IDA

INSTITUTE FOR DEFENSE ANALYSES

The UXO Discrimination Study at the Former Camp Sibert

Shelley Cazares
Michael Tuley
Michael May

January 2009

Approved for public release;
distribution is unlimited.

IDA Document D-3572
Log: H 08-000974

This work was conducted under contract DASW01-04-C-0003, Task AM-2-1528, for the SERDP/ESTCP. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of that Agency.

© 2008, 2009 Institute for Defense Analyses, 4850 Mark Center Drive, Alexandria, Virginia 22311-1882 • (703) 845-2000.

This material may be reproduced by or for the U.S. Government pursuant to the copyright license under the clause at DFARS 252. 227-7013 (NOV 95).

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-3572

The UXO Discrimination Study at the Former Camp Sibert

Shelley Cazares
Michael Tuley
Michael May

PREFACE

This report was prepared for the Director, Environmental Security Technology Certification Program, under a task titled "ESTCP/SERDP: Assessment of Traditional and Emerging Approaches to the Detection and Identification of Surface and Buried Unexploded Ordnance."

CONTENTS

Executive Summary ES-1
1. Introduction 1-1
1.1 Detailed Objectives 1-2
1.2 Demonstration Motivation 1-3
1.3 General Approach 1-3
1.4 Limitations 1-5
2. Methods 2-1
2.1 Selection of Site 2-1
2.1.1 Former Camp Sibert: History and Characteristics 2-3
2.1.2 Geolocation Survey Control Points 2-4
2.2 Magnetometer Survey 2-5
2.3 Selection of Areas 2-5
2.4 Clearance of Geophysical Prove Out 2-6
2.5 Excavation of Sample Area 2-7
2.6 Generation of Seed Plan 2-8
2.7 Emplacement of Seeds 2-9
2.8 Data Collection in Survey Mode 2-11
2.8.1 The Mag Array Instrument 2-12
2.8.2 The EM61 Array Instrument 2-12
2.8.3 The GEM Array Instrument 2-13
2.8.4 The EM61 Cart Instrument 2-15
2.8.5 The Berkeley UXO Discriminator Instrument 2-16
2.8.6 Mag-and-Flag 2-17
2.9 Anomaly Detection 2-18
2.9.1 Selecting Detection Thresholds 2-18
2.9.2 Applying Detection Thresholds 2-21
2.10 Cued List Generation 2-21
2.11 Data Collection in Cued Mode 2-23
2.11.1 The EM63 Cued Instrument 2-23
2.11.2 The GEM Cued Instrument 2-24
2.12 Master List Generation 2-25
2.12.1 Southeast 1 Area 2-26
2.12.2 Southeast 2 Area 2-41

2.12.3 Southwest Area 2-42
2.13 Selection of Survey Data at Locations on the Master List 2-44
2.14 Discrimination 2-44
2.14.1 Data Inversion 2-45
2.14.2 Ranked Dig List Generation 2-46
2.14.3 Dig Threshold Selection 2-48
2.14.4 Training and Test Sets 2-50
2.15 Excavation 2-52
2.16 Assignment of Ground Truth 2-53
2.17 Survey Detection Scoring 2-54
2.18 Discrimination Scoring 2-56
3. Selected Results and Discussion 3-1
3.1 Detection 3-1
Finding 1: Survey sensors detected almost all munitions, leading to excellent detection performance. 3-1
Finding 2: Data collected from the EM61 Array were often noisy due to the bouncing motion of the towed vehicle over the ground during data collection. 3-3
3.2 Discrimination 3-4
3.2.1 Detected items that cannot be analyzed 3-4
Finding 3: All "Can't analyze" locations must be dug. 3-4
Finding 4: A principled, documented method for identifying "Can't analyze" locations has not yet been agreed upon. 3-5
Finding 5: Once "Can't analyze" locations were dug, discrimination performance was typically very good for all remaining locations. 3-8
3.2.2 Commercially available or production instruments and software 3-9
Finding 6: Commercially available and production instruments and software provided good discrimination performance. 3-10
Finding 7: For survey instruments, cooperative inversions led to a slightly lower number of unnecessary digs. 3-13
Finding 8: Much of the discriminating power seen at Camp Sibert is due to size-based features. 3-14
Finding 9: Mag-and-flag led to a large number of unnecessary digs. ... 3-17
3.2.3 Frequency-domain EMI instruments 3-18
Finding 10: The GEM Array and SAIC custom software led to good discrimination performance. 3-19
Finding 11: High-density, cued GEM data had some discriminating power, but led to a large number of unnecessary digs, even with cooperative inversions. 3-19

3.2.4 Advanced instruments and software 3-21
Finding 12: High-density, cued EM63 data led to good discrimination performance, especially with cooperative inversions. 3-22
Finding 13: The multiple-axis BUD instrument provided high-SNR data from a single location leading to excellent discrimination performance. 3-24
Finding 14: The advantage to active learning could not be fully demonstrated at Camp Sibert. 3-25
Finding 15: The advantage to semi-supervised learning could not be demonstrated at Camp Sibert. 3-28
3.2.5 Dig Threshold 3-30
Finding 16: In some cases, a higher confidence in digging munitions could be achieved with only a few more unnecessary digs when using quantitative methods to set the dig threshold. 3-30
3.3 Limitations of Analysis 3-33
3.4 Lessons Learned 3-34
4. Conclusion 4-1
4.1 Detection 4-1
4.2 Discrimination 4-1
Acronyms GL-1
References Ref-1
Appendix A UXO Discrimination Study: Blind Seed Plan for Camp Sibert, AL A-1
Appendix B: UXO Discrimination Study: Results for Camp Sibert, AL B-1
Appendix C: UXO Discrimination Study: Anomalies That Could Not Be Analyzed C-1

TABLES

2.1 Available Survey Control Points in the Vicinity of Site 18 of the former Camp Sibert 2-4

2.2 Minimum expected responses for 4.2″ mortars at depths of 11 times the mortar diameter and selected detection thresholds for the Array instruments 2-20

2.3 Anomaly Sources for Master List 2-26

3.1 Detection Performance of Survey Instruments over the Test Set 3-2

3.2 Comparing the number of anomalies detected with the Mag Array that SAIC and Sky Research, Inc. could and could not analyze 3-7

3.3 Comparing the number of anomalies detected with the Mag Array that SAIC and SIG could and could not analyze 3-7

3.4 Comparing the number of anomalies detected with the Mag Array that Sky Research, Inc. and SIG could and could not analyze 3-8

FIGURES

2.1 Flowchart of the Discrimination Study. 2-2

2.2 Aerial photograph of Camp Sibert with the original target point (yellow circle) and areas initially surveyed with the magnetometer version of the MTADS (light green shading). 2-5

2.3 Aerial photograph of Camp Sibert with the original target site (yellow circle), the three selected test areas (large black shapes), geophysical prove out (dark blue square), mag-and-flag area (small black square), and sample excavation area (small red square). 2-6

2.4 Photographs of portions of the selected test areas of Camp Sibert. The left photograph looks south, showing a portion of the SW area in the foreground and the geophysical prove out in the middle of the picture. The right photograph shows a portion of the SE1 area. 2-7

2.5 Photographs of an intact 4.2″ mortar and a splayed half round seeded at Camp Sibert. 2-8

2.6 Example of an intended seed location.. 2-9

2.7 Depth distributions of intact mortars seeded in the GPO. 2-11

2.8 Depth distributions of the intact mortars seeded in the test areas. 2-11

2.9 Sketch of the three overlapping sensor coils mounted on the EM61 Array instrument. 2-13

2.10 A photograph of the EM61 Array collecting data at Camp Sibert. 2-13

2.11 Sketch of the three sensor coils and position sensors (MB1, MB2, MR, and IMU) mounted on the GEM Array instrument 2-14

2.12 Photograph of the GEM Array collecting data at Camp Sibert. 2-15

2.13 Photograph of the EM61 Cart collecting data at Camp Sibert. 2-16

2.14 Photograph of the BUD collecting data at Camp Sibert. 2-17

2.15 Predicted (red and blue lines) and measured responses (◊ and ×) for data collected by the Mag Array instrument from 4.2″ mortars in the Camp Sibert test pit and GPO. 2-20

2.16 Photograph of the EM63 Cued collecting data at Camp Sibert. 2-24

2.17 Sketch of the template grid for collecting data with the GEM Cued 2-25

2.18 Photograph of the GEM Cued collecting data at Camp Sibert. 2-25

2.19 Generating the master list consisted of six steps in the Southeast 1 area. 2-27

2.20 A cartoon mapping of the Array anomaly lists. 2-29

2.21 Step 1.a. Black circles represent 0.3 m radius halos around each Array anomaly. 2-29

2.22 Step 1.b. The centroid of each group (black stars) is calculated over all anomalies belonging to the group. 2-30
2.23 Step 1.c. Large black circles represent 1 m radius halos around each group centroid. 2-30
2.24 Step 1.c. (cont.) Two centroids have been labeled as "clustered" because they are within 2 m of each other. 2-31
2.25 Step 1.d. Large dashed black circles represent 1 m radius halos around each centroid that was (1) labeled as "not clustered" using the "2 m" quantitative criterion in the previous substep and (2) composed of more than one anomaly detected by the same instrument (more than one dot of the same color). 2-31
2.26 Step 1.d. (cont.) One centroid has been relabeled as "clustered" based on visual analysis of the collected data. 2-32
2.27 Step 1.e. All centroids labeled as "not clustered" (black stars) are included on the Array master list. 2-32
2.28 A cartoon mapping of the EM61 Cart anomaly list. 2-34
2.29 Step 3.a. Purple circles represent 0.3 m radius halos around each EM61 Cart anomaly. 2-35
2.30 Step 3.b. The centroid of each new group (purple stars) is calculated over all EM61 Cart anomalies belonging to the group. 2-35
2.31 Step 3.c. Gray and black stars represent original group centroids labeled as "clustered" and "not clustered," respectively, during generation of the Array master list in Step 1. 2-36
2.32 Step 3.c. (cont.) Large purple circles represent 1 m radius halos around each new centroid... 2-36
2.33. Step 3.c (cont.) Two new centroids have been labeled as "clustered" because they are within 2 m of another centroid. 2-37
2.34 Step 3.d. Large dashed purple circles represent 1 m radius halos around each new centroid that was (1) labeled as "not clustered" using the 2 m quantitative criterion in the previous substep and (2) composed of more than one EM61 Cart anomaly (more than one purple dot). 2-37
2.35 Step 3.d (cont.) One centroid has been relabeled as "clustered" based on visual analysis of collected data. 2-38
2.36 Step 3.e. All centroids labeled as "not clustered" (black and purples stars) are included on the Array/M&F/Cart master list. 2-38
2.37 Generating the master list consisted of four steps in the SE2 area. 2-42
2.38 Generating the master list consisted of four steps in the SW area. 2-43
2.39 A cartoon example of a ranked dig list. 2-48
2.40 A cartoon example of a ranked dig list, with locations categorized based on their likelihood of containing clutter versus munitions. 2-49
2.41 A cartoon example of a ranked dig list, with those locations that could not be analyzed inserted into the middle of the list. 2-49

2.42 A cartoon example of a ranked dig list, with those locations that could not be analyzed appended to the end of the list. 2-50
2.43 Master list locations at Camp Sibert. Locations included in the Training Set are shown in red (munitions) and green (clutter), and locations included in the Test Set are shown in black. 2-51
2.44 Scoring the detection performance of a survey instrument and the discrimination performance of an instrument/algorithm combination. 2-55
2.45 Plotting the operating point for an instrument/algorithm combination at the demonstrator's dig threshold..... 2-58
2.46 Generating a ROC curve for an instrument/algorithm combination..... 2-59
2.47 Generating a ROC curve for an instrument/algorithm combination..... 2-60
2.48 Generating a ROC curve for an instrument/algorithm combination..... 2-61
2.49 Generating a ROC curve for an instrument/algorithm combination..... 2-61
2.50 Sketches of two ROC curves. 2-63
3.1 ROC curve for the EM61 Cart instrument and the UXAnalyze software with IDL extension..... 3-5
3.2 ROC curve for the Mag Array instrument and software used by SAIC. 3-6
3.3 ROC curve for the Mag Array instrument and software used by Sky Research, Inc. 3-6
3.4 ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. 3-9
3.5 ROC curve for the Mag Array instrument and the UXAnalyze software..... 3-11
3.6 ROC curve for the EM61 Array instrument and the UXAnalyze software..... 3-12
3.7 ROC curve for the EM61 Cart instrument and the UXAnalyze software. Commercial contractors performed the discrimination analysis. 3-12
3.8 ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. 3-13
3.9 ROC curve for cooperative inversions based on the EM61 Array and Mag Array instruments and software based on a multidimensional classifier..... 3-14
3.10 ROC curve for the EM61 Array and software based on a multidimensional classifier. 3-15
3.11 ROC curve for the EM61 Array and software based on size-based features only. 3-16
3.12 Histograms of a sized-based feature extracted from clutter (green) and munitions (red). 3-17
3.13 ROC curve for the Mag Array and the UXAnalyze software..... 3-18
3.14 ROC curve for the GEM Array and SAIC custom software..... 3-19
3.15 ROC curve for the GEM Cued and software based on a multidimensional classifier. 3-20

3.16 ROC curve for independent inversions based on the GEM Cued and Mag Array instruments with a multidimensional classifier. 3-21
3.17 ROC curve for the EM63 Cued and software based on a multidimensional classifier. 3-23
3.18 ROC curve for cooperative inversions based on the EM63 Cued and Mag Array instruments and software based on a multidimensional classifier. 3-23
3.19 ROC curve for the BUD instrument in cued mode and software based on a multidimensional template matcher. 3-24
3.20 ROC curve for the BUD instrument in survey mode and software based on a multidimensional template matcher. 3-25
3.21 ROC curve for inversions based on the EM61 Array and Mag Array instruments and software based on a multidimensional classifier. 3-27
3.22 ROC curve for inversions based on the EM61 Array and Mag Array instruments and software based on a multidimensional classifier with active learning. 3-27
3.23 ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. 3-29
3.24 ROC curve for the EM61 Array instrument and software based on a multidimensional classifier with semi-supervised learning. 3-29
3.25 ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. 3-32
3.26 ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. 3-33

EXECUTIVE SUMMARY

INTRODUCTION

The Fiscal Year 2006 Defense Appropriations Bill contained funding for the "Development of Advanced, Sophisticated, Discrimination Technologies for UXO Cleanup" in the Environmental Security Technology Certification Program (ESTCP). The discrimination demonstration carried out at the former Camp Sibert near Gadsden, AL, was in direct response to the congressional language. The high-level goal of the demonstration was to assess the capability of discrimination algorithms, developed under the Strategic Environmental Research and Development Program (SERDP) and refined under ESTCP, to reliably determine which detected items could be left safely in the ground and which had to be dug. A 2003 Defense Science Board study noted that as much as 75% of current UXO cleanup costs might be associated with digging up non-hazardous scrap [7]. Obviously, the development, validation, and acceptance of reliable discrimination technologies that would allow nonhazardous items to remain in the ground has the potential to significantly reduce UXO clearance costs or to allow more areas to be cleared for the same amount of funding.

The intent of the demonstration was to evaluate on a live site those algorithms that had proven successful in previous testing, principally at engineered test sites. Another important goal was to involve the regulatory community early in the design of the demonstration in an effort to better understand what might be required if detected items were actually to be left in the ground. This report, prepared by the Institute for Defense Analyses (IDA), provides the detailed results of the demonstration.

OBJECTIVES AND APPROACH

The objectives of this demonstration were to

1. Test and validate detection and discrimination capabilities of currently available and emerging technologies on real sites under operational conditions.
2. In cooperation with regulators and program managers, investigate how discrimination technologies can be acceptably implemented in cleanup operations.

Camp Sibert was selected as a demonstration site because it met a number of desired characteristics. Historical records showed that Camp Sibert was likely to be contaminated with only one type of munition, the 4.2″ mortar; terrain and geology were relatively benign; the landowners were amenable to the demonstration; and the Army Corps of Engineers had ongoing clearance actions at Camp Sibert that were able to provide needed support to this effort. To improve the likelihood that a statistically significant number of munition items were detected and dug during clearance operations, IDA developed a seed plan, and 140 previously fired, inert 4.2″ mortars were buried on the demonstration site prior to data collection. The test areas were surveyed using five different data-collection instruments. IDA created a "master anomaly list" that included the locations of anomalies detected by one or more data-collection instruments. In addition, high-density "cued" data were collected using 3 data-collection instruments at 200 locations on the master anomaly list.

The data-collection team then excavated items from the ground at each location on the master anomaly list. Based on the excavated items, the Program Office assigned ground truth labels to each location, with some locations assigned the label of "munition" and other locations assigned the label of "clutter." IDA then separated the locations into a Training Set and Test Set.

The Program Office distributed the collected data and the master anomaly list to each demonstration team. The demonstrators also received the ground truth labels for all locations in the Training Set, but remained blind to the ground truth labels for all locations in the Test Set. The demonstrators performed a geophysical inversion on the data encompassing each anomaly to produce a feature vector and then used the ground-truth labels in the Training Set to optimize their data processing algorithms. The algorithms estimated the probability or likelihood that a location contained clutter only. The demonstrators applied their optimized algorithms to the data in the Test Set while remaining blind to the ground truth labels. The demonstrators created a "ranked dig list" by arranging the Test Set locations according to their estimated probability or likelihood of being clutter. The demonstrators also specified a "dig threshold" that could be applied to the ranked dig list, such that it was likely that all locations on the ranked dig list above the dig threshold could be left safely in the ground. For cases where the data did not support a reliable inversion, the associated location was designated "Can't analyze" and those anomalies were appended to the bottom of the list as items to be dug.

IDA scored each demonstrator's ranked dig list and dig threshold by comparing the "dig/do not dig" labels assigned to each location in the Test Set to its ground truth

label. The discrimination performance of each instrument/algorithm combination was summarized with the metrics Pd (probability of detection, or the fraction of munitions labeled as "dig") and FP (false positives, or the number of unnecessary digs). IDA also revisited the choice of dig threshold by retrospectively testing every possible value, calculating Pd and FP, and plotting these metrics against each other to form a receiver operating characteristic (ROC) curve. These ROC curves and the statistics drawn from them lead to the key findings from this demonstration.

FINDINGS

- *Once "Can't analyze" locations were dug, discrimination performance was usually good for all remaining locations*—A large majority of the tested instrument and algorithm combinations demonstrated very good discrimination performance for those locations that could be analyzed. That is, the demonstrator's dig threshold led to a large reduction in FP while Pd remained at or near 1.00.
- *Commercially available instruments and software often led to good discrimination performance*—Cesium vapor magnetometer array data and EM61 Mk2 array and cart data were successfully processed using the commercial UXAnalyze software package; the result was good discrimination performance (Pd near unity with significant FP reduction).
- *The multiple-axis Berkeley UXO Discriminator (BUD) instrument provided high-signal-to-noise-ratio data from a single location leading to excellent discrimination performance in both cued and survey modes*—The dig threshold applied to the BUD ranked dig list would have resulted in fewer than 25 of about 200 potential FP items being dug.
- *Much of the discriminating power seen at Camp Sibert is due to size-based features*—The 4.2″ mortar was substantially larger than much of the clutter found on the site. While multifeature classifiers provided some improvement over size-based classifiers, size was a sufficient discriminant to allow identification of a large percentage of non-munitions items.
- *Mag-and-flag led to a large number of unnecessary digs*—While mag-and-flag detected all munitions items in the 100′ x 100′ grid it surveyed, the overall background alarm rate was twice that of the magnetometer array before discrimination processing and a factor of 15 larger after discrimination processing.
- *Although all "Can't analyze" locations must be dug and can constitute a significant percentage of the dig list, a principled, documented method for identifying "Can't analyze" locations has not yet been agreed upon*—Anomalies for which data did not allow an inversion of sufficient quality for

discrimination obviously must be dug. However, different demonstrators judged as "Can't analyze" greatly different numbers of anomalies detected by the same instruments in the same areas. An objective of ongoing efforts is to understand the causes for an inability to successfully invert collected data and to suggest quantitative measures for declaring an anomaly as "Can't analyze."

CONCLUSIONS AND LESSONS LEARNED

The major conclusion that can be drawn from the Camp Sibert demonstration is that successful discrimination is possible on a live site using currently available sensors and software. By adjusting the dig threshold, most of the submitted dig lists would have resulted in significantly fewer digs while still removing all the 4.2″ mortars in the survey area. Although this was a very benign site, it was important to establish that current technology was successful in even that environment.

Apart from the findings and conclusions regarding performance that have been drawn from this demonstration, we have learned a number of lessons that will be used to guide the planning and conduct of follow-on discrimination demonstrations:

- "Can't analyze" items should not be part of the ranked dig list. Instead, they should be appended to the bottom of the list and scored as a group for retrospective ROC curve analysis.
- The Program Office should provide the demonstrators a standard template for ranked dig lists so that data arrive in a consistent fashion to ease scoring.
- A single geographic monument should be used for all data-collection activities, and that monument should be resurveyed as part of the setup process. If multiple monuments must be used, their absolute positions should be checked against each other.
- The schedule should be arranged to provide more time for quality assurance on sensor data sets before moving forward to the detection phase. In the Sibert case, motion noise problems in the southwest area due to furrows in the ground should have been recognized and dealt with early.
- Demonstrators should develop and apply specific, principled, quantitative criteria to determine what anomalies should be declared "Can't analyze."

1. INTRODUCTION

The Fiscal Year 2006 Defense Appropriations Bill contained funding for the "Development of Advanced, Sophisticated, Discrimination Technologies for UXO Cleanup" in the Environmental Security Technology Certification Program (ESTCP). The discrimination demonstration carried out at the former Camp Sibert near Gadsden, AL, was in direct response to the congressional language. The high-level goal of the demonstration was to assess the capability of discrimination algorithms, developed under the Strategic Environmental Research and Development Program (SERDP) and refined under ESTCP, to reliably determine which detected items could be left safely in the ground and which had to be dug. A 2003 Defense Science Board study noted that as much as 75% of current UXO cleanup costs might be associated with digging up non-hazardous scrap [7]. Obviously, the development, validation and acceptance of reliable discrimination technologies and algorithms has the potential to significantly reduce UXO clearance costs or to allow more areas to be cleared for the same amount of funding. This demonstration represents an initial step along the path to UXO discrimination validation and acceptance.

The intent of the demonstration was to evaluate on a live site those algorithms that had proven successful in previous testing, principally at engineered test sites. Another important goal was to involve the regulatory community early in the design of the demonstration in an effort to better understand what might be required if detected items were actually to be left in the ground.

Under a task titled "ESTCP/SERDP: Assessment of Traditional and Emerging Approaches to the Detection and Identification of Surface and Buried Unexploded Ordnance," the Institute for Defense Analyses (IDA) was assigned the responsibility to assist ESTCP in planning, carrying out, and scoring the discrimination demonstration. IDA's principal functions were to assist in site selection, provide seed emplacement locations and burial procedures, create a master anomaly list, develop scoring protocols, score demonstrators' detection and discrimination results, and provide a comprehensive final report describing the demonstration. This final technical report serves as an adjunct to the summary final report produced by ESTCP [15].

1.1 DETAILED OBJECTIVES

The discrimination study demonstration plan lays out the detailed objectives of this demonstration:

1. Test and validate detection and discrimination capabilities of currently available and emerging technologies on real sites under operational conditions.
2. In cooperation with regulators and program managers, investigate how discrimination technologies can be acceptably implemented in cleanup operations.

Within each of these two overarching objectives are several technical sub-objectives:

- Test and evaluate capabilities by demonstrating and evaluating individual sensor and software technologies, as well as processes that combine these technologies. Compare advanced methods to existing practices and validate the pilot technologies for the following:
  - Ability to detect UXO.
  - Ability to identify features that distinguish scrap and other clutter from UXO.
  - Ability to reduce false alarms (items that could be left in the ground that are incorrectly classified as UXO) while maintaining a probability of detection (Pd) of UXO that is acceptable to all.
  - Ability to identify sources of uncertainty in the discrimination process and to quantify their impact to support decision-making, including issues such as impact of data quality due to how data are collected.
  - Ability to quantify the overall impact on risk arising from the capability to clear more land more quickly for the same investment.
  - Ability to address the issues of a dig/no-dig decision process and the related quality-assurance/quality-control issues.
- Understand the applicability and limitations of the pilot technologies in the context of project objectives, site characteristics, and suspected munition contamination.
- Collect high-quality, well documented data to support the next generation of signal-processing research.

This report discusses a subset of these points. The remaining points are discussed in the summary final report produced by ESTCP [15].

1.2 DEMONSTRATION MOTIVATION

A 2003 Defense Science Board (DSB) study on UXO cleanup technologies pointed out that in a typical clearance action, more than 99% of the items dug could have been left in the ground [7]. It also noted that reducing the false-alarm rate from greater than 99% to a lower, yet still relatively high, number could still save much of the cost of clearance actions.

Significant progress has been made in discrimination technology as a result of SERDP and ESTCP funding. To date, however, testing of these approaches has been primarily limited to artificially constructed test sites with only limited application at live sites. Acceptance of discrimination technologies requires demonstration of system capabilities at live UXO sites under real-world conditions. Any attempt to declare detected anomalies to be harmless and requiring no further investigation will require the demonstration to regulators of not only individual technologies, but an entire decision-making process. This discrimination study was the first phase in a continuing effort that will span several years. A follow-on demonstration at a more challenging site is already in the initial planning stage.

The importance of live-site testing is that the distribution of the items in the ground before testing is realistic for both UXO and clutter items. While extremely valuable, areas such as the Standardized UXO Test Sites [13] will always be somewhat artificial because both UXO and clutter items have been emplaced in accordance with preconceived notions of how they should be distributed in type, size, and depth, as well as location. Although it is usually necessary in live-site testing to seed the area with appropriate UXO to ensure sufficient munitions to provide reasonable statistics, the in situ clutter and any in situ UXO types are, by definition, “real” for that site.

1.3 GENERAL APPROACH

The Program Office, in conjunction with IDA and the Discrimination Study Advisory Panel, selected Camp Sibert for the study because it met a number of desired characteristics. Namely, historical records showed that Camp Sibert was likely to be contaminated with only one type of munition, the 4.2″ mortar. Data-collection teams initially surveyed Camp Sibert with a magnetometer array and the initial survey results were used to select test areas for the study, as well as an area for the geophysical prove out (GPO). The purpose of the GPO was to confirm that all data-collection instruments were properly functioning—that is, they were able to detect all known munitions to the

desired depth. To that end, an exhaustive excavation was performed to clear the GPO of all metallic items before seeding the GPO and collecting data.

IDA generated a plan to emplace previously fired 4.2″ mortars (seeds) throughout the test areas and GPO (Appendix A). The site-support contractor (Parsons) followed this plan and emplaced the seeds as directed. The emplacement team took great care to seed the items at least 3 m away from each other and from other magnetic anomalies, as previous work has shown that current discrimination technologies cannot reliably analyze multiple, closely spaced items with overlapping signatures [1], [13].

Next, the data-collection team surveyed the test areas and GPO using five different data-collection instruments. In addition, IDA selected 200 locations at which the team collected high-density "cued" data using three data-collection instruments. The Program Office team selected detection thresholds for the survey instruments and confirmed the validity of these thresholds using the GPO. As was recognized at the beginning of the study, different survey instruments resulted in different anomaly detection lists. That is, many items were detected by all instruments, some items were detected by more than one but not all instruments, and some items were detected by a single instrument only. IDA developed documented methods for reconciling the differences between individual instrument's anomaly lists to produce a single "master anomaly list."

The site-support contractor then excavated items from the ground at each location on the master anomaly list. Based upon the excavated items, the Program Office assigned ground-truth labels to each location, with some locations assigned the label of "munition" and other locations assigned the label of "clutter." IDA then separated the locations on the master anomaly list into a Training Set and a Test Set.

The Program Office distributed the collected data and the master anomaly list to each demonstration team. The demonstrators also received the ground-truth labels for all locations in the Training Set, but remained blind to the ground-truth labels for all locations in the Test Set. The demonstrators used the data and ground-truth labels in the Training Set to optimize their inversion routines and discrimination algorithms. Inversion routines are used to fit the data collected around each location on the master anomaly list to a model to estimate parameters of the buried target. Discrimination algorithms are used to estimate the likelihood or probability that a buried target is clutter based on its estimated parameters. The demonstrators then applied their optimized processes to the data in the Test Set while remaining blind to the ground-truth labels. The demonstrators

created a "ranked dig list" by arranging the locations in the Test Set according to their estimated probability or likelihood of being clutter. The demonstrators also specified a "dig threshold" that could be applied to the ranked dig list, such that it was likely that all locations on the dig list above the dig threshold could be left safely in the ground.

IDA scored each demonstrator's ranked dig list and dig threshold by comparing the dig/no-dig labels assigned to each location in the Test Set to its ground-truth label. IDA summarized the discrimination performance of each instrument/algorithm combination with the metrics Pd (probability of detection, or the fraction of munitions labeled as "dig") and FP (false positives, or the number of unnecessary digs). IDA also revisited the choice of dig threshold by retrospectively testing every possible value. For each possible value of the dig threshold, IDA calculated Pd and FP and plotted these metrics against each other to form a receiver operating characteristic (ROC) curve. The ROC curves and the statistics drawn from them lead to the key findings from this demonstration. They are discussed in detail in the Selected Results and Discussion section of this report.

1.4 LIMITATIONS

As a first demonstration involving a number of data-collection instruments and discrimination algorithms employed at a live site, this effort had a number of limitations:

- The primary limitation was the need to seed Camp Sibert with munitions to obtain reasonable discrimination statistics. In an ideal demonstration, the area tested would be sufficiently large that valid statistics could be gained simply from recovered intact UXO. In that case, a potentially artificial distribution of UXO density, depths, and orientations is not a concern. In this demonstration, however, only one intact UXO item was found in the approximately 20 acres that were excavated. Furthermore, this item was found in the area of high anomaly density that was excavated during the initial stages of the study to help understand the distribution of local anomaly types. Thus, to collect data from enough recovered intact UXO for valid statistics, a very large area of the site would have to be tested. The cost of excavating all anomalies detected in such a large area would have been prohibitive. Thus, seeding was required in this demonstration, resulting in a potentially artificial distribution of UXO density, depths, and orientations. This is a limitation that is unlikely to ever be overcome in scientific testing because of funding constraints.
- A second, but planned, limitation of this effort was that the single, large UXO type was expected to be present on the site—a 4.2″ mortar round—and was the only type of seed emplaced. Although this situation can occur, as

evidenced by Camp Sibert, it is not typical. The plan in this case was to evaluate the capabilities of current instruments and discrimination algorithms in a benign, but live, site. The excellent performance in this demonstration points the way to subsequent demonstrations where site topography, geology, and target types become progressively more challenging.

- A third limitation was unexpected. The original plan had been to exhaustively excavate all anomalies that exceeded the detection threshold of all instruments. It was hoped that such an excavation would remain within the budget, allowing approximately 2,000 anomalies to be dug. However, motion noise in the GEM array and Mag array in portions of the southwest test area led to a number of anomalies that were not correlated with EM61 array or EM61 cart anomalies and were judged highly unlikely to arise from real objects. Thus, dozens of those anomalies were removed from the master anomaly list and were not intrusively investigated or included in the scoring process.

2. METHODS

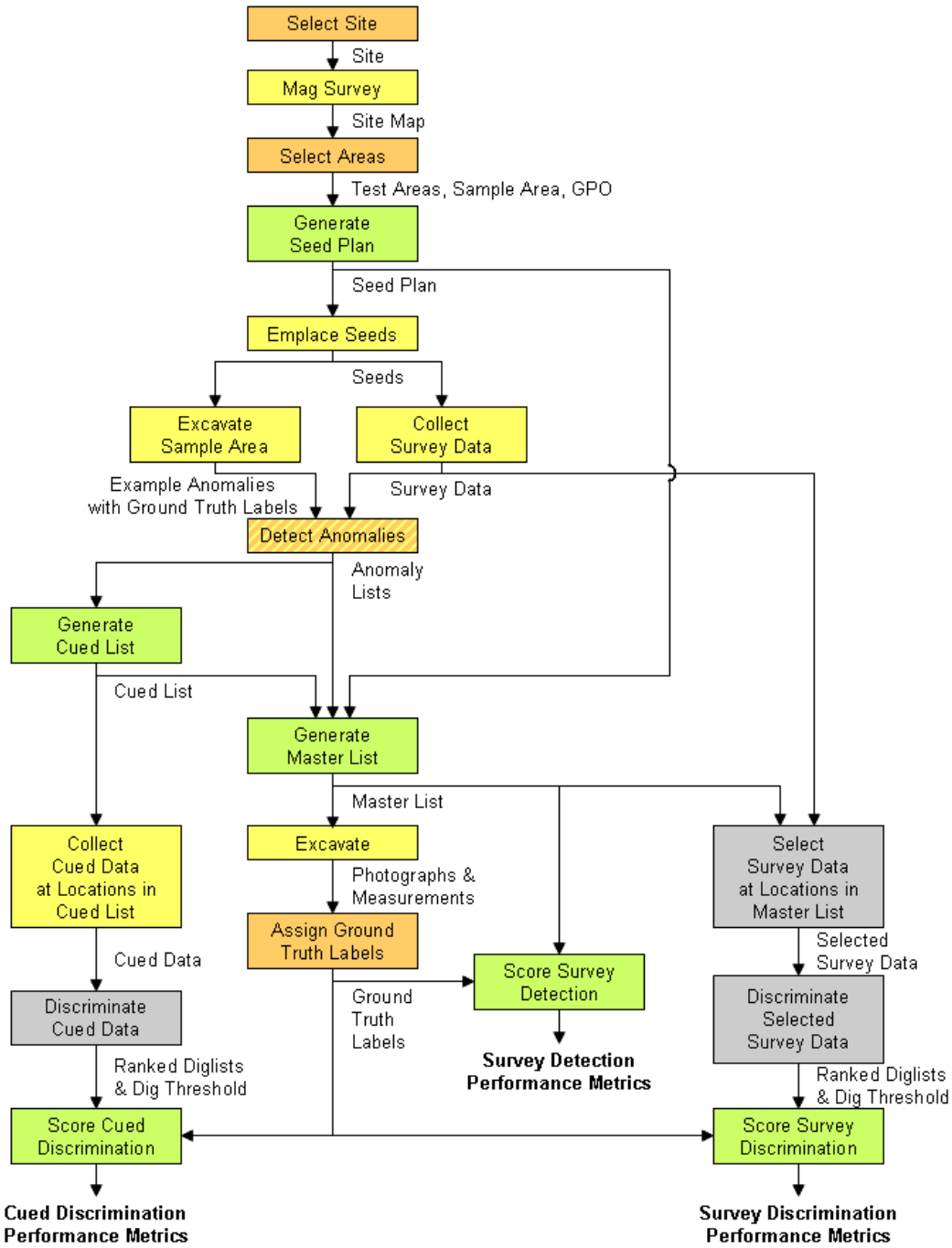
This chapter describes the process used to select a site for the study, select particular areas of the site as test areas, emplace seed targets in the test areas, collect data from the test areas, select anomalies from the collected data, provide collected data and a master list of selected anomalies to the discrimination demonstrators, and score the results of the demonstrators' discrimination outputs. Figure 2.1 shows a flowchart of the process that was followed. While the ultimate results of the scoring process were metrics describing the discrimination performance, those results came at the end of a relatively complicated, but carefully structured, process. This chapter expands and explains each box on the flowchart.

2.1 SELECTION OF SITE

The Program Office selected the site in close coordination with the Discrimination Study Advisory Group. As this study was the first attempt to demonstrate extensive discrimination on a live site, the Program Office made a conscious decision to select a site where challenges aside from discrimination were minimized. Furthermore, the Advisory Group discouraged the Program Office from selecting a practice bomb site because these sites are thought to be of minor interest to the UXO discrimination community. Other characteristics were sought as well, including:

- Benign topography and land cover.
- Relatively benign geology.
- A single-use site or a site containing no more than two munition types.
- Relatively large munitions (not 20 mm or 40 mm) but not practice bombs.
- A site that could provide anomaly densities of approximately 100 per acre to minimize overlapping signatures and provide approximately 2,000 targets to be dug in approximately 20 acres.

Benign topography and land cover were sought to allow the Multi-sensor Towed Array Detection System (MTADS) to be used as a data-collection instrument, along with the more typical commercial instrument based on an EM61-Mk2 sensor mounted on a cart. Use of these two types of instruments allowed comparison of high-quality array data versus carefully collected commercial survey data.



Box	Step carried out by
Orange:	Program Office
Yellow:	Data-collection team
Green:	IDA
Gray:	Discrimination demonstrators

Figure 2.1: Flowchart of the Discrimination Study.

As geologic features containing magnetic soil and rock can severely degrade magnetometer performance, relatively benign geology was sought to allow use of the magnetometer version of the MTADS instrument. High-quality magnetometer data were needed to select test areas for the study, as well as to assess the added benefit in discrimination performance resulting from the use of cooperative inversions. Cooperative inversions occur when magnetometer data are used to constrain inversions based on electromagnetic induction (EMI) data.

Based on the results from the Standardized UXO Test Sites, the Program Office recognized the difficulty of performing discrimination on sites containing several different types of munitions (ranging from 20 mm rounds to 155 mm artillery shells), along with clutter items of all sizes. Therefore, the Program Office decided to focus on sites with one or at most two expected munitions types. In addition, because of the well-known difficulties in surveying very small munitions [13], the Program Office sought a site where the expected munition was at least as large as a 60 mm mortar.

Finally, the Program Office sought a site with a sufficient density of anomalies to limit the survey area to a manageable size (approximately 20 acres), given the amount of funding available for excavations (of approximately 2,000 anomalies), but not so dense that there would be an abundance of overlapping signals. Previous work has shown that current data-collection instruments and discrimination algorithms do not allow reliable inversion and discrimination of clusters—multiple, closely spaced anomalies with overlapping signals [1]. Data were collected for a number of clusters in the selected site. The demonstration teams did not attempt to process the data collected from the clusters as part of this study, but these data are available for processing as part of future SERDP tasks.

2.1.1 Former Camp Sibert: History and Characteristics

After visiting a number of potential sites and polling the Advisory Panel, the Program Office selected the former Camp Sibert for the study. Camp Sibert met most of the desired characteristics and had a number of advantages, including:

- Camp Sibert was a single-use site, having been a training site for the use of 4.2″ mortars during World War II.
- Ongoing clearance activities were already occurring in a portion of Camp Sibert near the areas of interest; the paperwork required for survey and clearance activities was already in place and Parsons, a commercial contractor, was already on site to provide surveying, emplacement, and digging services.

- The area of interest was owned by a single landowner who was amenable to the survey and clearance efforts.
- A suitably large portion of the area of interest provided sufficiently benign land cover and geology to allow the collection of high-quality EMI and magnetic data.

Camp Sibert is a formerly used defense site about 8 miles southwest of Gadsden, AL. The portion of the site chosen for the study is within what is denoted as "Site 18." It is currently privately owned land predominantly used for hunting. A lodge is on the site (built on top of the mortar training aim point), and much of the site is regularly cultivated and planted to raise crops that attract animals to be hunted.

2.1.2 Geolocation Survey Control Points

As shown in Table 2.1, there were five control points in the vicinity of the surveyed areas that could be used as reference positions for differential global positioning system (GPS) measurements. The Program Office directed all participants to use Point 189 for their reference, as it was reasonably situated for all the measurement areas and had been used for the initial magnetometer survey during site selection. However, a number of the participants chose to use different monuments, including the Parsons team, which emplaced seed items in the GPO and the three survey areas.

Table 2.1: Available survey control points in the vicinity of Site 18 of the former Camp Sibert.

Point	Latitude	Longitude	Northing (m)	Easting (m)
	NAD 83		UTM Zone 16N	NAD 83
165	33° 54' 05.22848" N	86° 09' 17.17042" W	3,751,550.813	578,146.300
166	33° 54' 06.61350" N	86° 09' 09.19992" W	3,751,595.159	578,350.654
189	33° 54' 03.19413" N	86° 09' 03.92590" W	3,751,490.960	578,486.975
354	33° 54' 39.30301" N	86° 08' 39.26633" W	3,752,608.379	579,111.040
355	33° 54' 39.99249" N	86° 08' 36.07590" W	3,752,630.298	579,192.793

In evaluating detection results for the Mag Array in the GPO, IDA noticed a bias in the positions that led to several missed targets, based on the Parsons emplaced positions. A check by the Parsons surveyor revealed that Point 189 was not at the advertised position listed in Table 2.1, but instead was 0.22 m north and 0.11 m west of its published location. Because of that, as part of the scoring process, data sets had to be adjusted to be aligned with a common coordinate system. Lessons learned from this were that all participants should use the same geolocation reference point and that point should be resurveyed before data collection begins.

2.2 MAGNETOMETER SURVEY

A data-collection team surveyed portions of Camp Sibert with the magnetometer version of the MTADS. Figure 2.2 shows an aerial photograph of Camp Sibert with the surveyed portions shaded in light green. The purpose of the magnetometer survey was to select the test areas for the study. The total area of the surveyed portions was 19.3 ha (47.4 acres), with 2.4 ha in the region marked "East," 3.2 ha in the region marked "North," 4.6 ha in the "Southeast," and 9.1 ha in the "Southwest."

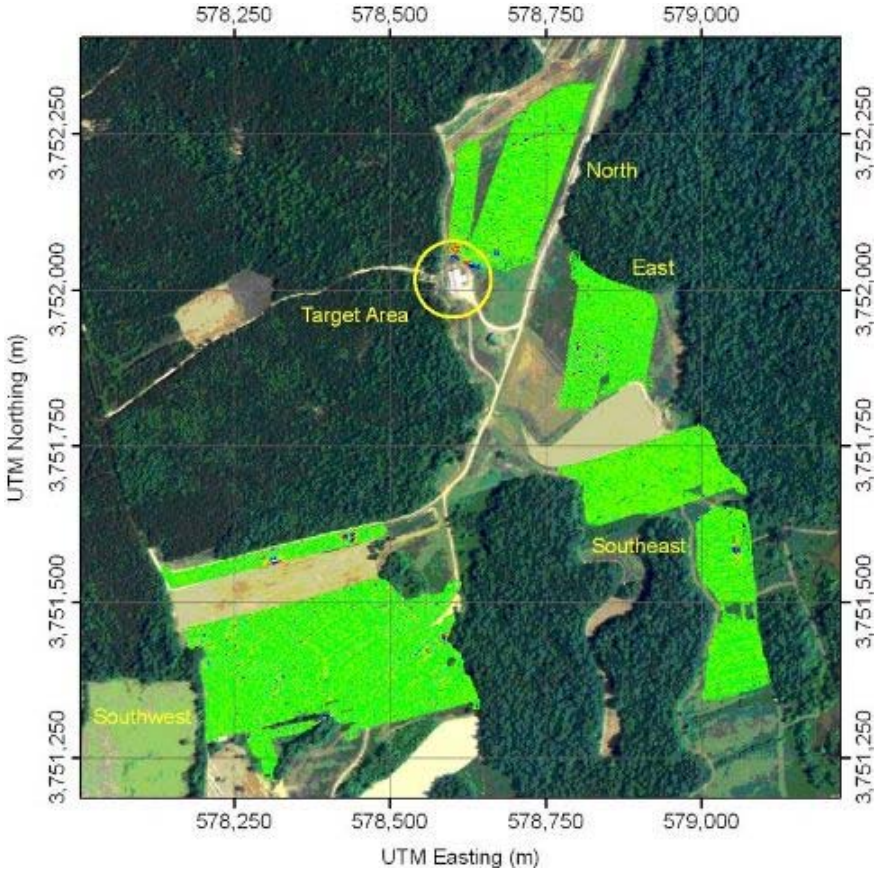


Figure 2.2: Aerial photograph of Camp Sibert with the original target point (yellow circle) and areas initially surveyed with the magnetometer version of the MTADS (light green shading).

2.3 SELECTION OF AREAS

Based on the results of the magnetometer survey, the Program Office selected three test areas for the study. Outlined in black in Figure 2.3, the test areas were designated Southwest (SW), Southeast 1 (SE1), and Southeast 2 (SE2). The total area of the three test areas was approximately 6 ha, and each area contained the desired density of magnetometer anomalies. Figure 2.4 shows photographs of two surveyed portions of

Camp Sibert. The right photograph shows a portion of the SE1 area. The left photograph looks south and shows a portion of the SW area in the foreground. Although not shown in the photograph, sections of the SW area had been previously plowed. Operating the data-collection instruments over the plowed furrows introduced significant motion noise into the EMI data. Although the land cover was generally benign, the right side of Figure 2.4 shows some plant growth that made surveying more difficult in those areas.

2.4 CLEARANCE OF GEOPHYSICAL PROVE OUT

The Program Office also used the results of the magnetometer survey to select an area for the GPO. The GPO was located adjacent to the southwest corner of the SW test area and is outlined in dark blue in Figure 2.3. It can also be seen behind the SW area in the left photograph in Figure 2.4.

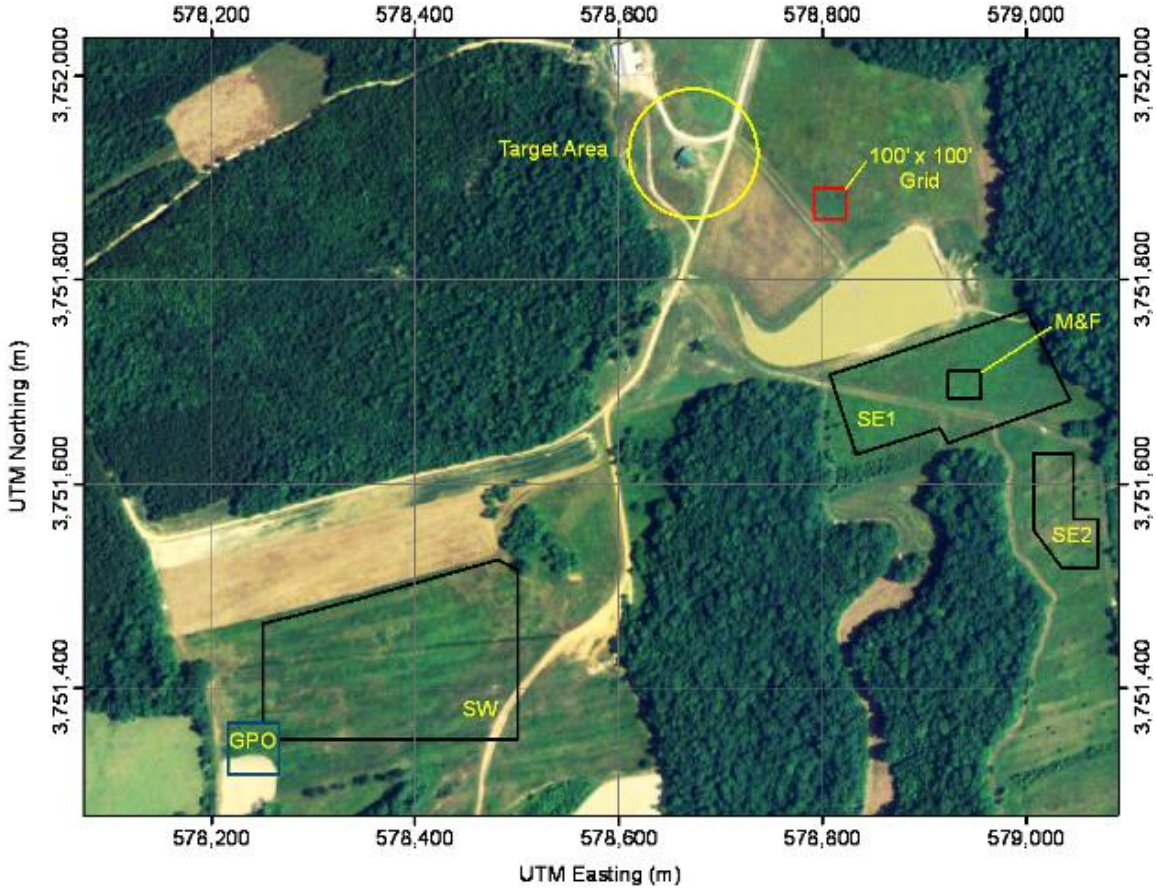


Figure 2.3: Aerial photograph of Camp Sibert with the original target site (yellow circle), the three selected test areas (large black shapes), geophysical prove out (dark blue square), mag-and-flag area (small black square), and sample excavation area (small red square).

Figure 2.4: Photographs of portions of the selected test areas of Camp Sibert. The left photograph looks south, showing a portion of the SW area in the foreground and the geophysical prove out in the middle of the picture. The right photograph shows a portion of the SE1 area.

The purpose of the GPO was to confirm the ability of the different data-collection instruments to detect all known munitions. It was therefore important to clear the GPO of all metallic objects before seeding. The seed emplacement team cleared the GPO of all anomalies previously identified in the magnetometer survey. Next, the team surveyed the GPO a second time using a typical commercial instrument based on an EM61-Mk2 sensor mounted on a cart. The team also cleared the GPO of any remaining anomalies identified in this second EMI survey.

2.5 EXCAVATION OF SAMPLE AREA

The Program Office also selected a 100′ × 100′ area for exhaustive excavation using the results of the magnetometer survey. This area, outlined in red in Figure 2.3, exhibited a high density of magnetic anomalies. The purpose of the excavation was to better understand the types, depths, orientations, and distributions of munitions at Camp Sibert. During the planning stage of the study, the Program Office intended that the results of the excavation would guide the emplacement of seeds in the three test areas and GPO. As the study progressed, however, it became clear that to remain on schedule, the emplacement of the seeds would have to begin before the excavation of the sample area was completed. In retrospect, this did not prove to be a problem, as only one intact munition (a 4.2″ smoke round) was excavated from the sample area. Even if the excavation been completed before seed emplacement, as originally intended, the seed emplacement would not have benefited greatly from the results.

2.6 GENERATION OF SEED PLAN

The magnetometer survey also aided in the development of a plan to seed the three test areas with intact rounds and the GPO with intact rounds and splayed half rounds. Seeded items were either intact inert 4.2″ rounds (previously fired at a location other than Camp Sibert) or splayed half rounds (4.2″ mortars than had been previously fired at Camp Sibert and had detonated). The intact rounds were obtained from the Montana National Guard. Figure 2.5 shows photographs of an intact mortar and a splayed half round.

Figure 2.5: Photographs of an intact 4.2″ mortar and a splayed half round seeded at Camp Sibert.

IDA generated the seed plan, which listed the intended locations of all seed items in both the GPO and the three test areas. Since all anomalies detected in the GPO had been previously cleared, the GPO provided a relatively clean region (other than magnetic geology at the northeast corner of the site) in which to emplace seed items. In contrast, the three test areas had *not* been previously cleared. Many large and small clutter items and some magnetic geology remained in these areas. When choosing the intended locations of the seed items, IDA attempted to avoid seeding an item near strong anomalies identified during the magnetometer survey because previous work showed that current data-collection instruments and algorithms cannot discriminate multiple, closely spaced items with overlapping signatures [1].

IDA provided the emplacement team with a list of intended burial parameters, including the intended locations, depths, and orientations of every item to be seeded at Camp Sibert. Intended locations were listed to 1 cm precision, intended depths to 10 cm precision, and intended azimuths were 30 degree precision. The emplacement team was instructed to use the same precision when emplacing the items but to precisely document the actual placement. Appendix A gives the intended burial parameters.

2.7 EMPLACEMENT OF SEEDS

The Parsons team emplaced each seed at or near its intended location. As part of the seed plan, the emplacement team inspected each intended seed location with a hand-held detector. The team removed any metallic objects that were found at the intended location, although no special efforts (e.g., sifting or expanding the hole) were made to find such objects. If, after removing all found metallic objects, a strong (>16 nT) magnetic anomaly not initially identified in the magnetometer survey was detected near the seed's intended location, then the emplacement team chose a different, yet nearby, location for seeding the item. Figure 2.6 shows an intended seed location in the SW area overlaid on the data collected from the initial magnetometer survey. The intended location was farther than 3 m away from all magnetic anomalies greater than 16 nT in strength (red and pink areas). Therefore, the intended location was considered suitable and the item was buried.

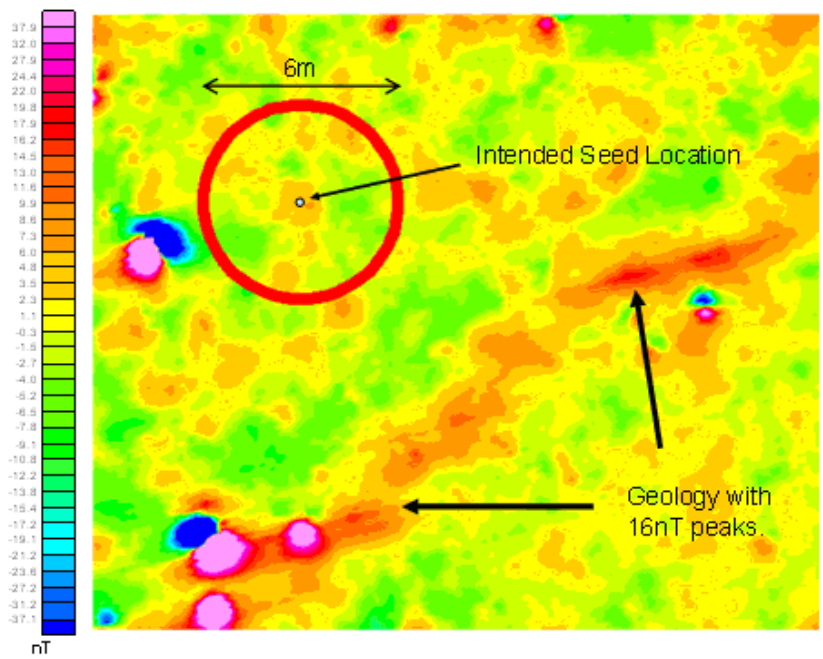


Figure 2.6: Example of an intended seed location. Note that the 5–8 nT variations are common in the southwest portion of the site.

After emplacing a seed item, the team recorded the item's actual burial parameters, including:

- The easting, northing, and depth coordinates for the nose, tail, and center of each item, with the depth determined by surveying a point on the edge of the hole to establish the elevation of the local ground surface. The depth to the center of the round was used to calculate depth distributions for the seed items.

- The dip orientation for the item, labeled as:
  - Up: Nose within 10 degrees of pointing straight up.
  - Down: Nose within 10 degrees of pointing straight down.
  - Sideways: Nose within 45 degrees (up or down) of being horizontal.
- A photograph of the item taken after it was put into place but before covering it with dirt, with the serial number of the item visible in the photograph and a ruler laid next to the item.

Thirty intact mortars and eight splayed half rounds were seeded in the GPO. Because the purpose of the GPO was to confirm the ability of the different data-collection instruments to detect munitions, some of the seeded items were buried in the GPO at depths close to or at 11 times their diameter, an Army Corps of Engineers (COE) guideline on the limits of detection performance for current survey instruments [13]. Figure 2.7 shows a histogram of the measured depths of the items seeded in the GPO. Five of the 30 intact mortars (20%) were seeded deeper than 8 times their diameter. In fact, 4 (13%) were seeded deeper than 10 times their diameter. Furthermore, two of these four deep mortars were unintentionally seeded in an area of high geologic noise, making their detection even more challenging than originally planned. Another purpose of the GPO was to collect data that the demonstrators could use to optimize their discrimination algorithms; therefore, the majority of the intact mortars were seeded at depths more typical of fired mortars: 21 of the intact mortars (70%) were seeded at depths shallower than 6 times their diameter.

The seeding philosophy in the three test areas differed from the seeding philosophy in the GPO. Most items are not found near their maximum depth, so what was felt to be a realistic depth distribution was chosen. In addition, because this study was intended to be a discrimination (rather than detection) study, and because previous work had indicated that accurate discrimination requires a high signal-to-noise ratio (SNR) in the collected data [2], proportionally fewer items were seeded at large depths in the test areas than in the GPO. Figure 2.8 shows a histogram of the depths of the items seeded in the three test areas. Of the 149 intact mortars seeded in the test areas, only 18 (12%) were seeded deeper than 8 times their diameter, with only 5 (3%) deeper than 10 times their diameter. In contrast, 124 intact mortars (83%) were seeded at depths less than 6 times their diameter.

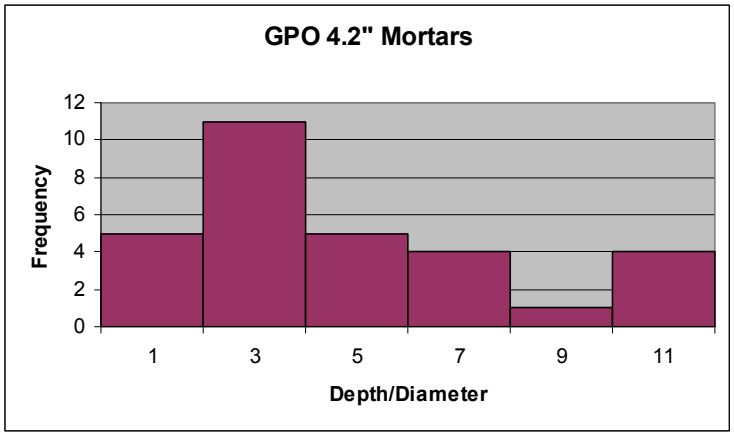


Figure 2.7: Depth distributions of intact mortars seeded in the GPO.

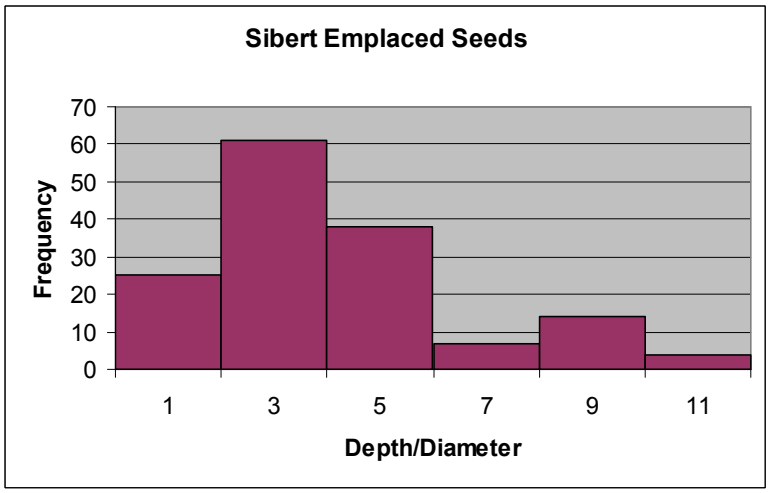


Figure 2.8: Depth distributions of the intact mortars seeded in the test areas.

2.8 DATA COLLECTION IN SURVEY MODE

The data-collection team surveyed the test areas with several different data-collection instruments. This section explains the motivation for selecting the instruments and briefly describes the sensor technology employed by each of the instruments. More detail can be found in the reports written by the data-collection team [8], [9], [11].

One goal of this study was to assess the performance of different discrimination algorithms. These algorithms are based on the assumption that the data collected around a detected anomaly originated from a dipole-like source, a standard model for UXO ([3], [11]). Other studies have shown that data quality has a large effect on the accuracy of inverting the collected data to fit the dipole models [2]. Therefore, a secondary goal of this study was to assess the performance of the discrimination algorithms when operating on data of different qualities.

Previous studies have demonstrated that the MTADS instruments can collect extremely high-quality survey data [13]. The MTADS, developed by the Naval Research Laboratory, consists of a specially designed tow vehicle with a low magnetic signature and one of three different sensor arrays. The Program Office selected the MTADS instruments to provide the "gold standard" survey data for this study. Because many commercial surveys use an EM61-Mk2 sensor mounted on a cart, this instrument was selected to provide the "typical commercial" survey data for this study. In addition, the Berkeley UXO Discriminator (BUD), an advanced instrument currently under development, was selected to provide the "next generation" survey data for this study. Finally, a typical mag-and-flag (M&F) survey was done over a 100′ × 100′ grid in the SE1 area. This area is outlined with a small black square in Figure 2.3. The M&F survey was performed so that the other technologies demonstrated in this study could be compared against a more traditional method that does not involve digital geophysical mapping—or subsequent discrimination processing.

2.8.1 The Mag Array Instrument [8]

The magnetometer version of the MTADS instrument, called the "Mag Array" throughout the remainder of this document, was used in the initial magnetometer survey of Camp Sibert. This instrument employs eight Geometrics 822A total field, Cs-vapor magnetometer sensors mounted in a linear array with 25 cm spacing. The distance of the sensors above the ground is also approximately 25 cm. The signals measured by the sensors are sampled at 50 Hz, leading to a down-track sample spacing of approximately 6 cm for the typical 3 m/s survey speed. A single real-time kinematic (RTK) GPS antenna is mounted over the center of the array and tracks the position of the sensors at a 5 Hz sampling rate. A base station receiver placed at a surveyed monument provides differential GPS (DGPS) corrections. Since total field sensors are used, an accurate measurement of the sensors' orientation (i.e., tilt) is not critical for accurate data inversion.

2.8.2 The EM61 Array Instrument [8]

Each sensor mounted on the time-domain EMI version of the MTADS instrument (called the "EM61 Array" for the remainder of this document) is a modification of the standard EM61-Mk2 sensor sold commercially by Geonics, Ltd. While the standard EM61-Mk2 sensor is based on a single 1 m × 0.5 m coil, the modified sensor is based on a 1 m × 1 m coil. Three overlapping 1 m × 1 m coils are mounted on the EM61 Array, as shown in Figure 2.9. The three transmitting coils are synchronized to provide as large a

magnetic moment as possible to maximize their sensitivity. The sensors pulse at 75 Hz but do internal stacking and provide an output at 10 Hz, leading to a down-track sample spacing of 15 cm for the typical 1.5 m/s survey speed. Accurate measurement of the orientation of the sensors is necessary for accurate data inversions, because these are vector sensors. Therefore, three RTK DGPS receivers are used to measure both the position and orientation of the sensors at 5 Hz. A Crossbow VG300 inertial measurement unit (IMU) also outputs the orientation of the sensors at 30 Hz. Figure 2.10 shows a photograph of the EM61 Array collecting data at Camp Sibert.

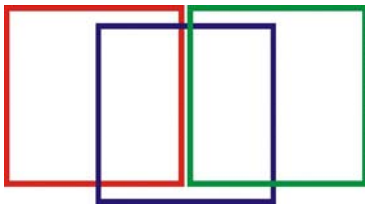


Figure 2.9: Sketch of the three overlapping sensor coils mounted on the EM61 Array instrument.

Figure 2.10: A photograph of the EM61 Array collecting data at Camp Sibert.

The EM61 Array has four sample gates and two receive coils (upper and lower). The instrument may be set up with four time gates from the lower coil or with three time gates from the lower coil and the first time gate from the upper coil. The discrimination demonstrators expressed a preference for the second option, since the height diversity provided by sampling the upper coil can improve EMI-only depth inversions.

2.8.3 The GEM Array Instrument [8]

The frequency-domain EMI version of the MTADS, called the "GEM Array" for the remainder of this document, consists of three Geophex Ltd. GEM-3 sensors, each approximately 1 m in diameter, arranged in a triangular configuration. Figure 2.11 shows

the placement of the sensors, as well as the RTK GPS antennas (labeled MB1, MB2, and MR) and the IMU. The GPS and IMU are identical to those used with the EM61 Array. In the figure, AVR1 and AVR2 are the vectors between the GPS antennas that allow resolution of the sensor array's pitch, roll, and yaw.

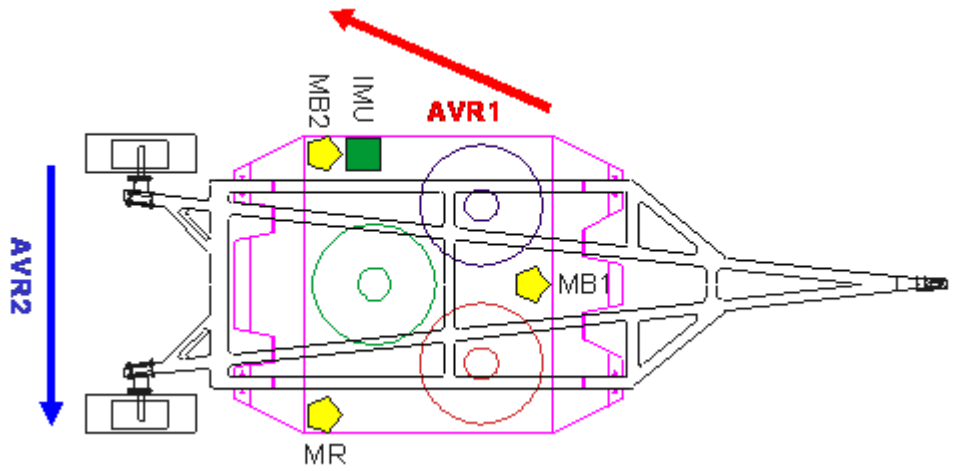


Figure 2.11: Sketch of the three sensor coils and position sensors (MB1, MB2, MR, and IMU) mounted on the GEM Array instrument [8].

As a frequency-domain EMI sensor, the GEM-3 uses bucking coils to null the primary field at the smaller, coaxial receive coil. For that reason, the transmit coils on the GEM Array cannot be fired simultaneously like the three coils in the EM61 Array, as the fields from the other coils in the GEM Array could potentially corrupt the received signals. The transmit coils in the GEM Array have been modified from those of the standard GEM-3 sensor to produce a significantly higher transmit moment, however. The base period for the sensor is 1/30 s. For the three-coil configuration and with settling time added, the effective sampling rate for each sensor coil is 9 Hz, leading to a down-track sample spacing of approximately 15 cm at the typical survey speed of 1.5 m/s, as well as a cross-track spacing of approximately 50 cm. Overlapping the survey lines can be used to reduce the cross-track spacing to approximately 25 cm. Figure 2.12 shows the GEM Array collecting data at Camp Sibert.

Figure 2.12: Photograph of the GEM Array collecting data at Camp Sibert.

2.8.4 The EM61 Cart Instrument [9]

The typical EMI instrument used in commercial surveys will be called the "EM61 Cart" for the remainder of this document. The EM61 Cart consists of a standard EM61-Mk2 sensor mounted on a two-wheel cart. This instrument employs a 1 m × 0.5 m receive coil mounted 30 cm above a second 1 m × 0.5 m coil that transmits as well as receives. As with the EM61 Array survey, the EM61 Cart survey was conducted in differential mode, with three time gates sampled on the lower coil and the first time gate sampled on the upper coil.

The operator of the instrument wears a backpack containing the sensor electronics and battery. The data-acquisition system records data (consisting of the three time gates for the lower coil and a single time gate for the upper coil) at a rate of 16 records per second and can store up to 1 million records. In typical commercial surveys, survey lines are often spaced 1 m apart. Because the purpose of this study was to collect high-quality data that could support discrimination, the operator was instructed to space his survey lines 0.5 m apart. Figure 2.13 shows the EM61 Cart collecting data at Camp Sibert.

Figure 2.13: Photograph of the EM61 Cart collecting data at Camp Sibert.

Parsons, the data-collection team that operated the EM61 Cart, employed a Trimble 5700 RTK DGPS system to track the position of the sensors. Figure 2.13 shows the GPS antenna mounted above the center of the sensor coils, the standard configuration. One disadvantage of the EM61 Cart relative to the EM61 Array is that the use of a single GPS receiver does not allow the orientation of the sensor to be measured; in addition, the tilting of the entire instrument as it is pulled over the ground can lead to errors in the measured position. A second disadvantage of the EM61 Cart relative to the EM61 Array is that the relative position from survey line to survey line is only as accurate as the GPS position measurements. The EM61 Array partly alleviates this problem by providing three cross-track samples with excellent relative position accuracy.

2.8.5 The Berkeley UXO Discriminator Instrument [11]

The BUD is a next-generation instrument whose design and construction were funded by ESTCP and SERDP. BUD was the only instrument that collected data in both survey and cued modes. As the BUD is still under development, its operation is slow. Therefore, the Program Office decided in advance that the BUD would survey the SE1 area only, rather than all three test areas. Figure 2.14 shows the BUD collecting data at Camp Sibert.

Figure 2.14: Photograph of the BUD collecting data at Camp Sibert.

The BUD consists of three orthogonal transmit coils and eight pairs of receive coils that are differenced to provide a gradiometer output. The eight pairs of receive coils are mounted diagonally across the upper and lower horizontal transmit coils to provide gradiometric samples along the three axes. In survey mode, the BUD detects the presence of an anomaly by pulsing only the horizontal transmit coils and assessing the return. If an anomaly is detected, the operators temporarily stop the cart and collect data in cued mode. In cued mode, the BUD pulses all three transmit coils to fully interrogate the source of the anomaly. The BUD samples at a rate of 250 kHz and has 35 sample gates logarithmically spaced from 153 to 1,387 µs. Because the BUD's transmit field is more spatially diverse than the transmit fields of other instruments, high-quality data supporting accurate inversions can be collected from a single BUD position. Therefore, the inversion of BUD data is not as affected by position errors as is the inversion of data collected from other instruments. Furthermore, because the BUD remains temporarily stationary while collecting data, more time is available for data stacking and motion noise is suppressed. This leads to an improved SNR, which in turn leads to more accurate data inversions.

2.8.6 Mag-and-Flag

M&F is a technique historically used in UXO clearance. An operator walks along survey lines with a hand-held magnetometer, sweeping the instrument back and forth

while listening for audio signals indicating that an anomaly is present. When the instrument signals an anomaly, the operator sweeps the instrument to determine the anomaly's exact location and plants a flag to indicate where to dig.

M&F is not used as often now as it was in the past because it relies upon the qualitative judgment of the operator and because there are no digital records that can be used for quality assurance/quality control (QA/QC) to document that the entire survey area was covered, or for documenting what sensitivity level was used for detection. Furthermore, because analog instruments principally depend on signal strength for target selection, small but shallow scrap items tend to provide a large number of false positives.

At Camp Sibert, Parsons conducted an M&F survey over a 100′ × 100′ grid in the SE1 area using a Schonstedt model GA-52Cx magnetic locator. This area is shown as a small black square in Figure 2.3. The operator placed flags at the locations of each detected anomaly. The positions of these flags were later measured and recorded using an RTK DGPS. All M&F locations were included in the list of locations to be excavated.

2.9 ANOMALY DETECTION

One purpose of this study was to automate, as much as possible, all data processing to eliminate the effects of operator judgment. To that end, the Program Office and Advisory Group agreed on the methods to select detection thresholds and then those thresholds were automatically applied to the collected data in order to detect anomalies.

2.9.1 Selecting Detection Thresholds

The purpose of the GPO was to ensure that the digital survey instruments were able to detect munitions to a depth of 11 times the munition diameter using a specific detection threshold. The "11×" rule is an empirically developed guideline created by the COE that specifies the depth to which modern magnetometer and EMI sensors are expected to detect metallic objects [13].

Often, the detection threshold for a digital geophysical instrument is selected as some multiplier of the instrument's noise floor as measured at the site (for example, 1.5 times the noise floor). In a clearance action where the type of munition to be detected is known in advance, however, setting the detection threshold as a function of the instrument's noise floor penalizes more sensitive instruments because more sensitive instruments also detect more false positives. Therefore, an alternative method was used to set the detection thresholds at Camp Sibert.

In this study, the detection thresholds were based on the smallest signal that the instruments could be expected to measure from a 4.2″ mortar, the only munition type expected at Camp Sibert. Given a reasonable model for the instrument's response and knowledge of the shape and material composition of the target, it is straightforward to calculate what signal the instrument could be expected to measure from the target as a function of the target's depth. Figure 2.15 shows a plot of the signal the Mag Array could be expected to measure from a 4.2″ mortar at different mortar depths. The red curve indicates the expected signal measured from the mortar when the mortar is in its most favorable orientation, that is, with its longitudinal axis parallel to earth's magnetic field. The blue curve is the expected signal based on the mortar's least favorable position, that is, with its longitudinal axis perpendicular to earth's magnetic field. The points overlaid on the curves are the signals measured by the Mag Array from 4.2″ mortars buried in a test pit and GPO at Camp Sibert. As expected, most data points lie on or between the two curves. The few outliers are likely due to slight inaccuracies in the mortars' burial depths or minor variations in the shape and material composition of the mortars. Similar curves for the EM61 Array and GEM Array can be found in [8], while a similar curve for the BUD can be found in [11].

Figure 2.15 shows that 12.1 nT is the smallest signal that the Mag Array could be expected to measure from a 4.2″ mortar buried at 11 times the mortar diameter (1.17 m). The Advisory Group agreed on detection thresholds that provide a 50% safety factor; therefore, the Program Office chose a Mag Array detection threshold of 6.1 nT. Note that the root mean square (RMS) noise in the GPO was slightly greater than 2 nT. Thus, a detection threshold calculated as 1.5 times the GPO's noise floor would have been 3 nT. A detection threshold of 3 nT would have significantly increased the number of anomalies detected by the Mag Array, none of which could have been 4.2″ mortars at depths of interest and all of which would have SNR values too low for accurate dipole inversion.

The Program Office selected detection thresholds for the EM61 Array and GEM Array using similar methods. Table 2.2, taken from [8], lists the minimum responses at 11 times the mortar diameter for each Array instrument. That is, these are the signals that each instrument could be expected to measure from 4.2″ mortars at a depth of 11 times the mortar diameter in the least favorable orientation. The table also lists the anomaly-detection thresholds determined for each instrument. The EM61 Array was configured with three time gates from the lower coil and the first gate from the upper coil. The signal sampled for detection was the first time gate in the lower coil (centered at 308 μs),

indicated as "S1" in the table. The GEM Array allows simultaneous transmission of multiple frequencies and then separates the received signal into a component that is in-phase (I) with the transmitted waveform and a component in phase-quadrature (Q), or shifted 90° in phase relative to the transmitted waveform. The Q responses are more immune to geology than the I responses and therefore provide better detection performance. Furthermore, experience has shown that the best detection results are given by the average of the Q responses from the five mid-range frequencies (270, 570, 1,230, 1,610, and 5,430 Hz), indicated as "Qave" in the table.

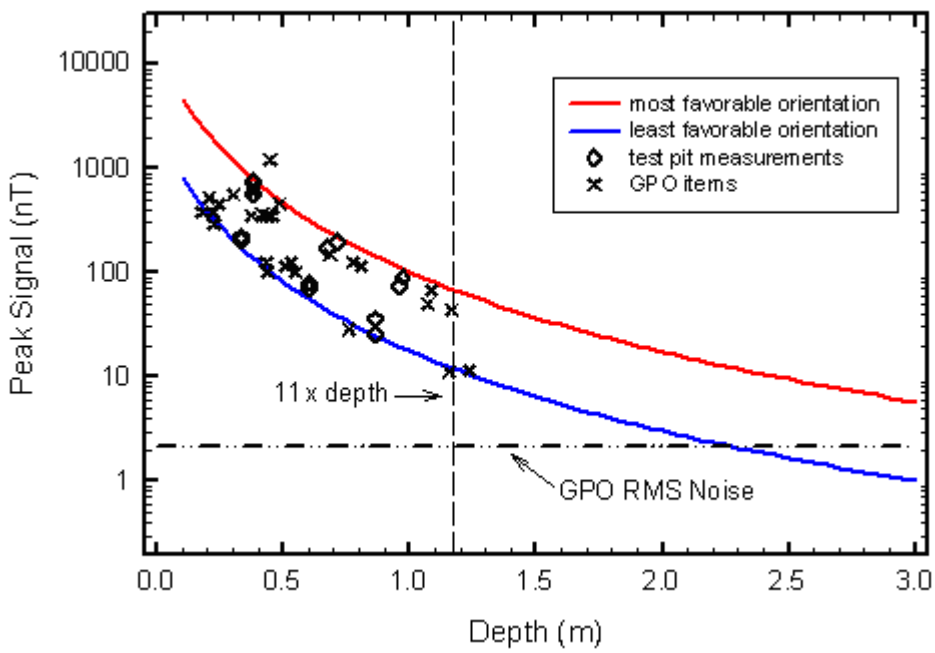


Figure 2.15: Predicted (red and blue lines) and measured responses (◊ and ×) for data collected by the Mag Array instrument from 4.2″ mortars in the Camp Sibert test pit and GPO.

Table 2.2: Minimum expected responses for 4.2″ mortars at depths of 11 times the mortar diameter and selected detection thresholds for the Array instruments [8].

Instrument	Minimum Response at 11×	Anomaly Detection Threshold
Mag Array	12.1 nT	6.1 nT
EM61 Array	51.6 mV, S1	25.8 mV, S1
GEM Array	2.6 ppm, Qave	1.3 ppm, Qave

The measured RMS noise floor in the GPO for the EM61 Array and GEM Array were 6.5 mV and 0.85 ppm, respectively. A detection threshold calculated as 1.5 times the noise floor would have been 9.6 mV for the EM61 Array, much lower than the 25.8 mV detection threshold set based on the smallest expected signal. Therefore an EM61 Array detection threshold selected based on the GPO noise floor would have resulted in a

large number of false positives. In contrast, a GEM Array detection threshold calculated as 1.5 times the noise floor would have been 1.3 ppm, identical to the GEM Array detection threshold set based on the smallest expected signal. This simply illustrates the known large difference in detection sensitivity between the EM61-Mk2 and GEM-3 sensor technologies.

As with the Array instruments, the detection threshold for the BUD was also based on the smallest expected signal. The data-collection team, rather than the Program Office, performed the calculations needed to select the detection threshold [11].

Finally, Parsons selected the detection threshold for the EM61 Cart [9]. Parsons was the commercial contractor hired to survey the test areas with the EM61 Cart instrument. The Program Office instructed Parsons to select the detection threshold for this instrument using its typical process because the purpose of the EM61 Cart survey was to collect "typical, commercial" data.

2.9.2 Applying Detection Thresholds

The Program Office team used the detection thresholds in Table 2.2 to declare anomaly detections in the Mag Array, EM61 Array, and GEM Array data. A computer routine based on quantitative criteria identified the areas where the data collected by an instrument exceeded the instrument's detection threshold. For each area where the collected data exceeded threshold, the data in that area were extracted and inverted using an appropriate inversion routine. For example, UXAnalyze, software funded by ESTCP and made part of Oasis Montaj, was used to invert the EM61 Array data. The MTADS Data Analysis System was used to invert the GEM Array data. For the Mag Array data, using a mixture of the two routines proved to be the most efficient method. The inversion routines returned estimates of the target's position (northing and easting), depth, and size. The routines also returned the "fit coherence," a measure of the ability of the routine to fit the collected data to a dipole model. Parsons used UXAnalyze to invert anomalies in the EM61 Cart data, and the BUD data-collection team used software developed in-house that employed similar methods to invert anomalies in the BUD data.

2.10 CUED LIST GENERATION

In contrast to the survey instruments, which collected data throughout the test areas for both detection and discrimination, some instruments collected high-density data at pre-determined locations for discrimination only. IDA created the "cued list," a list of these pre-determined locations, based on the anomalies detected by the EM61 Array.

EM61 Array anomalies, rather than Mag Array or GEM Array anomalies, were used because two of the three cued instruments were time-domain EMI instruments, like the EM61 Array. The procedure was as follows:

1. IDA labeled every EM61 Array anomaly as "clustered" if its estimated position (easting and northing) was within 2 m of another anomaly's position, otherwise it was labeled as "not clustered." Anomalies labeled as "clustered" were removed from further consideration for the cued list since these anomalies were likely to represent multiple, closely spaced items.
2. IDA labeled every EM61 Array anomaly as "associated with a seed" if its position was within 0.5 m of a seeded item's position; otherwise, it was labeled as "not associated with a seed."
3. IDA chose 200 EM61 Array anomalies for the cued list:
    a. Forty anomalies were randomly chosen from all anomalies that met the following criteria: (1) They were labeled as "not clustered" and (2) they were labeled as "associated with a seed." These anomalies were chosen to ensure that cued data were collected from many munitions. Approximately 1/3 of the 40 anomalies were randomly chosen from each of the three test areas.
    b. Eighty anomalies were randomly chosen from those anomalies that met the following criteria: (1) They were labeled as "not clustered," (2) they were labeled as "not associated with a seed," (3) their fit coherence was greater than or equal to 0.7, and (4) the estimated size of the buried item was greater than or equal to 0.04 m. These anomalies were chosen to ensure that cued data were collected from many large clutter items. Approximately 1/3 of the 80 anomalies were randomly chosen from each of the three test areas.
    c. Eighty anomalies were randomly chosen from those anomalies that met the following criteria: (1) They were labeled as "not clustered," (2) they were labeled as "not associated with a seed," (3) their fit coherence was greater than or equal to 0.7, and (4) the estimated size of the buried item was between 0.02 m and 0.04 m, inclusive. These anomalies were chosen to ensure that cued data were collected from many medium-sized clutter items. Again, approximately 1/3 of the 80 anomalies were randomly chosen from each of the three test areas.

Of the 200 locations on the cued list, 22 were later identified as "clusters" during the generation of the master list, once anomalies detected by other instruments were taken into account. That is, they were likely to represent multiple, closely spaced items. The remaining 178 locations were identified as "single targets." That is, they were likely to

represent one item only. The demonstration teams included *only* the single target locations in their discrimination analysis. While the cluster locations were not analyzed as part of this study, their data are available for future ESTCP and SERDP projects.

2.11 DATA COLLECTION IN CUED MODE

One goal of this study was to assess the added benefit in discrimination performance resulting from the use of high-density data collected at evenly spaced intervals along the ground. To that end, the data-collection team collected data in "cued mode"; that is, they collected high-density data at predetermined locations listed on the cued list. The BUD was one instrument used to collect cued data. This section explains the motivation for selecting the two other cued instruments and briefly explains the sensor technology employed by each of them. More detail can be found in the reports written by the data collection teams [11], [14], [19].

2.11.1 The EM63 Cued Instrument [19]

The EM63, a time-domain EMI sensor manufactured by Geonics, Ltd., is intended to extend the time period and number of time gates over those available with the EM61-Mk2 sensor. The instrument employs a 1 m × 1 m transmit coil and three vertically displaced coaxial 0.5 m × 0.5 m receive coils. Sky Research, Inc., modified a standard EM63 instrument to be more stable and to provide precise position and orientation data. The modified EM63 instrument employs 26 geometrically spaced time gates whose center times range from 180 μs to 25 ms. The modified instrument will be called the "EM63 Cued" for the remainder of this document. Figure 2.16 shows the EM63 Cued collecting data at Camp Sibert.

The EM63 Cued lowers the transmit coil from 40 cm (used in the standard EM61 Cart) to 25 cm above the ground in order to improve sensitivity. A Leica Robotic Total Station tracks a retro-reflector on the cart, while a Crossbow AHRS 400 IMU is used to track the sensor's position and orientation with a high data rate. Cued data are collected in dynamic mode, where the cart is pushed slowly over a 3 m × 3 m tarp, with lines marked every 30 cm in the north-south direction and with three east-west lines, one across the center of the target and two at 50 cm spacing on either side of the center. At the nominal 0.4 m/s survey speed, this provides 10 cm data-point spacing along each line. Data are also collected over the center of the anomaly with the cart stationary but pitched back and forth in the north-south and east-west directions.

Figure 2.16: Photograph of the EM63 Cued collecting data at Camp Sibert. The instrument collected data along a grid drawn on a tarp that was laid over the cued location.

2.11.2 The GEM Cued Instrument [14]

The final instrument used to collect cued data was a standard handheld GEM-3 instrument, a frequency-domain EMI system, called the "GEM Cued" for the remainder of this document. Before the GEM Cued collected data, a 1 m × 1 m template was centered over each target. The template included a grid of points with 25 cm spacing, as well as four other points positioned far enough from the target to be at background, as is shown in Figure 2.17. Holes were drilled in the template at each of these 29 different points. Paint was sprayed through the holes to mark the desired data locations. Finally, the GEM Cued was sequentially placed over each painted mark and held stationary as data were collected—first over the four background marks, then over each of the 25 grid marks, and finally over the first of the four background marks once more to assess sensor drift. Figure 2.18 shows the GEM Cued collecting data at Camp Sibert.

The GEM allows the operator to select the transmit frequencies, with 30 Hz as the lowest available frequency and 96 kHz as the highest. For this study, 10 logarithmically related frequencies from 30 Hz to 30,030 Hz were used: 30, 90, 150, 330, 690, 1,470, 3,090, 6,510, 13,950, and 30,030 Hz. For quality control during the collection, the operator displayed one of the center frequencies (6,510 Hz or 13,950 Hz) and monitored Q-channel variation while holding the instrument stationary. As normal variation should be less than 0.5 ppm, the operators were instructed to reset the sensor if the variation reached several parts per million.

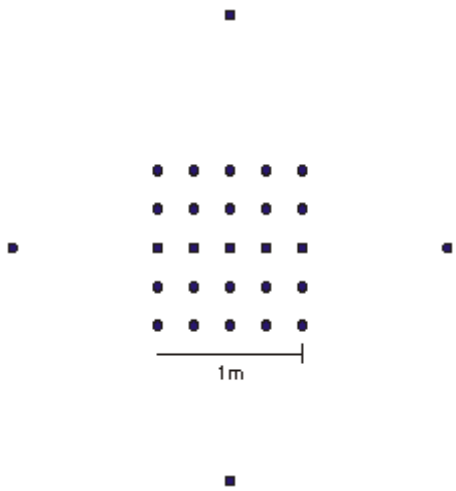


Figure 2.17: Sketch of the template grid for collecting data with the GEM Cued [14].

Figure 2.18: Photograph of the GEM Cued collecting data at Camp Sibert.

2.12 MASTER LIST GENERATION

Different survey instruments (e.g., GEM Array and EM61 Cart) detected anomalies at different apparent locations. IDA generated a "master anomaly list" or master list to reconcile these differences. The master list is a list of locations from which the demonstration teams were instructed to select survey data to input into their

discrimination algorithms. For the following reasons, different procedures were used to generate the master list in each of the three test areas:

- As noted earlier, the BUD is under development, and its operation is slow. Therefore, the Program Office decided in advance that the BUD would survey the SE1 area only, rather than all three test areas. The SE1 master list thus included anomalies detected by the BUD, but the SE2 and SW master lists did not.
- The Program Office decided in advance that the M&F survey would take place over a 100′ × 100′ section of the SE1 area only, rather than a section in each test area. Therefore, the SE1 master list included anomalies detected by the M&F operator, but the SE2 and SW master lists did not.
- The Mag Array collected very noisy data in the SW area due to the high level of magnetic geology. The GEM Array also collected noisy data in the SW area. Therefore, the locations on the SW master list were not formed from anomalies detected by either the Mag Array or GEM Array. In contrast, the locations on the SE1 and SE2 master lists were formed from anomalies detected by all survey instruments that collected data in those areas.

Table 2.3 summarizes the data sources used to generate the locations on the master list.

Table 2.3: Anomaly sources for master list

	Anomalies detected by					
Master List	**BUD**	**M&F**	**MAG Array**	**GEM Array**	**EM61 Array**	**EM61 Cart**
SE1	✓	✓	✓	✓	✓	✓
SE2	—	—	✓	✓	✓	✓
SW	—	—	—	—	✓	✓

2.12.1 Southeast 1 Area

The initial plan was to generate the SE1 master list in a single step using anomalies detected by all six instruments that surveyed this area: the GEM Array, the EM61 Array, the Mag Array, the EM61 Cart, the BUD, and the M&F operator. As the study progressed, however, it became apparent that to remain on schedule, the excavation team would have to begin excavating items from the ground before anomalies were selected in all data collected by all survey instruments. Therefore, the master list in the SE1 area was generated in six distinct steps, as outlined in Figure 2.19 and described below. In each step, an intermediate version of the master list was generated from those anomalies that had already been detected in the collected data. The excavation team began recovering items at the locations listed on the intermediate master list while

anomaly detection continued for other survey instruments. Note that for all the figures in this section, the data are simulated and for the purpose of illustration only.

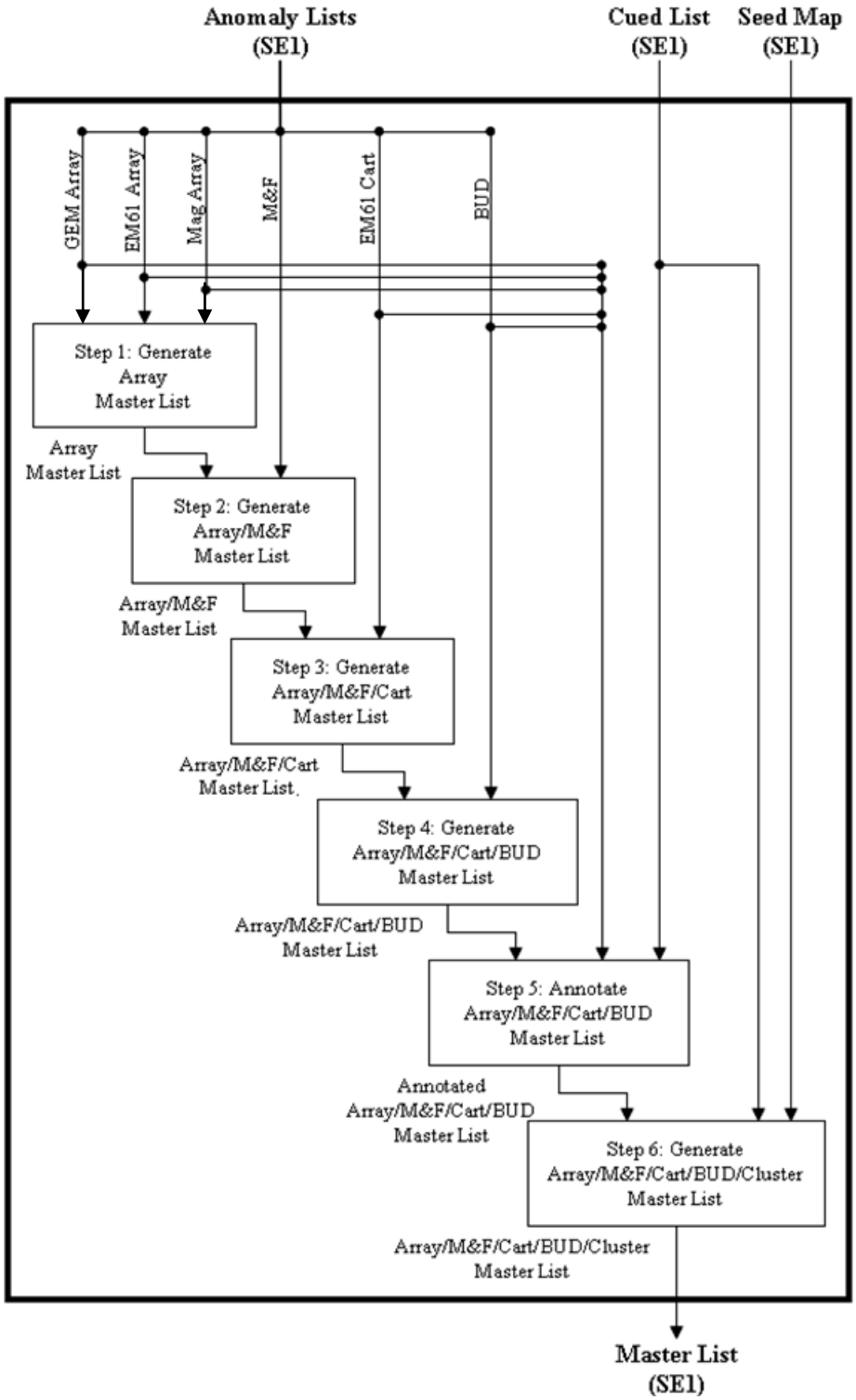


Figure 2.19: Generating the master list consisted of six steps in the Southeast 1 area.

Step 1: Generate Array Master List. IDA combined the GEM Array, EM61 Array, and Mag Array anomaly lists to generate the Array master list (See Figures 2.20–2.27) as follows:

a. IDA formed groups of individual Array anomalies. Figure 2.20 shows a cartoon mapping of array anomalies. Array anomalies formed a group if they were within 0.6 m of each other. The small black circles in Figure 2.21 represent 0.3 m radius halos centered on each anomaly. Anomalies formed a group if these halos touched or intersected. (Note that if Anomaly A was within 0.6 m of anomaly B, and anomaly B was within 0.6 m of Anomaly C, then all three anomalies belonged to the same group, even if Anomaly A was not within 0.6 m of Anomaly C. In this way, some groups consisted of chains of anomalies, such as Group 2 in Figure 2.21.)

b. IDA calculated the centroid location of each group. The easting coordinate of the centroid was calculated as the average over the easting coordinates of all anomalies belonging to the group. The northing coordinate of the centroid was calculated similarly. The black stars in Figure 2.22 represent group centroids.

c. IDA labeled every centroid as "clustered" if it was within 2 m of another centroid; otherwise, it was labeled as "not clustered." The large black circles in Figure 2.23 represent 1 m radius circular halos centered on each centroid. In Figure 2.24, centroids labeled as "clustered" have halos that touch or intersect. Two centroids are labeled as "clustered," and three centroids are labeled as "not clustered." "Clustered" centroids must be identified because they are likely to represent multiple, closely spaced items.

d. The Program Office visually analyzed the Array data to relabel a subset of centroids as "clustered" or "not clustered." This subset of centroids consisted of those that were (1) labeled as "not clustered" based on the "2 m" quantitative criterion in substep c and (2) composed of more than one anomaly detected by the same instrument. Centroids in this subset were relabeled as "clustered" if the Program Office believed that they were likely to represent multiple, closely spaced items based on visual analysis of the Array data. The large dashed black circles in Figure 2.25 represent 1 m radius circular halos centered on each of the two centroids that were analyzed visually. In this example, one of the two centroids was relabeled as "clustered" based on visual analysis, as is shown in Figure 2.26.

e. Finally, IDA included all centroids labeled as "not clustered" in the Array master list. The black stars in Figure 2.27 represent these centroids. Those centroids labeled as "clustered," either by the "2 m" quantitative criterion in substep c or by visual analysis in substep d, were not included in the

Array master list because current methods cannot accurately invert overlapping signatures from multiple closely spaced items [1].

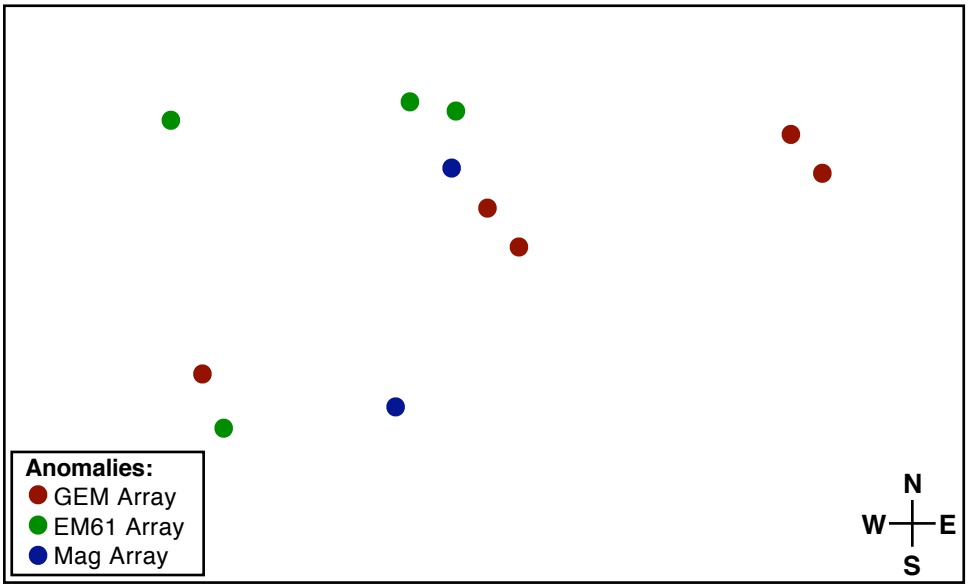


Figure 2.20: A cartoon mapping of the Array anomaly lists. Individual GEM Array, EM61 Array, and Mag Array anomalies are shown as red, blue, and green dots, respectively.

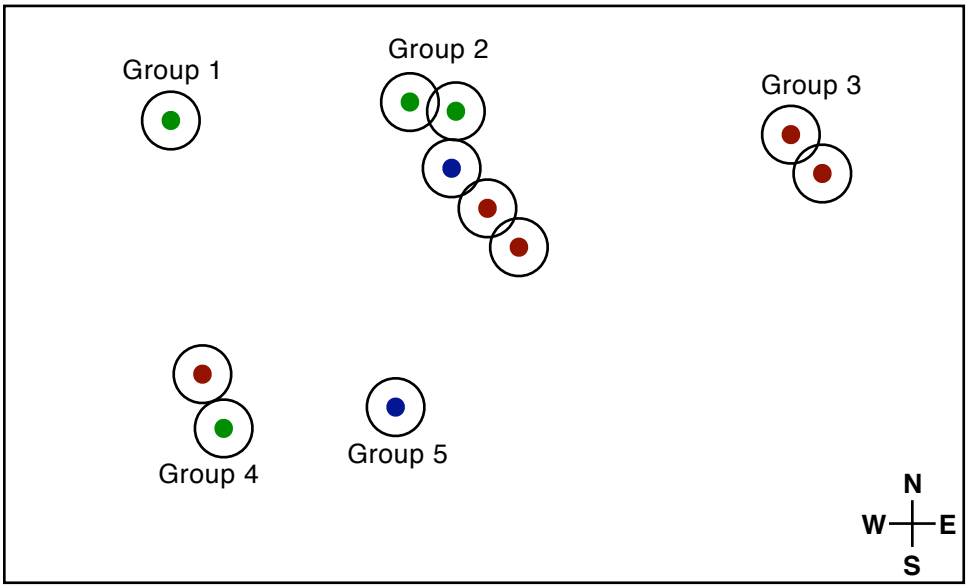


Figure 2.21: Step 1.a. Black circles represent 0.3 m radius halos around each Array anomaly. Anomalies form a group if they are within 0.6 m of each other (i.e., if their halos touch or intersect).

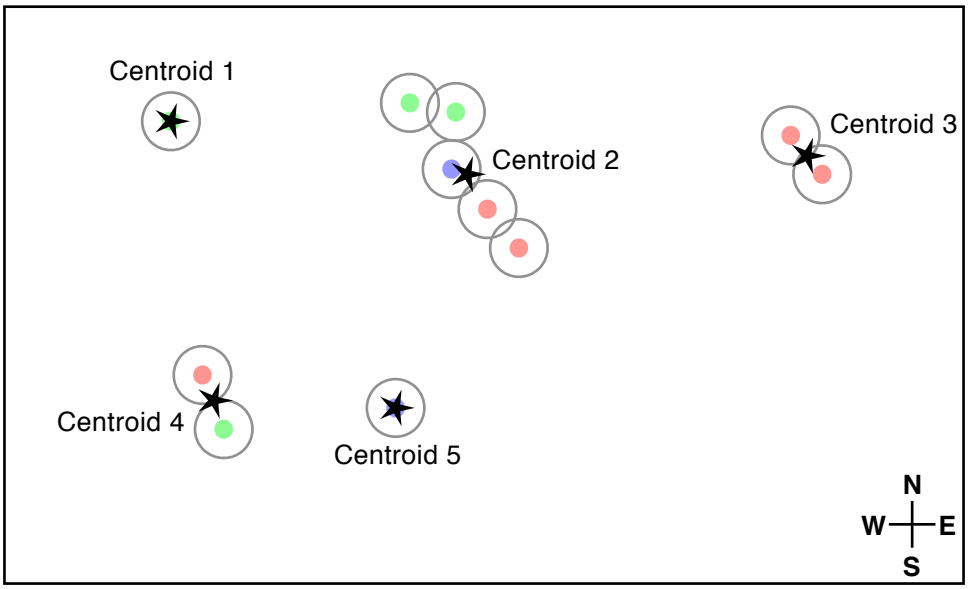


Figure 2.22: Step 1.b. The centroid of each group (black stars) is calculated over all anomalies belonging to the group.

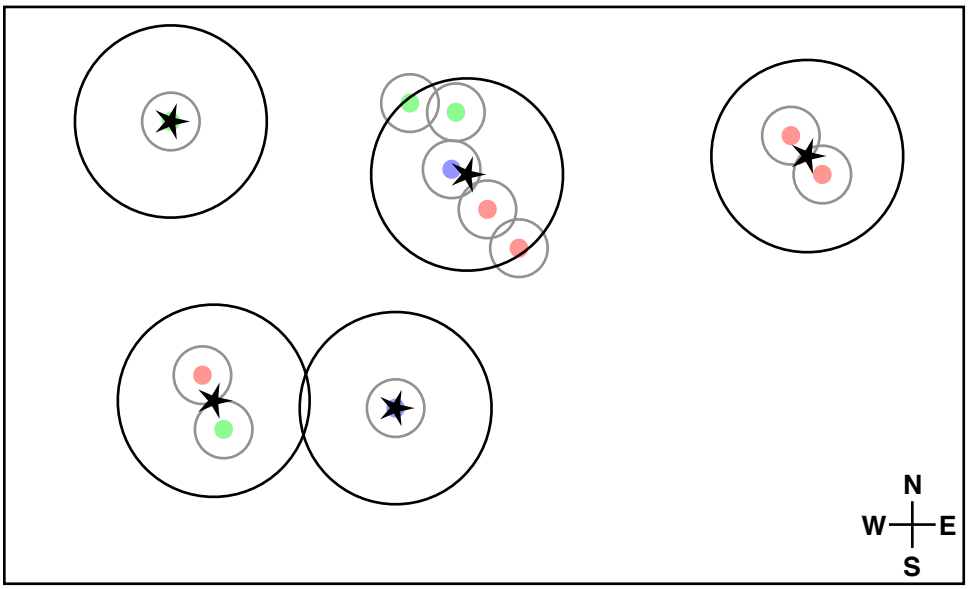


Figure 2.23: Step 1.c. Large black circles represent 1 m radius halos around each group centroid. Centroids must be labeled as "clustered" if they are within 2 m of another centroid (i.e., their halos touch or intersect); otherwise as "not clustered."

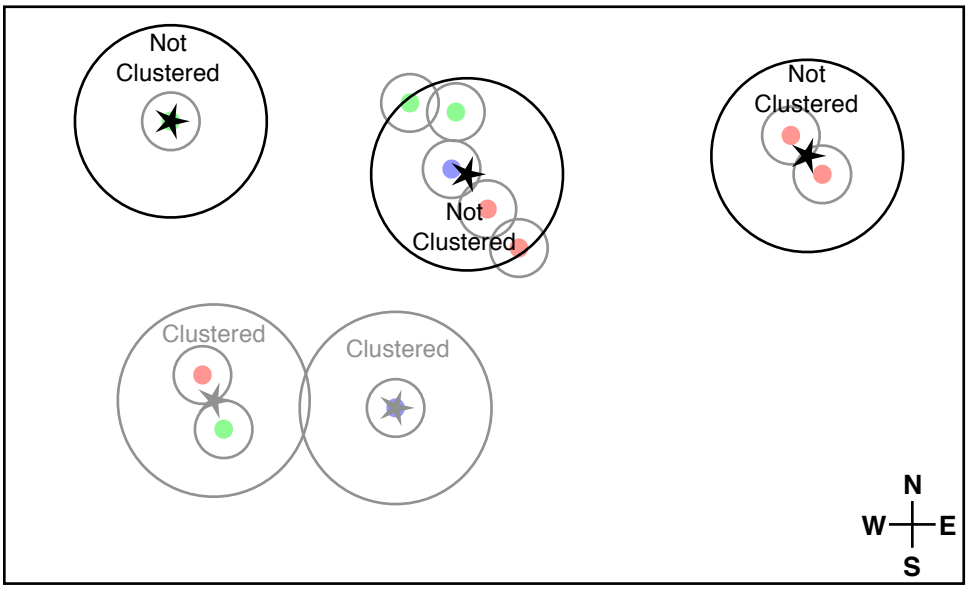


Figure 2.24: Step 1.c. (cont.) Two centroids have been labeled as "clustered" because they are within 2 m of each other.

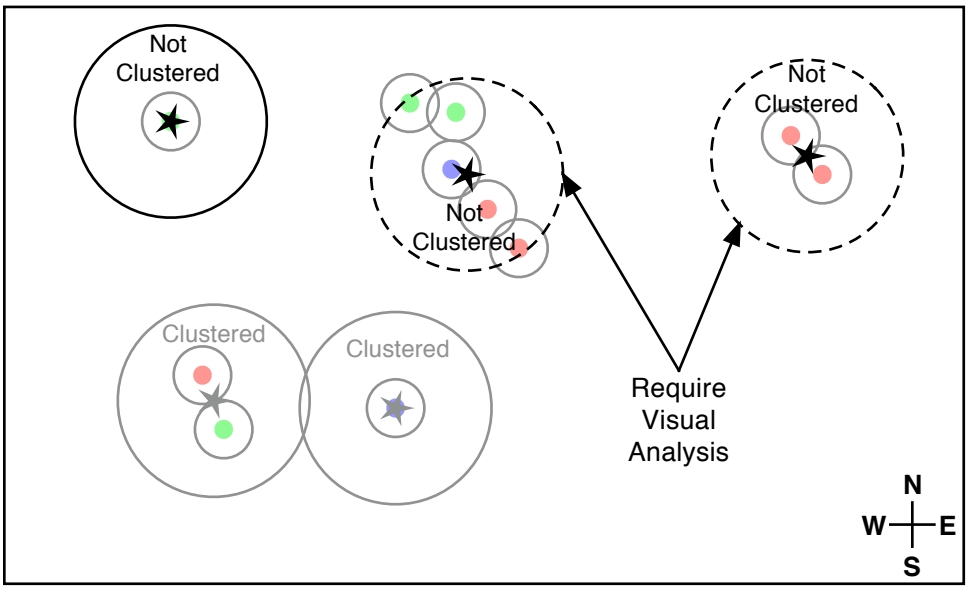


Figure 2.25: Step 1.d. Large dashed black circles represent 1 m radius halos around each centroid that was (1) labeled as "not clustered" using the "2 m" quantitative criterion in the previous substep and (2) composed of more than one anomaly detected by the same instrument (more than one dot of the same color). These centroids must be relabeled as "clustered" or "not clustered" based on visual analysis of the surrounding Array data.

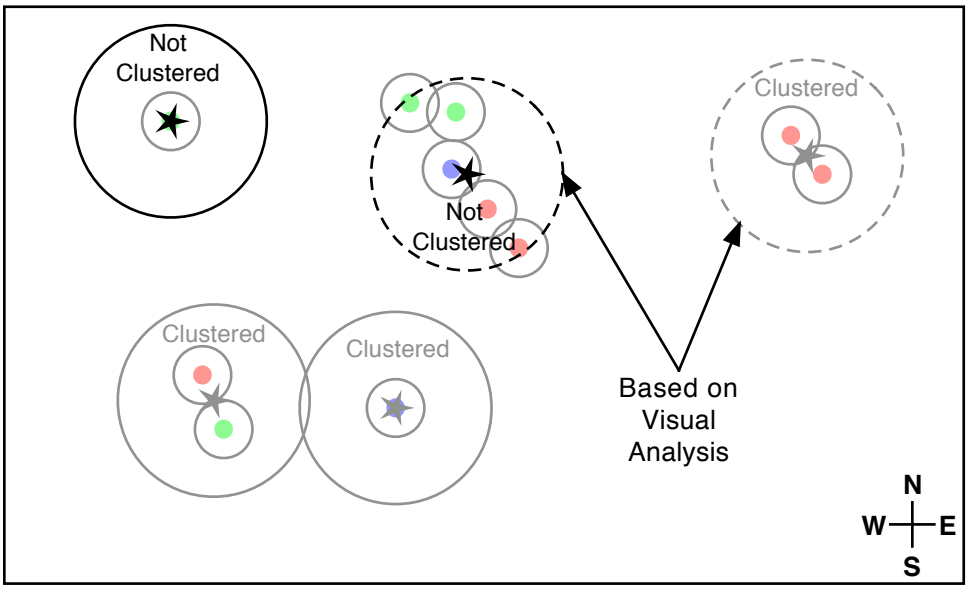


Figure 2.26: Step 1.d. (cont.) One centroid has been relabeled as "clustered" based on visual analysis of the collected data.

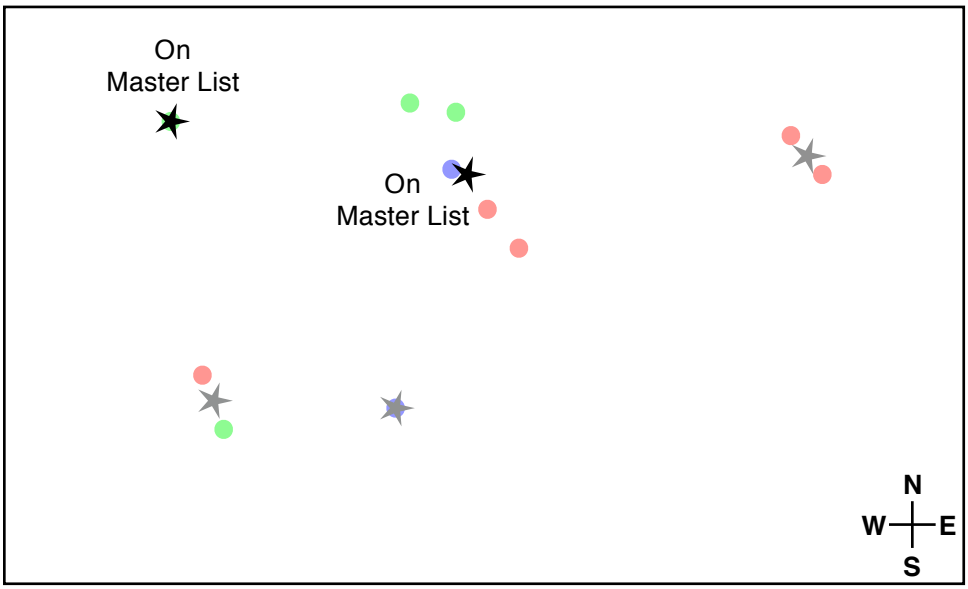


Figure 2.27: Step 1.e. All centroids labeled as "not clustered" (black stars) are included on the Array master list.

Step 2: Generate Array/M&F Master List. In the second step of creating the SE1 master list, IDA combined the M&F anomaly list with the Array master list to create the Array/M&F master list. (Note that the M&F survey was performed over only a 100′ × 100′ section of the SE1 area, and therefore an M&F anomaly list exists for the SE1 area only.)

a. The data collection team determined which M&F anomalies corresponded with anomalies detected by the Array instruments.

b. IDA appended to the Array master list those M&F anomalies that did *not* correspond with any Array anomaly.

Step 3: Generate Array/M&F/Cart Master List. IDA combined the EM61 Cart anomaly list with the Array/M&F master list to form the Array/M&F/Cart master list, as shown in Figures 2.28–2.36:

a. IDA formed new groups out of EM61 Cart anomalies. Figure 2.28 shows a cartoon mapping of the EM61 Cart anomaly list. Seven anomalies are shown. EM61 Cart anomalies formed a new group if they were within 0.6 m of each other. (Again, some groups were formed from chains of anomalies.) The purple circles in Figure 2.29 represent 0.3 m radius circular halos centered on every EM61 Cart anomaly. Anomalies that belong to the same group have halos that touch or intersect. Five groups are shown.

b. IDA calculated the centroid location of each new group. The purple stars in Figure 2.30 represent new group centroids.

c. IDA labeled new centroids as "clustered" if they were within 2 m of (1) another new centroid or (2) an original centroid formed during generation of the Array master list in Step 1. Otherwise, new centroids were labeled as "not clustered." Figure 2.31 shows a cartoon mapping of the new and original centroids. Purple stars represent new centroids, gray stars represent original centroids labeled as "clustered" in Step 1, and black stars represent original centroids labeled as "not clustered" in Step 1. In Figure 2.32, large purple circles represent 1 m radius circular halos centered on every new centroid. The large black and gray circles represent similar halos centered on every original centroid. New centroids within 2 m of another centroid (either new or original) have halos that touch or intersect other halos. In Figure 2.33, two new centroids are labeled as "clustered," and three new centroids are labeled as "not clustered."

   To avoid confusing the excavation team, original centroids were not relabeled in this step. At this point in the study, the excavation team was already excavating items from the ground at locations in the Array master list—that is, the original centroids labeled as "not clustered."

d. The Program Office visually analyzed the EM61 Array data to relabel a subset of new centroids as "clustered" or "not clustered." (The EM61 *Cart* data were not analyzed visually because the Program Office was not in direct possession of this data.) This subset of new centroids consisted of those that were (1) labeled as "not clustered" based on the 2 m quantitative criterion of the previous substep and (2) composed of more than one EM61 Cart anomaly. New centroids in this subset were labeled as "clustered" if the Program Office believed that they represented multiple

closely spaced items based on visual analysis of the surrounding EM61 Array data. In Figure 2.34, the large dashed purple circle represents 1 m radius circular halos centered on the two new centroids that were analyzed visually. In this example, one of the two centroids was "clustered" based on visual analysis, as shown in Figure 2.35.

e. Finally, IDA included on the Array/M&F/Cart master list all centroids (both new and original) labeled as "not clustered," as well as all M&F locations identified in Step 2. Since the labels of the original centroids had not changed, all locations on the Array master list (i.e., all original centroids labeled as "not clustered" in Step 1) were also included in the Array/M&F/Cart master list. These locations are shown as black stars in Figure 2.36. The Array/M&F/Cart master list also included the new centroids created from EM61 Cart anomalies and labeled as "not clustered." These locations are shown as purple stars in Figure 2.36. In contrast, all centroids (both new and original) that were labeled as "clustered" were not included in the Array/M&F/Cart master list. These locations are shown as either pink or gray stars in Figure 2.32.

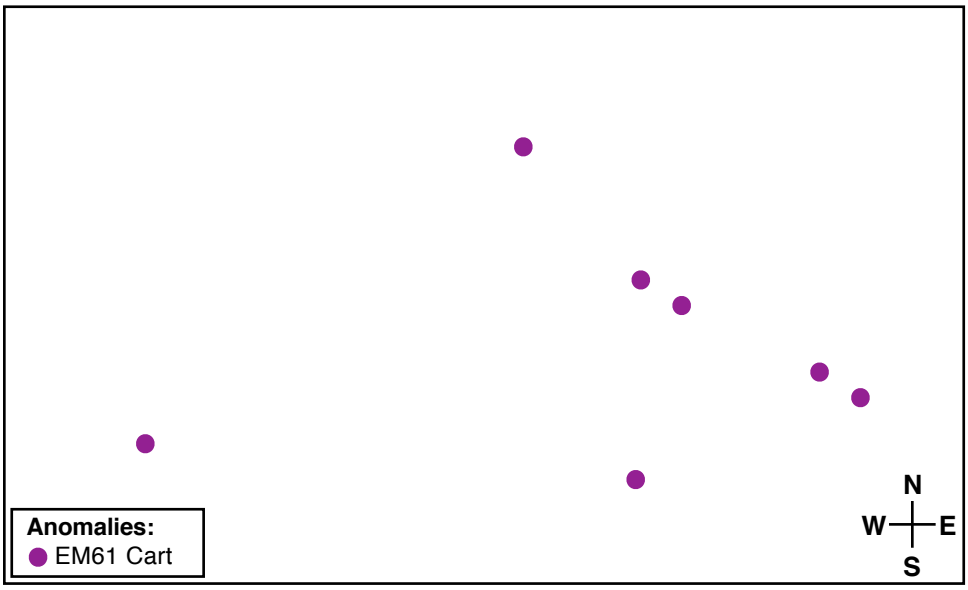


Figure 2.28: A cartoon mapping of the EM61 Cart anomaly list. Individual EM61 Cart anomalies are shown as purple dots.

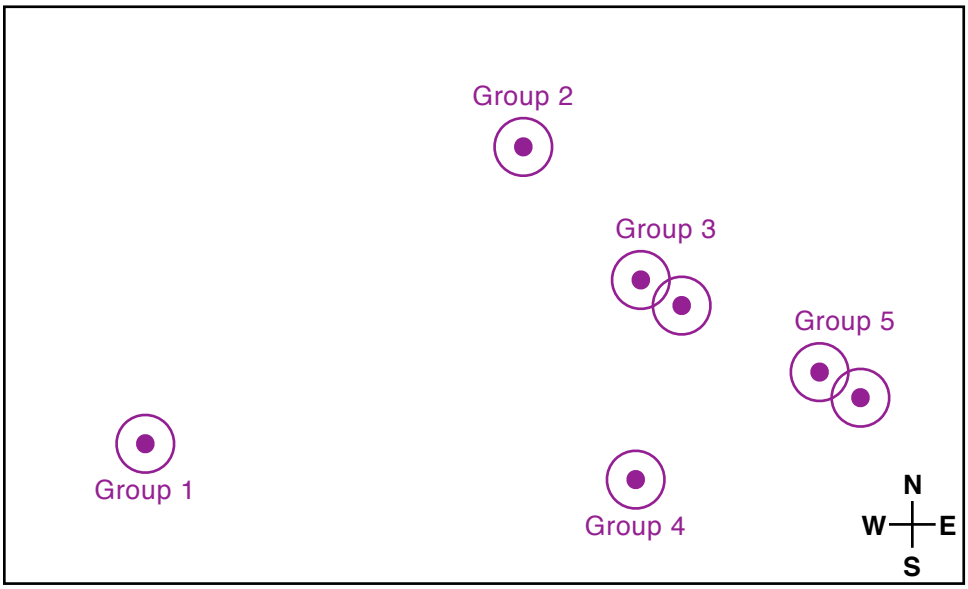


Figure 2.29: Step 3.a. Purple circles represent 0.3 m radius halos around each EM61 Cart anomaly. EM61 Cart anomalies form a new group if they are within 0.6 m of each other (i.e., their halos touch or intersect).

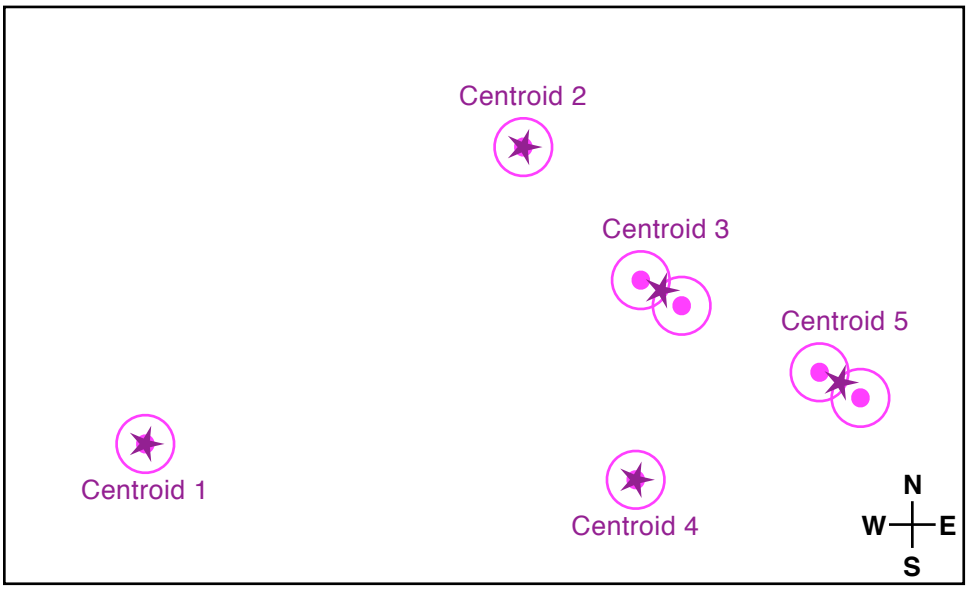


Figure 2.30: Step 3.b. The centroid of each new group (purple stars) is calculated over all EM61 Cart anomalies belonging to the group.

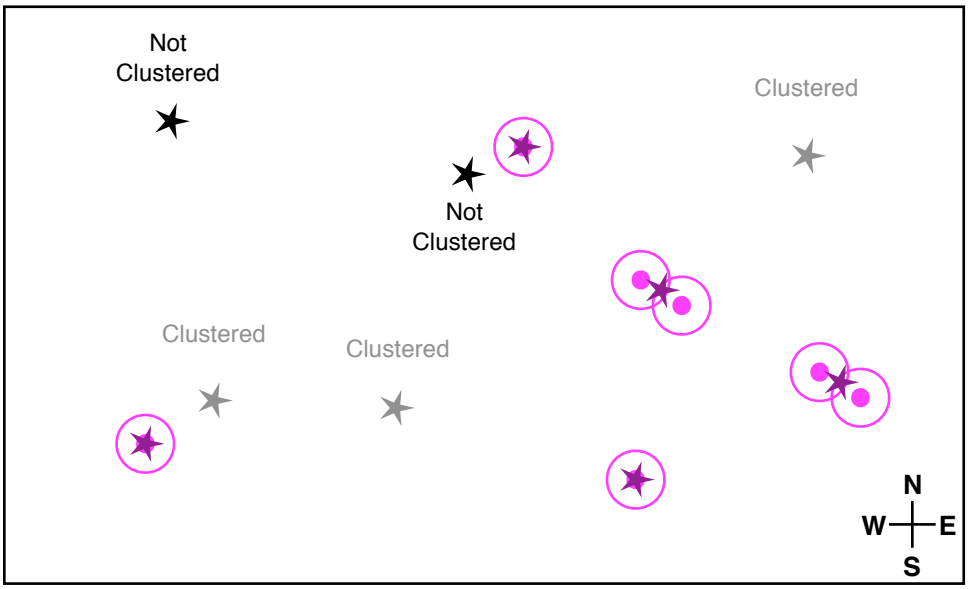


Figure 2.31: Step 3.c. Gray and black stars represent original group centroids labeled as "clustered" and "not clustered," respectively, during generation of the Array master list in Step 1.

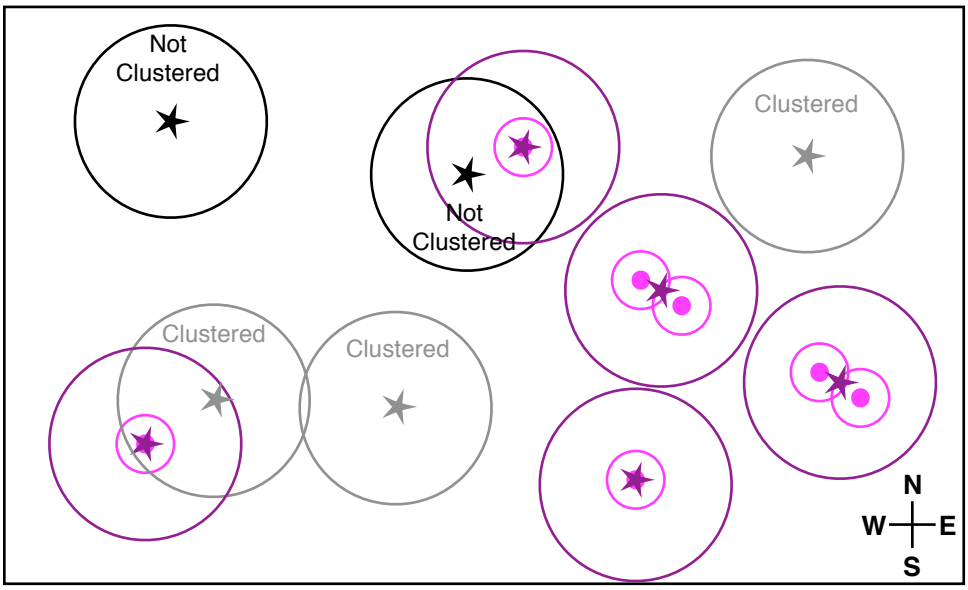


Figure 2.32: Step 3.c. (cont.) Large purple circles represent 1 m radius halos around each new centroid. Large black and gray circles represent similar halos around each original centroid calculated during generation of the Array master list in step 1. New centroids must be labeled as "clustered" if they are within 2 m of another new centroid or of an original centroid (i.e., their halos touch or intersect other halos); otherwise labeled as "not clustered." Original centroids are not relabeled.

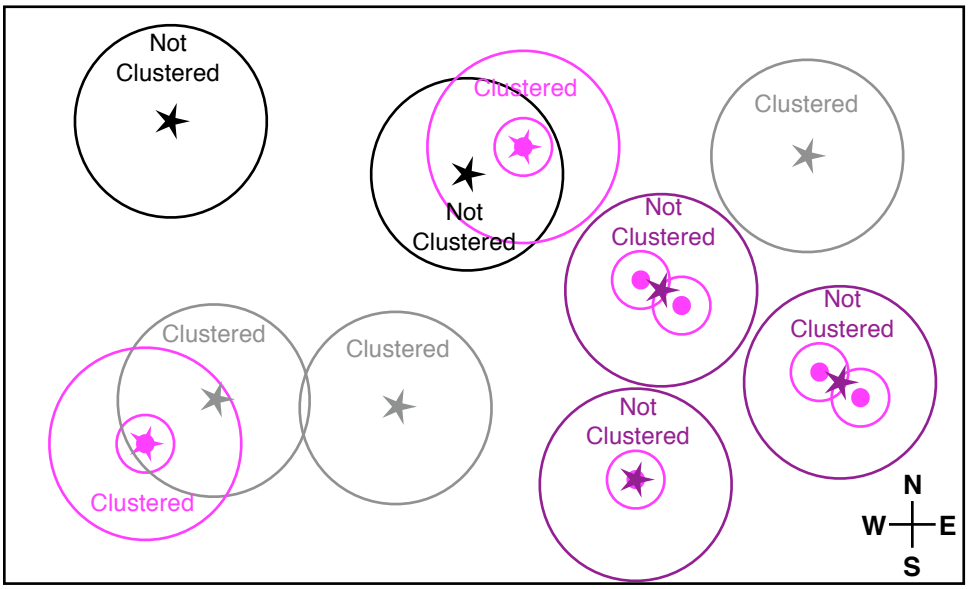


Figure 2.33. Step 3.c (cont.) Two new centroids have been labeled as "clustered" because they are within 2 m of another centroid.

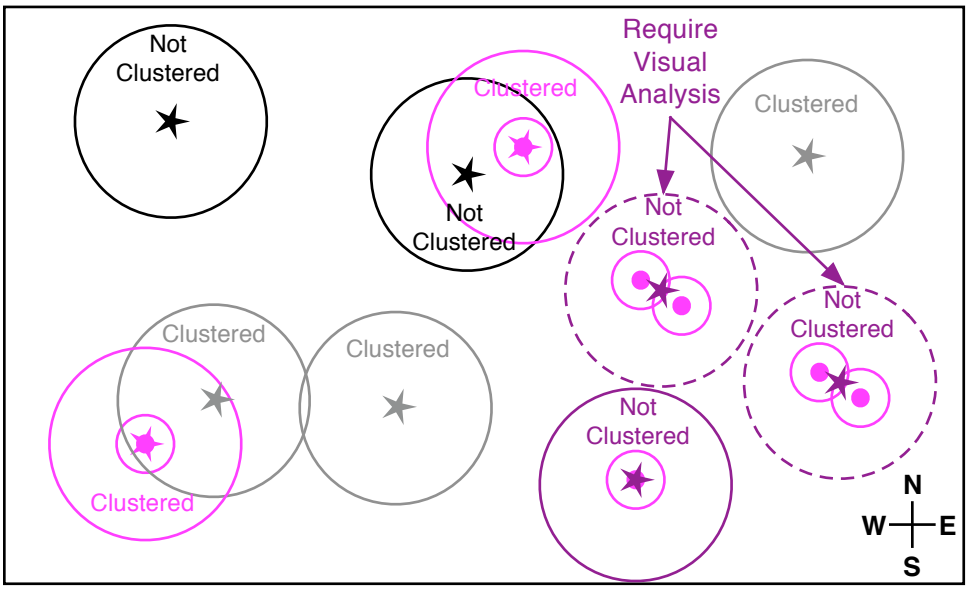


Figure 2.34: Step 3.d. Large dashed purple circles represent 1 m radius halos around each new centroid that was (1) labeled as "not clustered" using the 2 m quantitative criterion in the previous substep and (2) composed of more than one EM61 Cart anomaly (more than one purple dot). These centroids must be relabeled as "clustered" or "not clustered" based on visual analysis of the surrounding EM61 Array data.

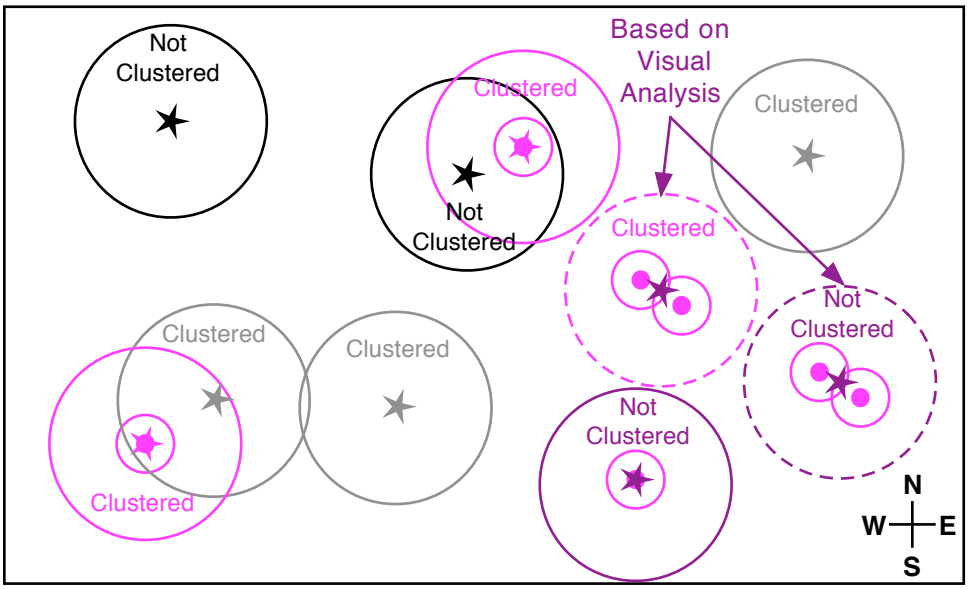


Figure 2.35: Step 3.d (cont.) One centroid has been relabeled as "clustered" based on visual analysis of collected data.

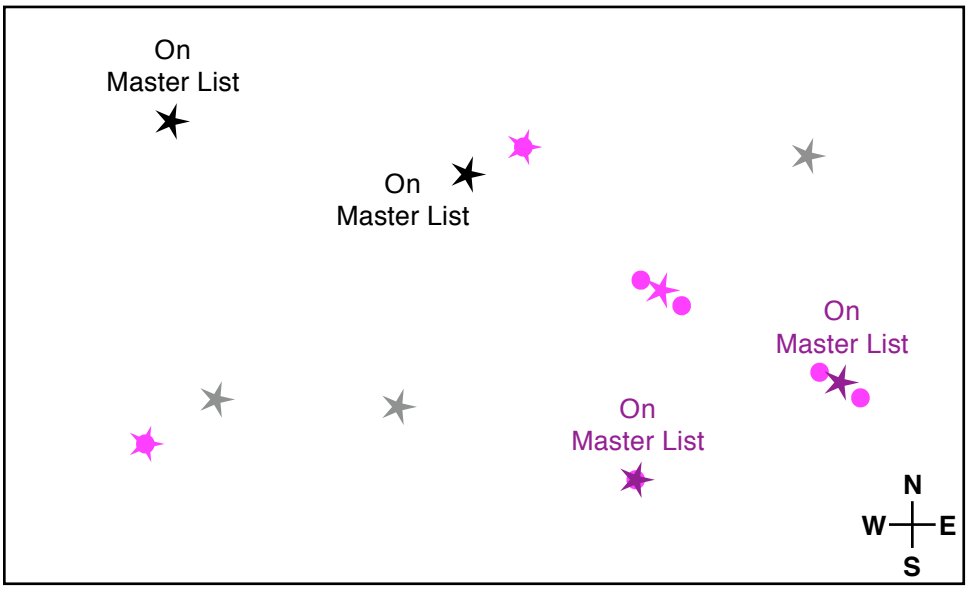


Figure 2.36: Step 3.e. All centroids labeled as "not clustered" (black and purples stars) are included on the Array/M&F/Cart master list.

Step 4: Generate Array/M&F/Cart/BUD Master List. In the fourth step of generating the SE1 master list, IDA combined the BUD anomaly list with the Array/M&F/Cart master list to form the Array/M&F/Cart/BUD master list. This step used a process similar to that used in Step 3 to generate the Array/M&F/Cart master list (compare Figures 2.28–2.36). (Note that because the BUD instrument surveyed only the SE1 area, a BUD anomaly list exists for that area only.) The following substeps were completed:

a. IDA formed new groups out of individual BUD anomalies using the 0.6 m separation criterion.

b. IDA calculated the centroid of each new group.

c. IDA labeled each new centroid as "clustered" or "not clustered" based on the 2 m separation criterion. That is, new centroids were labeled as "clustered" if they were within 2 m of either (1) another new centroid or (2) an original centroid created during generation of the Array or Array/M&F/Cart master lists. Otherwise, new centroids were labeled as "not clustered."

d. The Program Office relabeled a subset of the new centroids based on visual analysis of the EM61 Array data, as the Program Office was not in possession of the BUD data.

e. IDA included on the Array/M&F/Cart/BUD master list all centroids (both new and original) labeled as "not clustered," as well as all M&F locations identified in Step 2. The initial plan was that this would be the final substep in generating the Array/M&F/Cart/BUD master list. As the study progressed, however, it became apparent that a sixth substep was needed.

f. In the final substep, IDA appended additional locations to the Array/M&F/Cart/BUD master list. These locations were items on the cued list that corresponded with BUD anomalies. (The data-collection team determined which items on the cued list corresponded with BUD anomalies.) This last substep was necessary because, due to mis-communication, the data-collection team did not initially include on the BUD anomaly list those BUD anomalies that occurred in the vicinity of the locations on the cued list.

Step 5: Generate Annotated Array/M&F/Cart/BUD Master List. IDA annotated the Array/M&F/Cart/BUD master list to note which locations on the list were associated with the digital survey instruments and the locations on the cued list:

a. IDA associated a location on the Array/M&F/Cart/BUD list with the GEM Array instrument if the location was within 0.6 m of a GEM Array anomaly. List locations were also associated with the EM61 Mag Array, EM61 Cart, and BUD instruments in a similar manner.

b. IDA associated a location on the Array/M&F/Cart/BUD master list with a location on the cued list if *either* at least one EM61 Array anomaly used to generate the master list location had also been used to generate the cued list location *or* if the data-collection team determined that the cued location corresponded with a BUD anomaly.

Of the 200 locations on the cued list, 178 were associated with a location on the Array/M&F/Cart/BUD master list. The remaining 22 cued locations (9 in the SE1 area and 13 in the SE2 area) were not associated with a location on the master lists and are dealt with further in Step 6.a.

Step 6: Generate Array/M&F/Cart/BUD/Cluster Master List. IDA appended to the Array/M&F/Cart/BUD master list those locations likely to represent multiple, closely spaced items (i.e., "clusters") to form the Array/M&F/Cart/BUD/Cluster master list:

a. IDA analyzed in more detail the nine locations on the cued list in the SE1 area that had not been associated with a location on the annotated Array/M&F/Cart/BUD master list in Step 5. These locations corresponded in space with groups of Array anomalies that had been labeled as "clustered" in Step 1; they had not been included on the Array master list. These locations were now appended to the annotated Array/M&F/Cart/BUD master list.

   Note that the master list was not annotated to note any associations between the appended "clustered" cued locations and the survey instruments because these associations were meant for locations likely to represent a single item only (i.e., "single targets"). That is, locations on the master list associated with a survey instrument are, by definition, single target locations.

b. IDA compared the intended locations of the seeded items to the locations on the Array/M&F/Cart/BUD master list. Of the 151 seeded items, only 149 were within 0.6 m of a location on the Array/M&F/Cart/BUD master list. IDA analyzed in more detail the remaining two seeded items, both of which were in the SE1 area. As in the previous substep, results showed that the locations of these two seeded items corresponded to Array anomalies that had been labeled as "clustered" in Step 1; therefore, they had not been included on the Array master list. These two locations were now appended to the annotated Array/M&F/Cart/BUD master list.

   Again, note that the master list was not annotated to note associations between the two appended "clustered" seed locations and the survey instruments because these associations were meant for single target locations only.

c. Last, the Program Office visually analyzed maps of the Array anomaly lists superimposed on the collected Array data and identified locations that were likely to represent multiple, closely spaced items (i.e., "clusters"). These locations were appended to the annotated Array/M&F/Cart/BUD master list to form the final master list for the SE1 area. The appended

locations were not associated with survey instruments because these associations were meant for single targets only.

2.12.2 Southeast 2 Area

As with the SE1 area, the initial plan was to generate the SE2 master list in a single step using anomalies detected by all four instruments that surveyed this area: the GEM Array, the EM61 Array, the Mag Array, and the EM61 Cart. (BUD and M&F surveys were not carried out in the SE2 area.) Due to scheduling pressures, however, the SE2 master list was generated in four distinct steps, as shown in Figure 2.37, and described below.

Step 1: Generate Array Master List. IDA combined the GEM Array, EM61 Array, and Mag Array anomaly lists to produce the Array master list, as shown in Figures 2.20–2.27. This step required the same substeps as in the SE1 area.

Step 2: Generate Array/Cart Master List. IDA combined the EM61 Cart anomaly list with the Array master list to form the Array/Cart master list, as shown in Figures 2.28–2.36. This step required the same substeps as in the SE1 area.

Step 3: Annotate Array/Cart Master List. IDA annotated the Array/Cart master list to note which locations on the list were associated with the digital survey instruments and the cued list. This step was similar to the corresponding step for the SE1 area, except that no locations on the SE2 list were associated with the BUD instrument, which did not survey the SE2 area.

Step 4: Generate Array/Cart/Cluster Master List. IDA appended to the Array/Cart master list those locations likely to represent multiple, closely spaced items (i.e., “clusters) to form the Array/Cart/Cluster master list. This step required the same substeps as in the SE1 area.

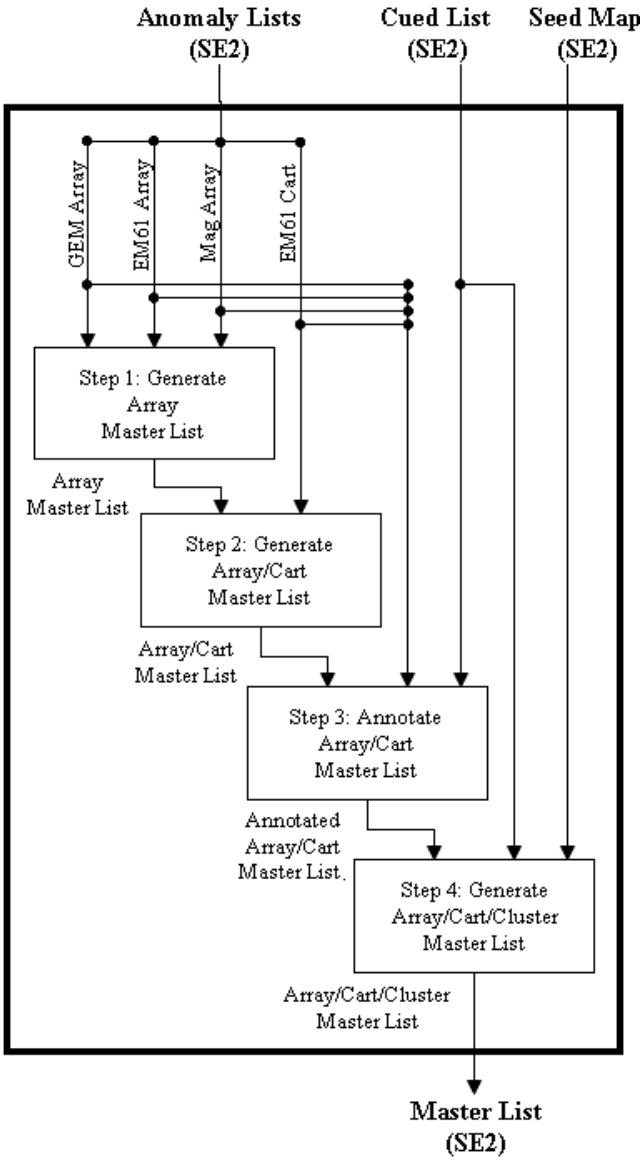


Figure 2.37: Generating the master list consisted of four steps in the SE2 area.

2.12.3 Southwest Area

As with the SE1 and SE2 areas, the initial plan was to generate the SW master list in one step using anomalies detected by all four instruments that surveyed this area: the GEM Array, the EM61 Array, the Mag Array, and the EM61 Cart. Once again, due to scheduling pressures, it was decided that the SW master list would be generated in four distinct steps, similar to what was done in the SE2 area. Figure 2.38 shows the steps used in this process, which are described below.

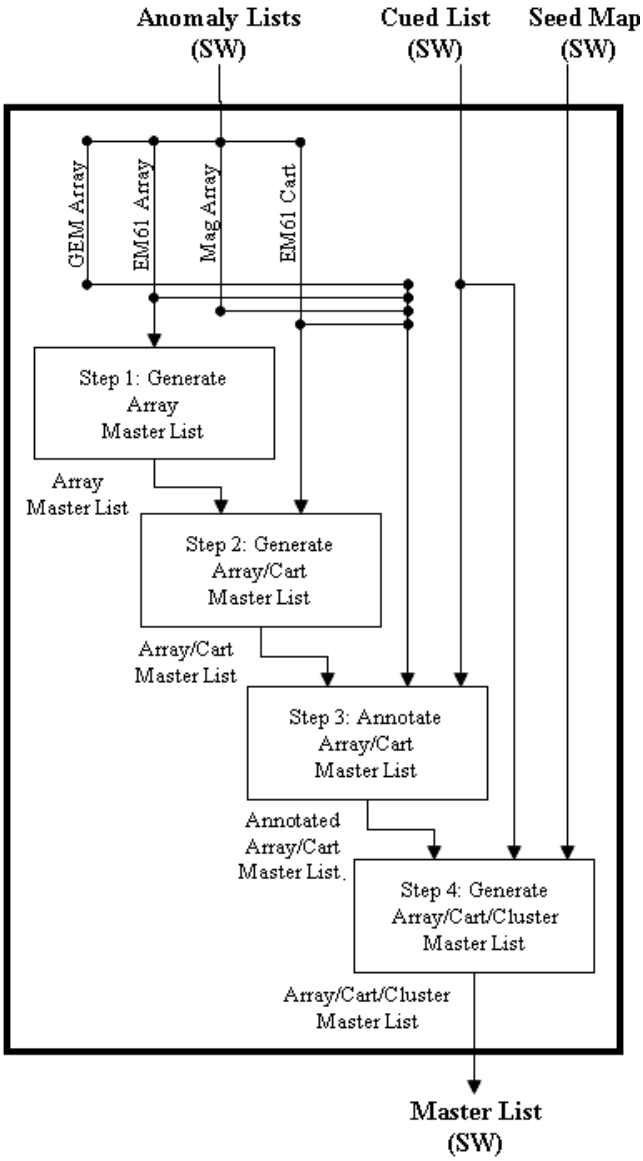


Figure 2.38: Generating the master list consisted of four steps in the SW area.

Step 1: Generate Array Master List. IDA generated the Array master list from anomalies detected by the EM61 Array only. Anomalies detected in the GEM Array and Mag Array data were not considered because many were likely caused by noise. The substeps used to form the Array master list were like those used in the SE1 and SE2 areas and described in Figures 2.20–2.27.

Step 2: Generate Array/Cart Master List. IDA combined the EM61 Cart anomaly list with the Array master list to form the Array/Cart master list, as is shown in Figures 2.28–2.36. This step required the same substeps as in the SE1 and SE2 areas.

Step 3: Annotate Array Cart/Cluster Master List. IDA annotated the Array/Cart master list to note which locations on the list were associated with

each instrument that surveyed the SW area: the GEM Array, EM61 Array, Mag Array, and EM61 Cart. This step required the same processing as in the SE2 area.

Step 4: Generate Array/Cart/Cluster Master List. IDA appended to the Array/Cart master list those locations likely to represent multiple, closely spaced items (i.e., "clusters") to form the Array/Cart/Cluster master list. This step required the same substeps as in the SE1 and SE2 areas.

For the remainder of this document, "master list" refers to the union of the final master lists generated for the SE1, SE2, and SW areas. Each location on the master list was assigned a unique Target ID number, ranging from 1 to 1,430. The master list consisted of 1,389 single target locations and 41 cluster locations. Each of the 1,389 single target locations was associated with 1 or more survey instruments (1,359 were associated with at least 1 digitized survey instrument, and 30 were associated with the M&F process only.) Furthermore, 178 of the 1,389 single target locations were associated with a location on the cued list and 149 were associated with a seeded item (the remaining 22 cued locations and the remaining 2 seeded items were included in the 41 cluster locations). The demonstration teams included *only* the 1,389 single target locations in their discrimination analyses. While the 41 cluster locations were not analyzed as part of this study, their data are available for future ESTCP and SERDP projects.

2.13 SELECTION OF SURVEY DATA AT LOCATIONS ON THE MASTER LIST

For each single target location on the cued list, the demonstration teams input all data collected at the location into their data-inversion routines, the first step in discriminating between a cued location highly likely to contain clutter and one likely to contain munitions. Discriminating *survey* data required an interim step, however. For each single target location on the master list, the demonstration teams first had to select a small region of survey data surrounding the location before inputting the selected data into their inversion routines. The demonstration teams selected the survey data based on subjective, visual analysis of collected data. This step represented one of the few subjective steps in the discrimination study.

2.14 DISCRIMINATION

For every combination of data-collection instrument and discrimination algorithm, demonstrators processed the data using the following steps: (1) Inversion, (2) Generation of a ranked dig list, and (3) Selection of a dig threshold. This section gives an

overview of each step. Details specific to any particular demonstrator can be found in the individual demonstrator's report ([4], [6], [11], [16], [18]).

2.14.1 Data Inversion

Demonstrators performed a geophysical inversion on the collected data to estimate parameters of the buried target. When analyzing data collected in cued mode, the demonstrators input all data collected at each location on the cued list into an inversion routine. Similarly, when analyzing data collected in survey mode, the demonstrators first identified all locations on the master list that were associated with the instrument. Then, for each of those locations, the demonstrators input into the inversion routine all data selected around the location. The purpose of the inversion routine was to fit the collected data to a dipole model. The underlying assumption of all discrimination algorithms used in this study is that the targets of interest (i.e., 4.2″ mortars) can be sensibly modeled as a two- or three-axis point dipole. Previous work has shown that this assumption holds fairly well in practice for all but large, shallow targets, where incident field variations over the target of interest cannot be ignored ([3], [11]).

Thus, the inversion problem reduces to determining the extrinsic and intrinsic parameters of the target of interest. Extrinsic parameters include the target's location (easting and northing), orientation and depth. Intrinsic parameters include characteristics of the target regardless of where the target is placed, such as a target's size and shape. UXO in general, and the 4.2″ mortar specifically, tend to be ferrous bodies of revolution with one large axis and two equal, smaller axes. In contrast, munitions debris and cultural artifacts are typically smaller and are not typically bodies of revolution. This difference in size and shape can be exploited during discrimination.

For EM61-Mk2 sensors, the potentially available intrinsic parameters include a set of the polarizabilities of the target along each of its three major axes in three time gates. While the strength of the polarizabilities is an indication of the target's size, their relative strength with respect to each other is an indication of the target's shape. For sensors that sample the received signal over a broader range of frequencies or temporal decays, such as the GEM-3, EM63, and BUD sensors, the intrinsic parameters include the three polarizabilities over a wider range of time. Once again, the polarizabilities indicate the target's size as well as shape and can provide information about such characteristics as material composition and wall thickness.

To accurately estimate the three principal polarizabilities of a buried target—that is, to estimate the target's shape as well as its size—the target must be illuminated and

sampled from three orthogonal directions. This immediately eliminates magnetometers from such an approach, because the illuminating field is the Earth's magnetic field, and there is no assurance that it will sufficiently illuminate all three axes (e.g., if a 4.2″ mortar axis were aligned with Earth's magnetic field, no information about the other two axes would be available). In the case of magnetometers, then, the only information that can be unambiguously extracted from the collected data is the magnetic moment, a parameter that indicates the target's effective size in the illumination direction only.

Some demonstrators performed cooperative inversions using EMI and magnetometer data. Although inversion of EMI data provides shape, as well as size, information, previous work has shown that inversions using EM61 data alone often provide poor depth estimates. Previous studies have also shown that magnetometer inversions lead to more accurate depth estimates [17]. Therefore, for all locations on the master list associated with *both* the EM61 Array *and* the Mag Array, demonstrators first inverted the magnetometer data to estimate the target's depth. Then, the demonstrators constrained the depth parameter of the EMI model to be the depth estimated from the magnetometer inversion. Next, the demonstrators inverted the EMI data using the depth-constrained EMI model to estimate the target's polarizabilities.

Other demonstrators performed individual inversions using the EMI and magnetometer data, but used parameters derived from both inversions in forming their feature vectors. Here, both the unconstrained EMI and magnetometer models were used. For all locations on the master list associated with the EM61 Array (regardless of whether they were also associated with the Mag Array), the demonstrators inverted the EM61 Array data using the unconstrained EMI model. Likewise, for all locations on the master list associated with the Mag Array (regardless of whether they were also associated with the EM61 Array), the demonstrators inverted the Mag Array data using the unconstrained magnetometer model. For all locations associated with *both* the EM61 Array and Mag Array, the demonstrators based their discrimination processing on both the EMI and magnetometer parameters.

For every data-collection instrument, the demonstrators selected a subset of the intrinsic parameters. Some demonstrators chose a very simple subset consisting of one parameter only, such as the magnetic moment or the principal polarizability. Other demonstrators chose more complex subsets consisting of multiple parameters, such as all three polarizabilities or the ratios between different polarizabilities. The demonstrators attempted to form one feature vector from the selected parameters for each location associated with the instrument.

For some locations, however, the data collected by the instrument suffered from position errors, poor data density, or low SNR to the extent that the demonstrators could not perform an accurate geophysical inversion and therefore could not form a useful feature vector from the estimated parameters. The demonstrators labeled these locations as "Can't analyze." Each demonstrator used different criteria for labeling these locations. While some demonstrators based their labels on quantitative criteria, such as the fit coherence or other measures of how well the collected data fit the dipole model, other demonstrators based their labels on subjective criteria, such as visual analysis of the collected data.

2.14.2 Ranked Dig List Generation

Demonstrators further analyzed every location associated with the instrument and for which a feature vector could be formed. To do this, the demonstrators input the feature vectors formed from these locations into a discrimination algorithm. The purpose of the discrimination algorithm was to estimate each location's likelihood of containing only clutter based on the location's feature vector. Different demonstrators used different algorithms for estimating the likelihood that a location contained only clutter. While some demonstrators used simple rule-based algorithms based on a quantitative threshold set using expert knowledge, others used more complicated algorithms based on statistical classifiers or template matchers.

The demonstrators formed one ranked dig list for each combination of data-collection instrument and discrimination algorithm. The ranked dig list consisted of a list of all locations associated with the instrument for which feature vectors could be estimated (i.e., those locations whose data could be analyzed). The demonstrators arranged the locations on the ranked dig list based on their likelihood of being clutter. Figure 2.39 shows a cartoon of a ranked dig list. The first location on the list is that location most likely to contain only clutter, based on the discrimination algorithm's quantitative interpretation of the feature vector estimated for the location. Conversely, the last item on the list is that location most likely to contain a munition. In a real-world scenario, the excavation team would begin recovering items from the locations at the bottom of the list and work its way up.

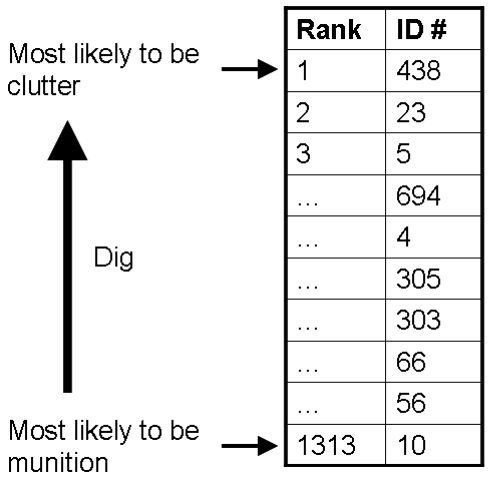


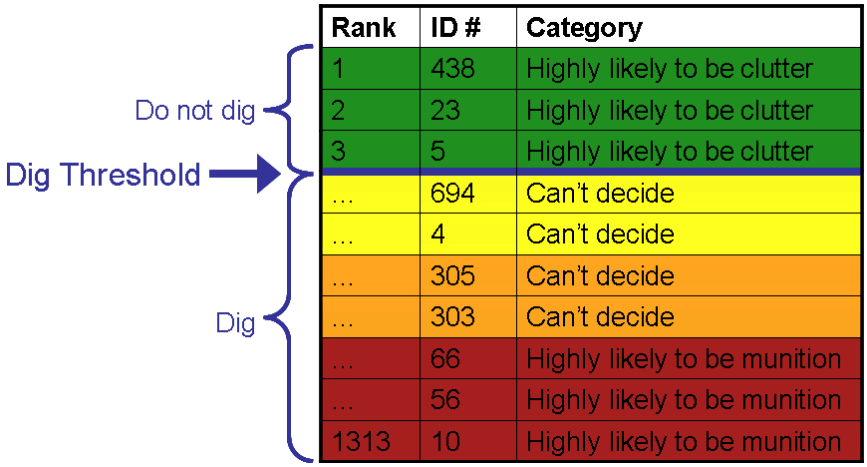
Figure 2.39: A cartoon example of a ranked dig list. Data are simulated and for illustration only.

2.14.3 Dig Threshold Selection

For each ranked dig list, demonstrators separated the locations on the list into four distinct categories. As shown in Figure 2.40, those locations near the top of the list were categorized as "highly likely to contain only clutter" (green); those locations near the bottom of the list were categorized as "highly likely to contain munitions" (red); and locations near the middle of the list were categorized as either "can't decide, but likely to contain only clutter" (yellow) or "can't decide, but likely to contain munitions" (orange). In our proposed real-world scenario, only those items that were highly likely to contain clutter could be left in the ground. Therefore, the boundary between the last green location and the first yellow location constitutes the dig threshold. In our proposed real-world scenario, the excavation team would begin recovering items from the locations at the bottom of the list and work its way up until it reached the dig threshold. Upon reaching the dig threshold, the excavation team would cease digging. That is, all locations below the dig threshold were assigned the label of "dig"; and all locations above the dig threshold were assigned the label of "do not dig."

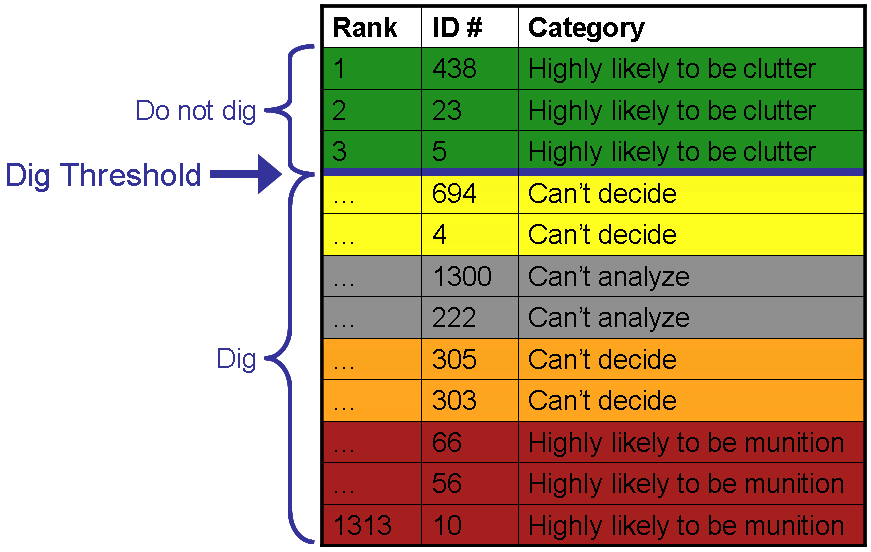
The ranked dig list shown in Figure 2.40 consists of all locations associated with a particular instrument for which the collected data could be inverted and feature vectors could be estimated. Some locations could not be analyzed, however, because their data did not support an accurate inversion. In a real-world scenario, these "Can't analyze" locations would have to be excavated since they possibly could contain munitions. The demonstrators were initially instructed to insert these locations into the ranked dig list between the two "Can't decide" categories. In this way, those locations with the highest likelihood of being either clutter or munitions occupied either end of the list and those

locations for which the label is most uncertain occupied the middle of the list (Figure 2.41). Because these "Can't analyze" locations fall below the dig threshold, they would be given the "dig" label and would be excavated in our proposed real-world scenario. As the study progressed, however, it became evident that the "Can't analyze" locations should be appended to the end of the list, rather than inserted into the middle (Figure 2.42). Doing so allowed the creation of more easily readable ROC curves during the discrimination scoring process. ROC curves are discussed in more depth in section 2.18: Survey Discrimination Scoring.



Rank	ID #	Category
1	438	Highly likely to be clutter
2	23	Highly likely to be clutter
3	5	Highly likely to be clutter
…	694	Can't decide
…	4	Can't decide
…	305	Can't decide
…	303	Can't decide
…	66	Highly likely to be munition
…	56	Highly likely to be munition
1313	10	Highly likely to be munition

Figure 2.40: A cartoon example of a ranked dig list, with locations categorized based on their likelihood of containing clutter versus munitions. Data are simulated and for illustration only.



Rank	ID #	Category
1	438	Highly likely to be clutter
2	23	Highly likely to be clutter
3	5	Highly likely to be clutter
…	694	Can't decide
…	4	Can't decide
…	1300	Can't analyze
…	222	Can't analyze
…	305	Can't decide
…	303	Can't decide
…	66	Highly likely to be munition
…	56	Highly likely to be munition
1313	10	Highly likely to be munition

Figure 2.41: A cartoon example of a ranked dig list, with those locations that could not be analyzed inserted into the middle of the list. Data are simulated and for illustration only.

	Rank	ID #	Category
Do not dig	1	438	Highly likely to be clutter
	2	23	Highly likely to be clutter
	3	5	Highly likely to be clutter
Dig Threshold →	...	694	Can't decide
	...	4	Can't decide
	...	305	Can't decide
	...	303	Can't decide
Dig	...	66	Highly likely to be munition
	...	56	Highly likely to be munition
	1313	10	Highly likely to be munition
	...	1300	Can't analyze
	...	222	Can't analyze

Figure 2.42: A cartoon example of a ranked dig list, with those locations that could not be analyzed appended to the end of the list. Data are simulated and for illustration only.

Three of the four demonstrators specified dig thresholds by subjectively balancing the ratio between the cost associated with leaving a munition in the ground versus the cost of unnecessarily digging a clutter item. The fourth demonstrator specified dig thresholds based on an equation that takes as its independent variable the ratio of these two costs [6]. Here, a stakeholder provides the value of the cost ratio, plugs it into the equation, and receives the value of the dig threshold. Specifying the cost ratio is equivalent to specifying the lower limit on the probability that a location is clutter. That is, all locations with probabilities of being clutter that are greater than or equal to this lower limit are labeled as "do not dig." For example, setting the dig threshold based on a probability equal to or greater than 96% that a location is clutter is mathematically equivalent to setting a dig threshold based on a cost ratio of 25 (i.e., leaving a munition in the ground is 25 times more costly than unnecessarily digging a clutter item). Similarly, setting the dig threshold based on an equal to or greater than 98% or 99% probability is equivalent to setting a dig threshold based on a cost ratio of 50 or 100, respectively.

2.14.4 Training and Test Sets

IDA separated the locations on the master list into a Training Set and a Test Set. The Program Office distributed the complete geophysical sensor data sets to the demonstrators. The demonstrators were given the ground-truth labels for the locations in the Training Set only; they remained blind to the ground-truth labels of the locations in the Test Set. The demonstrators used the ground truth in the Training Set to optimize their discrimination algorithms and methods for selecting dig thresholds. Once that was

done, they applied the optimized algorithms to the locations in the Test Set to form ranked dig lists and then selected dig thresholds using their optimized methods. For each instrument/algorithm combination, the demonstrators submitted the ranked dig list and the dig threshold for scoring. Results were scored over the Test Set only.

The Training Set consisted of the locations and identifications of all seeded items in the GPO, as well as the locations and identifications of all single targets on the master list that fell within predetermined subareas of the SE1, SE2, and SW test areas. In conjunction with the Program Office, IDA selected the subareas for the Training Set. Figure 2.43 shows a map of the SE1, SE2, and SW test areas (the GPO is not shown). All single target locations on the master list that were included in the Training Set are shown in red (munitions) and green (clutter), and all remaining master list locations (those included in the Test Set) are shown in black. Of the 1,359 single target locations on the master list associated with at least 1 digital survey instrument, 208 were assigned to the Training Set (of which 30 were munitions) and 1,151 were assigned to the Test Set (of which 119 were munitions). From the master list, 178 of the single target locations were also included on the cued list. Of these 178 cued locations, 28 were included in the Training Set (8 of which were munitions), and 150 (34 of which are munitions) were included in the Test Set. In each of the three test areas, a contiguous subarea was chosen for the Training Set to mirror what would likely occur in our proposed real-world scenario.

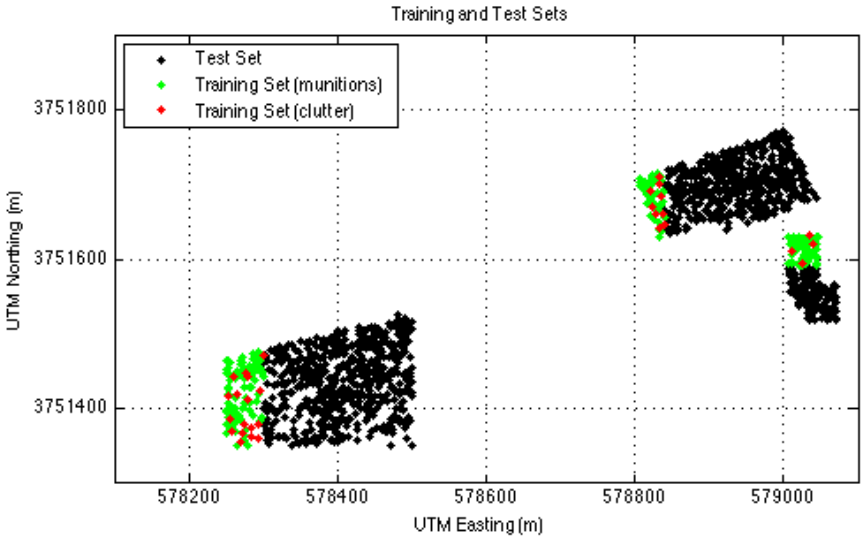


Figure 2.43: Master list locations at Camp Sibert. Locations included in the Training Set are shown in red (munitions) and green (clutter), and locations included in the Test Set are shown in black.

While three of the demonstration teams used supervised learning techniques to optimize their discrimination algorithms, one demonstration team experimented with two other optimization methods: semi-supervised learning and active learning. In traditional supervised learning, data in the Training Set are assigned ground-truth labels, and the labeled data are used to optimize the discrimination algorithms. In semi-supervised learning, however, the discrimination algorithms are optimized based on labeled data in the Training Set, *as well as* unlabeled data in the Test Set. Thus, an algorithm trained with semi-supervised methods exploits context in the Test Set data during optimization. This results in a more conservative estimate of the probability that a location contains clutter. In active learning, the Training Set is *not* determined in advance. Instead, all data initially remain unlabeled, and a set of information-theory metrics is used to determine which locations could benefit the optimization of the algorithm the most if ground-truth labels were assigned. Items are excavated from these locations, ground-truth labels are assigned, and the algorithm is optimized based on those ground-truth labels. The process then iterates several times until the information-theory metrics note that little further benefit can be gained by digging further items [5].

2.15 EXCAVATION

The excavation team recovered *all* items buried at locations specified in the master list. The purpose of the excavation was to obtain information that could be used to assign ground-truth labels to each location on the master list. The master list consisted of two types of locations: single targets and clusters.

Single target locations were likely to contain one item only. IDA provided the excavation team with a list of the estimated positions (easting, northing, and depth) of every single target location. The estimated easting and northing positions were the group centroids calculated during the generation of the master list. The estimated depths were the values that resulted from fitting the collected data to a dipole model during anomaly detection. The excavation team recovered all metallic items found at the specified locations. For each recovered item, the excavation team measured its exact position (easting, northing, and depth with respect to the elevation of the surface of the hole). The team also noted a description of the item (e.g., “UXO,” “splayed half round,” “wrench,” “horseshoe,” etc.) and took a photograph of each item.

Cluster locations were likely to contain multiple, closely spaced items. For each cluster location, IDA provided the excavation team with a set of four easting/northing coordinates. The four coordinates represented the vertices of a square, approximately 2 m

× 2 m, circumscribing the cluster location. The excavation team recovered all metallic items buried within the four vertices. The team noted the coordinates and descriptions of each recovered item and photographed each recovered item. Although none of this information was used in this study, it is available for future ESTCP or SERDP projects. Details of the excavation can be found in [16].

2.16 ASSIGNMENT OF GROUND TRUTH

The Program Office assigned one ground-truth label to each single target location on the master list based on the descriptions and photographs of each item recovered from the location. The purpose of the ground-truth labels was to score the discrimination performance of each instrument/algorithm combination used by the demonstrators.

During the initial stages of the study, the Program Office, in conjunction with the Advisory Panel, decided that a single target location would be labeled as "munition" if it met any of the following criteria:

- UXO was recovered from the location.
- An item that the general public could confuse with UXO was recovered from the location (such an item left in the ground could resurface in the future, causing great unease in the local community).
- A metallic item of the same size and aspect ratio as a 4.2″ mortar was recovered from the ground, because current data-collection instruments and algorithms could not be expected to discriminate such items from true 4.2″ mortars.

Conversely, a single target location would be labeled as "clutter" if it did not meet any of the criteria listed above.

Once the excavation was complete, it became apparent that the only locations meeting the criteria for "munitions" were those locations containing seeded UXO. No locations contained either an item that the general public could confuse with UXO or a metallic object of the same size and aspect ratio as a 4.2″ mortar.

Note that the excavation results did not always confirm that each single target location contained one item. In a number of cases, more than one item was recovered from the same single target location (e.g., several small pieces of munitions scrap), a magnetic rock was found (described by the excavation team as "hot rocks"), magnetic dirt was found (described as "hot soil"), or nothing was found (described as "no contact"). Each of these locations was still included in the single target data set, and each

location was assigned a single ground-truth label, with "munition" taking precedence over "clutter."

Ground-truth labels were assigned to 1,388 of the 1,389 single target locations on the master list. A ground-truth label could not be assigned to Target ID #321, since no measurements or photographs were taken during excavation. Therefore, this location was not included in the scoring process.

2.17 SURVEY DETECTION SCORING

Although the main goal of this study was to assess the discrimination performance of each instrument/algorithm combination, IDA also assessed the detection performance of each data-collection instrument used in survey mode.[1] Only those locations assigned to the Test Set were included in the scoring process. IDA scored the detection performance of each survey instrument by comparing the ground-truth label of each location on the master list to whether or not the instrument was associated with the location. In general, an instrument was associated with a location on the master list if an anomaly detected by the instrument was within 0.6 m of the location (see Section 2.12). The white box in Figure 2.44 summarizes the detection-scoring process. A true positive (TP) was a location on the master list that was assigned a ground truth label of "munition" *and* was associated with the instrument during generation of the master list (i.e., the instrument detected at least one anomaly within 0.6 m of the location). A false negative (FN) was a location that was assigned a ground truth label of "munition" *but* was *not* associated with the instrument. A false positive (FP) was a location that was assigned a ground-truth label of "clutter" *but* was associated with the instrument. True negatives (TN) were not counted.

To summarize the detection performance of each survey instrument, IDA calculated the probability of detection (Pd) and the false-alarm rate (FAR).

---

[1] Because data were collected by the cued instruments at the predetermined locations on the cued list, IDA did not assess the detection of each instrument used in cued mode.

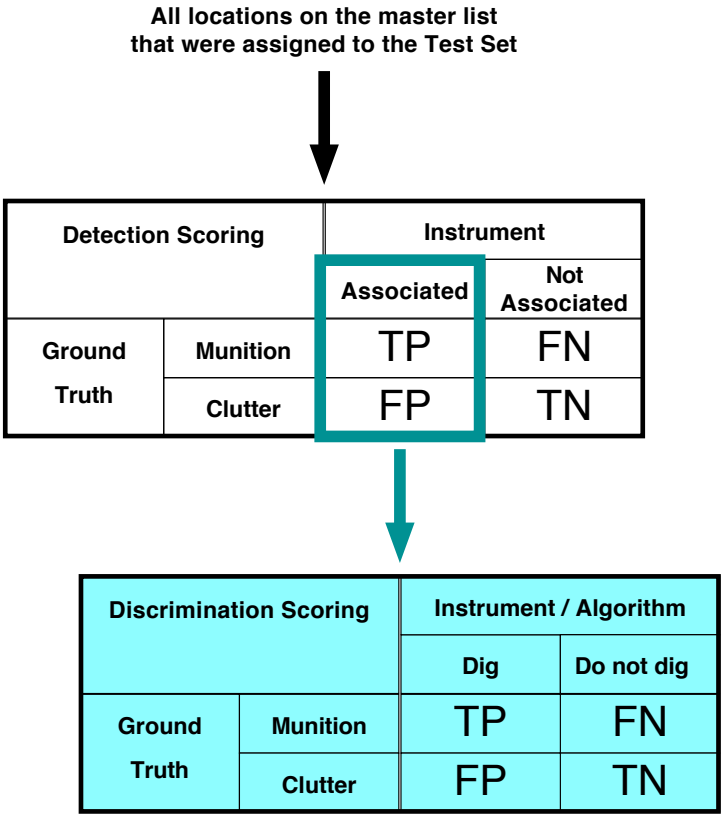


Figure 2.44: Scoring the detection performance of a survey instrument and the discrimination performance of an instrument/algorithm combination. All locations on the master list that were assigned to the Test Set were included in the detection-scoring process. All locations on the master list that were assigned to the Test Set and that were associated with the instrument were included in the discrimination scoring process.

Pd is the fraction of "munition" locations on the master list that were associated with the instrument. Pd is calculated as the ratio of the number of "munition" locations on the master list that were associated with the instrument (TP) to the total number of "munition" locations on the master list (TP + FN): Pd = TP/(TP + FN). Due to the very high cost of leaving a munition in the ground, the UXO community desires instruments with Pd values at or near 1.00. To assess statistically how near or far a Pd value is from the desired 1.00, the 95% confidence interval was calculated around Pd based on the exact binomial distribution [10].

Note that Pd is only an *estimate* of the fraction of munitions detected by the instrument because an exhaustive clearance was not done at Camp Sibert. The excavation team recovered items only at locations specified on the master list, and although unlikely, the possibility remains that other munitions existed at locations other than those on the master list. If these items exist and had been identified and factored into the scoring process, the instruments' Pd values could have been somewhat lower than what is

reported in this document. Nevertheless, while the purpose of the seeding program was to ensure that a sufficient number of UXO would be present to provide high-confidence statistics on Pd, seeding also guaranteed that even if a few existing munitions were missed, Pd statistics would be only marginally affected.

FAR is the number, per unit area, of "clutter" locations on the master list that were associated with the instrument. That is, FAR is an estimate of the number of unnecessary digs per unit area: FAR = FP/Area. In the absence of discrimination algorithms, all anomalies detected by an instrument must be dug. A high FAR suggests that many of these anomalies turned out to be clutter and therefore that many of the digs were unnecessary. Therefore, due to the cost of unnecessarily digging clutter, the UXO community desires instruments with FAR values as low as possible. When an instrument is used in conjunction with discrimination algorithms, however, all anomalies detected by the instrument are inverted and input into the discrimination algorithm so that the algorithm can label the anomalies as "dig" or "do not dig." In theory, it is possible that the algorithm can label many, or even all, of the clutter anomalies as "do not dig," thereby reducing the number of unnecessary digs for the instrument/algorithm combination with respect to the instrument on its own. Thus, an instrument with a high FAR can still be useful when used in conjunction with a discrimination algorithm.

Like Pd, FAR is only an *estimate* of the number of unnecessary digs per unit area, because an exhaustive clearance was not done at Camp Sibert. The locations on the master list associated with an instrument are only a subset of the anomalies detected by that instrument. Many anomalies detected by each instrument were "clustered" (i.e., they were too close in space to other anomalies) and were therefore not acknowledged in the scoring process. It is likely that many of these "clustered" anomalies represented clutter items. Had these anomalies been factored into the scoring process, the instruments' FAR values would likely have been higher than what is reported in this document.

2.18 DISCRIMINATION SCORING

The main goal of this study was to assess the discrimination performance of each instrument/algorithm combination. To that end, IDA scored the discrimination performance of each instrument/algorithm combination by comparing the ground-truth label of each location on the master list associated with the instrument to the "dig/do not dig" label assigned to the location during the discrimination process. Only those locations associated with the instrument and assigned to the Test Set were included in the scoring process. The blue box in Figure 2.44 summarizes the discrimination-scoring process. A

true positive (TP) is a location (associated with the instrument) that was assigned the ground-truth label of "munition" *and* was assigned the label of "dig" during the discrimination process. A false negative (FN) was a location (associated with the instrument) that contained a munition *but* was assigned the label of "do not dig." A false positive (FP) was a location (associated with the instrument) that contained clutter *but* was assigned the label of "dig." A true negative (TN) was a location (associated with the instrument) that contained clutter *and* was assigned the label of "do not dig."

As in detection scoring, IDA calculated Pd and FAR to summarize the discrimination performance of each instrument/algorithm combination.

The probability of detection (Pd) is the fraction of "munition" locations on the master list (associated with the instrument) that were labeled as "dig" during the discrimination process. That is, in terms of discrimination scoring, Pd is an estimate of the fraction of detected munitions that were dug: Pd = TP/(TP + FN). Due to the safety hazard of leaving a munition in the ground, the UXO community desires instrument/algorithm combinations with Pd values at or near 1.00. The 95% confidence interval around Pd was estimated using the exact binomial distribution [10].

The number of unnecessary digs (FP) is the number of "clutter" locations on the master list (associated with an instrument) that were labeled as "dig" during the discrimination process. In other words, FP is an estimate of the total number of unnecessary digs. Although the FAR metric was used for detection scoring, the FP metric is used for discrimination scoring because FP can be more easily translated into the dollars saved by using discrimination algorithms to reduce the number of unnecessary digs.

Due to the high cost associated with unnecessary digs, the UXO community desires instrument/algorithm combinations with FP values as low as possible. The main goal of UXO discrimination is to reduce FP as much as possible while still retaining Pd values at or near 1.0.

Figure 2.45 shows a cartoon of Pd plotted versus FP. The point on the graph illustrates the discrimination performance of an instrument/algorithm combination when the demonstrator's dig threshold is applied to the ranked dig list (the dark blue line in Figure 2.42). The 95% confidence interval around Pd is drawn through the [FP, Pd] point (gray bar). Because Pd is plotted on the vertical axis, the 95% confidence interval around Pd is drawn as a vertical bar. The vertical axis runs from zero to one because Pd is a fraction. The horizontal axis ranges from zero to the maximum possible FP value.

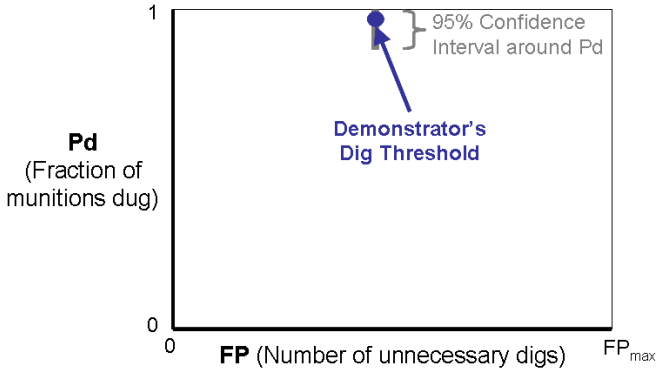


Figure 2.45: Plotting the operating point for an instrument/algorithm combination at the demonstrator's dig threshold. The probability of detection (Pd) and the number of false positives (FP) are calculated and plotted as a point (dark blue dot). The 95% confidence interval around Pd is drawn through the point (gray bar). Data are synthesized and for illustration only.

The plot of Pd versus FP can be used to revisit the choice of dig threshold. To analyze this choice, IDA applied every possible dig threshold to the ranked dig list, calculated the resulting Pd and FP values, and plotted those values as points. Figure 2.46 shows a cartoon plot of Pd versus FP for every possible value of the dig threshold (black dots). For each point, the 95% confidence interval around Pd is drawn through the point (gray bars).[2] Together, the points form a ROC curve.

The ROC curve shows the instrument/algorithm's maximum possible Pd and FP values in the upper right corner. Referring to Figure 2.42, a dig threshold could have been applied at the top of the ranked dig list such that all locations on the ranked dig list (i.e., all locations detected by the instrument) would have fallen below the dig threshold and would have been labeled as "dig." In such a case, *all* munitions detected by the instrument would have been dug, resulting in the maximum possible Pd of 1.0. However, all clutter items detected by the instrument would have also been dug. Thus, calculating the maximum possible value of FP for an instrument/algorithm combination during discrimination scoring is equivalent to calculating the FP for the instrument on its own during detection scoring. The purpose of the discrimination algorithm is to reduce FP from this maximum value while still maintaining a Pd at or near 1.00.

---

2 Note that the 95% confidence intervals were calculated for each point independently using the exact binomial distribution without any adjustments for multiple comparisons. Therefore, one cannot infer that 95 times out of 100, every point on the ROC curve will, simultaneously, lie within its 95% confidence interval. That is, one cannot infer that 95 times out of 100, the entire ROC curve will lie within the band generated by "smearing" the individual 95% confidence intervals [13].

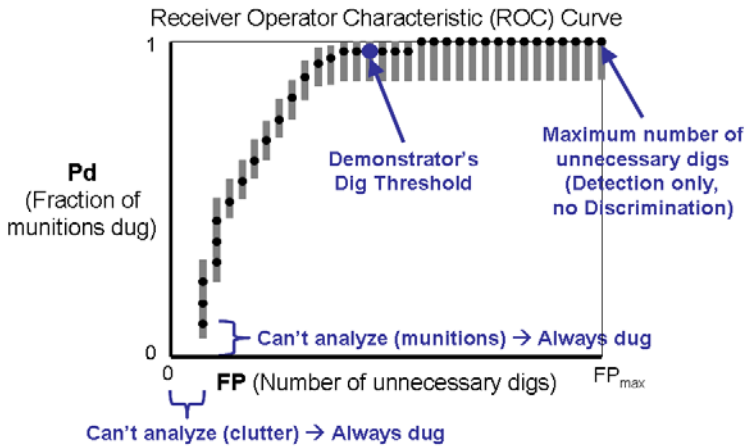


Figure 2.46: Generating a ROC curve for an instrument/algorithm combination. For every possible value of the dig threshold, Pd and the number of FPs are calculated and plotted as a point (small black dots). The 95% confidence interval around Pd is drawn through each point (gray bars). Together, the points form a ROC curve. The ROC curve cannot touch the lower left corner of ROC space if some locations cannot be analyzed.

The ROC curve also shows the instrument/algorithm's minimum possible Pd and FP values in the lower left corner. Referring again to Figure 2.42, a threshold could have been applied to the ranked dig list such that all locations on the ranked dig list *that could be analyzed* (those *not* gray) would have been labeled as "do not dig." In contrast, all locations that could *not* be analyzed (those colored gray and appended to the end of the ranked dig list) must, by definition, always be dug. That is, a dig threshold *cannot* be applied to the ranked dig list in the region of the list populated by the locations that cannot be analyzed. Therefore, if some of these "Can't analyze" locations are munitions, they will contribute to the Pd value, and Pd will never be zero. Furthermore, if some of the "Can't analyze" locations are clutter, they will contribute to the FP value, and FP will never be zero. Thus, a ROC curve that does not touch the [0, 0] origin of ROC space indicates that some locations could not be analyzed.

IDA analyzed the ROC curve to revisit the choice of dig threshold. As can be seen in Figure 2.46, an instrument/algorithm combination could potentially lead to both a high Pd and high FP (a dig threshold near the top of the dig list) or both a low Pd and low FP (a dig threshold closer to the "Can't analyze" items). Choosing the dig threshold is a critical step in UXO discrimination because the choice of dig threshold determines where the instrument/algorithm's performance lies along the ROC curve. Therefore, IDA identified what would have been the "best case scenario" dig threshold and compared its performance to the performance of the demonstrator's chosen dig threshold, illustrated by the dark blue dot on Figures 2.45–2.50.

IDA defined the best case scenario dig threshold in two ways. First, the best case scenario dig threshold is that which would have resulted in the largest possible reduction in FP while Pd remained at 1.00. That is, the cost of unnecessary digs would have been minimized while all munitions would have been dug. Figure 2.47 illustrates this dig threshold with a light blue dot. Second, the best case scenario dig threshold is that which would have resulted in the largest possible reduction in FP while Pd remained at 0.95. That is, the cost of unnecessary digs would have been minimized while 95% of munitions would have been dug, leaving 5% of munitions (the most difficult to find) in the ground. This dig threshold is denoted with a pink dot in Figure 2.47.

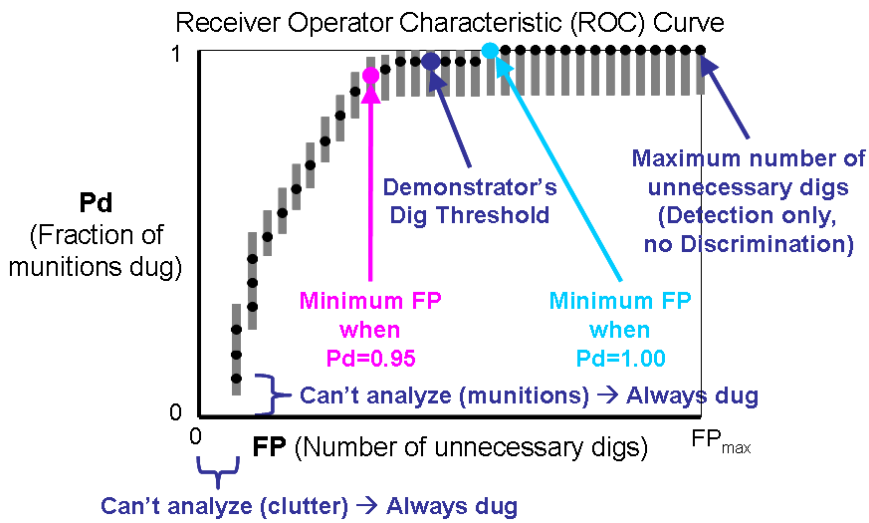


Figure 2.47: Generating a ROC curve for an instrument/algorithm combination. The dark blue dot denotes the discrimination performance resulting from the demonstrator's chosen dig threshold. In contrast, the light blue and pink dots denote the performance of two retrospectively chosen dig thresholds, each of which can be described as a best case scenario.

Next, IDA noted the location on the ranked dig list of every possible dig threshold in accordance with Figure 2.42. Figure 2.48 shows an example of a ROC curve with individual points colored according to the category in which the dig threshold fell. Note that by definition, the dark blue dot (the demonstrator's chosen dig threshold) separates the green and yellow dots just as the dark blue line in Figure 2.42 separates the green and yellow locations on the ranked dig list.

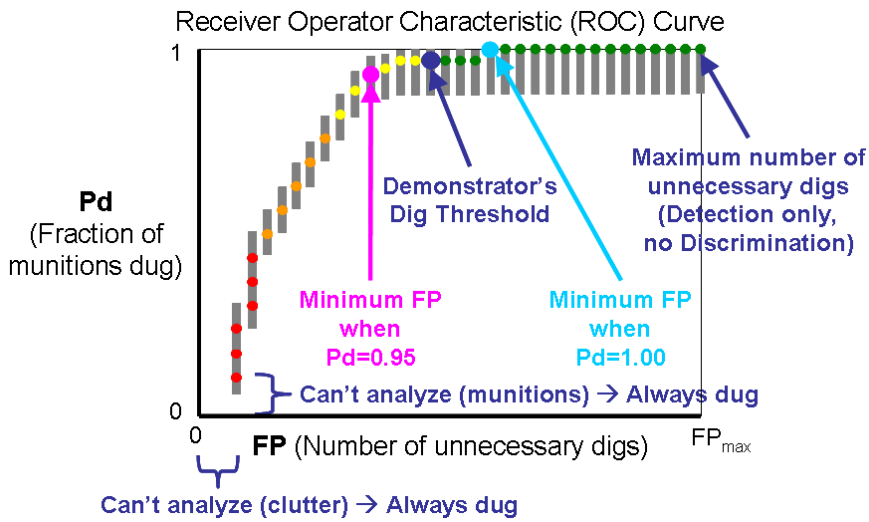


Figure 2.48: Generating a ROC curve for an instrument/algorithm combination. Points of the ROC curve are colored in accordance with Figure 2.42:
Green: "Highly likely to be clutter only"
Yellow: "Can't decide [but likely clutter only]"
Orange: "Can't decide [but likely munitions]"
Red: "Highly likely to be munitions"

Finally, IDA adjusted the horizontal axis of the ROC curve such that the ROC curves for all instrument/algorithm combinations could be plotted on the same scale. Figure 2.49 shows a ROC curve with the horizontal axis ranging from zero to "Overall FPmax," a value at least as large as the largest number of clutter items detected by a survey instrument.

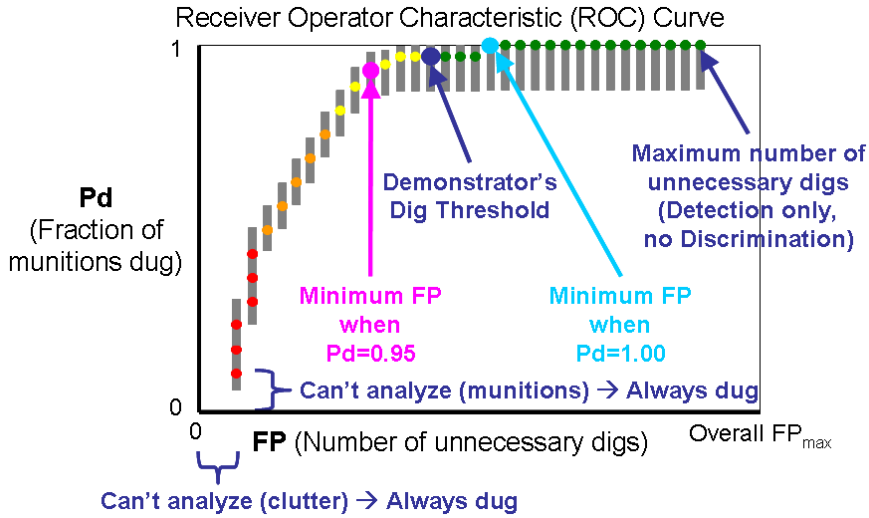


Figure 2.49: Generating a ROC curve for an instrument/algorithm combination. The horizontal axis is rescaled such that the number of FPs ranges from zero to an arbitrary yet consistent value greater than the number of FPs associated with any survey instrument. This allows for easier comparison between different instrument/algorithm combinations.

Figure 2.49 is an example of the ROC curves generated for each instrument/algorithm combination in this study. A ROC curve can be analyzed visually to quickly assess the performance of the instrument/algorithm combination:

- The dark blue dot can be used to assess the performance of the instrument/algorithm combination when the demonstrator's dig threshold is applied to the ranked dig list. A dark blue dot at or near 1.00 indicates that all or almost all munitions were dug. Furthermore, a dark blue dot much further to the left of the upper right end of the ROC curve indicates a large reduction in unnecessary digs with respect to the instrument used alone.
- The light blue and pink dots can be used to assess the performance of the instrument/algorithm combination, retrospectively, when the best case scenario dig thresholds are applied to the ranked dig list. By definition, the pink dot has a Pd of 0.95. A pink dot much further to the left than the upper right end of the ROC curve indicates that the dig threshold could have been adjusted to achieve a large reduction in unnecessary digs while leaving only 5% of munitions in the ground. Similarly, by definition, the light blue dot has a Pd of 1.00. A light blue dot much further to the left of the upper right end of the ROC curve indicates that the dig threshold could have been adjusted to achieve a large reduction in unnecessary digs even when all munitions were dug.
- The shape of the ROC curve can be used to assess the algorithm's ability to accurately discriminate between the two types of items.

In many traditional discrimination problems, the shape of the ROC curve can be described quantitatively as the area under the ROC curve. A ROC curve with a sharp angle near the upper left corner of ROC space has a large area under its curve and indicates that most clutter items were arranged higher on the ranked dig list than most munitions. That is, the algorithm estimated high likelihoods of being clutter for most clutter items and low likelihoods of being clutter for most munitions, because the feature vectors estimated for clutter and munitions overlapped little in multidimensional feature space.

Regarding the UXO discrimination problem, however, a ROC curve can indicate good discrimination performance even without a large area under its curve. This is the case because a munition incorrectly left in the ground (a false negative) is considered a much greater hazard than a clutter item unnecessarily dug (a false positive). Figure 2.50 shows cartoon sketches of two ROC curves. In the left sketch, the ROC curve exhibits a large area under its curve, with a very sharp angle near the upper left corner of ROC space. Even in retrospect, however, there exists no dig threshold that could reduce FP

while Pd remains at 1.00. In contrast, the ROC curve in the right sketch exhibits a smaller area under its curve and a much shallower angle. Yet proper selection of the dig threshold could lead to a large reduction in FP while Pd remains at 1.00. Thus, although a large area under the ROC curve is evidence of an algorithm's ability to discriminate between clutter and munitions, the true test of an algorithm's utility *in the UXO community* is its ability to reduce the number of unnecessary digs while still digging all munitions.

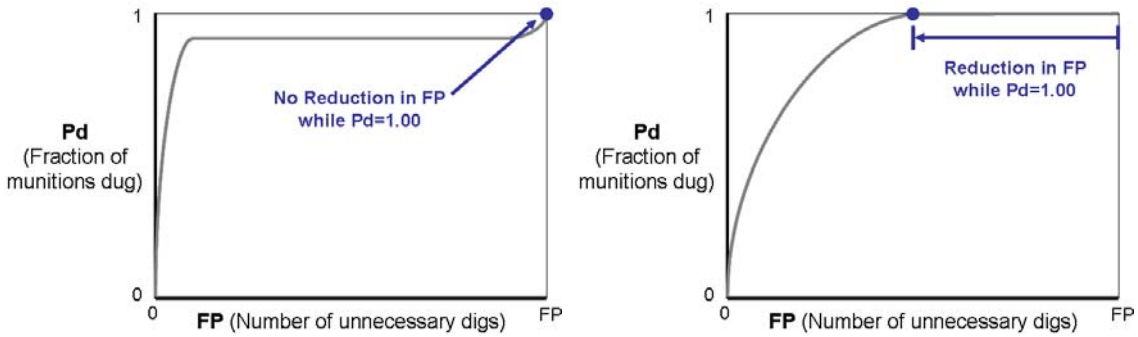


Figure 2.50: Sketches of two ROC curves. In the left sketch, the ROC curve exhibits a very large area under its curve, indicating a strong ability to discriminate between clutter and munitions. However, no dig threshold can be selected that would have led to a large reduction in unnecessary digs while digging all munitions. In contrast, the ROC curve in the right sketch exhibits a smaller area under its curve, indicating the inability to discriminate between a larger subset of clutter and munitions. However, a dig threshold can be selected that would have led to a large reduction in unnecessary digs while digging all munitions. Therefore, the instrument/algorithm combination described by the right ROC curve more closely addresses the needs of the UXO community.

ROC curves for each instrument/algorithm combination (either cued or survey) were the final scoring products resulting from the UXO discrimination study at the former Camp Sibert.

3. SELECTED RESULTS AND DISCUSSION

A total of 317 instrument/algorithm combinations were scored over different subareas at Camp Sibert. The performance of each combination is included in Appendix B, which exists in electronic form as a DVD accompanying this document. The remainder of this chapter discusses the major findings of the UXO Discrimination Study and illustrates them with selected results from Appendix B.

3.1 DETECTION

In this document, detection performance is considered "good" if Pd is at or very near 1.00 (with a 95% confidence interval that includes 1.00), indicating that the instrument detected all or almost all munitions.

An instrument's FAR is not considered in the definition of "good" detection performance. In most traditional UXO clearance operations, in which only detection is performed (i.e., no discrimination), "good" detection performance is characterized by a high Pd and a low FAR—all munitions are dug with few unnecessary digs—because in the absence of discrimination algorithms, all items detected by an instrument must be dug. A high FAR would indicate that many of those items turned out to be clutter and therefore that many of the digs were unnecessary. When used in conjunction with a discrimination algorithm, however, all items detected by an instrument are inverted and input into the discrimination algorithm, so that the items can be labeled as "dig" or "do not dig." In theory, it is possible that the algorithm could label many of or even all the clutter items as "do not dig," thereby reducing the number the unnecessary digs for the instrument/algorithm combination compared with the instrument on its own. Therefore, an instrument with a high FAR can still be useful in conjunction with discrimination algorithms, such as in this study.

Finding 1: Survey sensors detected almost all munitions, leading to excellent detection performance.

Data were collected in survey mode using the GEM Array, EM61 Array, Mag Array, EM61 Cart, and BUD instruments. (Note that the BUD instrument was tested in both cued and survey modes.) Table 3.1 summarizes the detection performance metrics of

these five survey instruments, as well as the M&F operation. Detection performance was scored over the Test Set only.

Table 3.1: Detection performance of survey instruments over the Test Set. The mag-and-flag (M&F) operator, the GEM Array, and the EM61 Array detected all munitions in their respective survey areas, each exhibiting a Pd of 1.00. The Mag Array, EM61 Cart, and BUD detected all but one munition, each exhibiting Pd values only slightly less than 1.00 (with 95% confidence intervals including 1.00).

Instrument	TP	FN	$\mathbf{Pd} = \frac{\mathbf{TP}}{\mathbf{TP}+\mathbf{FN}}$ *[95% CI]*	FP	Surveyed Area (acres)	$\mathbf{FAR} = \frac{\mathbf{FP}}{\mathbf{Area}}$ (per acre)
M&F	4	0	1.00 *[0.40, 1.00]*	45	0.2*	225.0
GEM Array	119	0	1.00 *[0.97, 1.00]*	760	16.8	45.2
EM61 Array	119	0	1.00 *[0.97, 1.00]*	615	16.8	36.6
Mag Array	118	1	0.99 *[0.95, 1.00]*	706	16.8	42.0
EM61 Cart	118	1	0.99 *[0.95, 1.00]*	428	16.8	25.5
BUD	56	1	0.98 *[0.91, 1.00]*	210	5.2**	40.4

* The mag-and-flag survey was done on only one 100′ x 100′ grid in the Southeast 1 area.
** The BUD instrument surveyed only the Southeast 1 area.

As is shown in Table 3.1, the M&F operator, the GEM Array, and the EM61 Array detected all munitions, leading to a Pd of 1.00. That is, every munition in the Test Set was within 0.6 m of at least one GEM Array anomaly and at least one EM61 Array anomaly.

In contrast, under the detection scoring rules employed in this demonstration, the Mag Array, EM61 Cart, and BUD detected (i.e., declared an anomaly within the munition detection halo) *all but one* munition in the Test Set, leading to Pd values of 0.99, 0.99, and 0.98, respectively. Note that in each of these three cases, the 95% confidence interval around Pd included 1.00 to two significant digits. Also, for each of these three instruments, an anomaly was detected very close to the location of the "missed" munition but slightly farther than the arbitrary distance threshold (0.6 m) used to associate a location on the master list with an anomaly during the generation of the master list.

The Mag Array and EM61 Cart both missed Target ID #998. That is, that munition was not associated with any Mag Array anomalies or any EM61 Cart anomalies during generation of the master list. BUD missed Target ID #170. Both munitions were

recovered at some of the deepest depths among all seeded items—Target ID #998 was recovered at a depth of 0.90 m (8.6 times the diameter), and Target ID #170 was recovered at a depth of 1.00 m (9.5 times the diameter). Not only is it likely that their depths made for a challenging data inversion, but also the large spatial extent of the anomaly could easily have included returns from small clutter pieces that biased the position estimates. However, the large spatial extent would make reacquisition for digging highly likely even with position errors somewhat greater than the 0.6 m criterion. In a well executed, practical field case, those munitions certainly would have been dug.

Finding 2: Data collected from the EM61 Array were often noisy due to the bouncing motion of the towed vehicle over the ground during data collection.

Table 3.1 shows that the EM61 Array exhibited 187 (44%) more FPs than the EM61 Cart: 615 versus 428. Some of this difference may be due to the increased transmit moment provided by the three synchronized transmit coils on the array. However, it is likely that most of the difference was due to the type of platform on which the sensors were mounted (a towed vehicle versus a hand-pulled cart) and to the differences in survey patterns. During data collection, a vehicle tows the EM61 Array's sensors. As the sensors bounce over ground irregularities, their heights and orientations change with respect to the surface of the ground. This leads to spurious peaks in the collected data that in this study were eventually scored as FPs. In contrast, the EM61 Cart's sensors are mounted on a cart that is pulled over the ground by an operator at a much slower speed than the array. If properly trained, as in this study, the operator's constant attention to and control over the cart should allow the sensors to maintain a more constant height with respect to the surface of the ground.

Most important in this case, however, likely was the ground condition in the SW area where much of the noise was seen. This area had been previously plowed, leaving a series of furrows in the ground. The EM61 Array surveyed the SW area in two orthogonal directions, one of which worsened the bouncing motion over the furrows, leading to a large amount of motion noise in North-South runs versus that seen in East-West runs. The GEM Array typically also suffers from motion noise. In this case, however, the GEM Array and EM61 Cart surveyed the SW area in a direction that did not lead to as much bouncing motion over the furrows as with the EM61 Array. Furthermore, although the Mag Array does not suffer as much from motion noise, it is particularly sensitive to magnetic geology. Coincidentally, the SW area had geologic features that created a great deal of noise in the Mag Array data.

3.2 DISCRIMINATION

In this section, the discrimination performance of different instrument/algorithm combinations is summarized and discussed. In general, discrimination performance is considered "good" if:

1. Pd is at or very near 1.00 (with a 95% confidence interval that includes 1.00), indicating that the instrument detected all or almost all munitions.
2. FP is much lower than the maximum possible FP (i.e., the FP value calculated for the instrument during detection scoring), leading to a large reduction in unnecessary digs relative to when the detection instrument is used alone.

3.2.1 Detected items that cannot be analyzed

For those locations on the master list that were associated with an instrument, demonstrators labeled the locations as "Can't analyze" if the data did not permit a geophysical inversion of sufficient quality to allow further analysis.

Finding 3: All "Can't analyze" locations must be dug.

According to the scoring protocol, all "Can't analyze" locations were assigned the discrimination label of "Dig," regardless of the dig threshold. Knowledge of ground truth allows us to revisit the wisdom of this protocol. Because many "Can't analyze" locations turned out to be clutter, they could have been labeled as "Do not dig" with no safety hazard. Some "Can't analyze" locations turned out to be munitions, however, and these locations could *not* have been labeled as "Do not dig" without creating a large safety hazard. We therefore conclude that in the absence of ground truth, all "Can't analyze" locations must be dug because of the large safety hazard of leaving a munition in the ground.

For example, Figure 3.1 shows the ROC curve for anomalies associated with the EM61 Cart detections and discriminated by the UXAnalyze software with IDL extension.[3] The curve does not reach the lower left corner of ROC space because some locations associated with the instrument were identified as "Can't analyze" and were therefore assigned the label of "Dig." Of the "Can't analyze" locations, 93 turned out to be clutter, once ground truth was known. These 93 locations were always scored as FP

---

[3] SAIC performed the discrimination analysis.

(unnecessary digs). Therefore, the minimum FP for the ROC curve is 93, rather than 0. Similarly, of the 118 munitions locations associated with the EM61 Cart, 18 (15.2%) could not be analyzed. These items were always scored as TP (necessary digs), and the minimum Pd is therefore 0.152, rather than 0.

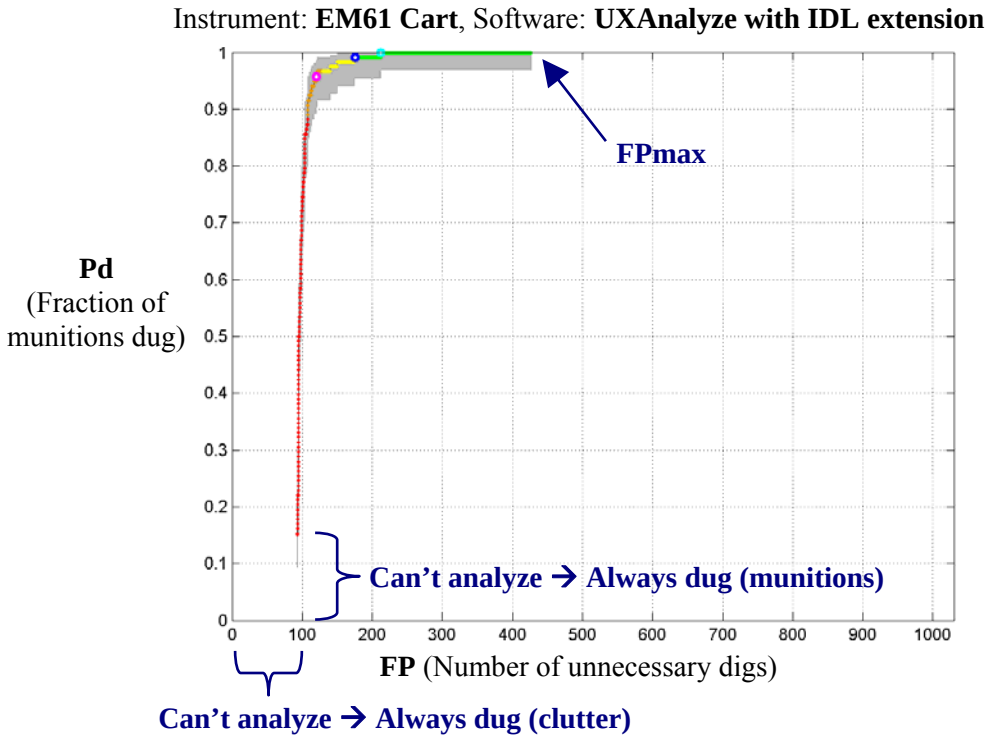


Figure 3.1: ROC curve for the EM61 Cart instrument and the UXAnalyze software with IDL extension. The curve does not reach the lower left corner of ROC space because some locations could not be analyzed. According to the scoring protocol, all locations that could not be analyzed were always labeled as "Dig," regardless of the dig threshold. Therefore, the minimum number of FPs is not zero because some locations that could not be analyzed were clutter and always scored as unnecessary digs. Similarly, the minimum Pd is also not zero because some locations that could not be analyzed were munitions.

Finding 4: A principled, documented method for identifying "Can't analyze" locations has not yet been agreed upon.

Each demonstrator used different criteria for labeling locations on the master list as "Can't analyze." As a result, even for a given set of sensor data, these locations differed from one another. Figures 3.2 and 3.3 show ROC curves for locations associated with the Mag Array. In Figure 3.2, the locations were discriminated by SAIC. SAIC labeled 198 locations as "Can't analyze," all of which turned out to be clutter. In comparison, Sky Research, Inc., discriminated the locations shown in Figure 3.3. Sky Research labeled 97 locations as "Can't analyze," 95 of which turned out to be clutter and 2 of which turned out to be munitions.

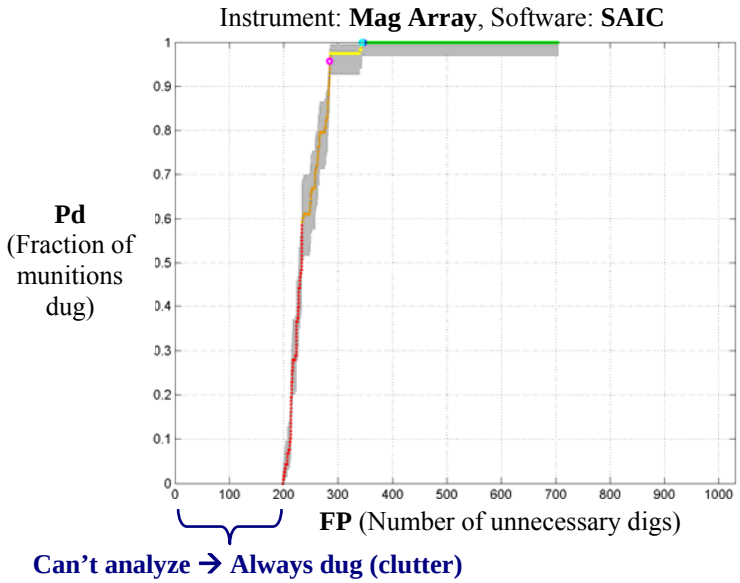


Figure 3.2: ROC curve for the Mag Array instrument and software used by SAIC. The curve does not reach the lower left corner of ROC space because some locations could not be analyzed. SAIC could not analyze 198 locations, more than the corresponding number for Sky Research, Inc., shown in Figure 3.3.

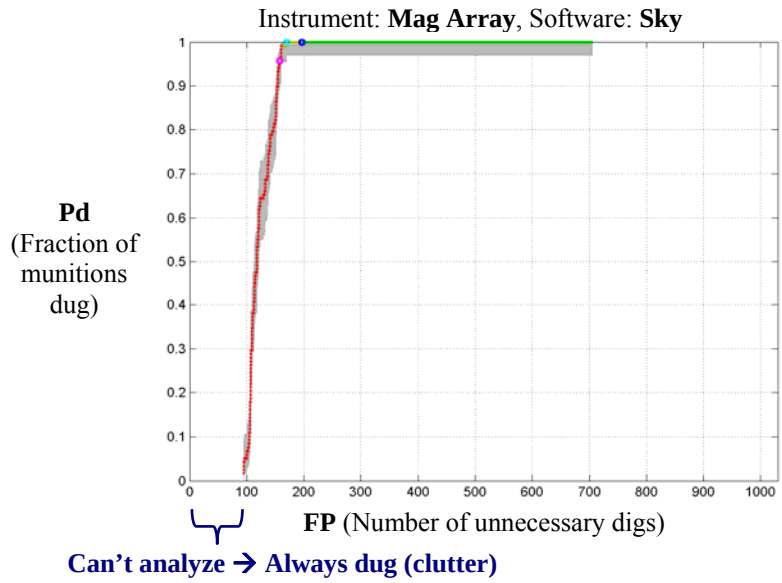


Figure 3.3: ROC curve for the Mag Array instrument and software used by Sky Research, Inc. The curve does not reach the lower left corner of ROC space because some items could not be analyzed. Sky could not analyze 97 locations, fewer than the corresponding number for SAIC, shown in Figure 3.2.

Table 3.2 compares the number of anomalies detected with the Mag Array that could and could not be analyzed by SAIC versus Sky Research, Inc. Either SAIC's criteria for determining that an anomaly could be analyzed were more conservative than Sky's criteria or there is a difference in performance related to anomaly data selection

and inversion. While 16% of all detected anomalies could be analyzed by Sky but *not* by SAIC, only 4% of all detected anomalies could be analyzed by SAIC but *not* by Sky.

Table 3.2: Comparing the number of anomalies detected with the Mag Array that SAIC and Sky Research, Inc. could and could not analyze. 16% of all detected anomalies could be analyzed by Sky Research, Inc. but *not* by SAIC, while only 4% of all detected anomalies could be analyzed by SAIC but *not* by Sky Research, Inc.

Mag Array		Sky		
		Can Analyze	**Can*not* Analyze**	**Total**
SAIC	**Can Analyze**	595 (72%)	31 (4%)	626 (76%)
	Can*not* Analyze	132 (16%)	66 (8%)	198 (24%)
	Total	727 (88%)	97 (12%)	824 (100%)

Along with SAIC and Sky, SIG also processed data collected by the Mag Array instrument. Tables 3.3 and 3.4 compare the number of detected anomalies that could and could not be analyzed by SIG versus SAIC and Sky, respectively. A larger percentage of anomalies could be analyzed by SAIC or Sky but *not* by SIG, compared to the percentage of anomalies that could be analyzed by SIG but *not* by SAIC or Sky. Appendix C shows similar tables for the other instruments used in this study.

Table 3.3: Comparing the number of anomalies detected with the Mag Array that SAIC and SIG could and could not analyze. 17% of all detected anomalies could be analyzed by SAIC but *not* by SIG, while only 1% of all detected anomalies were vice versa.

Mag Array		SIG		
		Can Analyze	**Can*not* Analyze**	**Total**
SAIC	**Can Analyze**	485 (59%)	141 (17%)	626 (76%)
	Can*not* Analyze	7 (1%)	191 (23%)	198 (24%)
	Total	492 (60%)	332 (40%)	824 (100%)

Table 3.4: Comparing the number of anomalies detected with the Mag Array that Sky Research, Inc. and SIG could and could not analyze. 31% of all detected anomalies could be analyzed by Sky Research, Inc., but *not* by SIG, while only 1% was vice versa.

Mag Array		SIG		
		Can Analyze	Can*not* Analyze	Total
Sky	Can Analyze	473 (57%)	254 (31%)	727 (88%)
	Can*not* Analyze	19 (2%)	78 (9%)	97 (12%)
	Total	492 (60%)	332 (40%)	824 (100%)

Finding 5: Once "Can't analyze" locations were dug, discrimination performance was typically very good for all remaining locations.

As discussed above, "Can't analyze" locations must always be labeled as "Dig," regardless of the dig threshold, because of the large safety hazard of leaving a munition in the ground. All other locations, however, can be analyzed by a discrimination algorithm and labeled as "Dig" or "Do not dig" based on the algorithm's output. For a large majority of the different instrument/algorithm combinations tested in this study, discrimination performance was good for those locations that could be analyzed. That is, the demonstrator's dig threshold led to a large reduction in FP while Pd remained at or near 1.00.

For example, Figure 3.4 shows the ROC curve for locations associated with the EM61 Array and discriminated by a multidimensional classifier.[4] The ROC curve does not reach the lower left corner of ROC space since some locations were labeled as "Can't analyze." Of the 119 munition locations, 8 (7%) could not be analyzed. Therefore, the minimum Pd is 0.07, rather than 0. Because 285 of the "Can't analyze" items were clutter, the minimum FP is 285, rather than 0.

[4] Sky Research, Inc. performed the discrimination analysis.

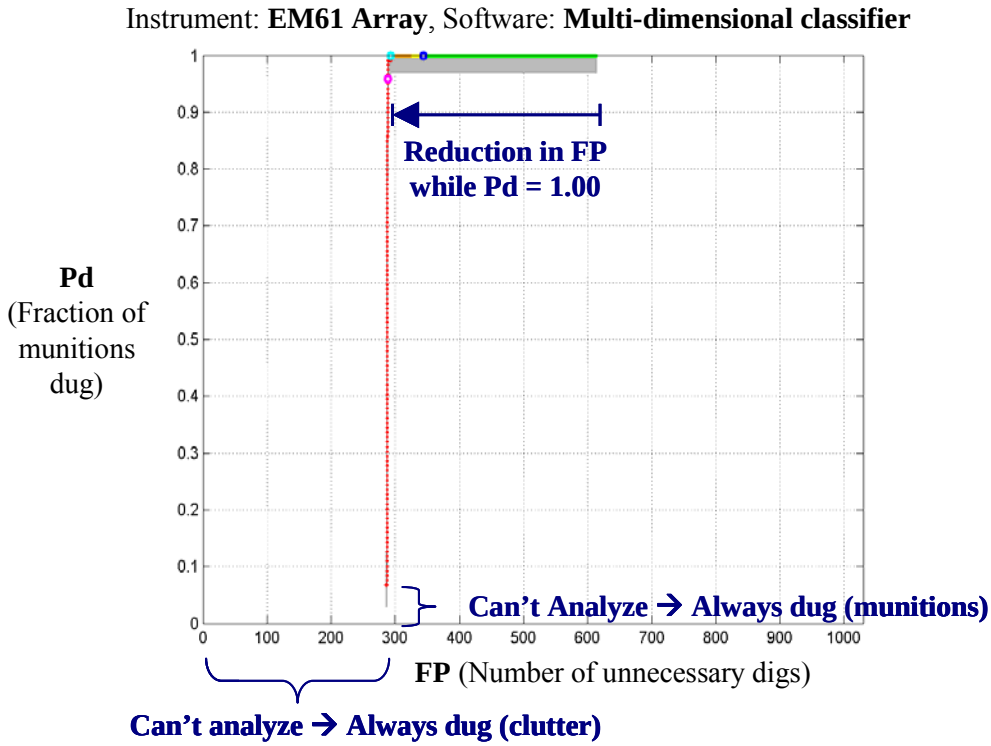


Figure 3.4: ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. The curve does not reach the lower left corner of ROC space because some locations could not be analyzed. The curve exhibits a perfect right angle. Furthermore, the demonstrator's dig threshold (dark blue dot) led to a reduction in the number of FPs by 271 while the probability of detection (Pd) remained at 1.00. An adjusted threshold (light blue dot) would have led to an even larger reduction in FPs while Pd remained at 1.00.

Had no discrimination been performed, all 734 locations associated with the EM61 Array instrument would have been labeled "Dig." Because 615 of these locations were clutter, the maximum FP is 615. Yet discrimination *was* performed. In fact, the demonstrator's dig threshold (dark blue dot) reduced the number of FPs from 615, the maximum possible, to 344, near the minimum possible. Thus, even though some locations could not be analyzed by the discrimination algorithm, use of the algorithm on the remaining locations reduced the number of FPs by 271.

We can revisit the choice of dig threshold. With knowledge of ground truth, the dig threshold could have been adjusted to reduce FPs even further while maintaining a Pd of 1.00 (light blue dot). By doing so, the discrimination algorithm would have performed even better. The perfect right angle of the ROC curve is further evidence of the algorithm's perfect ability to discriminate between munition and clutter locations. In this case, once the dig threshold is adjusted to achieve the largest possible reduction in FP while maintaining a Pd of 1.00 (light blue dot), the threshold *cannot* be adjusted further

to achieve an even greater reduction in FP, even at the expense of allowing Pd to drop to 0.95 (pink dot). This occurs because there is a complete lack of overlap in multidimensional space between the discriminating features extracted from the clutter and the munition locations.

3.2.2 Commercially available or production instruments and software

One commercially available instrument (EM61 Cart) and two custom-built platforms with commercial sensors (Mag Array) or modified commercial sensors (EM61 Array) were used to survey the site. Different demonstrators used different software to discriminate the locations on the master list associated with each of these instruments. Some demonstrators performed the discrimination using a simple one-dimensional analysis of a *single* feature extracted from the locations during data inversion. Other demonstrators used multidimensional classifiers to discriminate the locations. A variety of different software was used, including some that is available commercially. An off-the-shelf form of UXAnalyze was used, as well as a version extended with IDL routines.

Finding 6: Commercially available and production instruments and software provided good discrimination performance.

Figures 3.5 and 3.6 show ROC curves for locations associated with the Mag Array and EM61 Array, respectively. SAIC performed the discrimination analysis using UXAnalyze software. In both cases, the demonstrator's dig threshold (dark blue dot) led to a large reduction in FP while Pd remained at 1.00. In fact, in the case of the Mag Array, analysis shows that the demonstrator's dig threshold was almost optimal. Even with knowledge of ground truth, the dig threshold could not have been adjusted (light blue dot) to reduce FP much further while maintaining a Pd of 1.00. In contrast, the dig threshold for the EM61 Array could be adjusted retrospectively (light blue dot) to give an even larger reduction in FP while Pd remains at 1.00. Furthermore, in each example, the ROC curve exhibits a sharp angle, evidence that the algorithm used by the software has high discriminating power. That is, the discriminating features extracted from the clutter locations overlap little in multidimensional space with the features extracted from the munition locations.

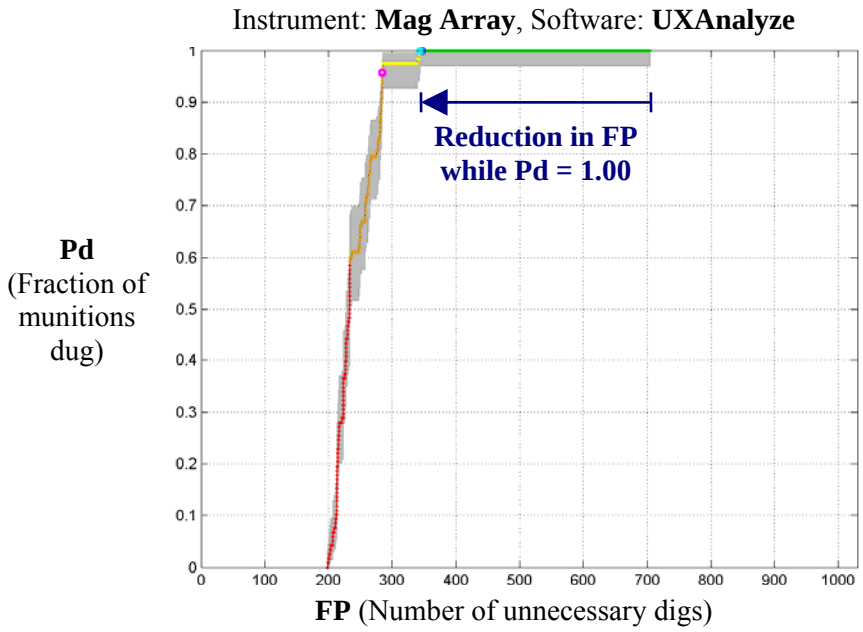


Figure 3.5: ROC curve for the Mag Array instrument and the UXAnalyze software. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs while the probability of detection (Pd) remained at 1.00. An adjusted threshold (light blue dot) could not have led to a much larger reduction in FP while Pd remained equal to 1.00.

Figure 3.7 shows a ROC curve for locations associated with the EM61 Cart instrument. Parsons (the commercial contractor hired by the Program Office to emplace seeds, collect data using the EM61 Cart instrument, and excavate all locations on the master list) performed the discrimination analysis using UXAnalyze software. The ROC curve shows a smaller reduction than many earlier examples in FP while Pd remained at 1.00. However, it is likely that the reduction in FP was smaller only because the maximum number of FPs was already quite low. As shown in Table 3.1, the Mag Array and EM61 Array had maximum FP values of 706 and 615, respectively, but the EM61 Cart had a much lower maximum FP of 428. Furthermore, when Pd was constrained to 1.00, the EM61 Cart's ROC curve showed a maximum FP of 298, lower than that shown by the Mag Array's ROC curve. Thus, commercially available instruments and software employed by commercial contractors can lead to good discrimination performance.

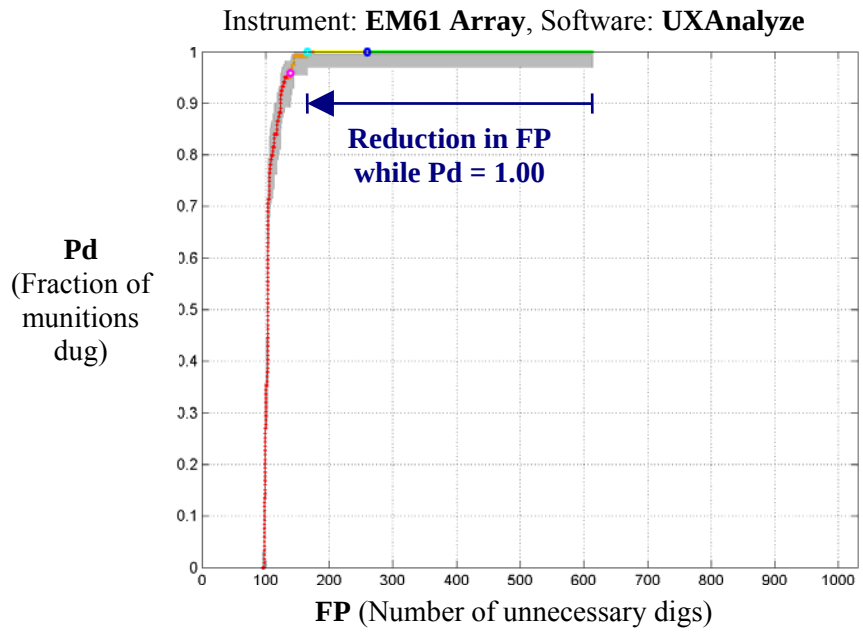


Figure 3.6: ROC curve for the EM61 Array instrument and the UXAnalyze software. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs while the probability of detection (Pd) remained at 1.00. An adjusted threshold (light blue dot) would have led to an even larger reduction in FP while Pd remained equal to 1.00.

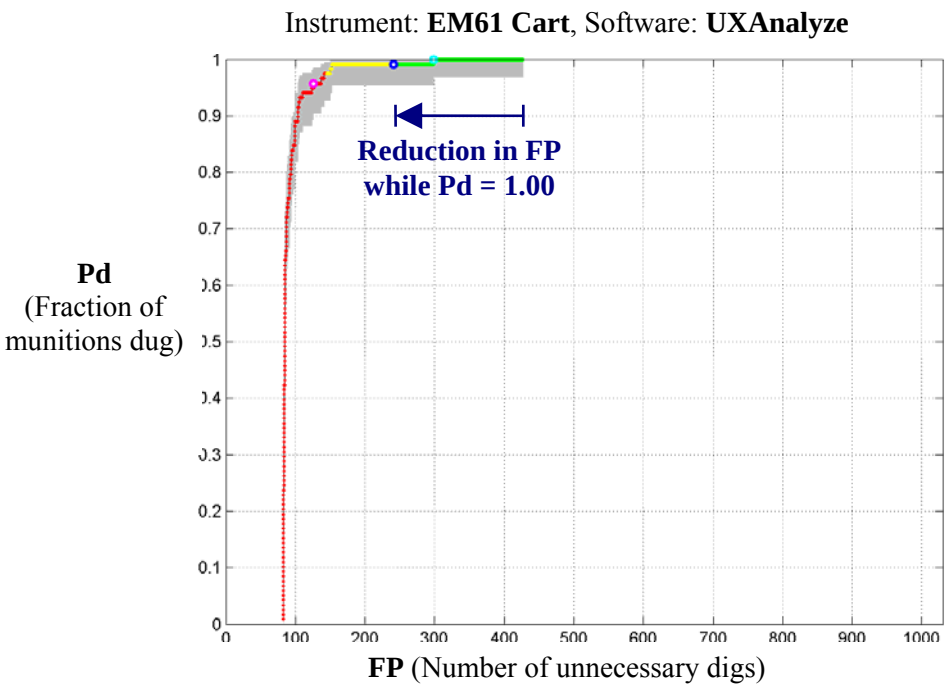


Figure 3.7: ROC curve for the EM61 Cart instrument and the UXAnalyze software. Commercial contractors performed the discrimination analysis. Their dig threshold (dark blue dot) led to a reduction in FPs while Pd remained near 1.00 (with a 95% confidence interval that includes 1.00). The reduction in FP is measured with respect to the maximum FP of 428. It is low compared with the maximum FP values shown in Figures 3.5 and 3.6.

Finding 7: For survey instruments, cooperative inversions led to a slightly lower number of unnecessary digs.

Figure 3.8 shows another ROC curve for locations associated with the EM61 Array. In this case, a multidimensional classifier was used to discriminate clutter versus munitions.[5] The ROC curve shows a large reduction in FP while Pd remains at 1.00. The ROC curve also shows a perfect right angle, indicating the algorithm's perfect ability to discriminate between clutter and munitions.

Instrument: **EM61 Array**, Software: **Multi-Dimensional Classifier**

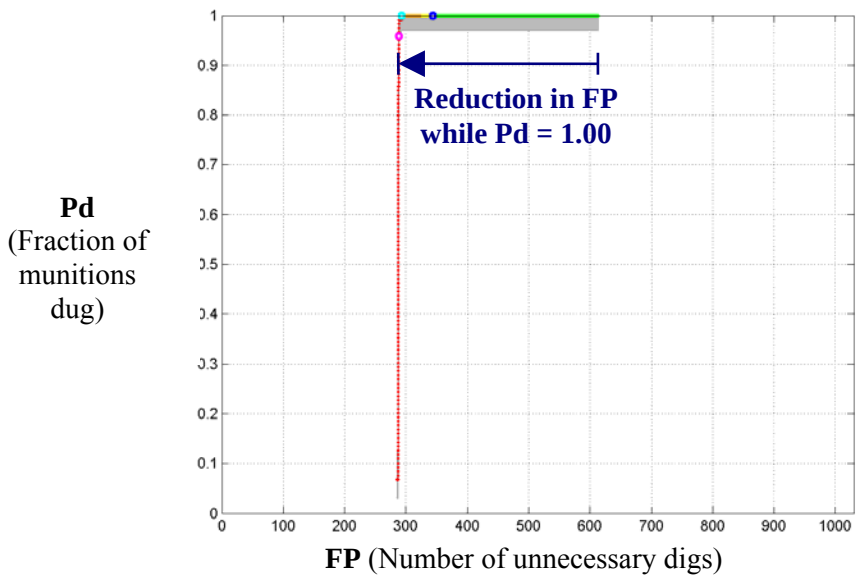


Figure 3.8: ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs while Pd remained at 1.00. An adjusted threshold (light blue dot) would have led to an even larger reduction in FPs while Pd remained at 1.00.

Figure 3.9 shows a similar ROC curve for locations associated with *either* the EM61 Array *or* the Mag Array. In this analysis, the demonstrators performed cooperative inversions for every location associated with *both* the EM61 Array *and* the Mag Array. In contrast, for every location associated with the EM61 Array only, the demonstrators inverted the EMI data using the unconstrained EMI model. Similarly, for every location associated with the Mag Array only, the demonstrators inverted the magnetometer data using the unconstrained magnetometer model.[6] Figures 3.8 and 3.9 show that cooperative

---

5 Sky Research, Inc. performed the discrimination analysis.

6 Ibid.

inversions led to an even larger reduction in FP while Pd remained at 1.00, compared with the EMI inversions alone. The reduction in FPs was much larger only because the maximum FP was much higher, however. The EM61 Array alone had a maximum FP of 615, but the cooperative inversion case had a much higher maximum FP of 862. But when Pd was constrained to 1.00, the EM61 Array alone had a maximum FP of 293, and the maximum FP for the cooperative inversion case was only slightly lower, at 230.

Instrument: **EM61 Array + Mag Array**, Software: **Multi-Dimensional Classifier**

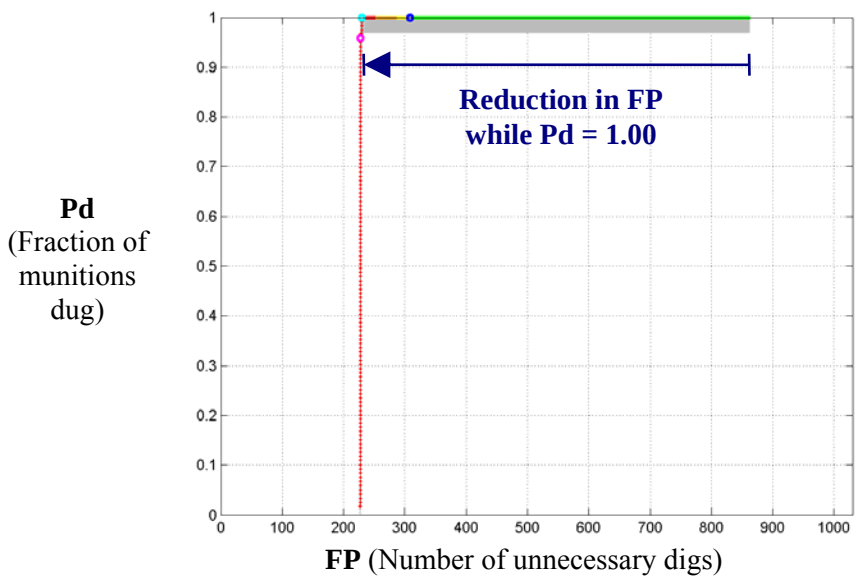


Figure 3.9: ROC curve for cooperative inversions based on the EM61 Array and Mag Array instruments and software based on a multidimensional classifier. Results show that while Pd is constrained to 1.00, cooperative inversions led to a lower number of unnecessary digs than inversions based on the unconstrained EMI model shown in Figure 3.8.

Finding 8: Much of the discriminating power seen at Camp Sibert is due to size-based features.

Historical records of Camp Sibert indicated that the only likely munition in the ground would be the 4.2″ mortar. This is a large item compared with much typical clutter, providing ample opportunity for discrimination algorithms to demonstrate their ability to reduce FP while maintaining a high Pd. As expected, only the 4.2″ mortar was found at the site, and this munition was indeed much larger than most of the surrounding clutter. Thus, size alone was a powerful discriminating feature at Camp Sibert.

Figures 3.10 and 3.11 show ROC curves of locations associated with the EM61 Array.[7] In each case, the maximum FP is 615, which is the number of clutter locations on the master list associated with the EM61 Array. The same demonstrator also used the same criteria to label locations as "Can't analyze" before applying the discrimination algorithms. The minimum FP and Pd are therefore the same in each case because the minimum FP is the total number of clutter locations labeled as "Can't analyze," and the minimum Pd is the fraction of munition locations labeled as "Can't analyze."

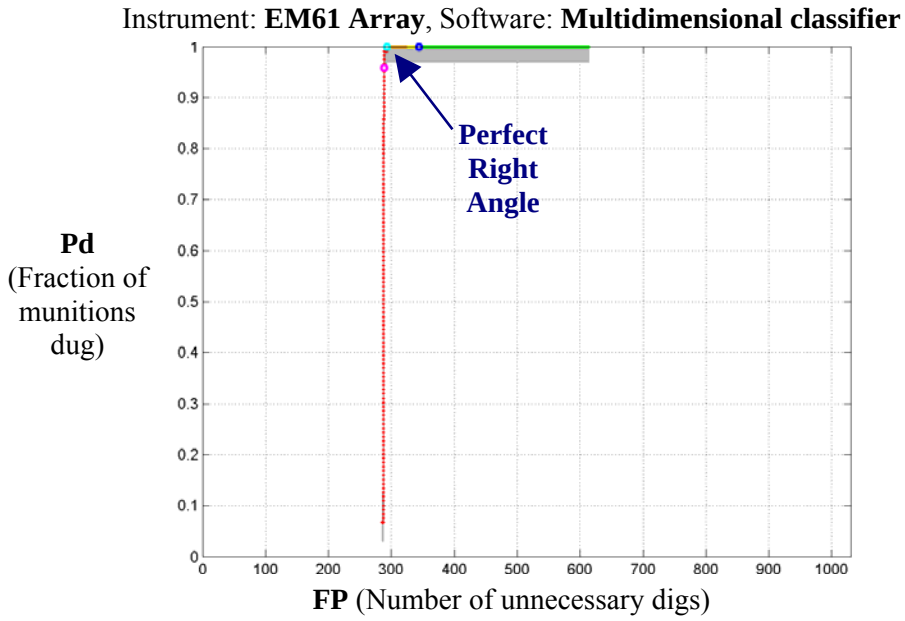


Figure 3.10: ROC curve for the EM61 Array and software based on a multidimensional classifier. The ROC curve exhibits the same maximum FP and Pd values as in Figure 3.11, since both ROC curves were based on locations associated with the EM61 Array. Similarly, the ROC curve exhibits the same minimum FP and Pd values as in Figure 3.11, since both ROC curves were based on the same demonstrator's definition of "Can't analyze." Unlike Figure 3.11, however, the curve exhibits a perfect right angle.

---

7 Sky Research, Inc. performed the discrimination analysis.

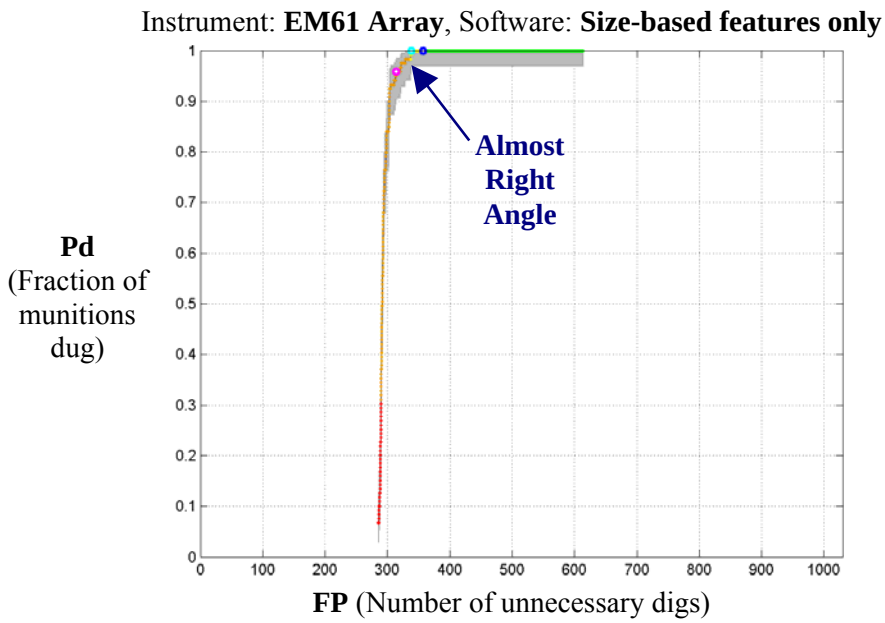


Figure 3.11: ROC curve for the EM61 Array and software based on size-based features only. The ROC curve exhibits the same maximum FP and Pd values as in Figure 3.10, since both ROC curves were based on locations associated with the EM61 Array. Similarly, the ROC curve also exhibits the same minimum FP and Pd values as in Figure 3.10, since both ROC curves were based on the same demonstrator's definition of "Can't analyze." Although the curve does not exhibit a perfect right angle, as in Figure 3.10, its angle is very sharp.

The two ROC curves differ only in shape: The ROC curve in Figure 3.10 exhibits a perfect right angle, while the ROC curve in Figure 3.11 exhibits a sharp, but not perfect, right angle. The difference in shape is due to the difference in discrimination algorithms. In Figure 3.10, locations were discriminated using a multidimensional classifier. The perfect right angle is evidence that the features extracted from the clutter and those extracted from munitions show no overlap in multidimensional space. In contrast, the locations in Figure 3.11 were discriminated based on size only. Of the parameters estimated from the data collected at these locations, only one parameter, the principal polarizability at the first time gate, which is related to target size, was used for discrimination. Although the ROC curve does not exhibit a perfect right angle, its angle is very sharp. The sharp angle is evidence that the single, size-based feature shows little overlap in one-dimensional space between clutter and munitions.

Figure 3.12 shows a one-dimensional histogram of this single discriminating feature extracted from clutter and munitions. Most munitions exhibit a polarizability greater than 400, and most clutter items exhibit a polarizability less than 400. Thus, a dig threshold set in the vicinity of 400 could label most munitions as "dig," leading to a high Pd, while labeling most clutter items as "Do not dig," leading to a low FP. Due to the

large safety hazard associated with leaving a munition in the ground, demonstrators chose to be conservative and set their dig threshold lower than 400, resulting in a smaller reduction in FP but a higher Pd.

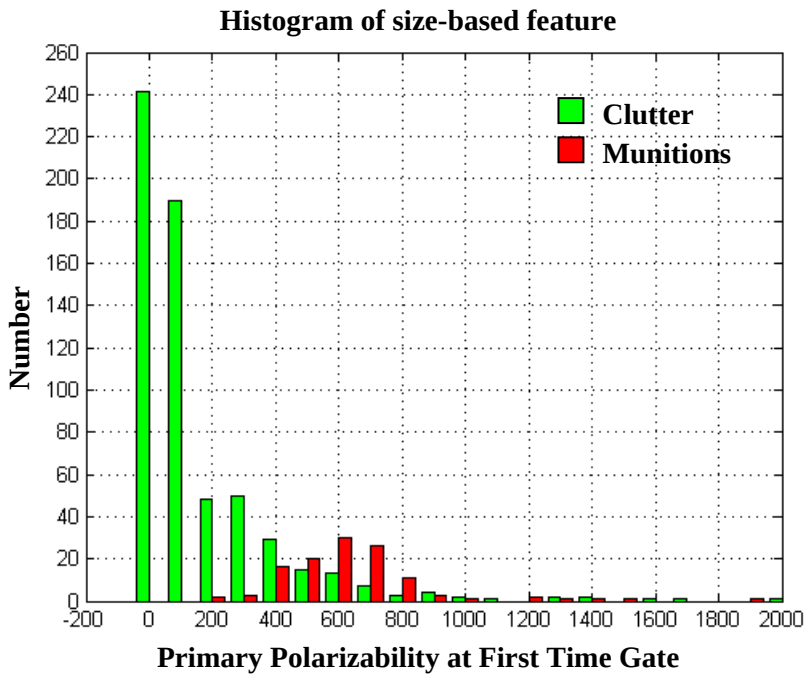


Figure 3.12: Histograms of a sized-based feature extracted from clutter (green) and munitions (red). The feature is calculated as the principal polarizability at the first time gate of the EM61 Array. Most munitions are larger than most clutter items.

Finding 9: Mag-and-flag led to a large number of unnecessary digs.

The performance of the M&F operator was compared to the Mag Array instrument in conjunction with discrimination algorithms. Figure 3.13 shows a ROC curve generated from locations on the master list associated with the Mag Array and discriminated using UXAnalyze. Only those locations within the 100′ × 100′ grid in which the M&F process took place are represented in the curve. Many fewer locations are represented in this curve than in the figures shown so far, and the resolution of the curve is much coarser. Furthermore, as fewer munition locations are represented in this curve, the 95% confidence intervals around Pd are very wide. The performance of the M&F process is superimposed on this curve (gray dot).

The Mag Array with UXAnalyze performed much better than the M&F process. The M&F operator detected all four munitions in the 100′ × 100′ grid, along with 45 clutter items. In contrast, the Mag Array detected 39 anomalies within the same grid, 27 of which resulted in locations on the master list (the remaining 12 anomalies were labeled

as "clustered" and not included in the scoring process). Of these 27 locations, 4 were munitions and the remaining 23 were clutter items. Application of the demonstrator's dig threshold led to a reduction in FP from 23 to 3. Analysis shows that the dig threshold could have been adjusted to eliminate all unnecessary digs save one while Pd remained at 1.00. These results confirm the results of previous work, which showed that M&F leads to a very large number of false positives [13].

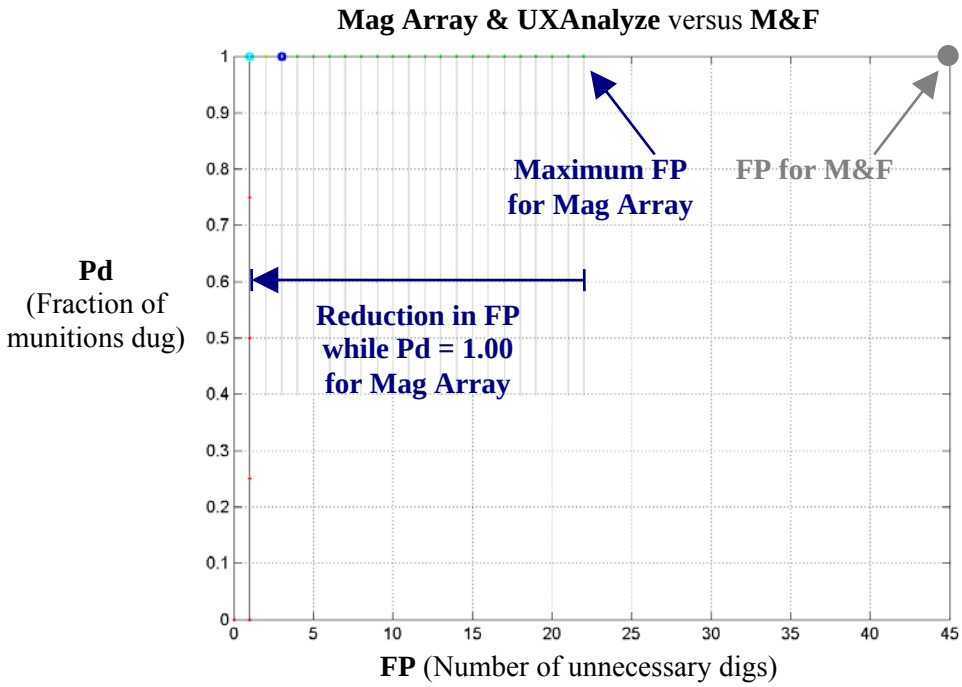


Figure 3.13: ROC curve for the Mag Array and the UXAnalyze software. The ROC curve is generated over only those locations in the master list associated with the Mag Array within the 100′ x 100′ grid in which the M&F process was done. The Pd and FP resulting from the M&F process is shown as a large gray dot. Discrimination with the Mag Array resulted in fewer unnecessary digs than the M&F process.

3.2.3 Frequency-domain EMI instruments

Two EMI instruments, the GEM Array and the GEM Cued, collected data in the frequency domain, rather than the time domain. Although both were frequency-domain EMI instruments, there were two large differences between them. First, the GEM Array was built on the same type of towed vehicle platform as the Mag Array and EM61 Array while the GEM Cued was a hand-held instrument. Second, the GEM Array collected data in survey mode, while the GEM Cued collected data in cued mode.

Finding 10: The GEM Array and SAIC custom software led to good discrimination performance.

Figure 3.14 shows the ROC curve for data collected by the GEM Array and discriminated using custom software developed by SAIC and written in IDL. The maximum FP is quite high. The GEM Array was a very noisy instrument, particularly in the SW area. However, the demonstrator's dig threshold led to a large reduction in unnecessary digs while Pd remained near 1.00 (with a 95% confidence interval that includes 1.00.) Our analysis shows that even if Pd had been constrained to 1.00, the dig threshold could be adjusted (light blue dot) to still lead to a large reduction in FP.

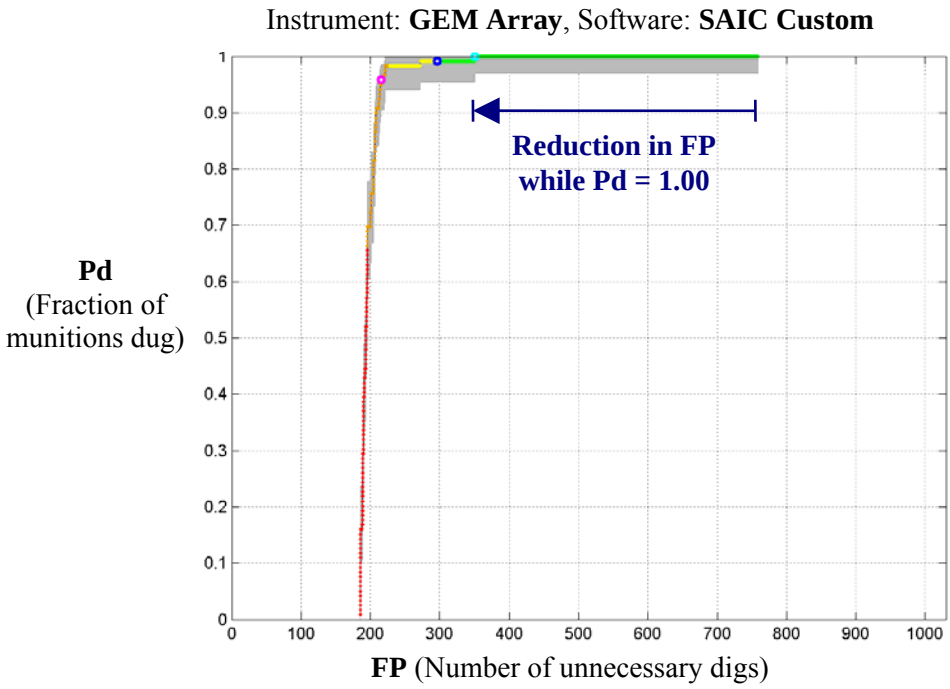


Figure 3.14: ROC curve for the GEM Array and SAIC custom software. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs while Pd remained near 1.00 (with a 95% confidence interval that includes 1.00). An adjusted threshold (light blue dot) would have also led to a large reduction in FP even when Pd remained equal to 1.00.

Finding 11: High-density, cued GEM data had some discriminating power, but led to a large number of unnecessary digs, even with cooperative inversions.

Figure 3.15 shows the ROC curve for data collected by the GEM Cued instrument and discriminated with a multidimensional classifier.[8] The shape of the ROC curve

[8] Signal Innovations Group performed the discrimination analysis.

suggests some discriminating power because the curve exhibits a large area underneath, remaining above the dashed line indicating the theoretical 50–50 chance of correct discrimination (i.e., a coin flip). Despite this discriminating power, the demonstrator's dig threshold (dark blue dot) led to only a small reduction in FP while Pd remained at 1.00. The poor performance of the GEM Cued can likely be attributed to the manner in which the data were collected. During data collection, the instrument was placed directly on the soil (see Figure 2.18), and the received signal was contaminated by a large in-phase component from the ground response.

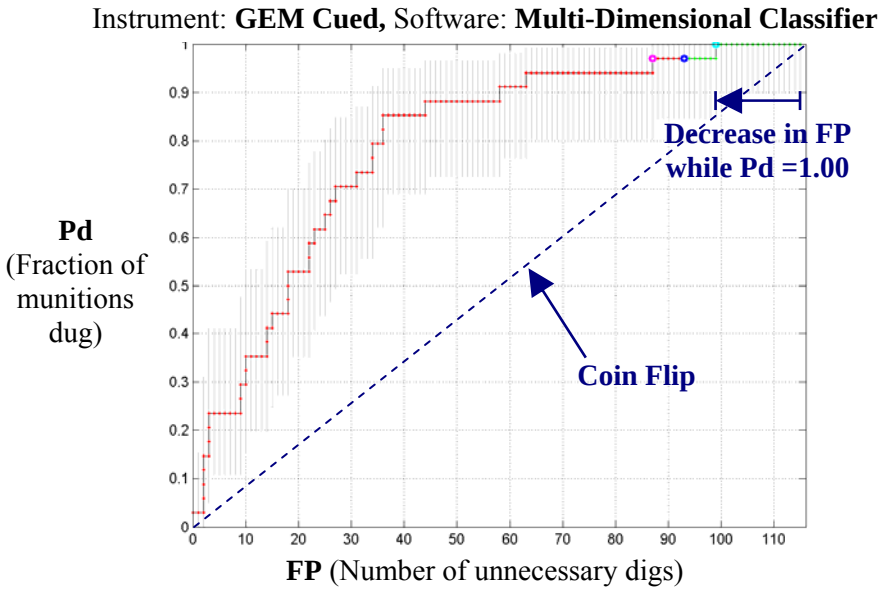


Figure 3.15: ROC curve for the GEM Cued and software based on a multidimensional classifier. The ROC curve lies above the dashed line that indicates the theoretical 50-50 chance of correct discrimination (i.e., a coin flip). However, the demonstrator's dig threshold (dark blue dot) led to a small reduction in FPs. An adjusted threshold (light blue dot) would have led to an even smaller reduction in FP while Pd was equal to 1.00.

Cooperative inversions were not performed on the GEM Cued data, but one demonstrator formed feature vectors using the outputs from independent inversions provided by the GEM cued and Mag Array sensors. For each cued location associated with the Mag Array, the demonstrators performed EMI inversions on the GEM Cued data and independently performed magnetometer inversions on the Mag Array data. The demonstrators then formed a feature vector from both EMI and magnetometer-related parameters. In contrast, for each cued location not associated with the Mag Array, the demonstrators performed EMI inversions only, and the feature vector contained only

EMI-related parameters. Figure 3.16 shows the ROC curve resulting from this analysis.[9] As with Figure 3.15, most of the ROC curve remains above the theoretical 50-50 chance of correct discrimination (dashed line). In this example, the demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs but only at the expense of a Pd significantly different from 1.00 (the 95% confidence interval does not include 1.00). Once again, our analysis shows that when Pd is constrained to 1.00, the reduction in FPs was extremely small.

Instrument: **GEM Cued + Mag Array**, Software: **Multi-Dimensional**

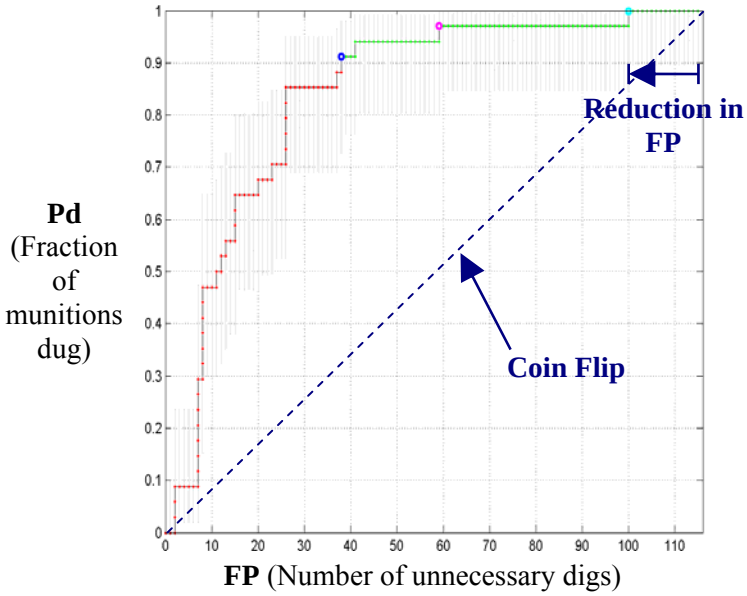


Figure 3.16: ROC curve for independent inversions based on the GEM Cued and Mag Array instruments with a multidimensional classifier. Most of the ROC curve lies above the dashed line that indicates the theoretical 50–50 chance of correct discrimination. The demonstrator-suggested dig threshold (dark blue dot) led to large reduction in FPs, but Pd was less than 1.00 (with a 95% confidence interval that does not include 1.00). An adjusted threshold (light blue dot) would have led to only a small reduction in FP while Pd was equal to 1.00.

3.2.4 Advanced instruments and software

Two advanced instruments were used at Camp Sibert. The EM63 Cued instrument collected data in cued mode, and the BUD instrument collected data in both cued and survey mode. As discussed above, a variety of software was used to discriminate the data.

---

9 Signal Innovations Group performed the discrimination analysis.

Some software was based on advanced techniques for optimizing the discrimination algorithms, such as semi-supervised learning or active learning.

Finding 12: High-density, cued EM63 data led to good discrimination performance, especially with cooperative inversions.

Figure 3.17 shows the ROC curve for data collected by the EM63 Cued and discriminated with a multidimensional classifier.[10] The demonstrator's dig threshold (dark blue dot) led to a large reduction in FP while Pd was near 1.00 (with a 95% confidence interval that includes 1.00). Analysis shows that with perfect hindsight, the dig threshold could have been adjusted (light blue dot) to give only a slightly smaller reduction in FPs while maintaining a Pd equal to 1.00. The excellent performance of the EM63 Cued is likely due to a number of factors. The sensor employs many more sample gates than the standard EM61-Mk2 sensor (26 versus 4), and the gates extend out much further in time than with the EM61-Mk2 (25 msec versus 1.3 msec). As the data are taken in cued mode, the instrument is pushed extremely slowly over each cued location, allowing time for data stacking, which results in a high SNR. Furthermore, the data were collected at a very high density using very careful geolocation, eliminating position-error noise.

Figure 3.18 shows a similar ROC curve, this time based on cooperative inversions.[11] The demonstrators performed cooperative inversions for every cued location associated with the Mag Array. In contrast, for every cued location not associated with the Mag Array, the demonstrators performed EMI-only inversions. The figure shows that the demonstrator's dig threshold (dark blue dot) led to a Pd of 1.00 and an even larger reduction in FP than was seen in Figure 3.17, in which EMI-only inversions were used. Adjusting the dig threshold (light blue dot) would have led to a further reduction in FP while maintaining a Pd equal to 1.00. In fact, the ROC curve exhibits a perfect right angle, indicating the algorithm's perfect discriminating ability.

---

[10] Sky Research, Inc. performed the discrimination analysis.

[11] Ibid.

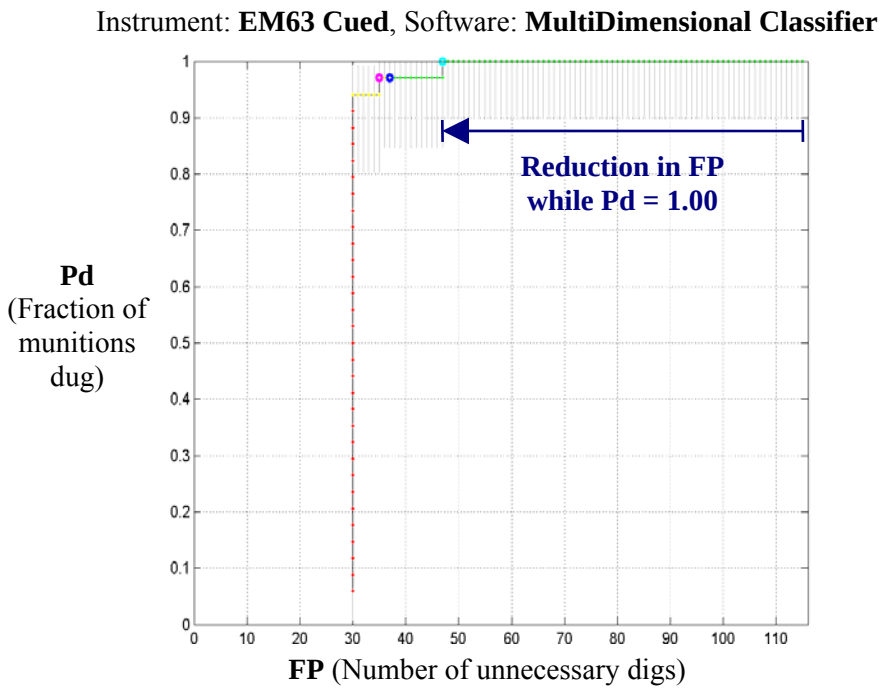


Figure 3.17: ROC curve for the EM63 Cued and software based on a multidimensional classifier. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs while Pd was near 1.00 (with a 95% confidence interval that includes 1.00). An adjusted threshold (light blue dot) would have led to only a slightly smaller reduction in FP while Pd remained equal to 1.00.

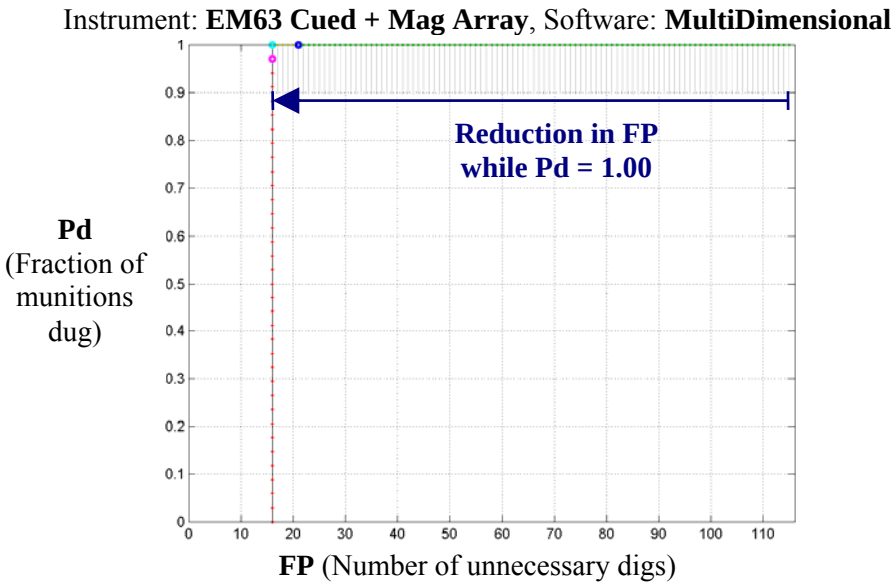


Figure 3.18: ROC curve for cooperative inversions based on the EM63 Cued and Mag Array instruments and software based on a multidimensional classifier. Results show that while Pd is constrained to 1.00, cooperative inversions led to an even larger reduction in FP than inversions based on the unconstrained EMI model, as shown in Figure 3.17.

Finding 13: The multiple-axis BUD instrument provided high-SNR data from a single location leading to excellent discrimination performance.

Figure 3.19 shows the ROC curve for data collected with the BUD instrument in cued mode and discriminated using a multidimensional template matcher.[12] Data from all cued items could be analyzed, so the minimum FP and Pd values were both zero. (Note the small red dot at the point [FP = 0, Pd = 0.00] on the ROC curve.) The demonstrator-suggested dig threshold (dark blue dot) led to a very large reduction in FP while Pd remained at 1.00. The dig threshold could have been adjusted even further to eliminate *all but one FP* while maintaining a Pd of 1.00. The ROC curve exhibits a perfect right angle, indicating the template matcher's perfect discriminating ability if the single FP object is discounted.

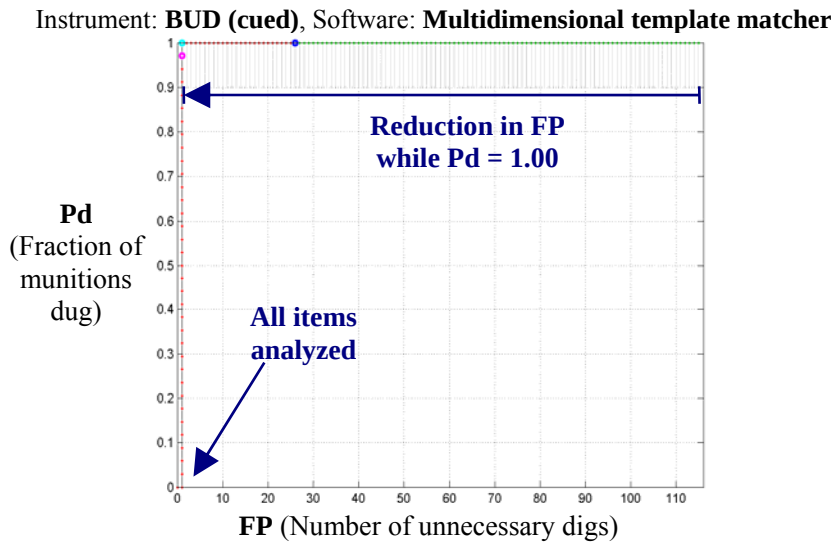


Figure 3.19: ROC curve for the BUD instrument in cued mode and software based on a multidimensional template matcher. The ROC curve reaches the lower left corner of ROC space because all locations were analyzed (note the existence of a red dot at FP = 0, Pd = 0.00). The demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs while Pd was 1.00. An adjusted threshold (light blue dot) could have eliminated all but one FP while Pd remained at 1.00. The curve exhibits a perfect right angle.

Figure 3.20 shows the ROC curve for locations on the master list that were associated with the BUD instrument in *survey* mode and discriminated using the same type of multidimensional template matcher.[13] Note that standard procedure for BUD in a survey mode is to declare a detection while moving and then to stop and collect

---

[12] Lawrence Berkeley National Laboratory performed the discrimination analysis.

[13] Ibid.

discrimination data. Thus, BUD operates as a cued sensor (albeit self-cued) even in survey mode.

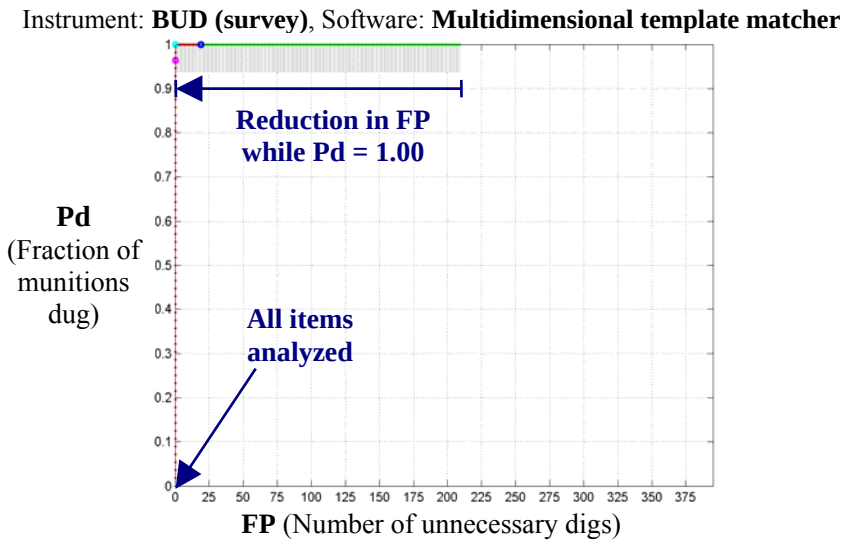


Figure 3.20: ROC curve for the BUD instrument in survey mode and software based on a multidimensional template matcher. Results are compiled over the SE1 area only because BUD surveyed only this area. The ROC curve reaches the lower left corner of ROC space because all locations were analyzed. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FPs while Pd was 1.00. An adjusted threshold (light blue dot) would have eliminated all FPs while Pd remained at 1.00. The curve exhibits a perfect right angle.

As the BUD instrument is under development, its operation is still slow. Therefore, the Program Office decided in advance that BUD would survey the SE1 area only. All locations on the master list associated with BUD could be analyzed, so the minimum FP and Pd values were both zero. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FP while Pd remained at 1.00—in fact, the dig threshold could have been adjusted further (light blue dot) to *eliminate all FPs* while Pd remained at 1.00. Once again, the ROC curve exhibits a perfect right angle, indicating perfect discriminating ability.

The excellent performance of the BUD is likely due to its more advanced design. Rather than having only one transmit and one receive coil, the BUD consists of three orthogonal transmit coils to provide strong illumination of the target in each axis and multiple receive coils to provide spatial diversity in the collected data. The illumination and receiver diversity mean that data do not have to be collected at multiple locations. Instead, data can be collected at a single point with a stationary platform to eliminate motion noise and allow for greater signal stacking. Furthermore, data from a single location can be inverted, reducing the inversion result's sensitivity to position-error noise.

Finding 14: The advantage to active learning could not be fully demonstrated at Camp Sibert.

One demonstrator used active learning to optimize the discrimination algorithms applied to inversions of both the EM61 Array and Mag Array data. For each location on the master list associated with *both* the EM61 Array *and* the Mag Array, the demonstrators performed EMI inversions on the EM61 Array data and independently performed magnetometer inversions on the Mag Array data. The demonstrators then formed a feature vector from both EMI and magnetometer-related parameters. In contrast, for each location on the master list associated with the EM61 Array but not the Mag Array, the demonstrators performed EMI inversions only, and the feature vector contained only EMI-related parameters. Similarly, for each location on the master list associated with the Mag Array but not the EM61 Array, the demonstrators performed magnetometer inversions only, and the feature vector contained only magnetometer-related parameters.

Figure 3.21 shows the ROC curve based on individual inversions from two sensors, joint feature vectors, and a multidimensional classifier.[14] The classifier was optimized over all labeled data in the Training Set (i.e., "initial" learning). The demonstrator's dig threshold (dark blue dot) resulted in a large reduction in FP while Pd remained at 1.00. Adjusting the dig threshold (light blue dot) could have resulted in a slightly larger reduction in FP while Pd remained at 1.00.

Figure 3.22 shows the ROC curve based on the same joint inversions and the same multidimensional classifier.[15] This time, however, the classifier was optimized using active-learning methods. The demonstrator's dig threshold (dark blue dot, hidden behind the light blue dot) led to a very similar reduction in FP (with Pd at 1.00) as was shown in Figure 3.21, in which active learning was not used. Adjusting the dig threshold (light blue dot, superimposed on the dark blue dot) could not have led to a further reduction in FP while Pd remained at 1.00.

---

[14] Signals Innovations Group performed the discrimination analysis.

[15] Ibid.

Instrument: **EM61 Array + Mag Array,** Software: **Supervised Learning**

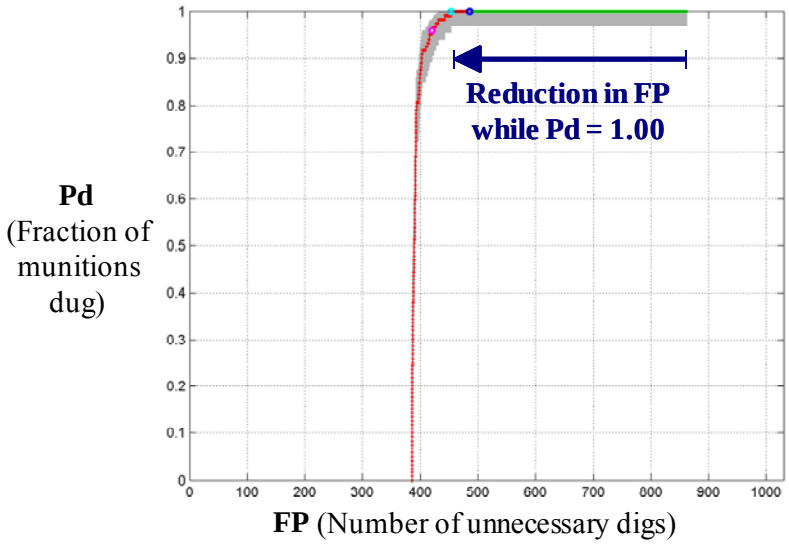


Figure 3.21: ROC curve for inversions based on the EM61 Array and Mag Array instruments and software based on a multidimensional classifier. The classifier was trained over all locations in the Training Set. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FP while Pd was 1.00; this reduction was similar to what is shown in Figure 3.22, in which active learning was used. The dig threshold could have been adjusted (light blue dot) to give a slightly larger reduction in FP while Pd remained at 1.00; this reduction was slightly larger than what is shown in Figure 3.22.

Instrument: **EM61 Array + Mag Array**, Software: **Active Learning**

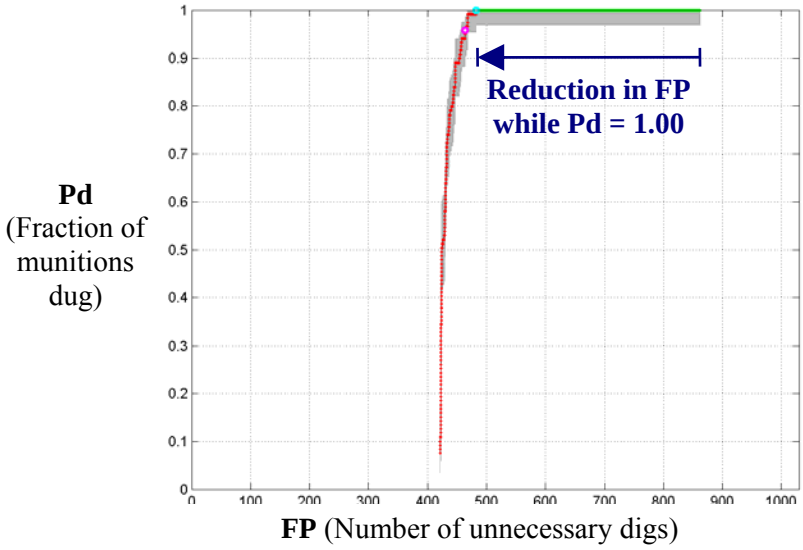


Figure 3.22: ROC curve for inversions based on the EM61 Array and Mag Array instruments and software based on a multidimensional classifier. The classifier was trained over locations that were actively chosen. The demonstrator's dig threshold (dark blue dot, hidden behind the light blue dot) led to a large reduction in FP while Pd was 1.00; this reduction was similar to what is shown in Figure 3.21, in which active learning was not used. Because the dig threshold could not have been adjusted (light blue dot, superimposed on the dark blue dot) to give any further reduction in FP while Pd remained at 1.00, the demonstrator's dig threshold was optimal.

Active learning and initial learning led to very similar discrimination results. This may be because Camp Sibert did not pose a large challenge to the initial learning algorithm, leaving the active learning algorithm little room to improve results. Future demonstrations at more challenging sites may be more informative as to the benefit of active learning in UXO discrimination problems. Note, however, that although active learning did not lead to improved discrimination performance, it did require a much smaller set of locations on which to train. The initial learning algorithm shown in Figure 3.21 used the approximately 200 locations in the Training Set for optimization. In contrast, the active-learning algorithm shown in Figure 3.22 used only 58 locations for optimization. Thus, in a real-world scenario, active learning may require the excavation of fewer locations (munitions and clutter) for algorithm optimization, resulting in fewer unnecessary digs in the training process.

Finding 15: The advantage to semi-supervised learning could not be demonstrated at Camp Sibert.

One demonstrator also used semi-supervised learning methods to optimize the discrimination algorithms. Figure 3.23 shows a ROC curve based on data collected by the EM61 Array instrument and discriminated using a multidimensional classifier.[16] The classifier was optimized using traditional supervised learning over the labeled Training Set. The demonstrator's dig threshold (dark blue dot) led to a large reduction in FP while Pd remained near 1.00 (with a 95% confidence interval including 1.00). In fact, analysis shows that the demonstrator's dig threshold was almost optimal. Even in retrospect, the dig threshold could not have been adjusted (light blue dot) to reduce FP much further while maintaining a Pd of 1.00.

In contrast, Figure 3.24 shows a ROC curve based on data collected by the same instrument and discriminated using the same type of multidimensional classifier.[17] In this case, however, the classifier was optimized using semi-supervised learning, using labeled data from the Training Set *as well as* unlabeled data from the Test Set. The demonstrator's dig threshold (dark blue dot) led to only a very small reduction in FP with a Pd of 1.00. However, adjusting the dig threshold (light blue dot) led to a very similar

---

[16] Signals Innovations Group performed the discrimination analysis.

[17] Ibid.

reduction in FP (while Pd remained at 1.00) as what was seen in Figure 3.23, in which semi-supervised learning was not used.

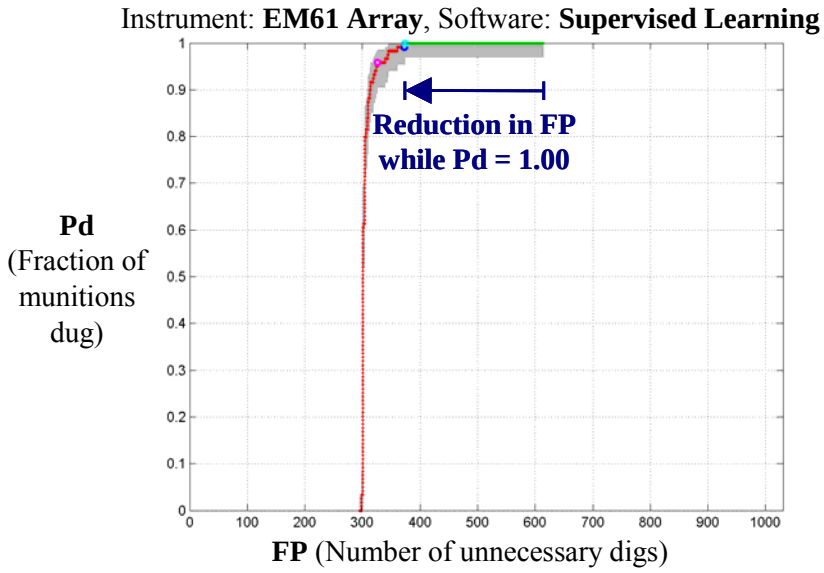


Figure 3.23: ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. The classifier was trained using a supervised learning protocol. Adjusting the dig threshold (light blue dot) could not have led to a much larger reduction in FP while Pd remained at 1.00. The demonstrator's dig threshold (dark blue dot) led to a much larger reduction in FP (while Pd remained near 1.00) than in Figure 3.22.

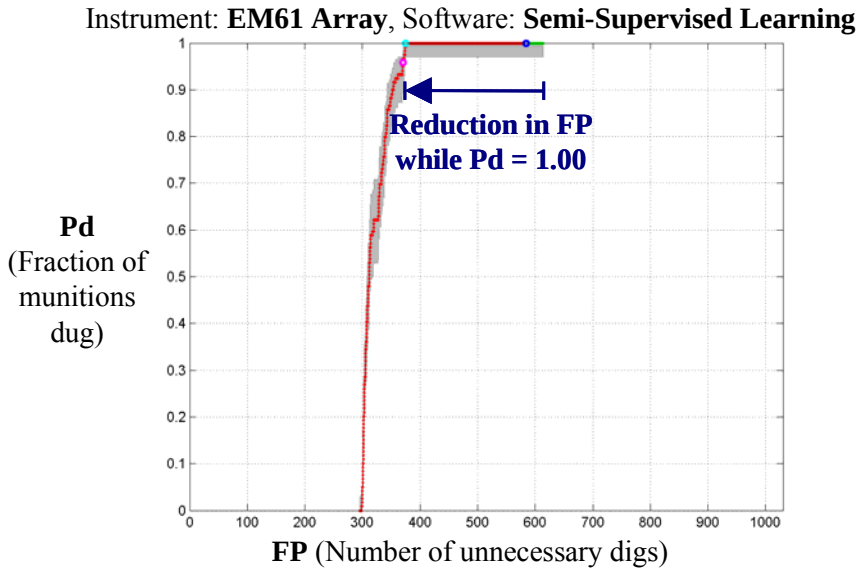


Figure 3.24: ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. The classifier was trained using a semi-supervised learning protocol. Adjusting the dig threshold (light blue dot) could have led to a large reduction in FP while Pd remained at 1.00, similar to what is shown in Figure 3.21 in which supervised learning was used. In contrast, the demonstrator's dig threshold (dark blue dot) led to a much smaller reduction in FP (while Pd remained at 1.00) than in Figure 3.21.

Note that in both Figures 3.23 and 3.24, the demonstrator's dig thresholds (dark blue dots) were quantitatively chosen so that locations were labeled as: "Do not dig" if their probability of being clutter was greater than or equal to 99%. As was discussed in the demonstrator's interim report [5], semi-supervised learning leads to very conservative estimates of a location's probability of containing clutter. Thus, in the semi-supervised learning case of Figure 3.24, very few locations exhibited estimated probabilities of being clutter that were greater than 99%. Therefore, very few probabilities were labeled "Do not dig," and as a result, many clutter items were dug unnecessarily. This explains why Figure 3.24 shows such a much smaller reduction in FP when the demonstrator's dig threshold is applied to the ranked dig list.

Thus semi-supervised learning did not lead to better results compared with the supervised learning approach. Once again, this may be because Camp Sibert did not pose a large challenge to the supervised algorithm, leaving the semi-supervised algorithm little room to improve. Future demonstrations at more challenging sites may provide more information on the benefit of semi-supervised learning in UXO discrimination problems.

3.2.5 Dig Threshold

As was shown in Figures 3.23 and 3.24, one demonstrator quantitatively selected dig thresholds such that locations were labeled as "Do not dig" if their probability of being clutter was greater than or equal to a threshold. This was equivalent to setting a threshold on the cost ratio comparing the cost of leaving a munition in the ground to the cost of unnecessarily digging a clutter item.

Finding 16: In some cases, a higher confidence in digging munitions could be achieved with only a few more unnecessary digs when using quantitative methods to set the dig threshold.

In theory, *any* estimate of the probability that a location is clutter can be used in conjunction with the cost-ratio equation described above. In this study, the demonstrator estimated the probabilities *quantitatively* using a discrimination algorithm. Note that in theory, the probability can be estimated in any manner: quantitatively (as done in this study), subjectively (using expert knowledge or a priori information taken from historical records), or even randomly. However, a subjectively or randomly estimated probability is likely to produce results that are neither as accurate nor as precise as a quantitatively estimated probability. Furthermore, since some discrimination algorithms are more suited to the UXO discrimination problem at this site than other algorithms, one algorithm may produce more accurate or precise quantitative estimates than another.

Conversely, any method to select the dig threshold can be used in conjunction with discrimination algorithms that quantitatively estimate the probability or likelihood that a location is clutter. Three demonstrators in this study used subjective methods to select the dig threshold, and one demonstrator used the cost-ratio equation to select the dig threshold quantitatively.

In either case, ROC curves allow us to judge both (1) the ability of an algorithm to estimate the probabilities or likelihoods than a location is clutter and (2) the ability of a selected dig threshold to classify the estimated probabilities/likelihoods. While the *shape* of a ROC curve (i.e., the sharpness of its angle, the area under its curve, the degree to which is lies above the 50–50 chance line, etc) depends upon the performance of the algorithm, the *color* of the ROC curve depends upon the suitability of the selected dig threshold. When generating a ROC curve, the dig threshold is stepped over the dig list, and Pd versus FP is plotted for each value of the dig threshold. Since no particular dig threshold has yet been chosen, the *shape* of the ROC curve is based solely on the probability estimates on which the curve is based or, rather, on the ability of the discrimination algorithm to estimate those probabilities accurately and precisely. In contrast, once a particular dig threshold has been chosen, the threshold is plotted on the ROC curve as a dark blue dot and the segment of the ROC curve to the upper right of the dot is colored in green. Thus, the *color* of the ROC curve is based on the selection of dig threshold.

As discussed above, Figure 3.23 shows the ROC curve for a demonstrator's dig threshold based on a 99% probability that a location is clutter.[18] The locations were associated with the EM61 Array instrument and discriminated based on a multidimensional classifier optimized using supervised learning. In comparison, Figures 3.25 and 3.26 show the very same ROC curves, this time using demonstrator's dig thresholds based on probabilities that were greater than or equal to 98% and 96%, respectively.[19] In all three figures, locations were associated with the same EM61 Array instrument; therefore, the maximum FP value is identical. Furthermore, locations were labeled as "Can't analyze" using the same criteria; therefore, the minimum FP and Pd values are also identical. Finally, those locations that could be analyzed were arranged

---

[18] Signal Innovations Group performed the discrimination analyses.

[19] Ibid.

into a ranked dig list based on the same multidimensional classifier optimized using the same supervised learning approach; therefore, the shape of the ROC curve is identical for the three figures.

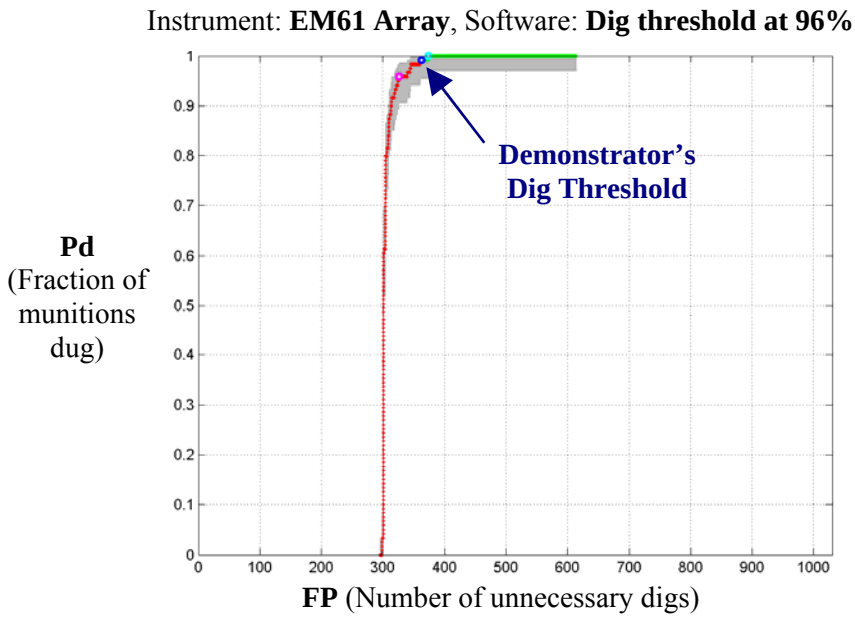


Figure 3.25: ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. The demonstrator's dig threshold (dark blue dot) was quantitatively chosen such that locations were labeled "Do not dig" if their probability of being clutter was greater than or equal to 96%. The shape of the ROC curve is identical to Figures 3.23 and 3.26, because all three figures are based on the same instrument and software. In contrast, the position of the demonstrator's dig threshold is slightly different than in Figures 3.23 and 3.26 because each of the three figures was based on different quantitative criteria for choosing the dig threshold.

The ROC curves in Figures 3.23, 3.25, and 3.26 differ *only* in the location of the dark blue dot, representative of the demonstrator's dig threshold, and the lengths of the segments colored in green. That is, as we raise the probability on which the dig threshold is based, the location of the demonstrator's dig threshold moves further to the upper right end of the ROC curve, indicating that fewer locations are labeled as "Do not dig." This happens because a location *must* exhibit a higher probability of being clutter before being labeled as "Do not dig," as the cost associated with leaving a munition in the ground is higher than the cost of unnecessarily digging a clutter item. However, although the dig thresholds in these three figures are not identical, they differ only slightly. In some cases, such as the example shown in the figures, so few locations had clutter probabilities between 96–98% and between 98–99% that the results of the discrimination differed little, regardless of which cost ratio was used to quantitatively select the dig threshold.

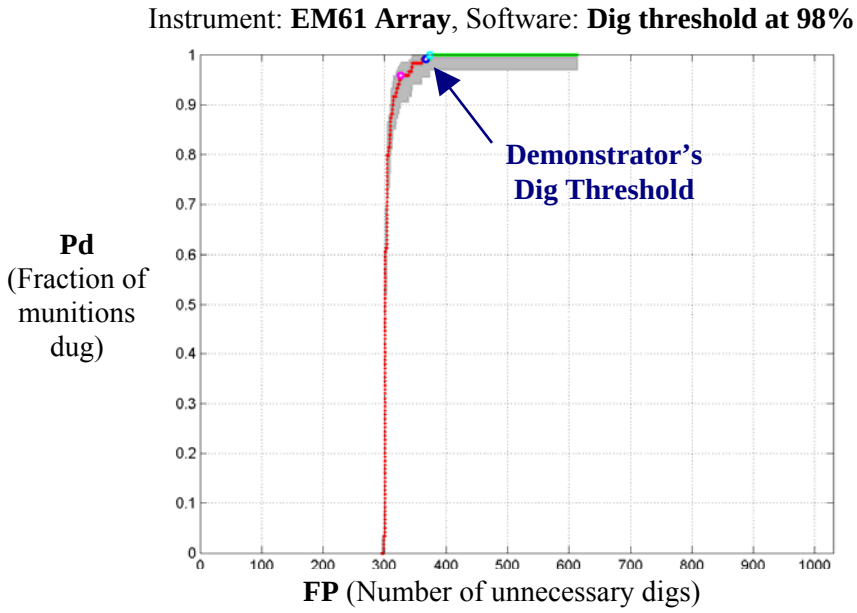


Figure 3.26: ROC curve for the EM61 Array instrument and software based on a multidimensional classifier. The demonstrator's dig threshold (dark blue dot) was quantitatively chosen so that locations were labeled "Do not dig" if their probability of being clutter was greater than or equal to 98%. The shape of the ROC curve is identical to Figures 3.23 and 3.26 because all three figures are based on the same instrument and software. In contrast, the position of the demonstrator's dig threshold is slightly different than Figures 3.23 and 3.26 because each of the three figures was based on different quantitative criteria for choosing the dig threshold.

3.3 LIMITATIONS OF ANALYSIS

The testing at Camp Sibert was intended to address a number of goals. Predominant was the desire to understand how the current generation of discrimination algorithms would perform on a site with a simple target set and whose topography, land cover, and geology allowed collection of high-quality digital geophysical data. Secondary goals included a desire to understand how various discrimination algorithms performed relative to each other, whether particular instruments or instrument combinations provided much better or much worse results than others, and what combination of algorithms and instruments performed best. Finally, there was a desire to understand whether semi-supervised learning or active learning could improve discrimination performance.

On the whole, the goals were remarkably well met by the designed demonstration. A large enough number of seed munitions were emplaced to provide decent statistics, in spite of no other intact munitions being found in the demonstration area. A large number of indigenous clutter items on the site provided sufficient anomalies that could be safely left in the ground to assess the capability of discrimination to reduce ultimate costs.

However, the demonstration did have limitations that narrowed the conclusions that could be drawn from the data.

A single-use site was intentionally chosen for a first demonstration of discrimination technology on a live site. In retrospect, that remains the correct choice. However, as illustrated in the histogram of Figure 3.12, the 4.2" mortar target was generally well separated in size from most of the competing clutter items. For this reason, size-based discrimination algorithms provided discrimination performance nearly as good as could possibly be achieved on this site. Hence, this demonstration did not allow us to fully assess the potential incremental value of the additional features. In addition, the simplicity of the site also did not allow a useful evaluation of semi-supervised and active learning algorithms. Furthermore, because size was the significant discrimination feature on this site, magnetometer performance was likely better relative to EMI sensor performance than it would have been on a more general site, although it will take a more challenging site to prove that thesis.

A final limitation of this demonstration and the related analysis is that it provides only an estimate of the detection performance of the sensors used. In the ideal experiment, the entire site would be carefully excavated to the deepest depth of interest, and all items recovered would be exhaustively cataloged. Because of very real funding limitations, complete excavation could not be done on this site and is unlikely to be done on any substantial live site in the future. Thus, in theory, UXO items could remain undetected on the site, although we consider the possibility highly unlikely.

3.4 LESSONS LEARNED

Distinct from the findings regarding performance that have been drawn from this demonstration, we have learned a number of lessons that will be used to guide the planning and conduct of follow-on discrimination demonstrations:

- Demonstrators should develop and apply specific, principled, documented criteria to determine what anomalies should be declared “Can’t analyze.”
- “Can’t analyze” items should not be part of the ranked dig list. Instead, they should be appended to the bottom and scored as a group for retrospective ROC curve analysis.
- The Program Office should provide the demonstrators a standard template for ranked diglists so that data arrive in a consistent fashion to ease scoring.
- The same monument should be used for all data-collection activities, and that monument should be resurveyed as part of the setup process. If multiple

monuments must be used, their absolute positions should be checked against each other.

- The schedule should be arranged to provide more time for quality assurance on instrument data sets before moving forward to the detection phase. In the Sibert case, motion noise problems in the SW area due to ground furrows could have been recognized and dealt with early.

4. CONCLUSION

The results described in this document show that successful discrimination is possible on a live site using currently available instruments and software. Specific findings from this demonstration are summarized below, grouped according to the stage of processing or the type of instrument or software to which they refer:

4.1 DETECTION

- Based on the arbitrary rules used to associate anomalies with UXO, survey sensors detected almost all munitions. In addition, for the few misses, given the proximity of an anomaly to the correct position and the spatial extent of the munitions' signatures, all UXO certainly would have been dug in a well executed, practical clearance action.
- Data collected from the EM61 Array were often noisy due to the bouncing motion of the towed vehicle over the ground during data collection.

4.2 DISCRIMINATION

Commercially available instruments and software:

- Commercially available instruments and software often led to good discrimination performance.
- For survey instruments, cooperative inversions led to a slightly lower number of unnecessary digs, even though the number of detected anomalies was much higher.
- Much of the discriminating power seen at Camp Sibert is due to size-based features.
- Mag & Flag led to a large number of unnecessary digs.

Advanced instruments and software:

- High-density, cued EM63 data often led to good discrimination performance, especially with cooperative inversions.
- The multiple-axis BUD instrument provided high-SNR data from a single location leading to excellent discrimination performance in both cued and survey modes.
- The advantage to active learning could not be demonstrated at Camp Sibert.

- The advantage to semi-supervised learning could not be demonstrated at Camp Sibert.

Dig threshold:

- The dig threshold could be set using objective, quantitative methods.
- In some cases, a higher confidence in digging munitions could be achieved with only a few more unnecessary digs when using quantitative methods to set the dig threshold.

Frequency-domain EMI instruments:

- The GEM Array and custom software led to good discrimination performance.
- High-density, cued GEM data had some discrimination power, but led to a large number of unnecessary digs, even with cooperative inversions.

"Can't analyze" locations:

- All "Can't analyze" locations must be dug, as some may be munitions.
- A principled, documented method for identifying "Can't analyze" locations has not yet been agreed on.
- Once "Can't analyze" locations were dug, discrimination performance was often good for all remaining locations.

As a first demonstration on a live site, it was important to establish these findings even in a site as benign as Camp Sibert, in which only a single, large munition was found. It is now important to conduct follow-up studies at more challenging sites, as this may (or may not) give more advanced instruments and software the opportunity to demonstrate their higher performance. The experience drawn and lessons learned from this demonstration can be applied to future demonstrations.

ACRONYMS

BUD	Berkeley UXO Discriminator
COE	Corps of Engineers
DSB	Defense Science Board
EMI	electromagnetic induction
ESTCP	Environmental Security Technology Certification Program
FAR	false alarm rate
FN	false negative
FP	false positives, or the number of unnecessary digs
GEM	frequency-domain EMI
GPO	geophysical prove out
GPS	Global Positioning System
I	in-phase
IDA	Institute for Defense Analyses
IMU	inertial measurement unit
M&F	mag-and-flag
MTADS	Multi-sensor Towed Array Detection System
Pd	probability of detection, or the fraction of munitions labeled as "dig"
Q	phase quadrature
QA/QC	quality assurance/quality control
RMS	root mean square
ROC	receiver operating characteristic
RTK	real-time kinematic
SAIC	Science Applications International Corporation
SE1	Southeast 1
SE2	Southeast 2
SERDP	Strategic Environmental Research and Development Program
SNR	signal-to-noise ratio
SW	Southwest
TN	true negative
TP	true positive
UXO	unexploded ordnance

REFERENCES

[1] AETC, "Adaptive and Interactive Processing Techniques for Overlapping Signatures," Technical summary report, AETC Inc., March 2006.

[2] B. Barrow, T. Bell, and N. Khadr, "Inversion of EMI Data Mapped with an IMU Positioning System," Orlando: The UXO/Countermine/Range Forum 2007, August 2007.

[3] C. Baum, *Detection and Identification of Visually Obscured Targets*, Philadelphia: Taylor and Frances, 1999.

[4] S. Billings, "Draft Demonstration Report: Data Modeling, Feature Extraction, and Classification of Magnetic and EMI Data," ESTCP Discrimination Study, Camp Sibert, AL, Project 200504: Practical Discrimination Strategies for Application to Live Sites, February 2008.

[5] L. Carin, L. Kennedy, X. Zhu, Y. Yu, D. Williams, "Preliminary Report on S10 Analysis of Sibert Data," revised September 2007.

[6] ———, "Final Report on SIG Analysis of Sibert Data," Signals Innovations Group, March 2008.

[7] Defense Science Board, "Report of the Defense Science Board Task Force on Unexploded Ordnance," Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics, December 2003.

[8] ESTCP, "Draft Technology Demonstration Data Report: ESTCP UXO Discrimination Study: MTADS Demonstration at Camp Sibert: Magnetometer/EM61 MkII, GEM-3 Arrays," ESTCP Project #MM-0533, April 2007.

[9] ———, "Data Collection Report: ESTCP UXO Discrimination Study: EM61-Mk2 Data Collection and Analysis at Camp Sibert, Gadsden, AL," June 2007.

[10] R. A. Hilgers, "Distribution-Free Confidence Bounds for ROC Curves," *Methods of Information in Medicine* 30: 96–101, 1991.

[11] Lawrence Berkeley National Laboratory, "Discrimination Report: ESTCP UXO Discrimination Study," ESTCP Project #MM-0437, December 2007.

[12] S. Macshassy and F. Provost, "Confidence Bands for ROC Curves: Methods and Empirical Study," *Proc 1st Workshop on ROC Analysis in AI,* ROCAI-2004, 2004.

[13] M. May and M. Tuley, "Interpreting Results from the Standardized UXO Test Sites," IDA Document D-3280, Alexandria, VA: Institute for Defense Analyses, January 2007.

[14] NAEVA, "Draft Technology Data Collection Plan: ESTCP UXO Discrimination Study: Cued GEM-3 Demonstration at Camp Sibert, Gadsden, AL," April 2007.

[15] H. Nelson, K. Kay, and A. Andrews, *ESTCP Pilot Program: Classification Approaches in Munitions Response,* Environmental Technology Certification Program, September 2008.

[16] Parsons, "Draft Report of Environmental Security Technology Certification Program (ESTCP) UXO Discrimination Study Support Activities," November 2007.

[17] L. R. Pasion, S. D. Billings, K. A. Kingdom, D. W. Oldeburg, N. Lhomme, J. Jacobson, "Cooperative Inversion of Time Domain Electromagnetic and Magnetometer Data for the Discrimination of Unexploded Ordnance," *J. Environmental and Engineering Geophysics* 13(3): 193–210, 2008.

[18] SAIC, "Interim Report: ESTCP Discrimination Pilot Program: SAIC analysis or Survey Data Acquired at Camp Sibert," ESTCP Project #MM-0210, March 2008.

[19] Sky Research, Inc., "Draft Demonstration Report for Geonics EM-63 Cued-Interrogation Data Collection, Processing and Archiving at Camp Silbert, Alabama," Project no. ESTCP MM-0504, February 2008.

[20] M. Tuley and M. May, "Draft Scoring Protocols for the ESTCP UXO Discrimination Study," Alexandria, VA: Institute for Defense Analyses, November 2006.

APPENDIX A
UXO DISCRIMINATION STUDY: BLIND SEED PLAN FOR CAMP SIBERT, AL

SITE PREPARATION AND HOLE CAMOUFLAGE

No specific site preparation will be done prior to seeding targets. Dig teams shall attempt to replace dirt in holes as completely as possible. No definite time for weathering-in is scheduled, but dig teams should spread grass seed on the filled-in holes.

QUALITY CONTROL

Each data collection demonstrator will submit a quality control plan to the Program Office for approval as part of his or her individual Demonstration Plan. Mr. Bob Selfridge, USACE, or his Program Office approved designee, will be the Quality Control Officer for the seeding of blind targets. He will be on-site and monitor the emplacement.

ANOMALY AVOIDANCE

Many areas designated for seeding may contain small metallic debris or be near magnetically active geology. The intent of the seed plan is to avoid geology and large (>16 nT) anomalies, but to allow seeding near smaller anomalies.

When inspecting a location prior to seeding, if an anomaly in the area is determined to be small in signal strength and size (horizontal extent), dig at that location. Any metallic objects found during the emplacement shall be removed from the site. However, no special effort (e.g., sifting or expanding the hole) shall be made to find and remove these objects.

If the intended location for a seed target is inadvertently near a large (>16 nT) anomaly, the emplacement team shall choose a different nearby location. However, the team shall take care not to move the seed target too close (within 6 m) to another seed target or near a large, slowly varying magnetic anomaly.

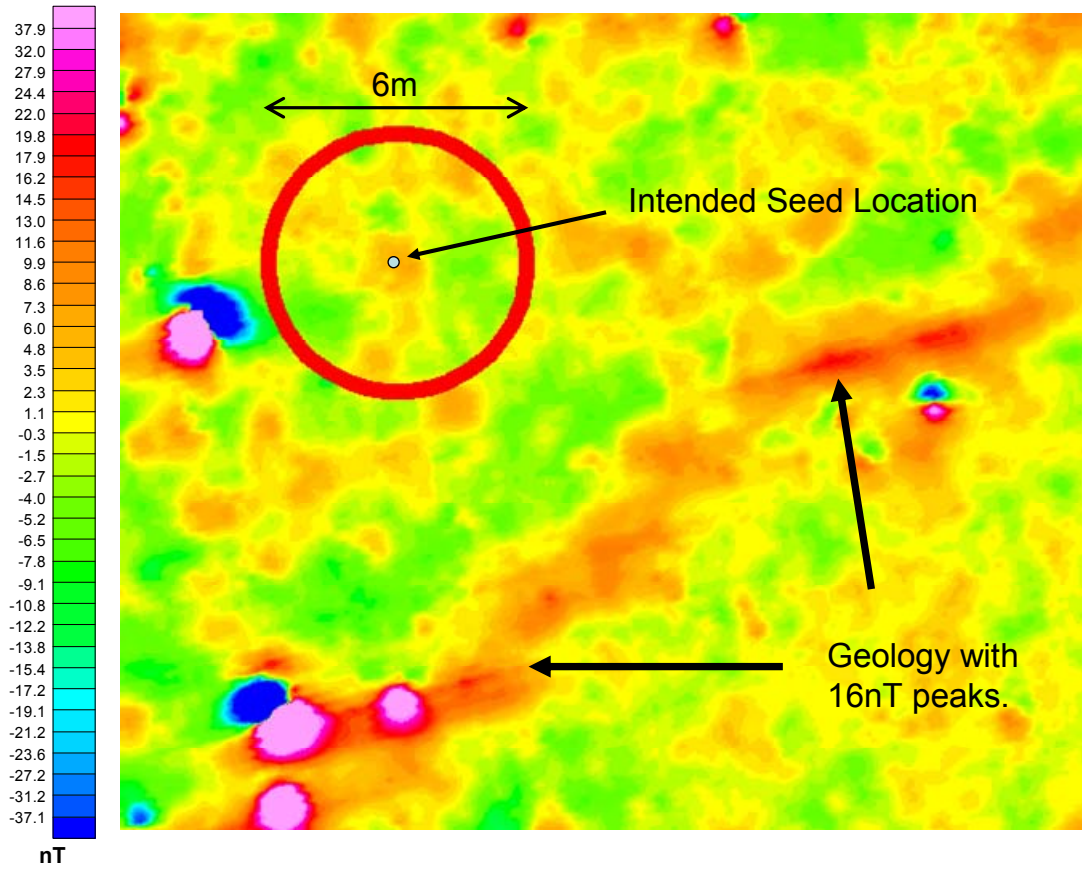


Figure A-1: An example of an intended seed location. Note that the 5–8 nT variations are common in the Southwest portion of the site. If the seed target does need to be moved because the magnetic map does not accurately reflect the true conditions on the ground, care shall be taken to avoid magnetic geology that may be near the intended site.

EMPLACEMENT OF SEED TARGETS

The seeding will be blind to all personnel completing a detection analysis on GPO data. Note that this is particularly relevant to Nagi Khadr of SAIC. While supporting the Program Office, Dr. Khadr should not see the ground truth until after he has marked his anomalies. Appropriate measures to protect the ground truth should be taken by the emplacement team.

The emplacement team will survey each seed target emplaced in the survey area and the vertices of a polygon enclosing the survey area. The reference point on the survey equipment should be physically placed within 1 cm of the location being surveyed. This study will attempt to reconstruct physical parameters of the buried targets such as depth, size, dip, and inclination. *It is critical for the success of this study that actual locations of the targets in the ground are surveyed as accurately and precisely as is feasible.* The

emplacement team shall dig in such a fashion that target migration (e.g., settling) after burial is minimized.

This emplacement plan is a *guide* for the emplacement team that describes the intended distribution of targets. The emplacement team should allow small deviations from the burial parameters listed in the seed plan (Table A-2) such as depth, dip, and inclination. This variation is desired and the exact parameters will be recorded by survey.

Table A-2 specifies the intended burial parameters. Locations shall be acquired to within 25 cm before digging begins. This is important to ensure anomaly avoidance. The depths are given to 10 cm precision and azimuths to 30 degrees; the same precision should be used when emplacing the targets. The burial depths were chosen assuming a round length of about 22″. If, for any reason, following the seed plan would result in a round not being completely buried, the emplacement team shall lower the depth until the round has 10 cm of overburden. The dip angles are specified as "up," "down," or "sideways." The emplacement team shall interpret these as follows:

- Up: nose within 10 degrees of pointing straight up.
- Down: nose within 10 degrees of pointing straight down.
- Sideways: nose within 45 degrees (up or down) of being horizontal. Note this is a large window and deviations from perfectly horizontal are desired. The emplacement team shall avoid burying all sideways targets exactly horizontal.

The accuracy afforded by the GPS system should be less than 2 cm for Easting and Northing and less than 4 cm for elevation. Locations will be surveyed relative to cm-level marker #189 (see Table A-1). Ferrous spikes should be driven into the ground at the vertices of the survey boundary to serve as fiducials. In addition, the vertices should also be marked with high-visibility non-metallic markers.

Field data that will be recorded during target emplacement will include:

- x, y, and z coordinates will be surveyed for the nose, tail, and center of each GPO target,
- the depth to target will be determined by surveying a point on the edge of the hole to establish the elevation of the local surface, and
- a photograph of each target after it is in place, but before covering it with dirt. The serial number of each item should be visible in the photograph. A ruler or similar scale will also be included in the picture.

This information will be organized by a unique target identification number[20] and reported to the Program Office. Coordinates should be reported in UTM (NAD83, Zone 16N). Center location, dip angle, azimuth, and other information will be calculated by the Program Office from the data recorded by the emplacement team.

The emplacement team will also:

- ensure all targets are marked with blue paint (inert),
- bury the targets and remove evidence of the intrusion to the extent practical. The team should carry a bag of grass seed to re-seed the disturbed earth. Some demonstrators will prove out in spring and this will allow some natural camouflage to regrow. Time and property constraints do not allow weathering-in or a full vegetation clearance, and
- mark and photograph the vertices of the GPO site with non-metallic, high visibility markers.

The Program Office will record the following data in a ground truth file that will be reported to analysis demonstrators:

- target serial number,
- munition type: 4.2″ mortar or splayed half-round for the GPO,
- northing and easting to target center,
- depth in cm from the local surface to the *center* of the object,
- dip angle: 0 degrees = sideways, +90 degrees = nose-up, and
- inclination from true north.

Table A-1: Available Survey Control Points in the Vicinity of Site 18 of the former Camp Sibert FUDS. 189 should be used for base stations; 165 may be used for QA/QC.

Point	Latitude	Longitude	Northing (m)	Easting (m)	Northing (US ft)	Easting (US ft)	HAE (NAD83 m)	Visually acquired?
	NAD83		UTM Zone 16N, NAD 83		Alabama State Plane East, NAD83			
165	33º 54' 05.22848" N	86º 09' 17.17042" W	3,751,550.813	578,146.300	1,237,596.221	558,630.983		Yes
189	33º 54' 03.19413" N	86º 09' 03.92590" W	3,751,490.960	578,486.975	1,237,387.109	559,746.706	134.835	Yes

[20] This identifier should match any electronic record (filename) of the location made with surveying equipment.

Control point 189 was used for the initial magnetometer survey. The control point is located on the eastern edge of the Southwest area on the top of a hill near a large tree. There are wooden stakes and survey tape marking the location. The monument is a piece of rebar.

Control point 165 is located at the Northwestern corner of the Southwest site on the Northeast corner of the road intersection. The monument is a piece of rebar marked with survey tape.

SEED PLAN TABLES

Table A-2a: Southwest mortar targets.

#	Item	Easting	Northing	Dip (nose direction)	Depth to Center (m)	Azimuth
42-032	4.2" mortar	578485.28	3751379.89	Sideways	0.5	270
42-033	4.2" mortar	578475.19	3751369.65	Down	0.4	30
42-034	4.2" mortar	578464.30	3751369.41	Sideways	0.2	60
42-035	4.2" mortar	578443.08	3751360.14	Down	0.4	210
42-036	4.2" mortar	578445.11	3751369.24	Sideways	0.4	270
42-037	4.2" mortar	578437.96	3751382.58	Sideways	0.6	210
42-038	4.2" mortar	578447.96	3751389.81	Sideways	0.1	60
42-039	4.2" mortar	578436.41	3751398.68	Sideways	0.3	180
42-040	4.2" mortar	578443.81	3751403.72	Sideways	0.6	300
42-041	4.2" mortar	578451.86	3751397.70	Up	0.5	270
42-042	4.2" mortar	578464.20	3751411.12	Sideways	0.3	270
42-043	4.2" mortar	578476.90	3751410.38	Sideways	0.1	0
42-044	4.2" mortar	578456.98	3751415.67	Down	0.4	120
42-045	4.2" mortar	578470.72	3751428.51	Down	0.4	0
42-046	4.2" mortar	578483.49	3751431.28	Sideways	0.3	90
42-047	4.2" mortar	578492.43	3751436.56	Down	0.3	270
42-048	4.2" mortar	578472.76	3751449.41	Sideways	0.3	30
42-049	4.2" mortar	578462.27	3751453.80	Down	0.3	150
42-050	4.2" mortar	578461.13	3751467.54	Down	0.5	330
42-051	4.2" mortar	578486.58	3751470.47	Sideways	0.2	300
42-052	4.2" mortar	578444.87	3751463.80	Up	0.5	300
42-053	4.2" mortar	578443.08	3751447.38	Down	0.5	60
42-054	4.2" mortar	578472.84	3751487.79	Sideways	0.2	240
42-055	4.2" mortar	578465.11	3751501.61	Up	0.4	270
42-056	4.2" mortar	578395.52	3751496.97	Sideways	0.3	180
42-057	4.2" mortar	578409.18	3751490.39	Down	0.6	330
42-058	4.2" mortar	578429.66	3751486.40	Down	0.5	270
42-059	4.2" mortar	578395.68	3751482.91	Sideways	0.1	180
42-060	4.2" mortar	578376.00	3751456.48	Sideways	0.6	240
42-061	4.2" mortar	578407.22	3751458.76	Down	0.3	300
42-062	4.2" mortar	578421.78	3751464.13	Down	0.8	300
42-063	4.2" mortar	578417.87	3751445.91	Down	0.4	90
42-064	4.2" mortar	578430.80	3751445.02	Up	0.8	150
42-065	4.2" mortar	578420.72	3751432.58	Sideways	0.7	150
42-066	4.2" mortar	578406.57	3751436.97	Up	1	150
42-067	4.2" mortar	578408.12	3751427.13	Up	0.3	300
42-068	4.2" mortar	578389.50	3751429.57	Down	0.4	240
42-069	4.2" mortar	578405.76	3751420.47	Sideways	0.2	150
42-070	4.2" mortar	578407.87	3751394.61	Up	0.4	150
42-071	4.2" mortar	578425.03	3751392.17	Sideways	0.8	210
42-072	4.2" mortar	578408.69	3751382.66	Sideways	0.3	90
42-073	4.2" mortar	578426.49	3751371.93	Sideways	0.2	270
42-074	4.2" mortar	578415.03	3751361.76	Sideways	0.1	270
42-075	4.2" mortar	578380.72	3751363.96	Up	0.5	240
42-076	4.2" mortar	578375.84	3751377.86	Sideways	0.4	330
42-077	4.2" mortar	578361.94	3751381.36	Up	0.5	180
42-078	4.2" mortar	578333.24	3751402.42	Down	0.4	300
42-079	4.2" mortar	578312.34	3751412.99	Sideways	0.7	210
42-080	4.2" mortar	578319.09	3751431.12	Up	0.5	60
42-081	4.2" mortar	578327.71	3751446.08	Up	0.4	60
42-082	4.2" mortar	578357.06	3751447.95	Sideways	0.1	0
42-083	4.2" mortar	578340.15	3751462.17	Sideways	0.1	240
42-084	4.2" mortar	578331.45	3751470.14	Up	0.4	330
42-085	4.2" mortar	578299.66	3751470.47	Sideways	1.1	300
42-086	4.2" mortar	578278.27	3751447.95	Sideways	0.4	60
42-087	4.2" mortar	578278.93	3751440.71	Down	0.5	210
42-088	4.2" mortar	578261.36	3751442.01	Up	0.3	210
42-089	4.2" mortar	578294.70	3751422.74	Down	0.6	30
42-090	4.2" mortar	578265.35	3751419.49	Sideways	0.2	30
42-091	4.2" mortar	578251.93	3751415.91	Down	1	330
42-092	4.2" mortar	578280.14	3751411.36	Sideways	0.5	240
42-093	4.2" mortar	578256.89	3751386.07	Up	0.4	90
42-094	4.2" mortar	578275.43	3751378.92	Up	0.3	300
42-095	4.2" mortar	578284.86	3751374.12	Up	0.5	270
42-096	4.2" mortar	578293.40	3751378.51	Down	0.8	300
42-097	4.2" mortar	578304.29	3751380.22	Up	0.4	30
42-098	4.2" mortar	578256.08	3751368.11	Up	0.8	270
42-099	4.2" mortar	578270.14	3751367.70	Down	0.5	330
42-100	4.2" mortar	578281.69	3751363.39	Down	0.5	180
42-101	4.2" mortar	578268.11	3751355.26	Up	0.5	330
42-102	4.2" mortar	578292.42	3751360.46	Sideways	1	120

Table A-2b: Southeast mortar targets.

#	Item	Easting	Northing	Dip (nose direction)	Depth to Center (m)	Azimuth
42-103	4.2" mortar	578821.25	3751694.30	Down	0.4	120
42-104	4.2" mortar	578832.06	3751710.60	Sideways	0.3	90
42-105	4.2" mortar	578831.58	3751701.40	Sideways	0.4	90
42-106	4.2" mortar	578840.13	3751701.57	Down	0.3	270
42-107	4.2" mortar	578843.84	3751711.41	Sideways	0.5	360
42-108	4.2" mortar	578834.80	3751684.62	Down	0.7	240
42-109	4.2" mortar	578823.35	3751669.62	Sideways	0.1	270
42-110	4.2" mortar	578827.38	3751661.55	Sideways	0.2	270
42-111	4.2" mortar	578837.71	3751661.39	Sideways	0.1	150
42-112	4.2" mortar	578844.48	3751656.23	Down	0.5	240
42-113	4.2" mortar	578839.32	3751647.03	Up	0.8	270
42-114	4.2" mortar	578832.22	3751640.57	Sideways	0.1	0
42-115	4.2" mortar	578852.55	3751648.64	Sideways	0.3	60
42-116	4.2" mortar	578852.39	3751658.00	Sideways	0.4	330
42-117	4.2" mortar	578855.94	3751642.19	Up	0.4	300
42-118	4.2" mortar	578868.52	3751645.09	Down	0.5	60
42-119	4.2" mortar	578875.95	3751654.29	Down	0.4	30
42-120	4.2" mortar	578861.10	3751664.62	Sideways	0.4	240
42-121	4.2" mortar	578858.52	3751704.47	Down	0.4	150
42-122	4.2" mortar	578865.62	3751714.80	Sideways	0.1	240
42-123	4.2" mortar	578875.46	3751709.63	Sideways	0.4	210
42-124	4.2" mortar	578877.72	3751677.52	Up	0.6	360
42-125	4.2" mortar	578887.08	3751675.10	Sideways	0.9	330
42-126	4.2" mortar	578880.79	3751663.81	Sideways	0.4	60
42-127	4.2" mortar	578891.76	3751652.68	Sideways	0.4	180
42-128	4.2" mortar	578906.60	3751655.58	Down	0.6	60
42-129	4.2" mortar	578918.54	3751657.35	Up	0.4	30
42-130	4.2" mortar	578893.05	3751705.76	Sideways	0.5	150
42-131	4.2" mortar	578880.47	3751720.12	Sideways	0.2	330
42-132	4.2" mortar	578888.69	3751723.51	Up	0.5	150
42-133	4.2" mortar	578895.63	3751716.73	Down	0.3	240
42-134	4.2" mortar	578903.22	3751712.38	Up	0.4	240
42-135	4.2" mortar	578914.03	3751689.14	Down	0.5	120
42-136	4.2" mortar	578925.80	3751692.69	Down	0.5	240
42-137	4.2" mortar	578930.32	3751685.43	Down	0.3	180
42-138	4.2" mortar	578918.87	3751671.55	Up	0.5	120
42-139	4.2" mortar	578937.58	3751673.33	Sideways	1	180
42-140	4.2" mortar	578939.36	3751666.07	Sideways	0.5	120
42-141	4.2" mortar	578943.88	3751656.06	Sideways	0.7	90
42-142	4.2" mortar	578924.19	3751647.67	Sideways	0.2	300
42-143	4.2" mortar	578959.04	3751656.55	Down	0.8	180
42-144	4.2" mortar	578955.98	3751667.52	Sideways	0.2	120
42-145	4.2" mortar	578949.36	3751675.75	Sideways	0.2	60
42-146	4.2" mortar	578947.91	3751686.08	Up	0.5	270
42-147	4.2" mortar	578944.20	3751700.11	Sideways	0.4	30
42-148	4.2" mortar	578942.75	3751713.67	Sideways	0.2	330
42-149	4.2" mortar	578932.42	3751721.57	Up	0.5	150
42-150	4.2" mortar	578929.19	3751729.96	Sideways	0.8	30
42-151	4.2" mortar	578925.16	3751741.58	Up	0.4	180
42-152	4.2" mortar	578943.39	3751747.71	Up	0.3	120
42-153	4.2" mortar	578954.36	3751752.39	Sideways	0.3	330
42-154	4.2" mortar	578945.01	3751734.32	Down	0.5	90
42-155	4.2" mortar	578954.85	3751729.64	Down	0.4	30
42-156	4.2" mortar	578912.57	3751715.76	Down	0.5	180
42-157	4.2" mortar	579006.48	3751674.78	Sideways	0.5	270
42-158	4.2" mortar	579022.29	3751707.70	Up	0.5	330
42-159	4.2" mortar	579003.09	3751693.98	Sideways	0.3	30
42-160	4.2" mortar	578980.50	3751688.17	Sideways	1.1	90
42-161	4.2" mortar	578983.57	3751725.77	Up	0.3	120
42-162	4.2" mortar	578960.66	3751726.09	Sideways	0.3	90
42-163	4.2" mortar	579003.25	3751755.13	Up	1	180
42-164	4.2" mortar	578970.02	3751712.54	Down	0.4	120
42-165	4.2" mortar	579029.23	3751624.76	Sideways	0.3	270
42-166	4.2" mortar	579038.11	3751620.08	Up	0.6	90
42-167	4.2" mortar	579024.39	3751579.91	Sideways	0.5	60
42-168	4.2" mortar	579032.62	3751584.75	Down	0.3	150
42-169	4.2" mortar	579018.58	3751563.77	Sideways	0.6	120
42-170	4.2" mortar	579013.26	3751553.12	Sideways	0.3	120
42-171	4.2" mortar	579033.43	3751541.02	Up	0.5	30
42-172	4.2" mortar	579023.75	3751542.15	Sideways	0.1	300
42-173	4.2" mortar	579056.66	3751522.95	Down	1	330
42-174	4.2" mortar	579067.15	3751560.06	Sideways	0.3	210
42-175	4.2" mortar	579042.46	3751566.84	Up	0.4	300
42-176	4.2" mortar	579011.64	3751611.69	Down	0.4	90
42-177	4.2" mortar	578986.64	3751745.87	Down	0.9	150
42-178	4.2" mortar	579003.26	3751732.11	Up	0.3	210
42-179	4.2" mortar	579016.83	3751727.33	Sideways	0.3	60
42-180	4.2" mortar	579030.78	3751694.27	Up	0.3	150
42-181	4.2" mortar	579005.17	3751703.44	Up	0.4	330

Table A-2c: Southwest splayed half-rounds.

#	Item	Easting	Northing	Orientation	Depth to Center (m)	Azimuth
HR-11	Splayed Half-Round	578456.95	3751512.20	Flat	0.1	NA
HR-12	Splayed Half-Round	578469.17	3751458.89	On edge	0.1	NA
HR-13	Splayed Half-Round	578419.82	3751455.24	Flat	0.2	NA
HR-14	Splayed Half-Round	578424.74	3751422.08	On edge	0.2	NA
HR-15	Splayed Half-Round	578371.90	3751418.11	Flat	0.3	NA
HR-16	Splayed Half-Round	578373.65	3751476.50	On edge	0.3	NA
HR-17	Splayed Half-Round	578417.92	3751376.70	Flat	0.4	NA
HR-18	Splayed Half-Round	578420.77	3751352.42	On edge	0.4	NA
HR-19	Splayed Half-Round	578384.28	3751404.94	Flat	0.4	NA
HR-20	Splayed Half-Round	578286.38	3751384.79	On edge	0.5	NA
HR-21	Splayed Half-Round	578304.94	3751427.79	Flat	0.5	NA
HR-22	Splayed Half-Round	578360.00	3751484.59	On edge	0.7	NA
HR-23	Splayed Half-Round	578493.28	3751405.74	Flat	0.8	NA

Table A-2d: Southeast splayed half-rounds.

#	Item	Easting	Northing	Orientation	Depth to Center (m)	Azimuth
HR-24	Splayed Half-Round	578825.21	3751681.47	Flat	0.1	NA
HR-25	Splayed Half-Round	578883.02	3751697.74	On edge	0.1	NA
HR-26	Splayed Half-Round	578904.14	3751734.44	Flat	0.2	NA
HR-27	Splayed Half-Round	578892.19	3751663.12	On edge	0.2	NA
HR-28	Splayed Half-Round	579017.34	3751678.18	Flat	0.3	NA
HR-29	Splayed Half-Round	578964.37	3751745.69	On edge	0.3	NA
HR-30	Splayed Half-Round	578959.01	3751689.78	Flat	0.4	NA
HR-31	Splayed Half-Round	578975.45	3751756.08	On edge	0.4	NA
HR-32	Splayed Half-Round	579039.50	3751521.36	Flat	0.7	NA
HR-33	Splayed Half-Round	579009.38	3751599.60	On edge	0.8	NA

Table A-3: Program Office Contact List

Name	Organization/Office	Role	Mailing Address	Phone Fax	Email
ESTCP Program Office Team					
Andrews, Anne	ESTCP/SERDP	MM Program Manager	ESTCP Program Office, 901 North Stuart Street, Suite 303, Arlington, VA 22203	P:703-696-3826 F:703-696-2114	anne.andrews@osd.mil
Khadr, Nagi	SAIC Inc.	Data Analyst	SAIC Inc. Advanced Sensors and Analysis Division 1225 S. Clark St. Arlington, VA 22202	P: 217-531-9026	nagi.khadr@saic.com
Marqusee, Jeff	ESTCP/SERDP	ESTCP Director/SERDP Technical Director	ESTCP Program Office, 901 North Stuart Street, Suite 303, Arlington, VA 22203	P:703-696-2117 F:703-696-2120	jeffrey.marqusee@osd.mil
May, Michael	Institute for Defense Analyses	ESTCP Support	Institute for Defense Analyses Science and Technology Division 4850 Mark Center Drive Alexandria , VA 22311	P: 703-578-2821	mmay@ida.org
Nelson, Herb	Naval Research Lab	ESTCP Discrimination Study Program Manager	Code 6110, Naval Research Lab, Washington, DC 20375-5342	P: 202-767-3686 F:202-404-8119	herb.nelson@nrl.navy.mil
Selfridge, Robert	U.S. Army Corps of Engineers, Huntsville	ESTCP Support	ATTN: CEHNC-ED-CS-G 4820 University Square Huntsville, AL 35816-1822	P: (256) 895-1887 F: (256) 895-1602	Bob.J.Selfridge@hnd01.usace.army.mil
Tuley, Mike	Institute for Defense Analyses	ESTCP Support	Institute for Defense Analyses Science and Technology Division 4850 Mark Center Drive Alexandria , VA 22311	P: 703-578-2825 F: 703-578-2877	MTuley@ida.org
Kaye, Katherine	ESTCP/SERDP Support, HGL	MM Program Manager Assistant	10001 Herding Row Columbia, MD 21046	P: 410-884-4447 F: 703-471-4180	kkaye@hgl.com
Camp Sibert Site Support					
Nivens, Gregory	Parsons		Parsons 4890 University Square, Suite 2 Huntsville, Alabama 35816	P: (256) 217-2523 (office) P: (256) 684-1526 (cell)	Gregory.Nivens@parsons.com
Cudney, Joe	Parsons			P: 678 969 2344 (work) P: 404 606 0347 (cell)	Joseph.cudney@parsons.com
Meacham, Kim	U.S. Army Corps of Engineers, Huntsville	Camp Sibert Project Engineer/Technical Manager	USAESCH ATTN: CEHNC-ED-CS-P-Meacham 4820 University Square Huntsville, AL 35816-1821	P: 256-895-1667	Kim.K.Meacham@hnd01.usace.army.mil
Smith, Michael	U.S. Army Corps of Engineers, Huntsville	CWM Safety	USAESCH ATTN: CEHNC-OE-S-Smith 4820 University Square Huntsville, AL 35816-1822	P: 256-509-8708	Michael.G.Smith@hnd01.usace.army.mil
Shott, Kenneth	U.S. Army Corps of Engineers, Huntsville	CWM Safety	USAESCH ATTN: CEHNC-OE-S-Shott 4820 University Square Huntsville, AL 35816-1822	P: 256-656-2405	Kenneth.D.Shott@hnd01.usace.army.mil
Walters, Wilson	U.S. Army Corps of Engineers, Huntsville	CWM Safety Supervisor	USAESCH ATTN: CEHNC-OE-CW-Walters 4820 University Square Huntsville, AL 35816-1822	P: 256-895-1290	Wilson.C.Walters@hnd01.usace.army.mil

APPENDIX B: UXO DISCRIMINATION STUDY: RESULTS FOR CAMP SIBERT, AL

Refer to the companion DVD for all detection and discrimination performance metrics.

For each instrument/algorithm combination, the following file types exist on the DVD:

- A series of *.tif files, each showing a figure of a ROC curve generated by adjusting the dig threshold from its minimum to maximum value. These figures can be viewed with graphics applications such as Microsoft Paint. Each *.tif file for a given instrument/algorithm combination is generated over a different geographical subregion of Camp Sibert—Southeast 1, the union of Southeast 1 and Southeast 2, the entire surveyed area, etc.
- A series of *.mat files, each including the detection performance metrics of the instrument on its own, as well as the discrimination performance metrics of the instrument/algorithm combination. These metrics include the data needed to generate the corresponding ROC curve. These files can be directly loaded into MATLAB. As with the *.tif files, each *.mat file contains metrics calculated over a different geographical subregion of Camp Sibert.
- A series of *.csv files, each including the data needed to generate the corresponding ROC curve. These files can be read by spreadsheet applications such as Microsoft Excel. As with the other file types, each *.csv file contains data calculated over a different geographical subregion.

The DVD also contains a README.doc file, describing in detail the naming conventions and metrics and data contained in each file type. The README.doc file can be read with Microsoft Word.

APPENDIX C: UXO DISCRIMINATION STUDY: ANOMALIES THAT COULD NOT BE ANALYZED

EM61 CART

Three demonstrators processed the anomalies detected with the EM61 Cart instrument: SAIC, Sky Research, Inc., and Parsons. Each demonstrator used his own techniques to determine the extent of anomaly data to be inverted, his own inversion algorithms, and his own criteria to determine which detected anomalies could and could not be analyzed further (i.e., which could and could not be successfully inverted and input into discrimination algorithms). Tables C.1–C.3 compare the number of detected anomalies that could and could not be analyzed by the three different demonstrators. Parsons could analyze the largest number of detected anomalies, 463 (85%) of 546, followed by SAIC at 435 (80%), and Sky Research, Inc. at 384 (70%). As shown in Table C.2, 11% of the detected anomalies could be analyzed by Parsons but *not* SAIC, while 6% could be analyzed by SAIC but *not* Parsons. In turn, 18% of the anomalies could be analyzed by SAIC but *not* Sky Research, Inc., while 9% could be analyzed by Sky Research, Inc. but not *SAIC*, as shown in Table C.1.

Table C.1: Comparing the number of anomalies detected with the EM61 Cart that SAIC and Sky Research, Inc. could and could not analyze. 18% of all detected anomalies could be analyzed by SAIC but *not* Sky Research, Inc., while 9% could be analyzed by Sky Research, Inc. but *not* SAIC.

EM61 Cart		SKY		
		Can Analyze	Can*not* Analyze	Total
SAIC	Can Analyze	337 (62%)	98 (18%)	435 (80%)
	Can*not* Analyze	47 (9%)	64 (12%)	111 (20%)
	Total	384 (70%)	162 (30%)	546 (100%)

Table C.2: Comparing the number of anomalies detected with the EM61 Cart that SAIC and Parsons could and could not analyze. 11% of all detected anomalies could be analyzed by Parsons but *not* SAIC, while 6% could be analyzed by SAIC but *not* Parsons.

EM61 Cart		Parsons		
		Can Analyze	Can*not* Analyze	Total
SAIC	Can Analyze	401 (73%)	34 (6%)	435 (80%)
	Can*not* Analyze	62 (11%)	49 (9%)	111 (20%)
	Total	463 (85%)	83 (15%)	546 (100%)

Table C.3: Comparing the number of anomalies detected with the EM61 Cart that Sky Research, Inc. and Parsons could and could not analyze. 18% of all detected anomalies could be analyzed by Parsons but *not* Sky Research, Inc., while 3% could be analyzed by Sky Research, Inc. but *not* Parsons.

EM61 Cart		Parsons		
		Can Analyze	Can*not* Analyze	Total
SKY	Can Analyze	365 (67%)	19 (3%)	384 (70%)
	Can*not* Analyze	98 (18%)	64 (12%)	162 (30%)
	Total	463 (85%)	83 (15%)	546 (100%)

EM61 ARRAY

SAIC and Sky Research, Inc. also processed anomalies detected with the EM61 Array. SIG processed these anomalies, as well. Tables C.4–C.5 compare the number of detected anomalies that could and could not be analyzed by the three different demonstrators. Of the 734 anomalies detected by the EM61 Array, SAIC analyzed the largest number (619 anomalies or 84%), followed by Sky Research, Inc. (441 anomalies or 60%) and SIG (439 anomalies or 60%). Furthermore, Table C.4 shows that 28% of all detected anomalies could be analyzed by SAIC but *not* Sky Research, Inc., while 3% of all detected anomalies could be analyzed by Sky Research, Inc. but *not* SAIC. Table C.5 shows a similar result. Finally, Table C.6 shows 9% of all detected anomalies could be analyzed by Sky Research, Inc. but *not* SIG, while a different 9% could be analyzed by SIG but *not* Sky Research, Inc.

Table C.4: Comparing the number of anomalies detected with the EM61 Array that SAIC and Sky Research, Inc. could and could not analyze. 28% of all detected anomalies could be analyzed by SAIC but *not* Sky Research, Inc., while 3% were vice versa.

EM61 Array		Sky Research, Inc.		
		Can Analyze	**Can*not* Analyze**	**Total**
SAIC	**Can Analyze**	416 (57%)	203 (28%)	619 (84%)
	Can*not* Analyze	25 (3%)	90 (12%)	115 (16%)
	Total	441 (60%)	293 (40%)	734 (100%

Table C.5: Comparing the number of anomalies detected with the EM61 Array that SAIC and SIG could and could not analyze. 29% of all detected anomalies could be analyzed by SAIC but *not* SIG, while 4% were vice versa.

EM61 Array		SIG		
		Can Analyze	**Can*not* Analyze**	**Total**
SAIC	**Can Analyze**	409 (56%)	210 (29%)	619 (84%)
	Can*not* Analyze	30 (4%)	85 (12%)	115 (16%)
	Total	439 (60%)	295 (40%)	734 (100%)

Table C.6: Comparing the number of anomalies detected with the EM61 Array that Sky Research, Inc. and SIG could and could not analyze. 9% of all detected anomalies could be analyzed by Sky Research, Inc. but *not* SIG, while a different 9% were vice versa.

EM61 Array		SIG		
		Can Analyze	**Can*not* Analyze**	**Total**
SKY	**Can Analyze**	373 (51%)	68 (9%)	441 (60%)
	Can*not* Analyze	66 (9%)	227 (31%)	293 (40%)
	Total	439 (60%)	295 (40%)	734 (100%)

EM61 ARRAY AND MAG ARRAY (COOPERATIVE OR JOINT INVERSIONS)

Three demonstrators processed the EM61 Array and Mag Array data in tandem, either through cooperative or joint inversions: SAIC, Sky Research, Inc., and SIG. The demonstrators were instructed to include on their ranked diglist all anomalies that were detected with *either* the EM61 Array *or* Mag Array. Tables C.7–C.9 compare the number

of these anomalies that could and could not be analyzed by the three different demonstrators. Of the 982 detected anomalies, SAIC could analyze the most (821 or 84%), followed by Sky Research, Inc. (753 or 77%) and SIG (596 or 61%). As shown in Table C.7, 15% of all detected anomalies could be analyzed by SAIC but not Sky Research, Inc., while 9% were vice versa. Similarly, Tables C.8 and C.9 show that SIG was more conservative than either SAIC or Sky Research, Inc.

Table C.7: Comparing the number of anomalies detected with either the EM61 Array or Mag Array that SAIC and Sky Research, Inc. could and could not analyze. 15% of all detected anomalies could be analyzed by SAIC but *not* Sky Research, Inc., while 9% of all detected anomalies were vice versa.

EM61 Array & Mag Array		SKY		
		Can Analyze	**Can*not* Analyze**	**Total**
SAIC	**Can Analyze**	669 (68%)	152 (15%)	821 (84%)
	Can*not* Analyze	84 (9%)	77 (8%)	161 (16%)
	Total	753 (77%)	229 (23%)	982 (100%)

Table C.8: Comparing the number of anomalies detected with either the EM61 Array or Mag Array that SAIC and SIG could and could not analyze. 24% of all detected anomalies could be analyzed by SAIC but *not* SIG, while 1% was vice versa.

EM61 Array & Mag Array		SIG		
		Can Analyze	**Can*not* Analyze**	**Total**
SAIC	**Can Analyze**	585 (60%)	236 (24%)	821 (84%)
	Can*not* Analyze	11 (1%)	150 (15%)	161 (16%)
	Total	596 (61%)	386 (39%)	982 (100%)

Table C.9: Comparing the number of anomalies detected with either the EM61 Array or Mag Array that Sky Research, Inc. and SIG could and could not analyze. 19% of all detected anomalies could be analyzed by Sky Research, Inc. but *not* SIG, while 3% were vice versa.

EM61 Array & Mag Array		SIG		
		Can Analyze	Can*not* Analyze	Total
SKY	Can Analyze	563 (57%)	190 (19%)	753 (77%)
	Can*not* Analyze	33 (3%)	196 (20%)	229 (23%)
	Total	596 (61%)	386 (39%)	982 (100%)

EM63 CUED

Only two demonstrators processed data taken with the EM63 Cued instrument: Sky Research, Inc. and SIG. Table C.10 compares the number of anomalies that could and could not be analyzed by each demonstrator. Of the 150 locations in which cued data was taken, SIG could process the data from 136 locations (91%), while Sky Research, Inc. could process the data from only 118 locations (79%). While the data from 14% of all locations could be analyzed by SIG but *not* Sky Research, Inc., the data from 2% of all locations could be analyzed by Sky Research, Inc. but *not* SIG.

Table C.10: Comparing the number of locations at which EM63 Cued data was taken that Sky Research, Inc. and SIG could and could not analyze. 14% of all detected anomalies could be analyzed by SIG but not Sky Research, Inc., while 2% were vice versa.

EM63 Cued		SIG		
		Can Analyze	Can*not* Analyze	Total
SKY	Can Analyze	115 (77%)	3 (2%)	118 (79%)
	Can*not* Analyze	21 (14%)	11 (7%)	32 (21%)
	Total	136 (91%)	14 (9%)	150 (100%)

OTHER INSTRUMENTS

SAIC, Sky Research, Inc., and SIG processed the anomalies detected by the Mag Array. These results are presented and discussed in chapter III of this document. In contrast, only SAIC processed anomalies detected by the GEM Array, while only SIG processed the GEM Cued data and only LBNL processed the BUD data. Since only one

demonstration team processed the anomalies detected by each of these instruments, no comparisons between demonstrators can be made.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE January 2009	2. REPORT TYPE Final	3. DATES COVERED *(From–To)* July 2006 – March 2008
4. TITLE AND SUBTITLE The UXO Discrimination Study at the Former Camp Sibert		5a. CONTRACT NUMBER DASW01 04 C 0003 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Shelley Cazares Michael Tuley Michael May		5d. PROJECT NUMBER 5e. TASK NUMBER AM-2-1528 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882		8. PERFORMING ORGANIZATION REPORT NUMBER IDA Document D-3572 Log: H08-000974
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Environmental Security Technology Certification Program 901 North Stuart Street, Suite 303 Arlington, VA 22203		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited. (30 June 2009)

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The Fiscal Year 2006 Defense Appropriations Bill contained funding for the "Development of Advanced, Sophisticated, Discrimination Technologies for UXO Cleanup" in the Environmental Security Technology Certification Program (ESTCP). The discrimination demonstration carried out at the former Camp Sibert near Gadsden, AL, was in direct response to the Congressional language. The high-level goal of the demonstration was to assess the capability of discrimination algorithms that had been developed under the Strategic Environmental Research and Development Program (SERDP) and refined under ESTCP to determine reliably which items could be left safely in the ground and which had to be dug. The intent of the demonstration was to evaluate on a live site those algorithms that had proven successful in previous testing, principally at engineered test sites. Another important goal was to involve the regulatory community early in the design of the demonstration in an effort to better understand what might be required if detected anomalies were actually to be left in the ground. This report provides a detailed record of the scoring of the detection and discrimination performance of all the demonstrators, sensors and algorithms involved in the pilot program.

15. SUBJECT TERMS

Unexploded ordnance (UXO), discrimination, detection, classification, magnetometers, electromagnetic induction sensors

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON Dr. Anne Andrews
a. REPORT Uncl.	b. ABSTRACT Uncl.	c. THIS PAGE Uncl.	SAR	144	19b. TELEPHONE NUMBER *(include area code)* 703-696-3826

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

For individual demonstrator results please contact the ESTCP Program Office at (703) 696-2124.

The Institute for Defense Analyses is a non-profit corporation that administers three federally funded research and development centers to provide objective analyses of national security issues, particularly those requiring scientific and technical expertise, and conduct related research on other national challenges.