived parameters do not permit a high confidence determination of TOI or not-TOI.
- GREEN: The final items in the list were those which the analyst was most certain do NOT correspond to TOI.

• THRESHOLD: A threshold is set at the point that separates all the items that the analyst can confidently classify as not-munitions from the other categories which must be investigated as potential targets of interest. This is indicated by the heavy black dashed line.

Rank	Anomaly ID	Category	Comment
n/a	2498	0	Can't analyze
1	247	1	
2	1114	1	High confidence munition
3	69	1	
…	…	2	Can't make a decision
…	…	3	
…	…	3	
…	…	3	
…	…	3	
…	…	3	High confidence non-munition
…	…	3	
…	…	3	
…	…	3	
N	…	3	

Threshold

Figure 3-15. Ranked anomaly list for classification.

3.2.9 Dig to Stopping Point

It is envisioned that once the anomaly list is constructed at a minimum all of the anomalies in the red part of the list, those identified as high confidence TOI, would be dug, as would all of those in the can't analyze category. In addition, some anomalies in the yellow part of the list may be dug to provide additional information to the classifier and to reduce uncertainly near the threshold. This information can be used to refine the decisions about the anomaly list ranking and classifier if necessary.

3.2.10 Feedback – Evaluate Seed and Dig Results and Adjust Dig List and Threshold

Although not a formal part of the Pilot Program demonstrations, there will be an opportunity to provide the site team feedback based on the initial anomaly list. Evaluating performance stepwise will likely be essential to applying classification to a site where all anomalies will not be dug. There will be two main sources of information to evaluate.

- The first is the performance on the seeded items. Correctly identifying all of the seeds as TOI will lend confidence that the detection and classification are working as intended. If seeds are misclassified, a failure analysis is indicated.

- The second is the identity of the dug items. Only the anomalies labeled green on the final anomaly list are high confidence non-TOI. All of the others must be dug. Once everything to the threshold has been dug, the dig results are useful in assessing how well the classifier performed.

There will be an opportunity to classify/dig/evaluate and reclassify in an iterative fashion until the site team ultimately decides they are "done," that is they are as confident that the cleanup objectives for the site have been met.

The final step will be to validate the assumptions that were used to formulate all of the decision criteria, from the anomaly selection criteria to the final thresholded ranked anomaly list. The dig results will provide a great deal of information that can be compared to the initial conceptual site model – including the depths and types of munitions. The estimated parameter values of the TOIs, how tightly the parameters of common munitions form clusters, and whether there are additional clusters beyond those attributed to the known munitions all serve to verify whether the design of the classification was appropriate.

4 PRODUCTS AND QUALITY CONSIDERATIONS

In the demonstrations, all of the anomalies that exceeded the detection anomaly selection criteria were dug, providing ground truth to assess the success of the classification process. To implement classification on a project and not dig everything, the evaluation of success will need to be made with the information that is available. To allow a site team to assess the veracity of the classification process as it was applied on the project, it will be essential that all decisions are justified and transparent and detailed supporting documentation should be provided.

The primary product is the ranked anomaly list with a threshold specified to indicate the anomalies that the analyst is confident are not TOI. To make this transparent, this product should be accompanied by a detailed discussion of how the data were collected and processed and how and why various decisions were made. The analyst's report should include sections on the following, which aggregate the main steps in the flow diagram in Section 3. This information need not necessarily all be compiled to a single report. In fact, it is likely that multiple reports produced in a step-wise fashion as the project proceeds will better support oversight. This will depend on the work flow of a specific project. Regardless, these points for quality control should be present.

4.1 Data collection procedures

Data collection procedures should be traceable to project objectives, driven mostly by the targets of interest. For classification, the requirements for survey data are more stringent than for detection only. The report should cover how and why line spacing, time decay recorded, and the like were specified and what quality control procedures were used to ensure that the data meets the requirements set.

For an advanced sensor returning to a cued location, the procedures for selecting this location and ensuring that the correct target is interrogated should be outlined. Procedures should be outlined for how and when an analyst decides that sufficient data have been acquired at each location.

For both, as in any geophysical data collection for munitions response, daily checks that the sensor is working properly should be performed and documented. We recommend the procedures outlined in the Geophysical System Verification approach (Ref. 21), where a handful of items at surveyed locations are encountered daily to measure the reproducibility of the expected signal strength and geolocation accuracy.

4.2 Detection

A rigorous process should be used to set the anomaly selection criteria for detection, tied to project objectives and the conditions at the site that will govern detectability. These factors include the types and expected depths of munitions, the capabilities of the survey sensor, and the site noise. How each of these factors play into the selection of the anomaly selection criteria for detection should be discussed quantitatively. Sensor response curves predict to what depth different munitions can be reasonably expected to be detected. (Ref. 22) Specifying a depth beyond which reliable detection is expected from physics-based modeling is not recommended.

An example of this process is shown in Figure 4-1. For the EM61 cart system, the item of interest with the lowest predicted signal at a particular depth is the 60-mm mortar. The predicted signal in gate 2 (reference Figure 3-5) for this target in its least favorable orientation is plotted in the figure along with a vertical line marking the depth of interest, in this case 45 cm. The anomaly selection amplitude threshold for this sensor system was set at 11 mV based on this curve. Also plotted on the figure is the observed noise at the site. As can be seen from the figure, the anomaly selection threshold is well above the measured noise so the anomaly selection process should be relatively unambiguous for this sensor system.

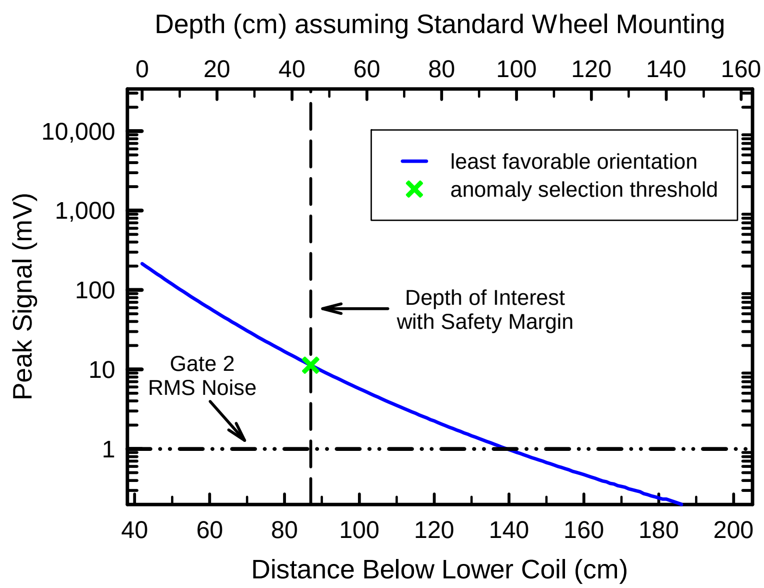


Figure 4-1. Predicted EM61 cart anomaly amplitude in gate 2 for a 60-mm mortar in its least favorable orientation. Also shown are the RMS noise measured at the site, the 45 cm depth used to set the threshold and the anomaly selection threshold used in the SLO demonstration.

There are software programs that can automatically identify threshold exceedances once the threshold has been specified. These programs are widely used and can be valuable, but automatic detection also can leave ambiguities, particularly for anomalies with large spatial area made up of many points above the threshold and in areas with overlapping signals from multiple buried items. In addition, some single items produce double-hump anomalies. When passed to a cued data collection system or a classifier, such ambiguous anomalies must be flagged as such and analyzed to determine whether one or multiple items are present.

The target-based amplitude threshold employed in setting the anomaly selection criteria in the pilot demonstration is an important component of the classification process. The number of threshold exceedances identified in the 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 anomalies exceeding the threshold 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.

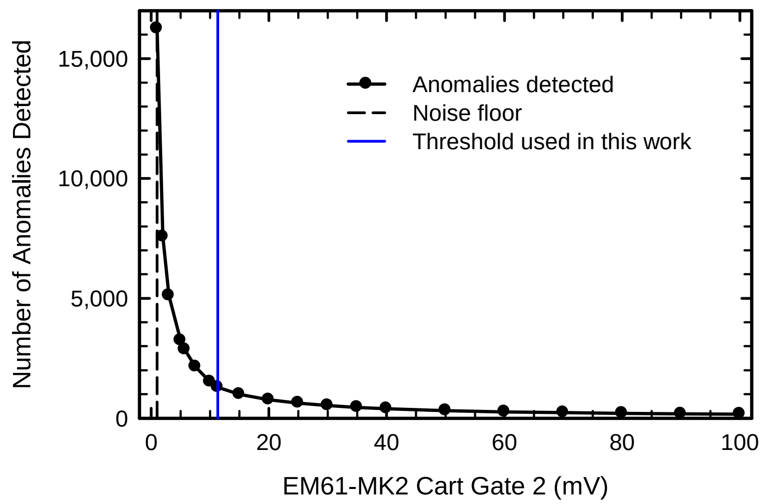
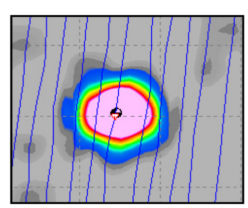


Figure 4-2. Number of EM61 cart threshold exceedances as a function of the anomaly selection threshold applied. Also plotted are the survey noise at this site and the threshold used for the SLO demonstration.

4.3 Model used for parameter estimation

Any application of classification in the near future is likely to use a dipole model. All of the forward models are similar, relying on the same physics. There are differences in the solvers, the mathematical algorithms used to find the minimum mismatch between model and data and determine when enough adjustments to the parameters values have been completed and the match is adequate.

Some of the questions that should be addressed in the documentation of the parameter estimation include: A data chip must be selected upon which to do the analysis. How is this done? If it is an automated process, does a human look at each one? What are the criteria for deciding a solution has been reached? Does a human examine the fit error and parameters for each target? What are the criteria for determining that reliable features cannot be extracted for a particular anomaly? How are these anomalies treated? How are multiple items identified and evaluated, as they pose a special challenge?

> A **data chip** is the subset of the geophyscial data that will be included in a single target analysis. For survey data, it will be the measured response from all of the locations in the vicinity of the anomaly that exceed the threshold and generally will extend to where the response returns to background.
>
> 

Ultimately, the documentation should include a summary of the analysis results for each anomaly. This summary should include the values of all of the parameters estimated, both intrinsic and extrinsic, the measure of goodness of fit of the model to the measured data, and a visual comparison of the measured data and the predicted signal from the best fit parameters. For the advanced sensors, where the dimensionality of the data is high, a simple representation is not possible. Any visual comparison will likely represent only a subset of the total data acquired for each anomaly, and what is plotted and why that representation is of value should be clearly described.

4.4 Results on the training set

The classifier may be trained using archival or site-specific data or a combination. The documentation should include graphs showing which parameters are important for this site. This may include a combination of fitted model-based and data-based parameters. Ideally, these parameters should be traceable to a physical characteristic that makes sense for the site. The report should include graphs showing results of applying the classifier on the training data, if used. It should also document any clusters of non-TOI that are present and their origin should be investigated by requesting more training data.

Figure 4-3 shows an example of how TOI cluster for two different sensors. On this site, the TOI are 105-mm projectiles, 37-mm projectiles, and M48 fuzes. The important features are the decay of the EM signal and the sum of the polarizabilities, which is a measure of item size. The panel on the left, which shows the EM61 data, indicates that these items do form identifiable clusters in these parameters, but there is considerable scatter. The panel on the right shows the same plot generated using TEMTADS data. Here the clusters for each TOI are in similar positions on the plot, but the clusters are tighter and better defined. These plots suggest that both data sets could be useful for classification but that the TEMTADS would be more successful, as it shows less overlap between the regions occupied by the TOI and the non-TOI. Clearly the more overlap there is between classes, the fewer anomalies will be recognized as high confidence non-TOI.

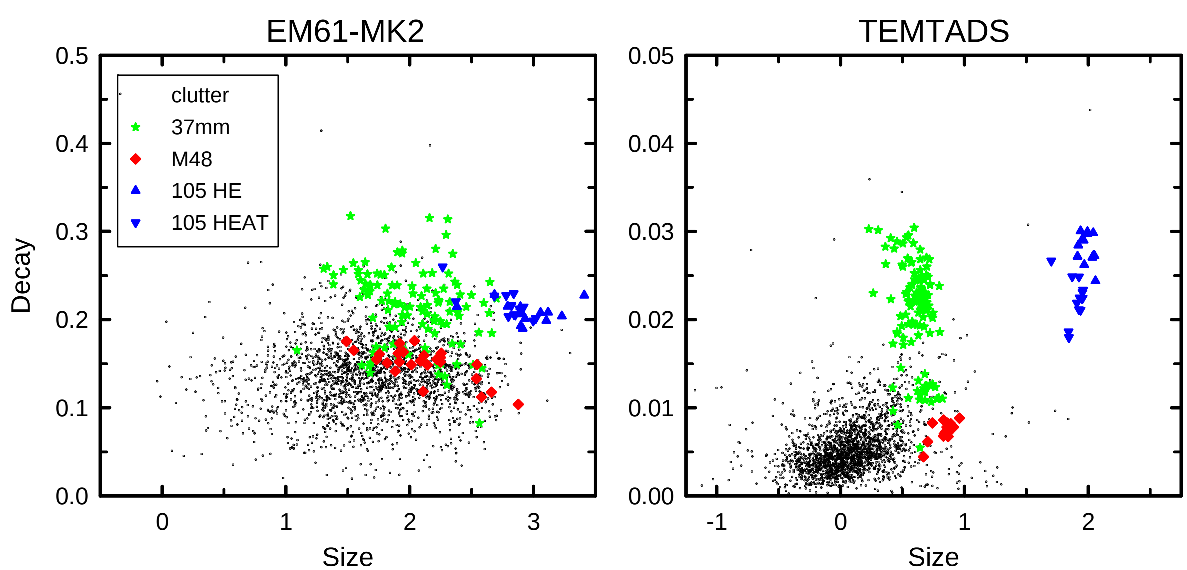


Figure 4-3. Plots showing the parameters derived from the EM61 and TEMTADS at Camp Butner using a dipole model.

In complex classification problems where there are a large number of variables, resulting in multidimensional feature space, such plots are not feasible. For munitions classification using EM61 data this is unlikely to occur. There are a limited number of parameters to explore and a few have generally dominated on live site demonstrations to date and they have been physically meaningful for the conditions at the site. When munitions of interest were large, then parameters related to object size were important. If, however, the TOI span a wide range of sizes of and clutter is of similar size, then size should be suspect and decay rate may be more relevant. This should all be explained in the report.

The advanced sensors on the other hand, collect data of much higher fidelity and classification is often accomplished using many more features than can conveniently be represented on a two-dimensional plot. Even in those cases, a pair of simplified features such as a size parameter and a decay parameter is often plotted for purposes of explanation.

The final outcome from the training step is selecting the threshold for deciding a TOI versus non-TOI. The plots in Figures 4-3 show how the target types separate in feature space but do not specify a TOI/non-TOI threshold. At least an initial threshold should be specified at this point – it could be updated based on results from the seeds or initial digging – and the report should provide a qualitative and quantitative discussion of what this threshold means in terms understandable to the site team. For example, for a statistical classifier the boundary may correspond to a likelihood of a missed target less than 0.01, under certain assumptions. Any assumptions should be carefully articulated. For a rules-based classifier the boundary may represent some distance in feature space from the TOI.

4.5 Feedback - Results on seeds and initial digging

In the pilot study, all anomalies were dug and the success of the classification analysis could be established in hindsight. This will not be possible for an implementation of classification in which all anomalies are not dug. Figure 4-4 is an example of the information that is available at the end of a demonstration, when all of the ground truth is known. It plots the results for all anomalies. The list is stepped through in rank order and the vertical axis represents when a TOI is correctly classified, both the seeds and any TOI encountered in the digging. The horizontal axis represents the number of clutter items encountered. In hindsight, we know that this analyst correctly identified nearly 2000 of the 2100 clutter items on the site, because we excavated all 2100 of them.

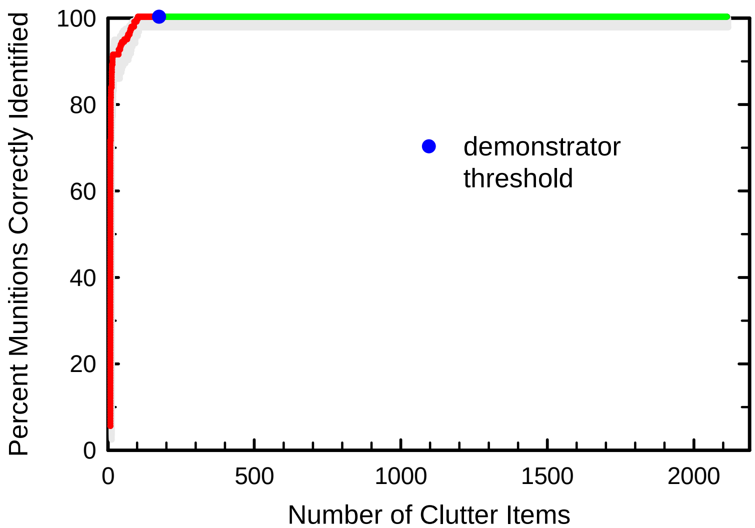


Figure 4-4. Performance from analysis of the TEMTADS cued data from Camp Butner, when full ground truth is known after all items have been dug.

On production site, the information would not all be available. Instead, the site team would be provided with the results of the initial pass at classification. This will include information on both the seeds and dig results from all anomalies in the TOI list. The report should address: Were any seeds missed at the threshold? What was missed? Were they in the anomaly list and where? Were live munitions found? At what rank were they in the list? Were the recovered items found consistent with the estimated parameters? Were there clusters in feature space other than TOI that have not been investigated? Figure 4-5 shows an example where only two munitions types were expected and, in addition to the clusters that correspond to the parameters of those munitions, an unexpected cluster is also seen, which should be investigated. It is possible that it represents commonly occurring clutter or unexpected munitions.

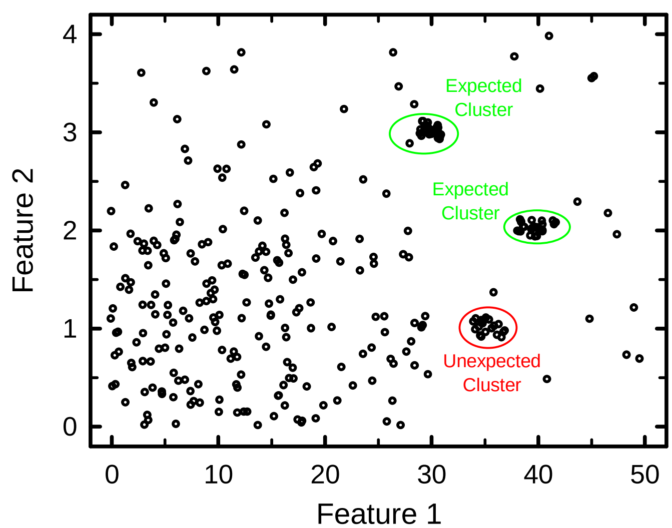
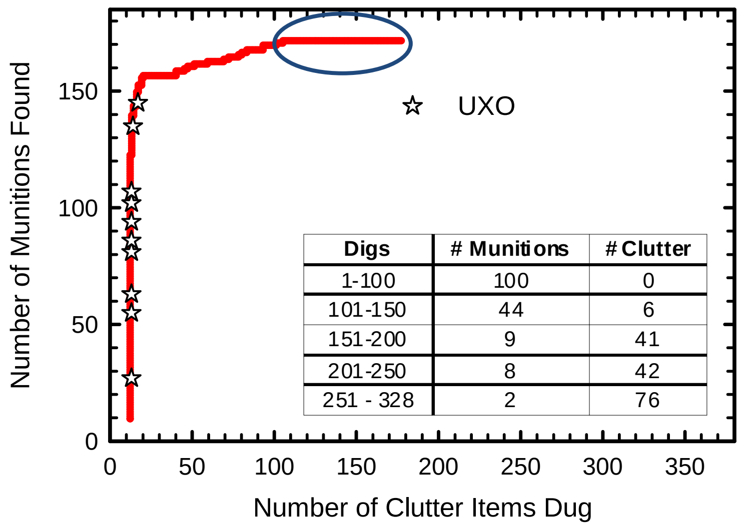


Figure 4-5. Example showing clusters: the green correspond to the expected munitions types and the origin of the red is unknown

Figure 4-6 is an example of how this feedback might be provided. No seeds were missed and seven UXO were found in this example – all were included in the high confidence TOI category. These are shown by the stars in Figure 4-6. The curve looks good, with a strong vertical rise, indicating that the classifier was good at identifying TOIs. The real UXO are evenly distributed among the high confidence TOI, indicating that they are well captured by the classifier. The plot appears to have plateaued, with the last ~100 anomalies in the TOI part of the ranked anomaly list having turned out to be non-TOI. These anomalies are circled in Figure 4-6. This suggests that the classifier set a conservative threshold. The site team must use this information to decide whether it is satisfied that the results meet their site cleanup objectives or additional digs are required.



Digs	# Munitions	# Clutter
1-100	100	0
101-150	44	6
151-200	9	41
201-250	8	42
251 - 328	2	76

Figure 4-6. Initial dig results from analysis of the TEMTADS cued data from Camp Butner. The last ~100 anomalies in the TOI list, circled, were not TOI.

Figure 4-7 shows another example, where the results are not so good. The top panel is the first attempt at classification. In this case, the contractor missed 4 seeds and the curve appears to still be climbing at the end of the TOI list. At this point, the quality officer must decide whether to let the contractor know a seed was missed and whether to reveal the type, or exactly which item was missed. The QC officer knows where the missed seeds are on the analyst's ranked anomaly list (i.e if they are near the threshold or way down into the non-TOI portion of the list). This may influence her decision.

In the case shown in Figure 4-7, the site team gave the contractor feedback that seeds were missed, but not which ones, and let the analysts go back and revisit the classifier rules and threshold. In the second pass, the classifier still missed two seeds. The site team must now decide how to proceed. The contractor may continue to refine the classifier/threshold until all the seeds are captured or the team may become convinced this sensor/classifier combination is not going to be successful at this site. In the latter case, the options include reanalysis and classification of the sensor data using another approach, collection of additional geophysical data, or a decision that all anomalies must be investigated to meet the site cleanup objectives.

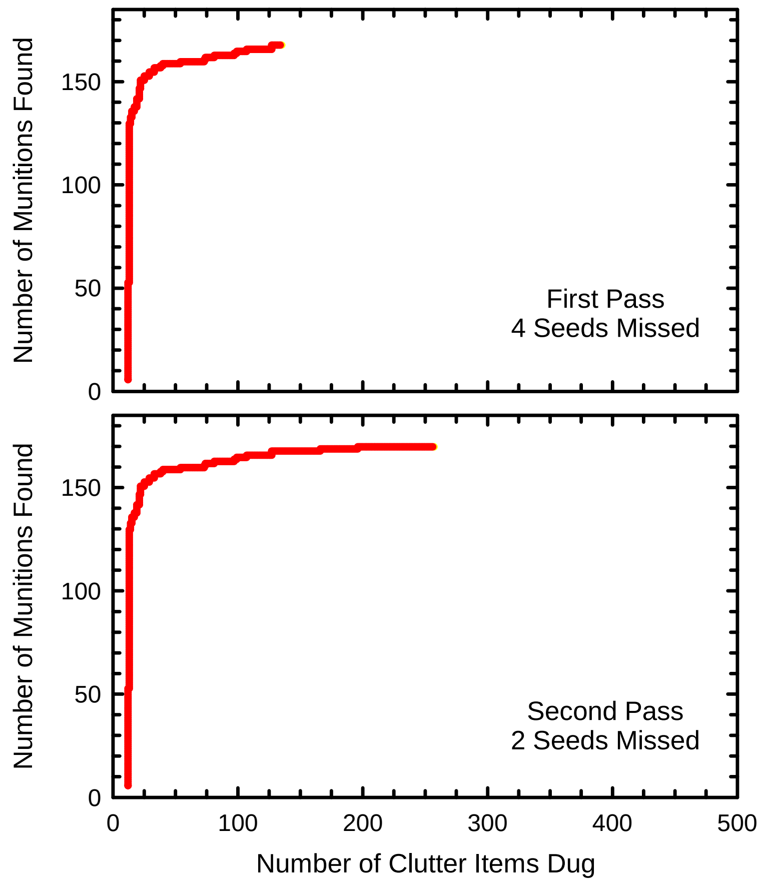


Figure 4-7. Dig Results from Camp Butner. The top panel shows an initial pass in which the contractor missed 4 seeds. The lower panel shows the results after the data were re-analyzed and still two seeds were missed.

In this example, the plots shown in Figure 4-7 were prepared after the anomalies marked as high confidence TOI were investigated. Of course the QC team knows the locations of the seeds as soon as they are emplaced and this check that the QC seeds were correctly classified could occur before any anomalies are dug.

4.6 Anomaly Resolution

The intent of classification is to remove targets of interest and avoid having to dig every piece of metal. This will present a new challenge in quality control for digging. How do you know you retrieved the right target? Others nearby will remain. The classification process will provide information about the size, shape, and depth predicted for the target that is not currently available. This information should be transmitted to the dig crews. A geophysicist should verify that the item recovered from the hole is consistent with the data interpretation. For each, a photo and details about retrieved item should be documented.

One option would be to measure each item recovered with the same sensor that was used to do the classification. Inversion of the data from a post recovery measurement would allow the QC officer to verify that the signatures match and the right item was recovered. This could be problematic for multiple items recovered from the same hole, as their exact relative orientation will likely be unknown. However, a multi-source solver should estimate consistent target parameters for the dominant objects before and after excavation.

There are a number of ways that this type of QC check could be implemented. It could serve as a real-time QC check for all items. Depending on resources, it could also be implemented for the first one hundred items dug up to build confidence in the process or applied only to those items where there is some uncertainty. Since the UXO will tend to cluster in feature space as like items with known signatures, this could be most informative in building confidence in the correct identification of the non TOI.

4.7 Process QC vs Validation Sampling

Validation sampling is not recommended for use in classification. It may sound appealing to randomly sample some of the anomalies classified as non-TOI as a quality check. However, it should be recognized that little confidence, in the statistical sense, is likely to be gained. The expectation from experience on many sites is that there will be few UXO – a handful in thousands of total anomalies is not uncommon. A random sample of a few percent of items in the green high confidence non-TOI category is unlikely to encounter a missed TOI, even if it were present.

Any validation sampling should be directed by carefully defined objectives. This will involve digging some items that were classified as non-hazardous. Examples include sampling regions of interest in feature space, no matter how they were classified, to confirm the parameters are physically reasonable or sampling clusters of unknown items. Of course, all the clusters identified as TOI are dug anyway. Many of these might be sampled in building the classifier as training data. The quality control for a well executed classification project should include sampling any clusters as part of the building the classifier. Additionally, confirming the inversion result in non-sampled areas of feature space including anomalies classified as non-hazardous can build confidence in the process.

For most projects, process QC will be based on the blind seed program. The QC team will always check that the seed items were placed on the anomaly list (anomaly selection threshold), correctly classified (feature extraction and classifier performance), and returned by the dig team (anomaly resolution). Successful completion of all these steps will bolster confidence in the end-to-end process.

5 REFERENCES

1. Defense Environmental Programs Annual Report to Congress, Fiscal Year 2009, http://deparc.xservices.com/do/home

2. Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, *Report of the Defense Science Board Task Force on Unexploded Ordnance*, November 2003, http://www.acq.osd.mil/dsb/reports/ADA419970.pdf

3. ESTCP, *ESTCP Pilot Program Classification Approaches in Munitions Response*, November 2008, http://serdp-estcp.org/Tools-and-Training/Munitions-Response/Classification-in-Munitions-Response/ESTCP-Pilot-Program-Classification-Approaches-in-Munitions-Response-Camp-Sibert-Alabama/ESTCP-Pilot-Program-Classification-Approaches-in-Munitions-Response-Camp-Sibert-Alabama

4. ESTCP, *ESTCP Pilot Program Classification Approaches in Munitions Response, San Luis Obispo, California*, May 2010, http://serdp-estcp.org/Tools-and-Training/Munitions-Response/Classification-in-Munitions-Response/ESTCP-Pilot-Program-Classification-Approaches-in-Munitions-Response-Camp-San-Luis-Obispo-California/ESTCP-Pilot-Program-Classification-Approaches-in-Munitions-Response-Camp-San-Luis-Obispo-California

5. ESTCP, *ESTCP Pilot Program Classification Approaches in Munitions Response, Camp Butner, North Carolina*, June 2011, http://serdp-estcp.org/Tools-and-Training/Munitions-Response/Classification-in-Munitions-Response/ESTCP-Pilot-Program-Classification-Approaches-in-Munitions-Response-Camp-Butner-North-Carolina/ESTCP-Pilot-Program-Classification-Approaches-in-Munitions-Response-Camp-Butner-North-Carolina

6. J. R. McDonald, H. H. Nelson and R. A. Jeffries, *MTADS* Live Site Demonstration, Pueblo of Laguna, 6 July – 7 August 1998, NRL/PU/6110--00-398.

7. U.S. Army Aberdeen Test Center, Aberdeen Proving Ground, MD, *Final Report for the Evaluation of Unexploded Ordnance (UXO) Detection Technology at the Standardized UXO Test Sites Aberdeen and Yuma Proving Grounds.* November 2007. DTC Project No. 8-CO-160-UXO-021 Report # ATC-9379. http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA474960

8. J.R. McDonald and H.H. Nelson. Naval Research Laboratory, *Man-Portable Adjuncts for the MTADS*, June 2003, http://www.serdp.org/content/download/5574/77395/file/UX-9811-FR-01.pdf

9. Department of Defense and Environmental Protection Agency, *Management Principles for Implementing Response Actions at Closed, Transferring and Transferred (CTT) Ranges*, (Principle 4), 2000, Retrieved June 28, 2011, from: http://www.epa.gov/fedfac/documents/uxo_principles.htm

10. US Environmental Protection Agency, Federal Facilities Restoration and Reuse Office *Munitions Response Guidelines*, 27 July 2010, http://www.epa.gov/fedfac/documents/docs/munitions_response_guidelines.pdf

11. USACE Publication Number: EM 1110-1-4009, *Engineering and Design - Military Munitions Response Actions*, 15 June 2007,. http://140.194.76.129/publications/eng-manuals/em1110-1-4009/toc.htm

12. Geometrics MetalMapper webpage, http://www.geometrics.com/geometrics-products/geometrics-electro-magnetic-products/metal-mapper/

13. ESTCP Project MR-200603, MetalMapper: A Multi-Sensor TEM System for UXO Detection and Classification, Fact Sheet. http://www.serdp.org/Program-Areas/Munitions-Response/Land/Sensors/MR-200603/MR-200603

14. ESTCP Project MR-200601, Combined Magnetometer/Electromagnetic Induction Array for Cued UXO Discrimination (TEMTADS), Fact Sheet. http://www.serdp.org/Program-Areas/Munitions-Response/Land/Sensors/MR-200601/MR-200601

15. ESTCP Project MR-200437, A Multisensor System for the Detection and Characterization of UXO (BUD), Fact Sheet. http://www.serdp.org/Program-Areas/Munitions-Response/Land/Modeling-and-Signal-Processing/MR-200437/MR-200437

16. ESTCP Project MR-201005, Man-Portable Vector EMI Sensor for Full UXO Characterization (MPV), Fact Sheet. http://www.serdp.org/Program-Areas/Munitions-Response/Land/Sensors/MR-201005/MR-201005/

17. ESTCP Project MR-201166, Handheld BUD Demonstration, Fact Sheet. http://www.serdp.org/Program-Areas/Munitions-Response/Land/Live-Site-Demonstrations/MR-201166/MR-201166

18. ESTCP Project MR-200807, Handheld Electromagnetic Induction Sensor for Cued UXO Discrimination (TEMHH) Fact Sheet. http://www.serdp.org/Program-Areas/Munitions-Response/Land/Sensors/MR-200807/MR-200807

19. ESTCP Project MR-201165, Advanced EMI Data Collection Systems' Demonstration, Fact Sheet. http://www.serdp.org/Program-Areas/Munitions-Response/Land/Live-Site-Demonstrations/MR-201165/MR-201165

20. B. Barrowes, F. Shubitidze, *Isolating and Discriminating Overlapping Signatures in Cluttered Environments*, SERDP Project MR-1664, Interim Report, May 2011, http://www.serdp.org/content/download/10458/129799/file/MR-1664-IR.pdf

21. *Geophysical System Verification (GSV): A Physics-Based Alternative to Geophysical Prove-Outs for Munitions Response*, July 2009, http://serdp-estcp.org/content/download/7426/94837/version/1/file/GeoSysVerif-July-09-FINAL.pdf

22. Naval Research Laboratory, *EM61-MK2 Response of Standard Munitions Items*, NRL/MR/6110--08-9155, October 2008, and Naval Research Laboratory, *EM61-MK2 Response of Three Munitions Surrogates*, NRL/MR/6110--09-9183, March 2009, http://serdp-estcp.org/Tools-and-Training/Munitions-Response/Geophysical-System-Verification/GSV-Response-Calculator

Appendix A

Classification Applied to Munitions Response: A Site Team Case Study

1 Introduction

1.1 Why Classification?

The Munitions Response Program (MRP) is charged with characterizing and, where necessary, remediating munitions-contaminated sites. During a cleanup, a site is typically mapped with a geophysical system, based on either a magnetometer or electromagnetic induction (EMI) sensor, and the locations of all signals above some detection criterion 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 95% of objects excavated during the course of a munitions response are found not to be munitions. Current technology, as it is traditionally implemented, does not provide a physics-based, quantitative, validated means to discriminate between 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 Unexploded Ordnance (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 items that pose no threat. If these items could be determined with high confidence beforehand to be nonhazardous, some of these expensive measures could be eliminated or they could be left unexcavated entirely.

The MRP 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. (Ref. 1) 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 MRP could be redirected to accelerate the cleanup of munitions response sites, reducing real risk more quickly. As an added benefit, fewer excavations will often lessen the environmental impact of a remediation project.

Classification is a process used to make a decision about the likely origin of a signal. Following a decade of research and development, classification technology has now been successfully demonstrated on several live sites under the Environmental Security Technology Certification Program (ESTCP). The Classification Pilot Program is validating the application of a number of recently developed classification technologies in a comprehensive approach to munitions response. The Pilot Program has thus far conducted a number of such demonstrations and envisions a series of continuing demonstrations at live sites representing a wide range of site conditions.

1.2 About this Document

Classification has matured to the point that it is ready for use on some production projects and it is likely that it will be proposed for use in the near future, even as additional demonstrations are on-going and understanding of the capabilities and limitations is evolving. Implementing classification will require site managers to evaluate data products that may be unfamiliar and make decisions in new ways. This document is intended to provide some guidance in this task; it describes an exercise in which the ESTCP Live Site Advisory Group acted as a mock site team as they evaluated the results of the demonstration at Pole Mountain conducted in summer 2011.

Even though the underlying demonstration at Pole Mountain involved complete validation for scoring, a real classification project would only have ground truth for those items required to be dug; decisions would have to be made in the absence of complete ground truth. For this exercise, the Advisory Group was only provided the ground truth that would be available in a real project.

In this document, the data products expected from a classification contractor will be presented and the reasoning behind the Advisory Group's decisions will be discussed. While this exercise should not be considered an exhaustive manual for the implementation of classification on munitions response projects, it will serve as a useful introduction to the concepts that will be encountered.

2 Site Description

The Pole Mountain Target and Maneuver Area (PMTMA) is a 62,448-acre site located east of Laramie, WY. The PMTMA was established in 1879 as the Fort D.A. Russell Wood and Water Reserve. The land status alternated between national forest and military reservation from 1897 to 1925. The Pole Mountain area has also been known as the Crow Creek Forest Reserve, Fort D.A. Russell Target and Maneuver Range, Fort Francis E. Warren Target and Maneuver Range, Pole Mountain Reservation, Pole Mountain Training Annex, and Warren Training Annex. It was extensively used before 1959 as a target and maneuver area by the Army, the Reserve Officers Training Corps, the Citizens Military Training Corps, various National Guard units, and the Department of the Air Force. The site is now part of Medicine Bow National Forest.

There are several Munitions Response Sites at the Pole Mountain FUDS; the ESTCP demonstration was conducted in the portion of the site referred to as the Bisbee Hill Maneuver Area. An aerial photo of the demonstration area is shown in Figure 2-1.

A variety of munitions have been reported as used at PMTMA. Physical evidence for the following items was discovered during the recent Remedial Investigation:

- Projectiles containing high explosive (HE) filler (37-mm to 155-mm, and 2.95-inch);
- Shrapnel projectiles (75-mm and 3-inch);
- 3-inch Stokes mortars (practice, fuzed); and
- 60-mm mortars containing HE filler.

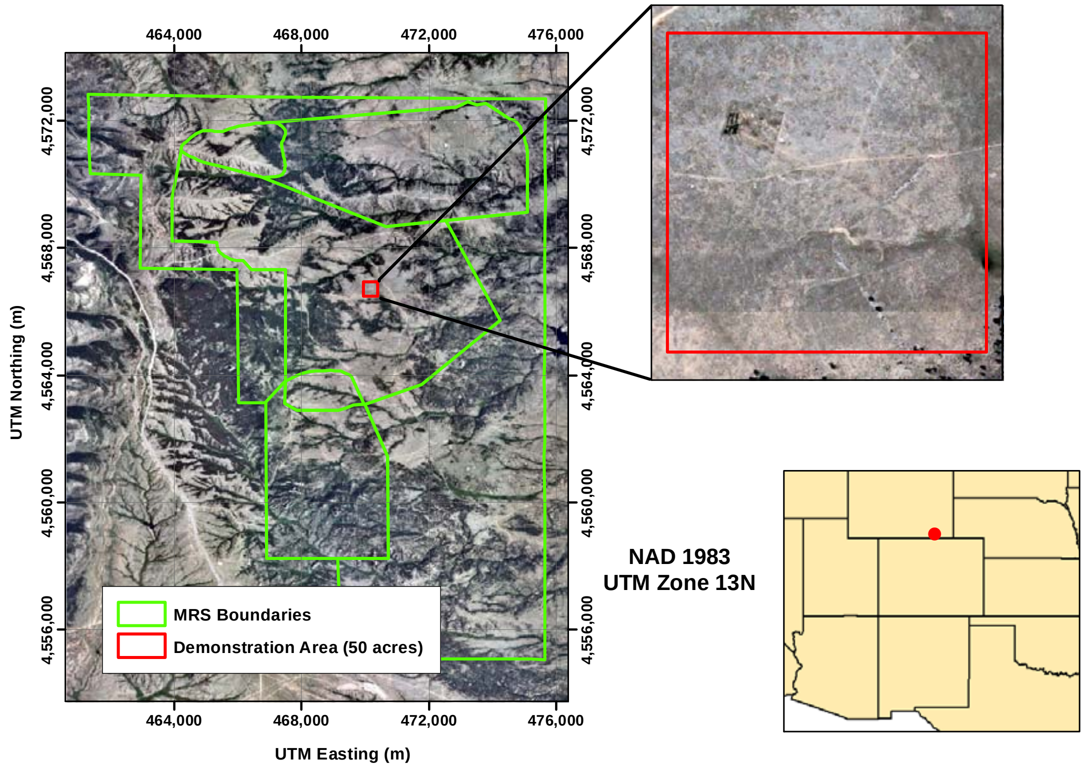


Figure 2-1. Aerial photo of part of the Pole Mountain Target and Maneuver Area showing the demonstration site

3 Project Overview

3.1 Project Objective

The project area is used for recreational purposes, primarily hiking and camping. No development is anticipated at the site since it is located in a National Forest. Based on these considerations, the site team objectives are to remove all 37-mm projectiles down to one foot (30 cm).

The detection sensor selected for use at this site was the EM61-MK2. This is the predominate sensor used for munitions response and is well suited for detection of small munitions. The site manager will not allow vehicular surveys so, even though the site is flat and open and therefore ideal for towed array surveys, the initial survey was conducted using a cart-mounted EM61-MK2 pulled by a geophysicist with a cm-level GPS system used for sensor location.

3.2 Anomaly Selection Threshold

The anomaly detection threshold for the EM61-MK2 detection survey using a 0.5-m lane spacing was set at 5.2 mV in gate 2 as shown in Figure 3-1. In addition to the objective of detecting all 37-mm projectiles to 30 cm, this threshold corresponds to detection of all 60-mm mortars to 60 cm, all 75-mm projectiles to 80 cm, and all 3-in stokes mortars to 85 cm. The survey noise measured at this site is also shown in this figure. The anomaly selection threshold is well above the measured noise so we anticipate good detection performance with minimal interference from survey noise.

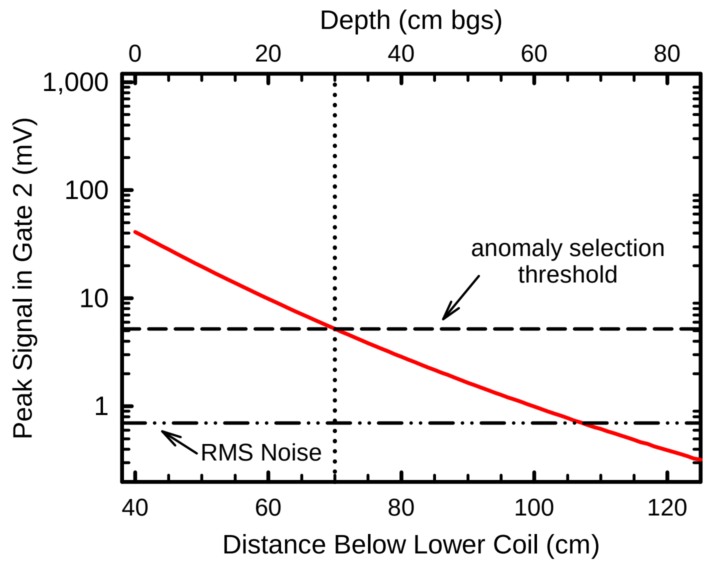


Figure 3-1. EM61-MK2 signal in gate 2 expected from a 37-mm projectile in its least favorable orientation (red line), the depth of interest at this site (vertical dotted line), and the resulting anomaly selection threshold (horizontal dashed line). Also shown is the survey noise measured at this site.

3.3 Project Workflow

After assembling any relevant historical documents and establishing the project objectives, the exercise was carried out as shown in Figure 3-2. It is likely that most implementations of classification for munitions response will follow a similar work flow (Ref. 3).

3.3.1 Instrument Verification Strip and Blind Seeds

The GSV process (Ref 4.) was followed for this project. This entails establishment of an Instrument Verification Strip and reliance on blind seeds. Inert versions of the munitions expected to be present on site were used as seeds. Since it is often difficult to obtain enough inerts, small Industry Standard Objects were used to supplement the inert seeds. Details of the seeds emplaced on the site are given in Table 3-1. The seeds were divided into QC seeds and extra TOI to ensure good statistics since the demonstration area is small and not likely to contain a significant number of TOI.

Table 3-1. Detail of the seed items emplaced for this project

Item	Number	Depth Range (cm)
37-mm projectile	43	15 – 30
57-mm projectile	10	20 – 35
60-mm mortar	41	30
75-mm projectile	25	20 – 40
3-in stokes mortar	1	30
small Industry Standard Object	40	15 – 25

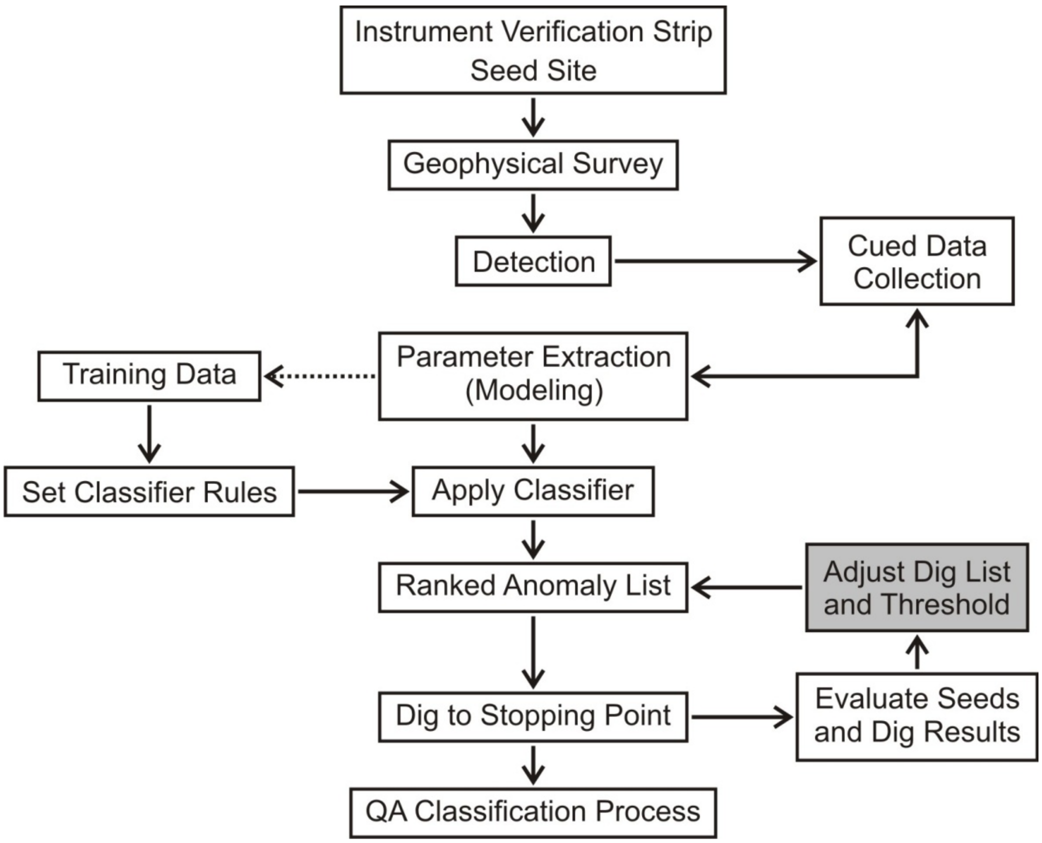


Figure 3-1. Outline of the workflow used for this exercise. No adjustment to the initial threshold was required on this site so that boxed is colored gray.

3.3.2 Geophysical Sensors

As discussed above, a cart-based EM61-MK2 sensor was used for the detection survey with cm-level GPS geolocation. The MetalMapper advanced sensor (Ref. 5) was used for cued data collection over all anomalies identified from the detection survey. More details of these data will be given in later sections.

3.3.3 Data Analysis and Anomaly Ordering

EM61-MK2 survey data were processed and anomalies identified using the UX-Process tools in Oasis montaj. Target parameters were extracted from the cued MetalMapper data using custom software and a statistical classifier was used to rank the anomalies. Each of these steps will be discussed in detail below.

3.4 Project Timeline

Although all ESTCP demonstration activities were conducted in the summer of 2011, the exercise was separated into two parts to simulate a multi-year effort. Approximately 25% of the area was surveyed and analyzed for the first part of the exercise, referred to here as Pole Mountain 1. The remaining 75% of the site was analyzed for Pole Mountain 2.

4 Detection Survey

An overview of the Pole Mountain 1 detection survey data, collected by two teams of geophysicists, is shown in Figure 4-1. The false color scale is set such that signal amplitudes below the 5.2 mV anomaly selection threshold are in grayscale and amplitudes above that are in color.

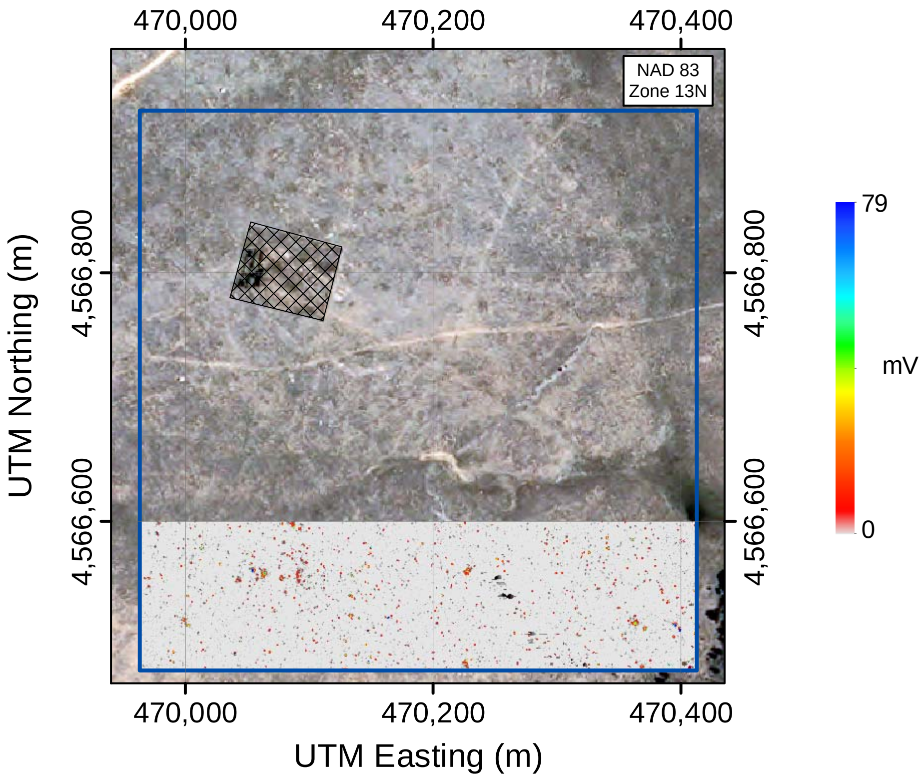


Figure 4-1. EM61-MK2 survey data in gate 2 from the first part of the demonstration site

4.1 IVS Performance

There are two elements to the twice-daily IVS checks for a detection survey: anomaly amplitude reproducibility and position reproducibility. The contractor provided cumulative plots for both elements for each item in the IVS for each data collection team. Example plots for a 37-mm projectile buried horizontally, across track at 15 cm are shown in Figures 4-2 and 4-3.

Figure 4-2 plots the measured amplitudes over this item along with lines marking the ± 20% variation from the mean allowed by the specifications. In general, there is good reproducibility with the standard deviation of the measurements less than 10% of the mean value. One morning measurement by team 2 fell just outside the specification however. The afternoon measurement by this team was within specification and the QC seed encountered by this team during the day had the expected signal amplitude so the data from that day were accepted.

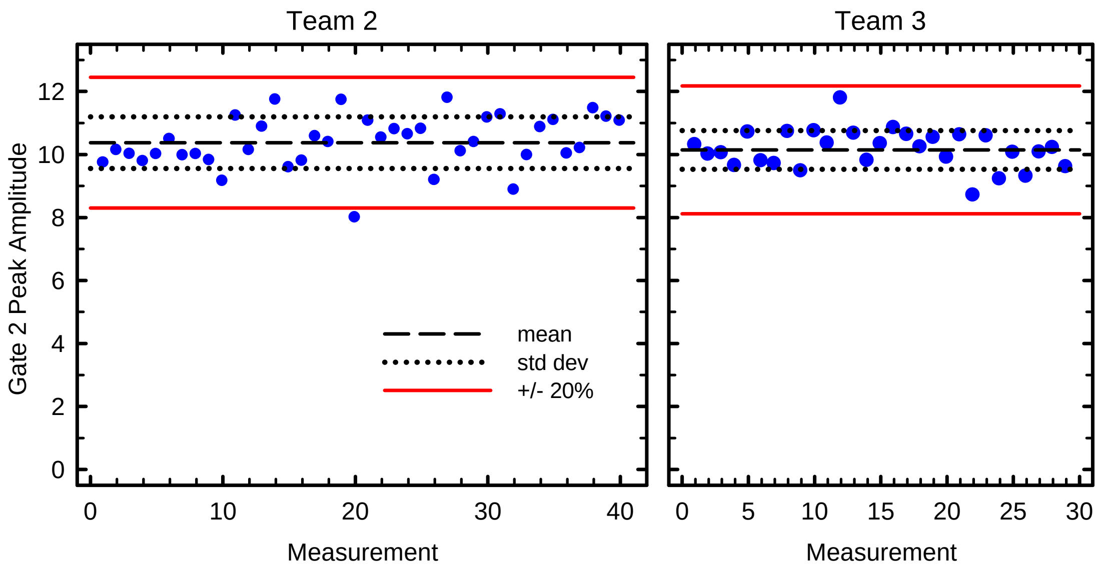


Figure 4-2. Observed amplitude reproducibility for the 37-mm projectile in the IVS for both EM61 survey teams

The corresponding position reproducibility is plotted in Figure 4-3. All measurements are well within the ± 50 cm specification. The IVS was surveyed in a north-south direction; this explains the tighter down-track pattern than that observed across-track.

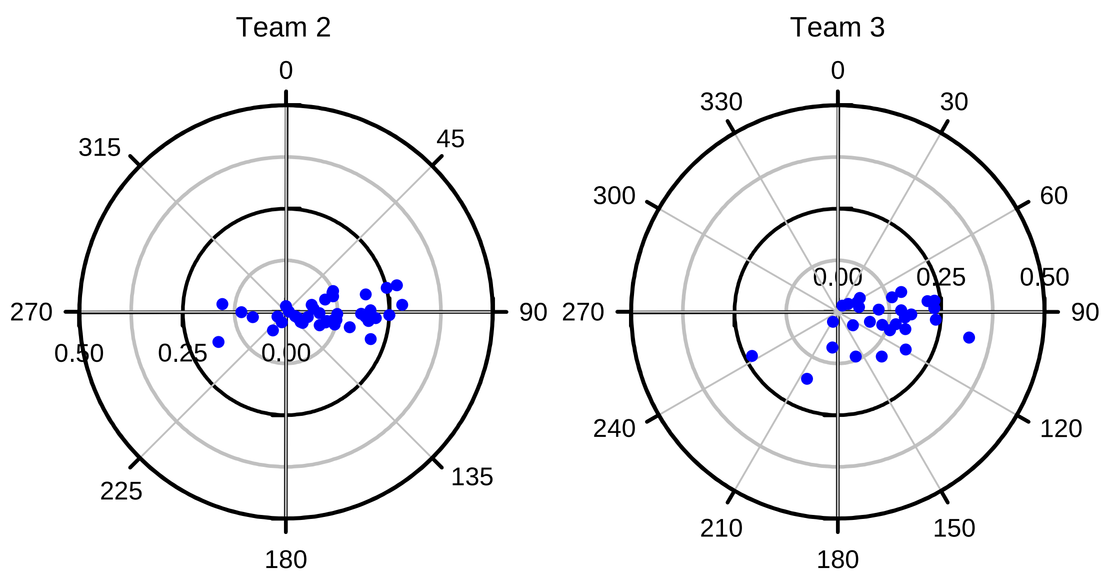


Figure 4-3. Observed position reproducibility for the 37-mm projectile in the IVS for both teams

4.2 **Anomaly Selection**

As shown in Figure 3-1, the measured survey noise at Pole Mountain is well below the anomaly selection threshold established from consideration of the targets of interest and the depths they are expected so little interference from noise was anticipated. Using the threshold of 5.2 mV in gate 2, 984 anomalies were selected for cued data collection.

The first QC check on detection is to make sure all the seeds were detected. Using a detection halo of 60-cm radius, all QC seeds were detected.

4.3 Other QC Considerations

As is usual for a munitions response project, there were a number of measurement density and survey coverage specifications for the EM61 survey data. The data collected was within specification on all these measures.

4.4 Summary

Table 4-1 summarizes the results of all the QC checks applied to the EM61-MK2 survey data and analysis. Based on these results, the site team judges the survey results acceptable and consents to the collection of MetalMapper cued data.

Table 4-1. Summary of QC checks on the EM61 survey data

Item	Result
IVS Amplitude Reproducibility	Acceptable
IVS Position Reproducibility	Acceptable
Total Anomalies Selected	984
QC Seeds Detected	100%
Measurement Density	Acceptable
Survey Coverage	Acceptable

5 Cued Data Collection

Cued MetalMapper data were collected over each of the 984 anomalies selected in the detection phase. As part of this data collection, the MetalMapper sensor was deployed over each item in the IVS twice each day.

5.1 IVS Performance

The QC checks performed on advanced sensor data are more stringent than those required for a detection only survey. Not only must we confirm that the sensor system is producing the expected signal levels over each IVS item, we must also ensure that the data are of sufficient quality for the feature extraction routines to be used successfully on the data. Accordingly, we report not the raw signal amplitudes but the results of geophysical inversion on the data collected over each point.

Figure 5-1 shows the IVS QC data associated with this survey. The derived polarizability decay curves for each of the five IVS items are shown along with the standard deviations of each set of curves. In most cases the standard deviation is a few percent of the derived polarizabilities and in all cases is within the 10% specification.

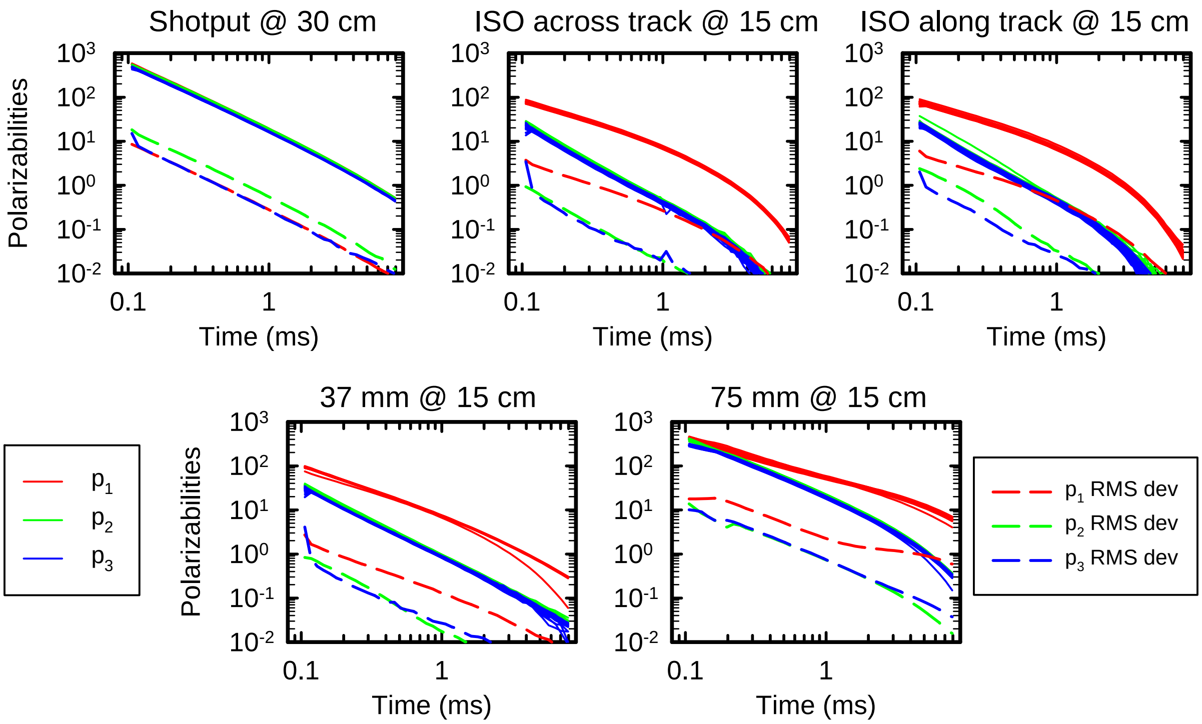


Table 5-1. Polarizability decay curves (solid lines) derived from twice-daily MetalMapper measurements over each of the Pole Mountain IVS items and their respective standard deviations (dashed lines). The data quality objective for this measurement required a standard deviation of less than 10% of the mean which would correspond to one major division lower on this log scale plot. This was true in all cases with some being closer to 1% of the mean.

In addition to confirming that the IVS results met our error tolerance, we can learn a good deal about the process from the plots in Figure 5-1. The shotput is nearly spherical so we expect three equal polarizabilities which is what we observe. The other four targets are cylindrical so we expect one large response and two smaller and equal responses. The total polarizability scales with target volume so we expect the 75 mm response to be roughly an order of magnitude larger than the 37 mm which is also what we observe.

5.2 **QC of Cued Data**

At the end of each day's data collection, the MetalMapper data from that day are transmitted to a QC geophysicist for analysis. The purpose of this step is to confirm that the data collected over each anomaly is complete and of sufficient quality to lead to reliable parameter extraction. The QC screen that this analyst uses contains a number of individual plots. An example from the original data collection over anomaly 912 is shown in Figure 5-2.

Notice four areas of the QC screen in Figure 5-2 that have been highlighted with colored boxes. There is poor agreement between the measured and modeled data from each of the transmit and receive pairs (green and blue decays in the orange box) and the trial inversion failed (red box on the bottom middle of the screen) leaving one of the estimated polarizabilities with an un-physical fall-off (purple box on the upper left). The data collection error that led to this is shown in the light blue

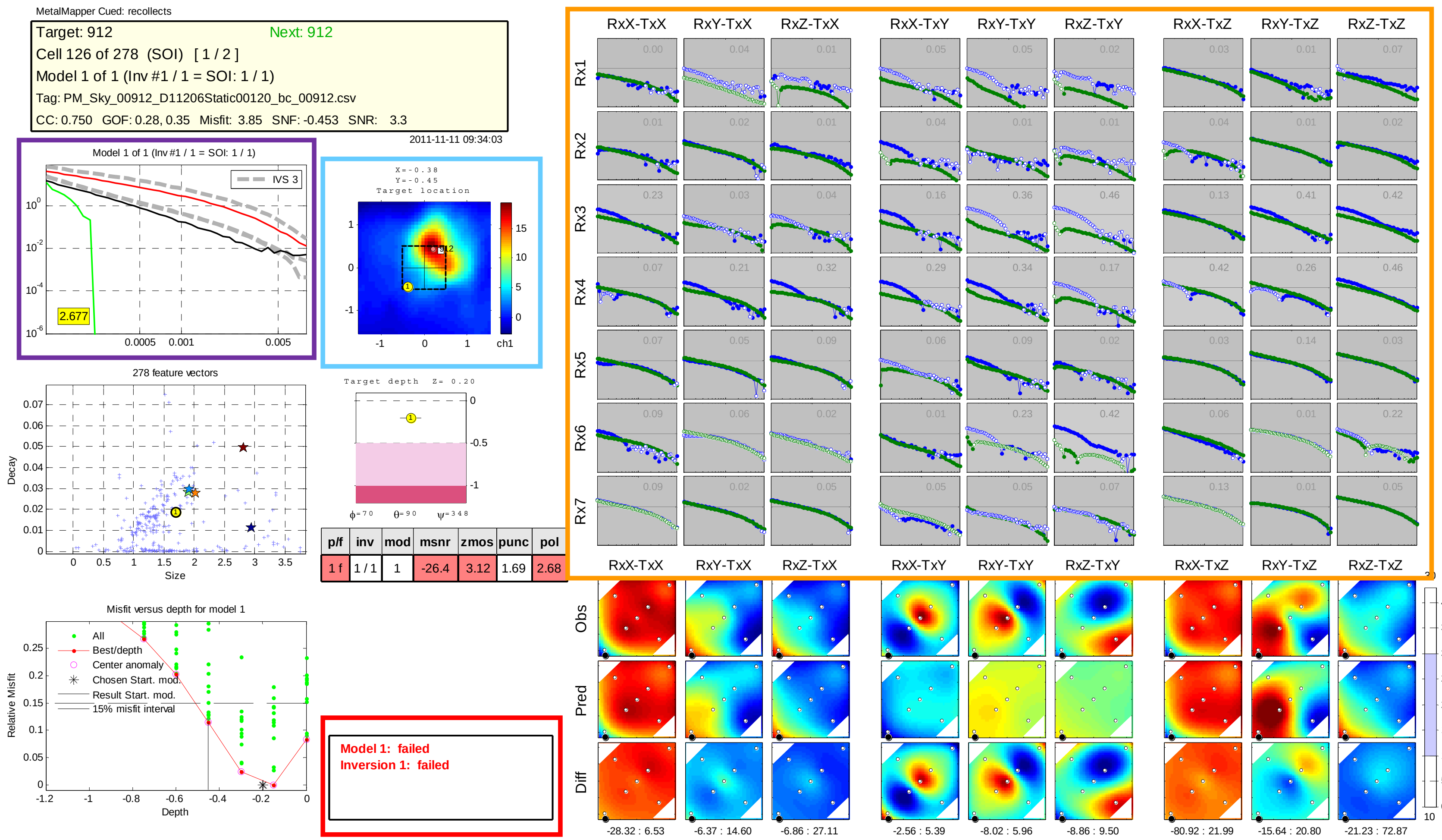
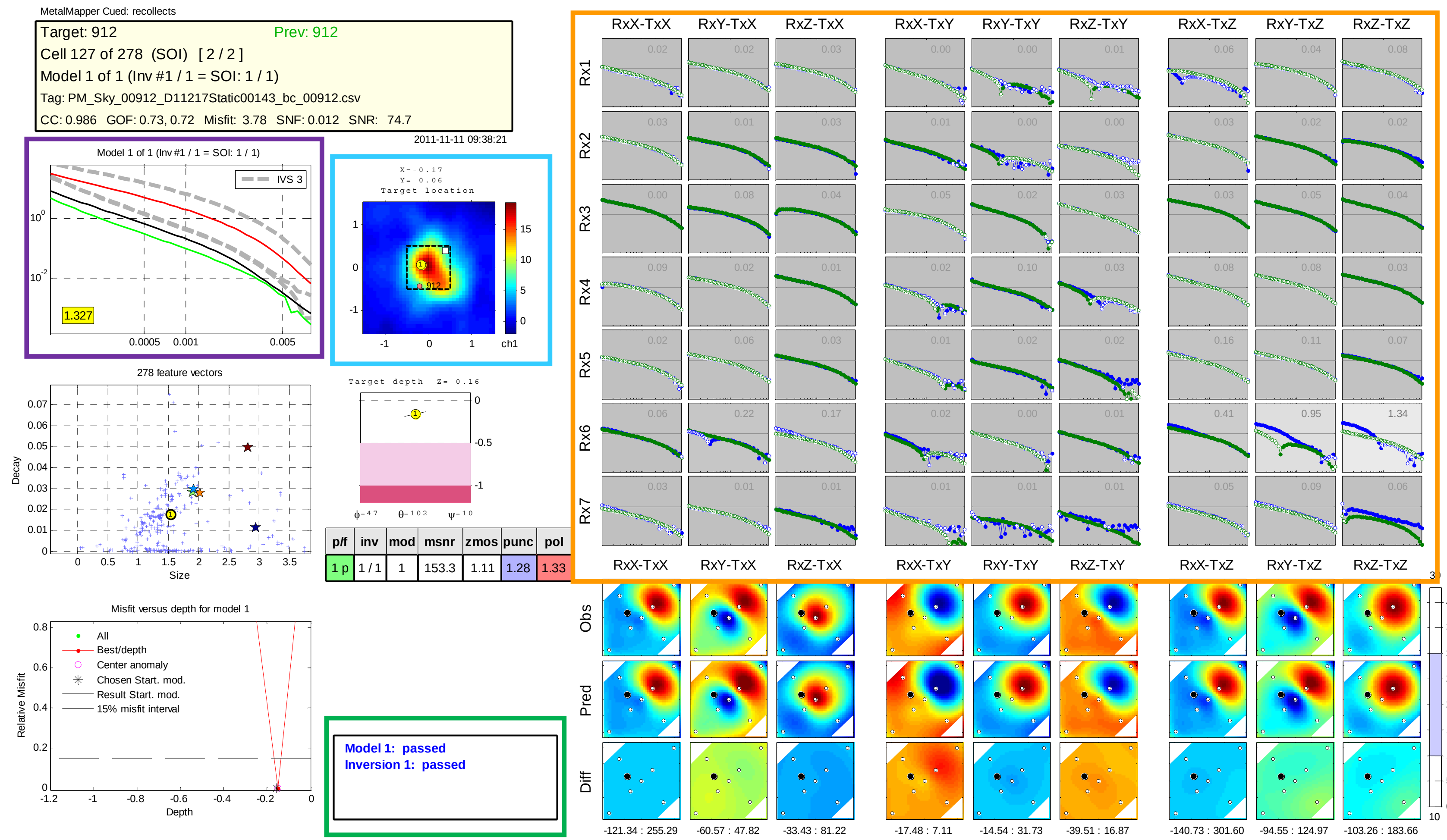


Figure 5-2. Analyst's QC screen for the near-real-time check of the MetalMapper data originally collected over anomaly 912.



p/f	inv	mod	msnr	zmos	punc	pol
1 p	1 / 1	1	153.3	1.11	1.28	1.33

Figure 5-3. Analyst's QC screen for the near-real-time check of the MetalMapper data recollected over anomaly 912.

box in the upper middle of the screen which plots the EM61 survey data in false color and the position of the MetalMapper by an outline. The MetalMapper was not centered over the object for data collection but was offset to the SW. In fact, it was offset more than the 0.4 m that our data quality objectives specify which leads the analyst to fail this measurement and place this anomaly on the list for recollection. The QC screen from the second collection is shown in Figure 5-3. In this case, the MetalMapper was positioned correctly and the trial inversion succeeded as shown in the green box on the bottom of the figure. These data were accepted by the QC analyst.

5.3 Summary

The MetalMapper data from the IVS were well within specifications. Data were collected over each anomaly on the detection list, those data were QC'd each evening, and the ~5% of collections that failed QC were recollected. Based on this, the site team judges the MetalMapper data acceptable and agrees to proceed to classification.

6 Classification

A statistical classifier was used for this project. All classification decisions were made using the full polarizability curves, the points in Figure 6-1; the full vector comprised of 3 polarizabilities times 42 measurement gates was submitted to the classifier. This is difficult to depict graphically so a simplified two-dimensional feature space is often used for illustration purposes. This space consists of a feature related to the object's size and a feature related to the decay of the polarizability. The sources of these two simplified features are shown in Figure 6-1. The total polarizability scales with the object's volume. The sum of the polarizabilities at the first time gate is used to calculate this feature. The decay feature corresponds to the ratio of the 29th gate to the first.

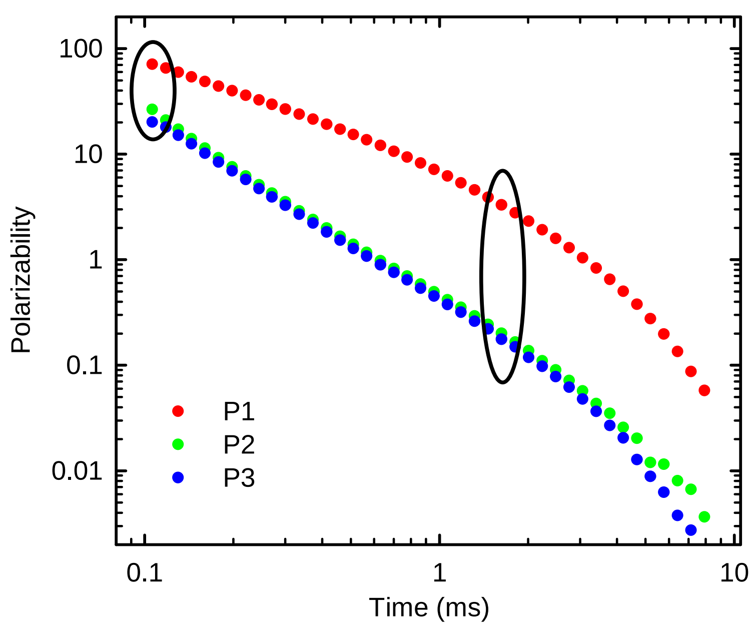


Figure 6-1. Example polarizability decay curves. The two areas used to construct the simplified feature space are marked with ellipses.

The simplified features derived from analysis of the MetalMapper cued data collected in the first phase of the Pole Mountain exercise are shown in Figure 6-2. Here, and in all subsequent plots, the size and decay features are based on the amplitude and decay of the total polarizability as illustrated in Figure 6-1.

This plot will be familiar to some as it is often used when attempting classification using EM61 survey data; these two simplified features are all that can extracted from those data. Munitions are generally found in the upper right region of a plot like this. Munitions are larger than the fragments found on most munitions sites and they are made of heavy-wall steel which leads to slower decays (larger ratio of late to early measurement gates).

It is important to keep in mind that this simplified feature space is used as an illustration only, not for classification. Classification using these features is very imprecise; it will become evident later that two items that are very close on this plot may look completely different when their complete polarizability decay curves are examined.

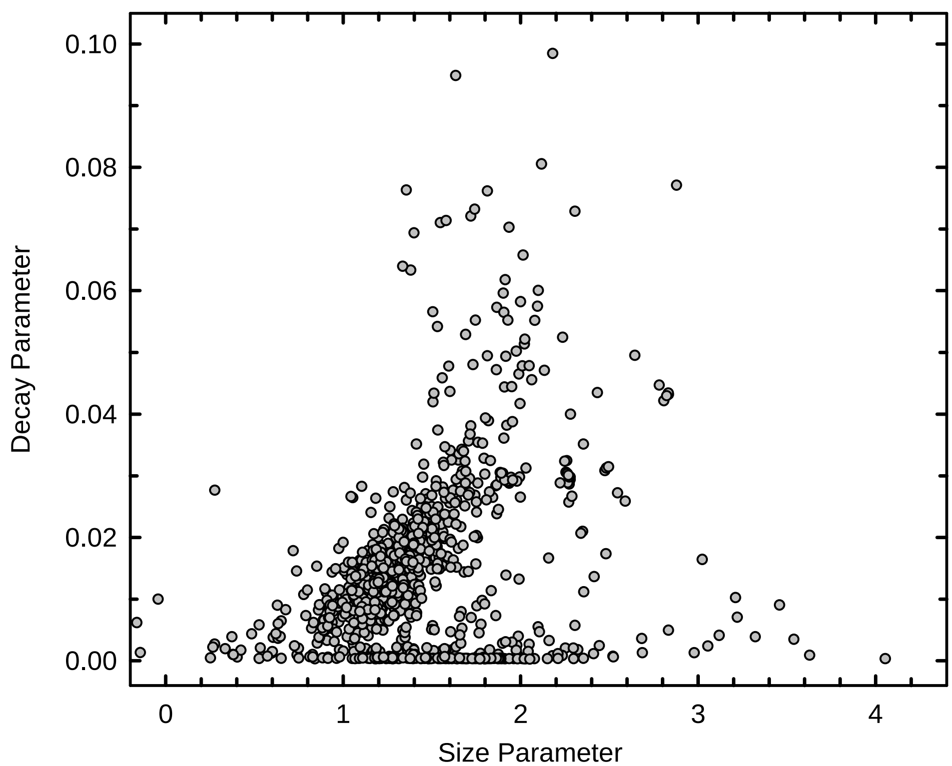


Figure 6-2. Simplified two-dimensional feature space for the anomalies from the first phase of the Pole Mountain demonstration. The simplified features are defined in the text.

6.1 **Classifier Training**

MetalMapper data were collected over a number of the munitions expected to be encountered at this site in the Instrument Verification Strip and the training pit provided for the data collection team. The simplified features derived from those measurements are shown plotted over the unknowns in Figure 6-3. As we expected, the features corresponding to targets of interest are found in the upper right portion of this feature space.

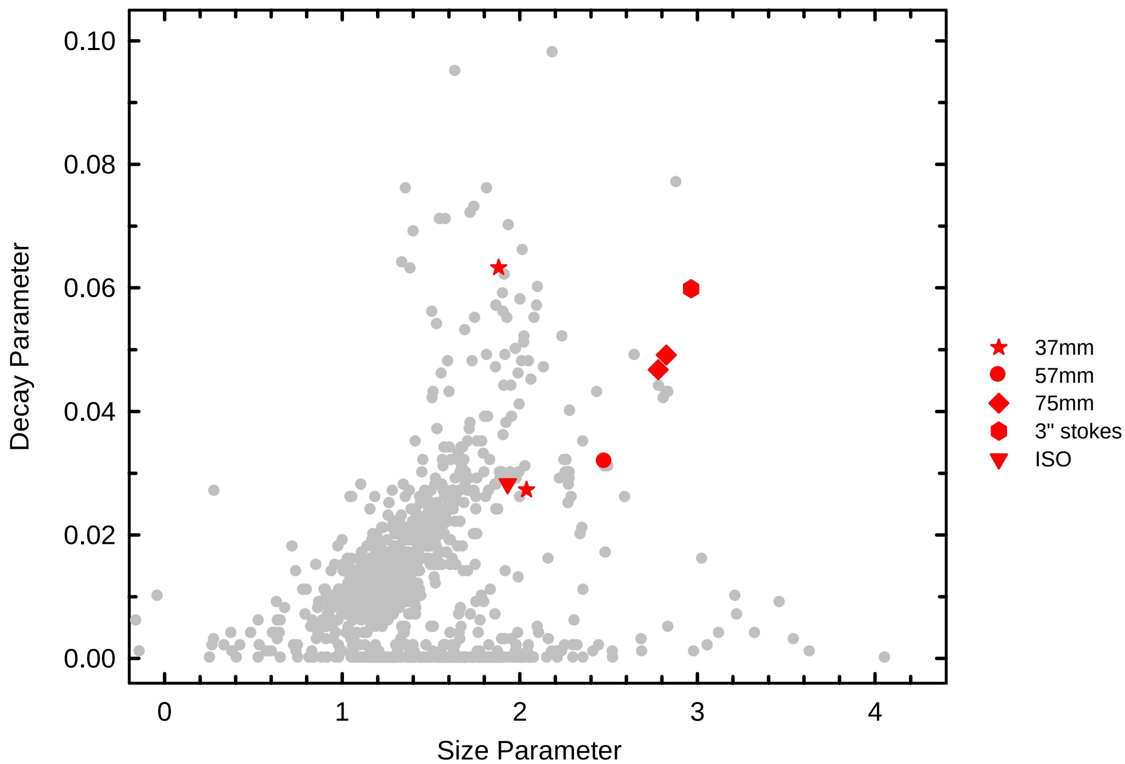


Figure 6-3. Simplified features for objects measured in the IVS and test pit plotted over those from the unknowns

One additional thing to note in Figure 6-3 is the two types of 37-mm projectiles used in the training. Projectiles with brass rotating bands exhibit a substantially slower decay that projectiles for which the bands have broken off. Apparently, one of each type was used for these measurements.

To allow the classifier to parse the anomalies into items corresponding to targets of interest and items corresponding to clutter, the identity of seventeen anomalies was requested for training purposes. These training requests, shown in Figure 6-4 as blue symbols, were not selected randomly; the rationale for each choice is discussed below.

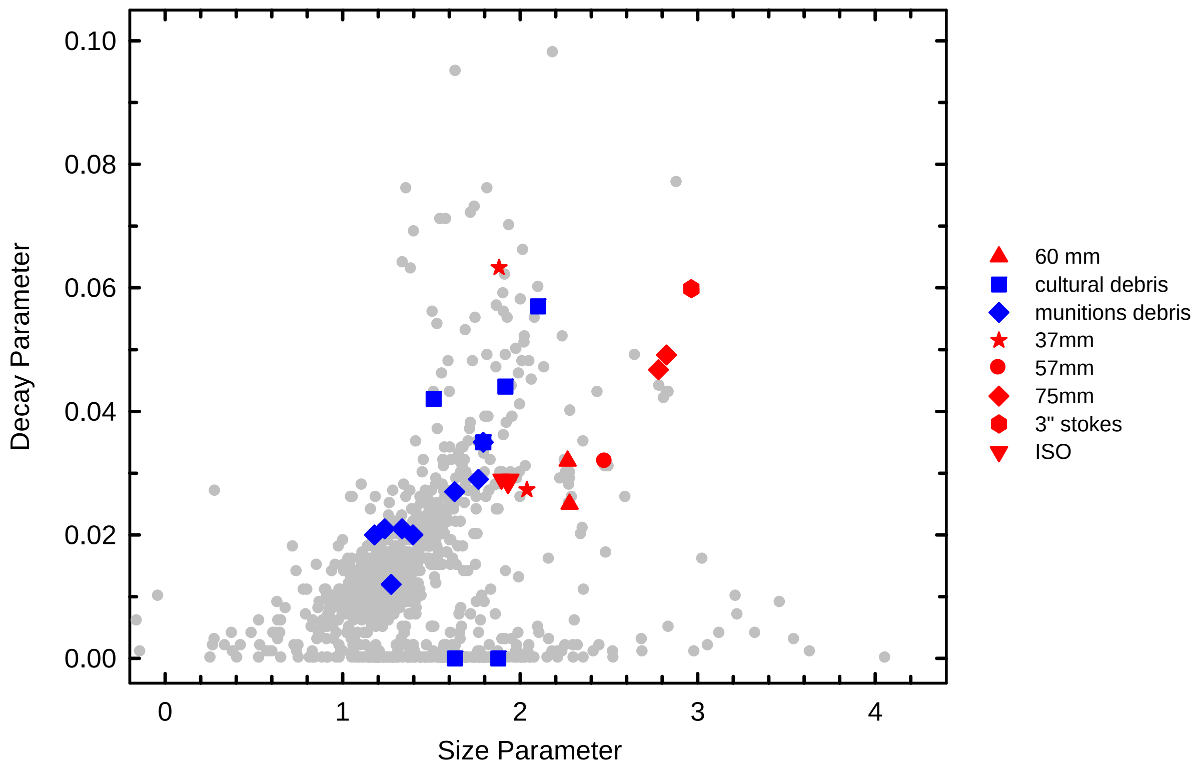


Figure 6-4. Classifier training data (blue symbols) requested in this exercise

6.1.1 **Known Clusters**

Careful examination of each of the complete polarizability decay curves of the unknowns allows the analyst to group a number of the anomalies into "clusters" whose polarizabilities indicate they arise from a common source. Six of these "clusters" have polarizability decay curves that match the items measured in the IVS or test pit. These six clusters are shown in Figure 6-5 with the cluster resulting from small ISOs highlighted.

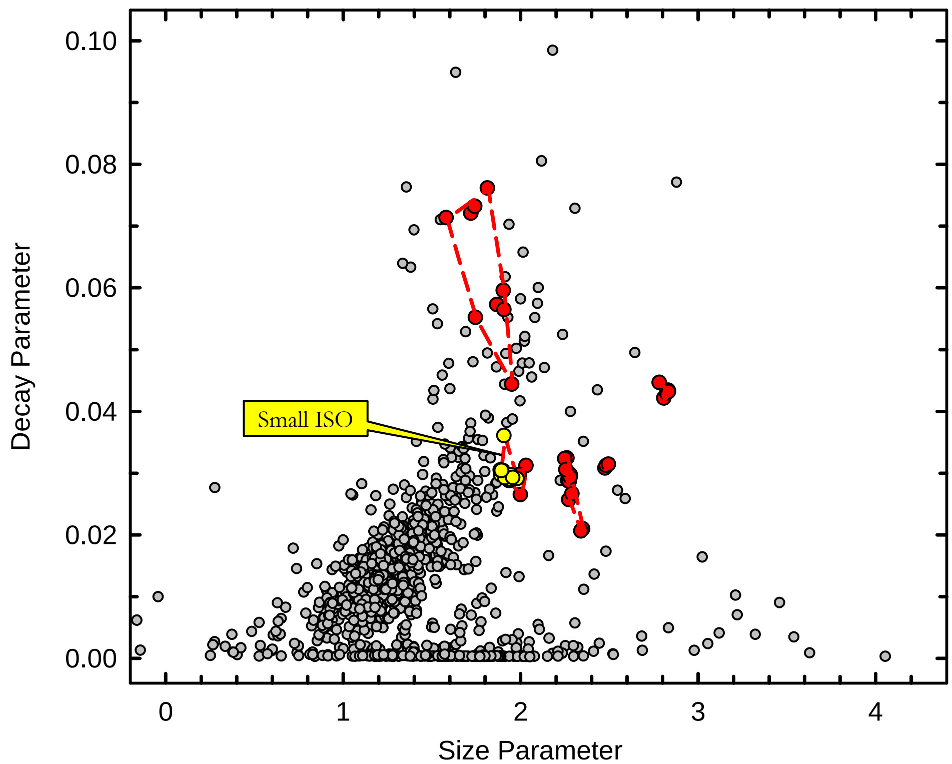


Figure 6-5. Six clusters known from either previous work or site-specific measurements. The cluster corresponding to small ISOs is marked.

The individual polarizability decay curves corresponding to the eleven items in the cluster labeled small ISO are shown in Figure 6-6. The derived polarizability decay curves for all eleven items match that observed for the small ISO in the IVS. One item, PM1-505, was requested as training data to confirm this assignment; the resulting dig photo is shown in the bottom right of Figure 6-6.

The other known clusters were handled in the same way. Their identities were determined from matches to previously surveyed items.

6.1.2 **Unknown Clusters**

Analogous to the "known" clusters shown above, a number of clusters of polarizability curves were observed that did not correspond to previously known objects. As with the "known" items, these clusters are determined by matches of the full polarizability decay curves, not from their positions on these plots. The first of these is highlighted in Figure 6-7.

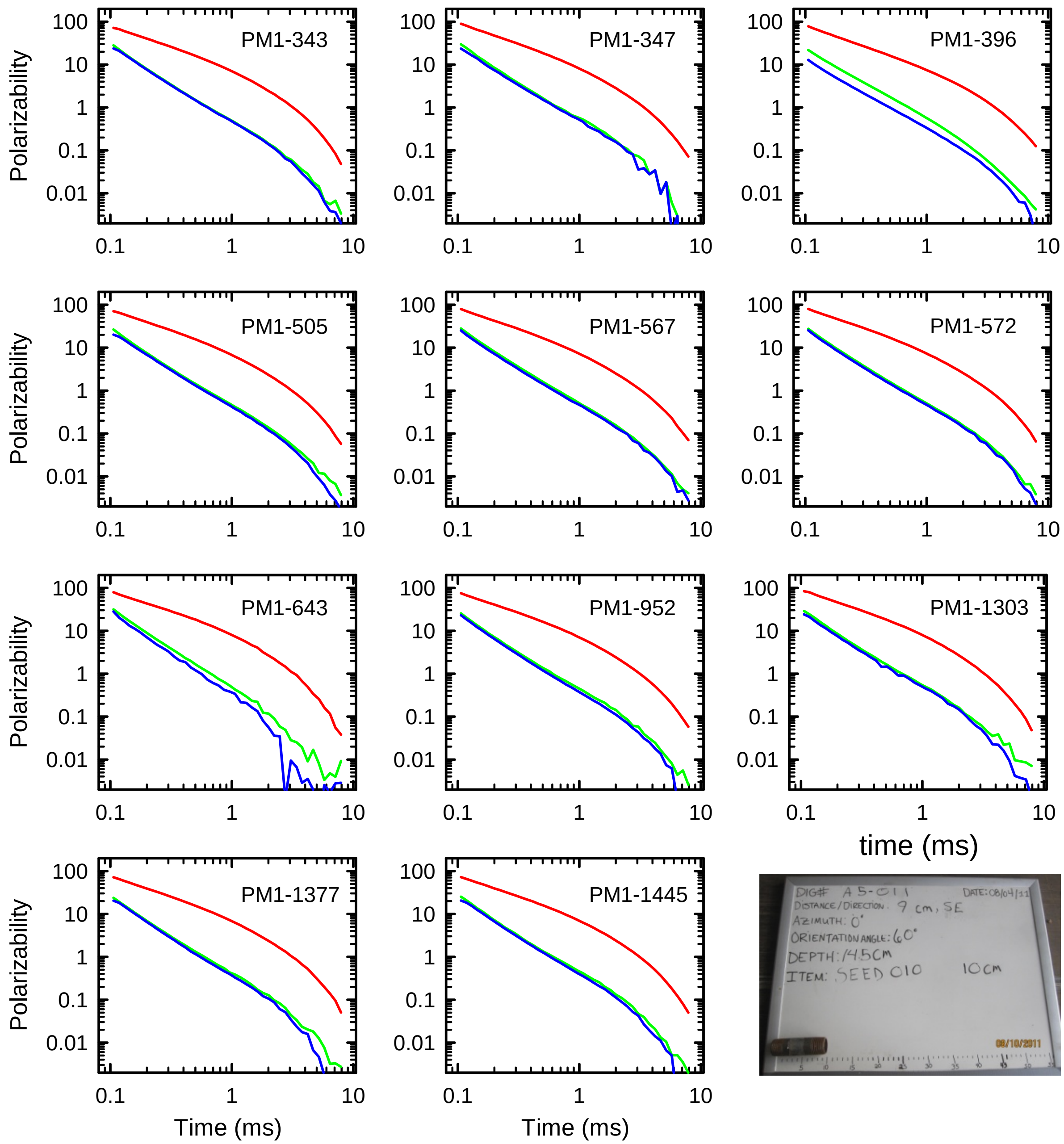


Figure 6-6. Polarizability decay curves for the eleven items identified as part of the small ISO cluster with the dig photo for anomaly PM1-505 shown

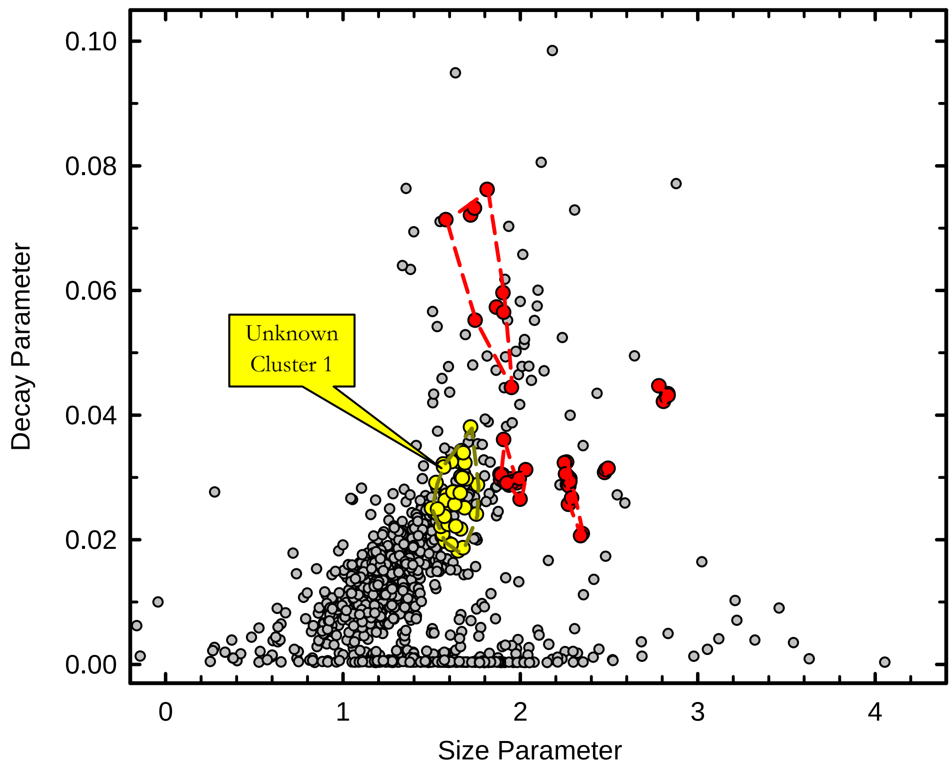


Figure 6-7. Unknown cluster number one

The polarizability decay curves for the first six of 39 items in this cluster are shown in Figure 6-8.

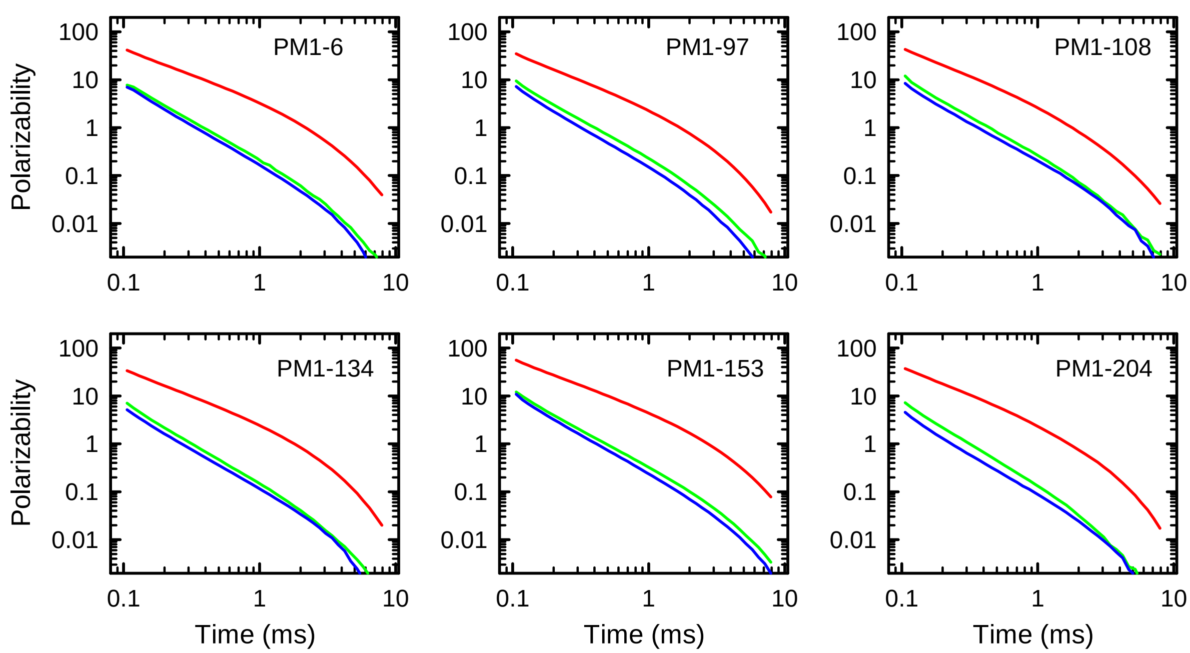


Figure 6-8. Polarizability decay curves for six examples from unknown cluster one

Each of the objects shown in Figure 6-8 is cylindrical (one large response and two smaller and approximately equal responses) but the overall magnitude of the polarizabilities is smaller than the

ISOs shown in Figure 6-6 so we expect these items to be smaller. The dig photo for anomaly PM1-153 is shown in Figure 6-9. As expected, this item is an approximately cylindrical small fragment so it, and all the items in this cluster, will be classified as high-confidence not munition.

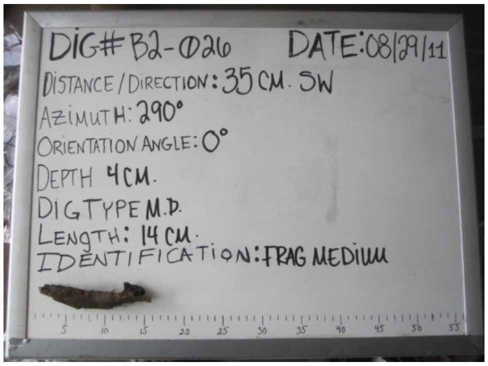


Figure 6-9. Dig photo for anomaly PM1-153

A second cluster located in feature space near where we expect to find munitions is highlighted in Figure 6-10 and the first six polarizability decay curves from this cluster are plotted in Figure 6-11.

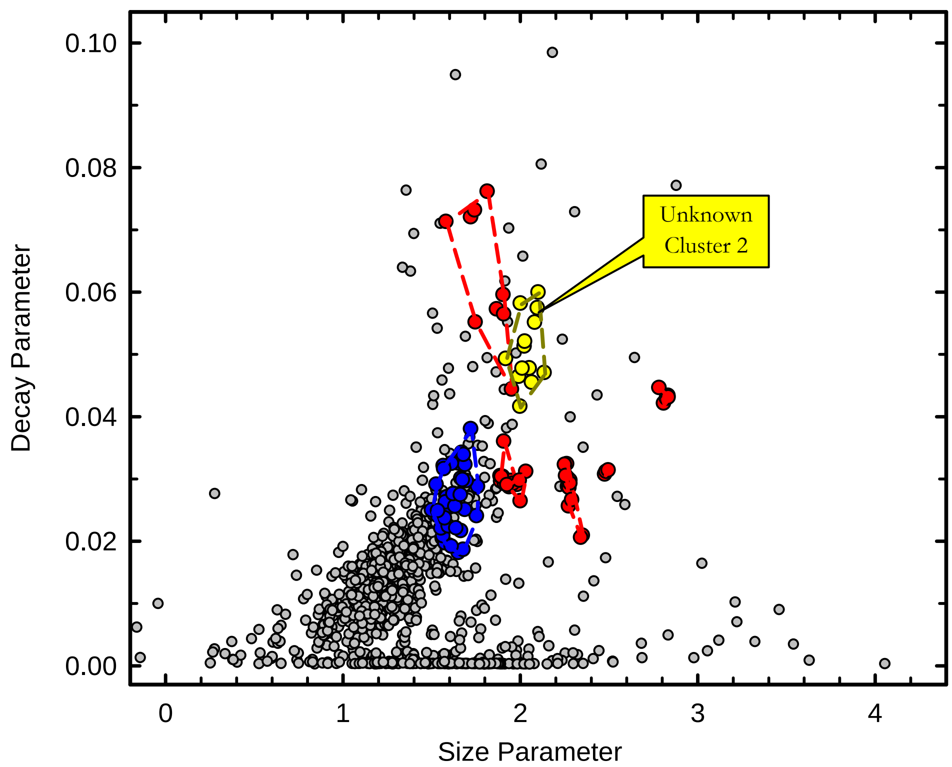


Figure 6-10. Unknown cluster two

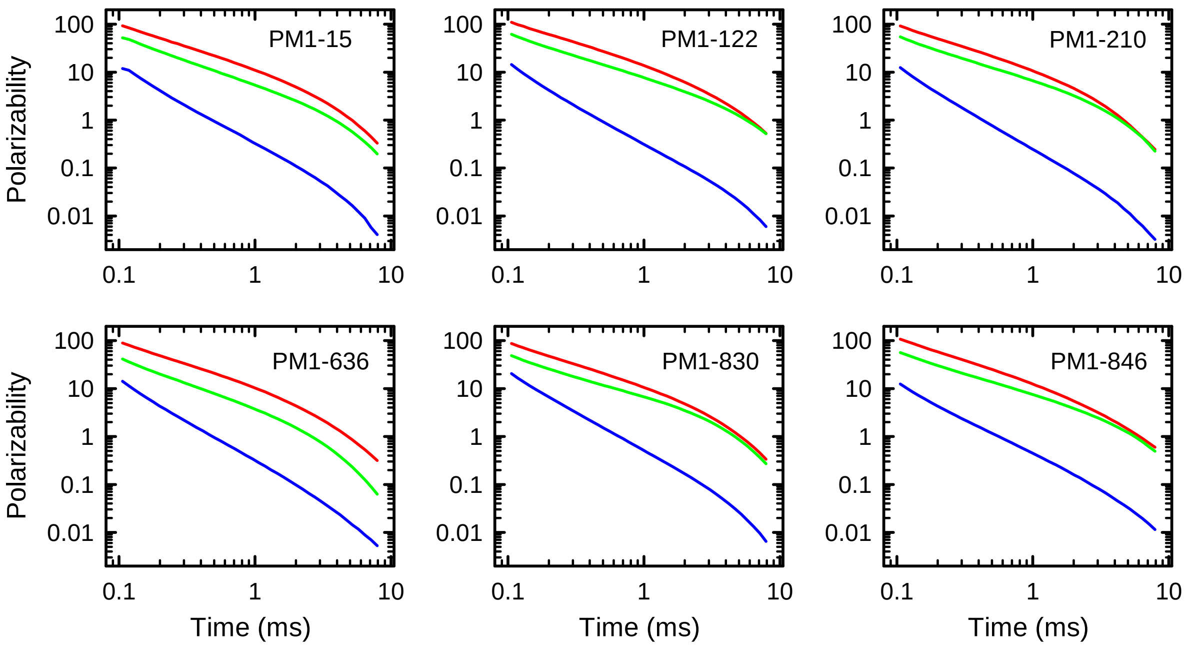


Figure 6-11. Polarizability decay curves for the first six examples of unknown cluster two

Unlike the curves presented so far, these items are more plate-like with two large polarizabilities and one smaller one. In many cases, the two larger polarizabilities are roughly, but not exactly, equal indicating the buried object is only roughly symmetric. The dig photo for one example, PM1-122, is shown in Figure 6-12.

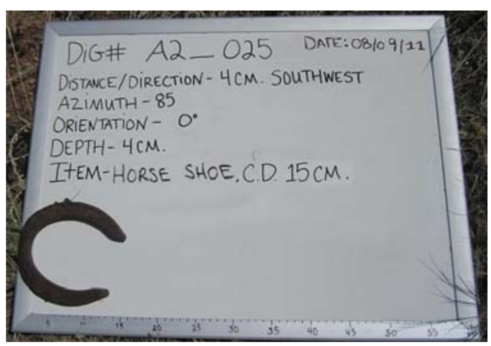


Figure 6-12. Dig photo of anomaly PM1-122

A final example of an unknown cluster is highlighted in Figure 6-13. These objects exhibit a decay parameter of near zero, indicating that they are very thin metal and unlikely to be munitions. There are a large number of these items on this site so we dug several examples to confirm this expectation. Example polarizability decay curves from this cluster are plotted in Figure 6-14. They correspond to plate-like objects with a very thin wall. An example dig result is shown in Figure 6-15 confirming our expectation.

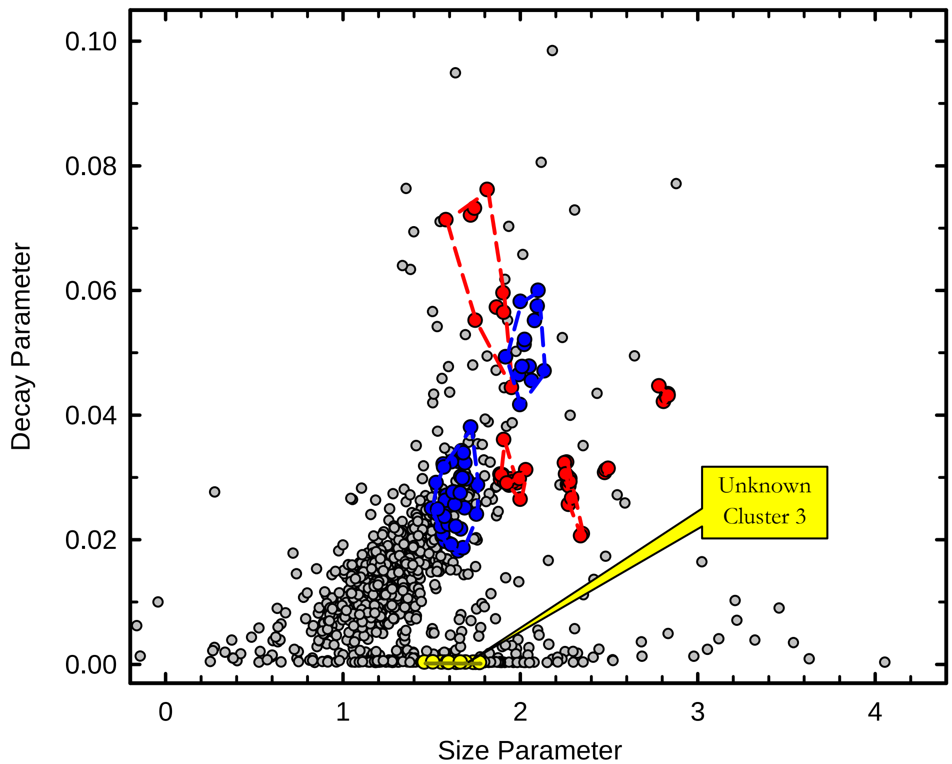


Figure 6-13. Unknown cluster three

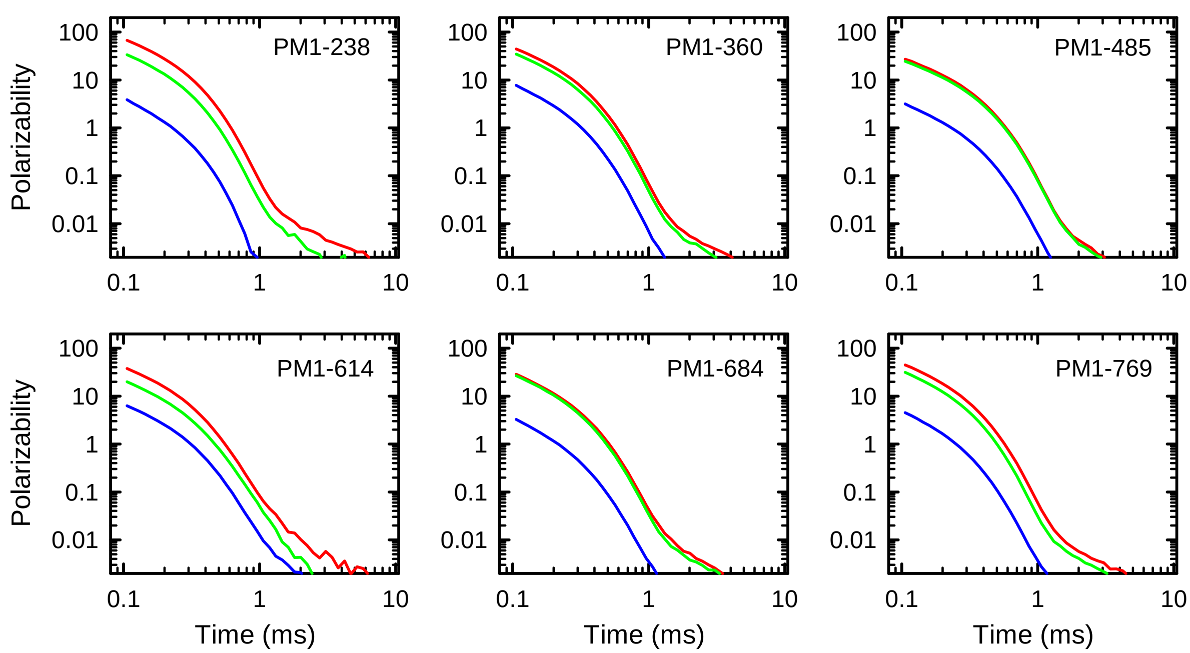


Figure 6-14. Polarizability decay curves for the first six examples of unknown cluster three

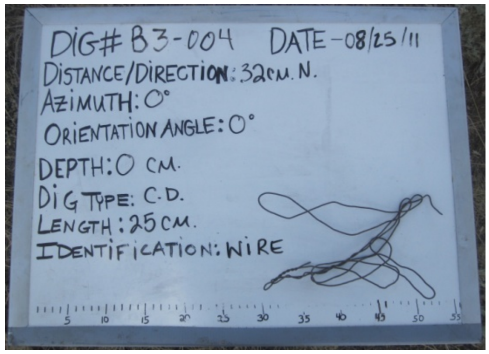


Figure 6-15. Dig photo for anomaly PM1-238

6.2 **Ranked Anomaly List**

At the completion of training our statistical classifier, we were able to make a decision about every anomaly in the first year of the exercise. Consequently, we constructed a ranked anomaly list as shown in Figure 6-16. We were conservative in our rankings; all anomalies between unknown cluster 1 and the known munitions were marked to be dug.

Rank	**Anomaly ID**	**Comment**
1	247	
2	1114	High confidence munition
3	69	
…	…	
…	…	
…	…	
…	…	
…	…	High confidence non-munition
…	…	
…	…	
…	…	
N	…	

Figure 6-16. Ranked anaomaly list resulting from this exercise

6.3 **Initial Dig Results**

Digging all anomalies marked as high-confidence munitions results in the partial receiver-operating-characteristic (ROC) curve shown in Figure 6-17. This curve, described in Ref. 3, results from plotting only those targets we have dug; those classified as high-confidence not munitions remain in the ground and their identity is unknown to us.

This partial ROC curve shows that we were quite efficient at identifying TOI; the initial portion of the curve is almost vertical. Additionally, the last ~75 items dug were clutter.

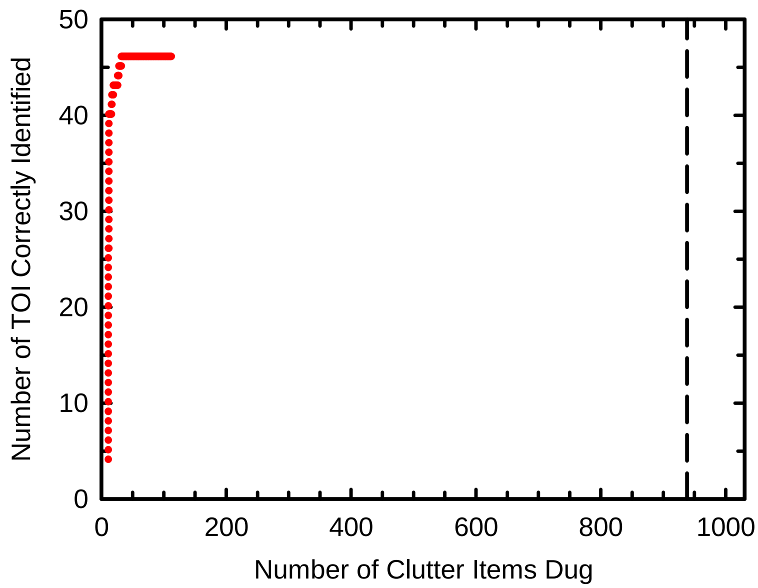


Figure 6-17. Partial ROC plot that results from digging all anomalies classified as high-confidence munitions on our initial ranked anomaly list. The dashed line represents the total number of anomalies minus the TOI dug.

As mentioned above, we emplaced QC seeds in conjunction with this project. All QC seeds were classified correctly; their positions on the ranked anomaly list can be seen in Figure 6-18.

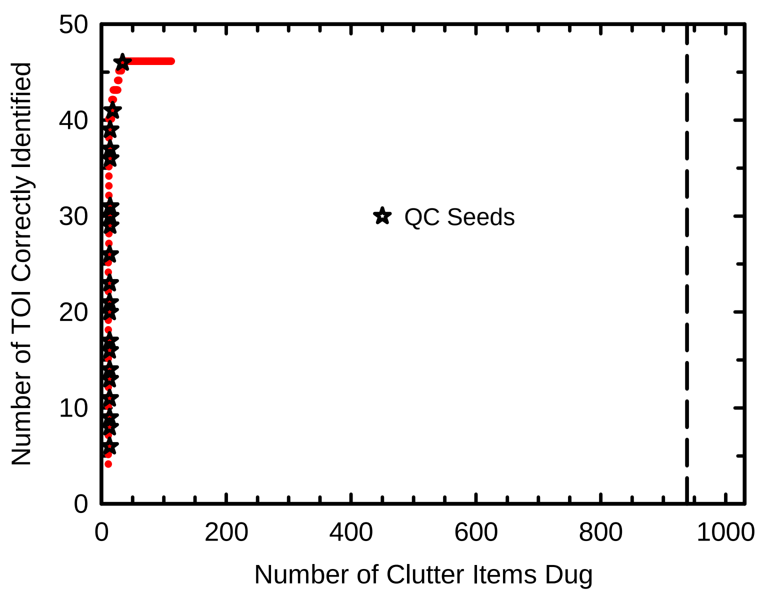


Figure 6-18. Position of the QC seeds on the partial ROC curve. The dashed line represents the total number of anomalies minus the TOI dug.

6.4 QA of the Process

Over 800 anomalies were ranked as high-confidence not munitions. To be comfortable leaving these anomalies in the ground, we have to devise a QA procedure for classification. In this case, we have adopted the following strategy to confirm the proper conduct of the classification procedures.

We have already had an opportunity to compare our classification results to the actual identities of those items that were classified as likely TOI. To extend this comparison to a sample of the items classified as likely clutter, we will randomly select some of the anomalies that were marked to be left in the ground, excavate them, and compare the excavated item to the polarizability decay curves derived from our cued measurements. Note, we are not expecting to find any remaining UXO by this procedure; UXO start out rare and we have made them much rarer through this process. We are attempting to gain confidence that the classification procedures were reliable from start to finish. We illustrate this process in Figure 6-19 which shows the location on the ranked anomaly list of the three randomly selected anomalies that we excavated.

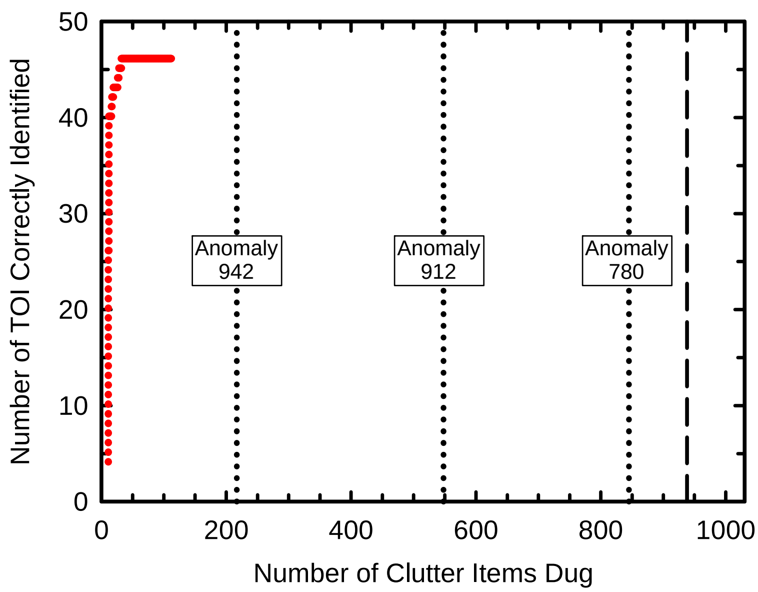


Figure 6-19. The positions of the three random QA digs

The polarizability decay curves corresponding to these three anomalies are shown in Figure 6-20.

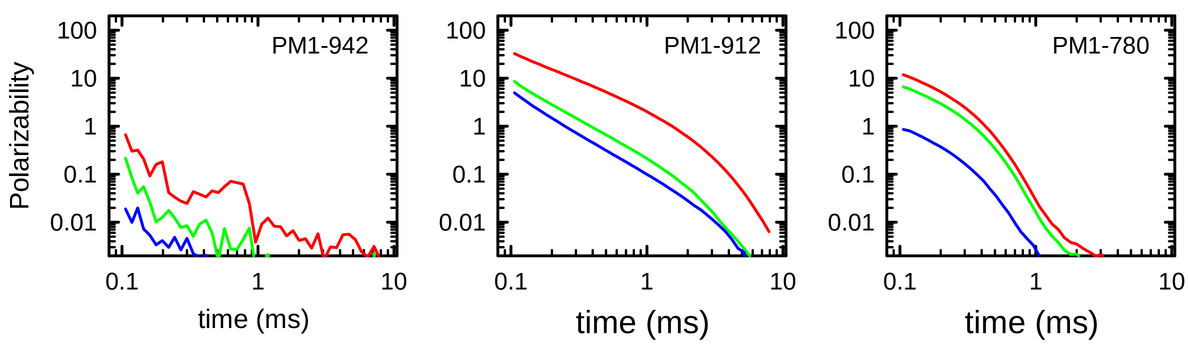


Figure 6-20. Polarizability decay curves for the first three QA items

The curves for anomaly 942 correspond to a small piece of metal, if anything. The polarizabilities are barely above the noise; this anomaly likely corresponds to a noise spike in the detection instrument. Anomaly 912's curves look very similar to the small frag pieces that made up unknown cluster 1 while those for anomaly 780 are a very good match for the baling wire that was unknown cluster 3. The excavation photos for these three anomalies are shown in Figure 6-21.

Figure 6-21. Dig photos for the three example QA digs

The results for anomalies 942 and 912 match our expectations. The baling wire that corresponds to anomaly 780 does not have the plate-like shape we expected based on the polarizability curves. When we query the UXO techs about this, they tell us they uncoiled the wire before taking the photograph so we could see how long it was. Based on these QA digs, we have confidence in the classification procedure employed and agree to stop digging at the recommended threshold.

6.5 **Year 2 of the Exercise**

MetalMapper data were collected on 1500 anomalies in the second part of the exercise. The simplified features derived from these data are compared to the features from year 1 in Figure 6-22.

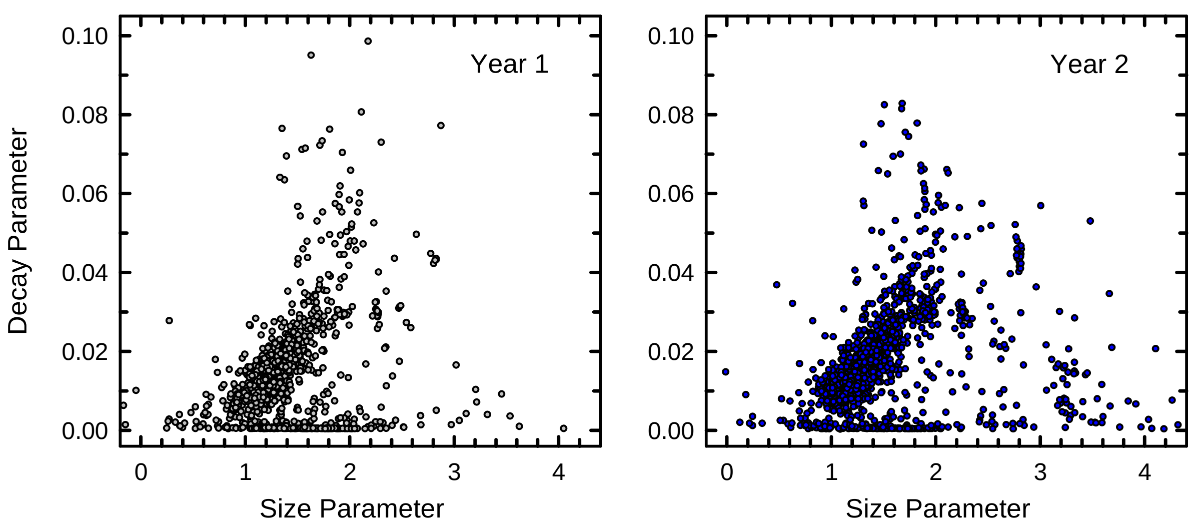


Figure 6-22. Comparison of the simplified features derived for the year 2 anomalies with those from year 1

The year 2 features are very similar to those from year 1. This gives us confidence that we have not come across a different distribution of targets and will allow us to use the classifier and decision thresholds from the first year's effort without modification for ranking the year 2 anomalies.

7 Summary

This case study was designed to illustrate some of the decisions that will be required of a site team when managing a classification-based munitions response project. Typical documentation that should be expected was presented and the thought processes behind approval were explored.

There are many commonalities between this exercise and a modern munitions response project. The Instrument Verification Strip and QC seeds are used in both cases, with more expected from them in the classification project. The main difference is the QA procedures required. In a traditional project, significant metal left in a grid is cause for a QA failure. Many metal objects are intentionally left behind as a result of classification so this traditional QA approach is no longer appropriate. One alternative was presented in this case study; others will surely be developed as the community gains more experience with these methods.

In a real project, the site team would have made their decisions and left the site. Since this was a demonstration project, we dug all anomalies for scoring purposes. The complete ROC curve for the analysis presented here is shown in Figure 7-1. As can be seen, 100% of the munitions were identified well before the analyst's threshold. We were correct to accept this threshold.

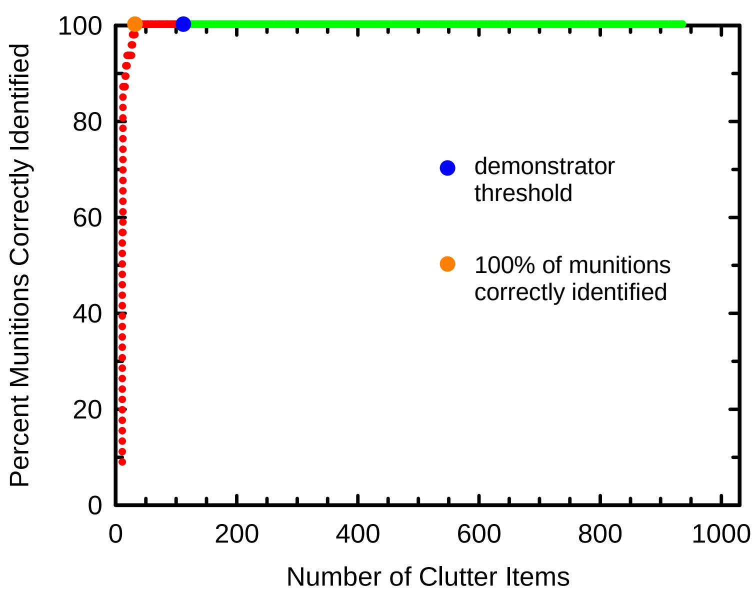


Figure 7-1. Complete ROC curve for the first year's analysis at Pole Mountain. These complete results are only available because we dug all the anomalies as part of the ESTCP demonstration.

8 References

1. Defense Environmental Programs Annual Report to Congress, Fiscal Year 2009, http://deparc.xservices.com/do/home

2. Office of the Under Secretary of Defense for Acquisition, Technology and Logistics, *Report of the Defense Science Board Task Force on Unexploded Ordnance*, November 2003, http://www.denix.osd.mil/mmrp/upload/uxo.pdf

3. *Implementing Advanced Classification on Munitions Response Sites: A Guide to Informed Decision Making for Project Managers, Regulators, and Contractors, December 2011*, http://serdp-estcp.org/content/download/12780/151578/version/1/file/Implementing_Classification_on_Munitions_Response_Sites_FR.pdf

4. *Geophysical System Verification (GSV): A Physics-Based Alternative to Geophysical Prove-Outs for Munitions Response*, July 2009, http://serdp-estcp.org/content/download/7426/94837/version/1/file/GeoSysVerif-July-09-FINAL.pdf

5. MetalMapper: A Multi-Sensor TEM System for UXO Detection and Classification, ESTCP Project 200603 Final Report, February 2011, http://serdp-estcp.org/content/download/9593/122667/file/MR-200603-FR.pdf.