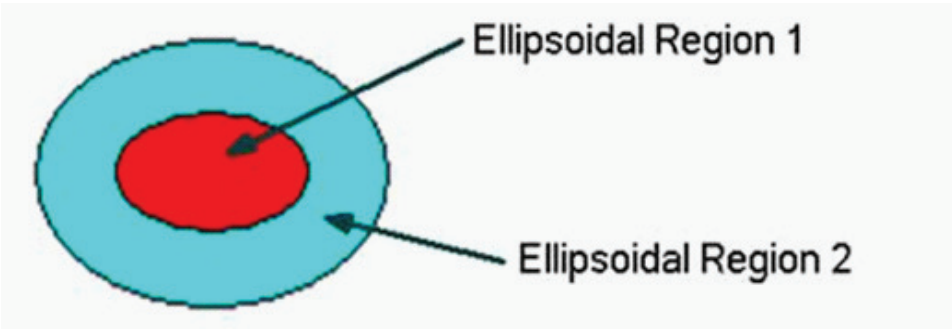
# ESTCP
# Cost and Performance Report

(MR-200811)




## LGP Discrimination and Residual Risk Analysis: Camp Sibert and Camp San Luis Obispo

June 2010

ENVIRONMENTAL SECURITY
TECHNOLOGY CERTIFICATION PROGRAM

U.S. Department of Defense

# COST & PERFORMANCE REPORT
# Project: MR-200811

## TABLE OF CONTENTS


Page

1.0 EXECUTIVE SUMMARY ..... 1

2.0 INTRODUCTION ..... 3
- 2.1 BACKGROUND ..... 3
- 2.2 OBJECTIVE OF THE DEMONSTRATION ..... 3
- 2.3 REGULATORY DRIVERS ..... 3

3.0 TECHNOLOGY ..... 5
- 3.1 TECHNOLOGY DEVELOPMENT ..... 5
- 3.2 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY ..... 5

4.0 PERFORMANCE OBJECTIVES ..... 7

5.0 SITE DESCRIPTION ..... 9
- 5.1 SITE SELECTION ..... 9
- 5.2 SITE HISTORY ..... 9
- 5.3 MUNITIONS CONTAMINATION ..... 10

6.0 TEST DESIGN ..... 11
- 6.1 CONCEPTUAL EXPERIMENTAL DESIGN ..... 11
- 6.2 SITE PREPARATION ..... 11
- 6.3 SYSTEM SPECIFICATIONS ..... 11
- 6.4 DATA COLLECTION PROCEDURES ..... 11
- 6.5 VALIDATION ..... 12

7.0 DATA ANALYSIS AND PRODUCTS ..... 13
- 7.1 DESCRIPTION OF DATA ..... 13
- 7.2 OVERVIEW OF PROCEDURES ..... 13
- 7.3 DEFINE TARGET POLYGONS AND ELLIPSES FOR EACH TARGET ..... 13
- 7.4 REMOVE CANNOT-ANALYZE TARGETS ..... 14
- 7.5 ATTRIBUTE EXTRACTION ..... 14
- 7.6 DESCRIPTION OF A MODELING AND RISK ANALYSIS ITERATION ..... 15
- 7.7 SAMPLING OF ADDITIONAL GROUND TRUTH ..... 18

8.0 PERFORMANCE ASSESSMENT ..... 21
- 8.1 CAMP SIBERT ..... 21
  - 8.1.1 EM ONLY ..... 21

**TABLE OF CONTENTS** (continued)

**Page**


8.1.2 MAG AND EM ..... 21
8.1.3 INVERSION FEATURES ..... 22
8.2 CAMP SLO ..... 23
8.2.1 ITERATION 1 ..... 23
8.2.2 ITERATION 2 ..... 24
8.2.3 SAMPLING OF GROUND TRUTH BETWEEN ITERATIONS ..... 25

9.0 COST ANALYSIS ..... 27
9.1 COST MODEL ..... 27
9.2 COST DRIVERS ..... 28
9.3 COST BENEFIT ..... 29

10.0 IMPLEMENTATION ISSUES ..... 33
10.1 COST OBSERVATIONS ..... 33
10.2 PERFORMANCE OBSERVATIONS ..... 33
10.3 SCALE-UP ..... 33
10.4 OTHER SIGNIFICANT OBSERVATIONS ..... 34

11.0 REFERENCES ..... 35

APPENDIX A POINTS OF CONTACT ..... A-1

## LIST OF FIGURES

**Page**


Figure 1. A target polygon ..... 13
Figure 2. Example of a cannot-analyze one blob ..... 14
Figure 3. A simple illustration of ellipsoidal rings for attribute extraction..... 15
Figure 4. Prediscriminator model of falling probability of UXO as a function of rank for a simple prediscriminator. ..... 16
Figure 5. Prediscriminator model of falling probability of UXO applied to blind data..... 17
Figure 6. LGP model of falling probability of UXO as a function of rank after LGP modeling. ..... 18
Figure 7. Region of selection of blind targets for sampling around an outlier UXO..... 19
Figure 8. ROC chart showing blind scoring for EM-only track..... 21
Figure 9. ROC chart showing blind scoring for combined track. ..... 22
Figure 10. ROC chart showing blind scoring for inversion track. ..... 22
Figure 11. ROC curve on blind data for Iteration 1 prioritized dig list..... 23
Figure 12. ROC curve on blind data for Iteration 2 prioritized dig list..... 24
Figure 13. Example ROC curve that illustrates cost savings due to skillful classification ..... 29
Figure 14. Conceptual cost model illustrating potential savings from a skillful UXO discrimination project assuming a stop-digging threshold 50% of the way through prioritized dig list. ..... 30

## LIST OF TABLES

**Page**

Table 1. Performance objectives summary for Camp Sibert. ..... 7
Table 2. Performance objectives summary for Camp SLO. ..... 7
Table 3. Comparison of Iteration 1 and Iteration 2 results ..... 25
Table 4. Cost model for a detection/discrimination survey technology. ..... 27

## ACRONYMS AND ABBREVIATIONS

| | |
|---|---|
| AUC | area under the curve (of an ROC curve) |
| BRAC | base realignment and closure |
| CFS | correlation-based feature selection |
| CWS | Chemical Warfare Service |
| DGM | digital geophysical mapping |
| DoD | Department of Defense |
| EMI | electromagnetic induction |
| ESTCP | Environmental Security Technology Certification Program |
| FUDS | formerly used defense site |
| GPS | Global Positioning System |
| HE | high explosive |
| HRR | Historical Records Review |
| IDA | Institute for Defense Analyses |
| LGP | Linear Genetic Programming |
| MAGMTADS | magnetometer MTADS array |
| MEC | munitions and explosives of concern |
| MMRP | Military Munitions Response Program |
| MRMR | maximum relevance minimum redundancy |
| MTADS | Multisensor Towed Array Detection System |
| $N_{fa}$ | number of false alarms |
| $P_{class}$ | probability of correct classification |
| QA | quality assurance |
| QC | quality control |
| RML | RML Technologies, Inc. |
| ROC | receiver operating characteristic |
| RTK | real-time kinematic |

| | |
|---|---|
| SAIC | Science Applications International Corporation |
| SERDP | Strategic Environmental Research and Development Program |
| SLO | San Luis Obispo |
| TOI | target of interest |
| UXO | unexploded ordnance |

# ACKNOWLEDGEMENTS

This work was funded by the Environmental Security Technology Certification Program (ESTCP) office, under MM-0811.

The work was performed by a joint team of personnel from RML Technologies, Inc. (RML) and Science Applications International Corporation (SAIC).

We gratefully acknowledge Dr. Herb Nelson, program manager of ESTCP's Munitions Response focus area, and Dr. Jeff Marqusee, ESTCP Director, for their vision, support, and guidance throughout the life of this project.

Finally, we thank C. Edward Dilkes for his financial support and patience over the years in the development of this technology.

The principal investigators for this project are Frank Francone, RML; Dean Keiswetter, SAIC; and Larry M. Deschaine, SAIC (year one only). Key technical contributors were Phillip Stanwood, RML; Stanley Wong, SAIC; and Tom Furuya, SAIC.

*Technical material contained in this report has been approved for public release.*
*Mention of trade names or commercial products in this report is for informational purposes only; no endorsement or recommendation is implied.*

*This page left blank intentionally.*

# 1.0 EXECUTIVE SUMMARY

This report describes a 2-year unexploded ordnance (UXO) classification demonstrating the application of the Linear Genetic Programming (LGP) Discrimination Process™ to the problem of UXO discrimination and residual risk analysis. In support of project objectives, we analyzed multisensor electromagnetic and magnetic data acquired at two live sites.

The objective of this project was to discriminate a variety of potentially hazardous munitions from items that may be safely left in the ground. At former Camp San Louis Obispo (SLO) the targets of interest (TOI) included 60 mm mortars, 81 mm mortars, 2.36-inch rockets, and 4.2-inch mortars. At former Camp Sibert, the lone TOI was a 4.2-inch mortar.

The LGP Discrimination Process™ begins with the digital geophysical mapping (DGM) from a site suspected of containing UXO. It then (1) extracts attributes from the DGM near potential targets that may be UXO, (2) uses LGP and the attributes to rank the potential targets in their order of likelihood of being UXO, and (3) applies statistical residual risk analysis to determine which of the ranked targets may be safely left in the ground as Not-UXO.

The attributes extracted for each target are analyzed by information-theoretic and statistical methods to reduce the attribute set to a handful of highly predictive attributes. Then, LGP is used to rank the "blind" targets as either UXO or Not-UXO using a small "training" set of targets for which ground truth was provided. Finally, statistical residual risk analysis is applied to the rankings and to the training ground truth to determine the stop-digging cutoff.

For data acquired at Sibert, 100% of the UXO and 89.6% of the non-UXO were correctly classified. For data acquired at SLO, the LGP process correctly classified 98.6% of the UXO and 35.9% of the non-UXO.

Finally, the intention in this project was to test an iterative process that would be very useful in actual Military Munitions Response Program (MMRP) site cleanups. It is based on the fact that DGM and ground truth do not come in all at once in actual cleanups. Accordingly, the first iteration of LGP rankings and risk analysis was used to sample further ground truth. That further ground truth would be used as the basis for additional LGP ranking and risk analysis. That process would have iterated until a stop-digging decision was reached. The goal of iteration was to improve the receiver operating characteristic (ROC) charts and to improve the accuracy of the stop-digging cutoff with additional ground truth.

For data acquired at Sibert, no iterations were required because the original classification was nearly perfect. At SLO, the sampling of additional ground truth for a second iteration of discrimination and risk analysis very significantly improved the performance of the technology over the first iteration by almost any metric. In other words, intelligently selecting which targets to "dig" and then rebuilding discrimination models using those new targets as training targets significantly improved UXO discrimination results and the accuracy of our residual risk assessment.

*This page left blank intentionally.*

# 2.0 INTRODUCTION

## 2.1 BACKGROUND

In 2003, the Defense Science Board observed: "The … problem is that instruments that can detect the buried UXOs also detect numerous scrap metal objects and other artifacts, which leads to an enormous amount of expensive digging. Typically 100 holes may be dug before a real UXO is unearthed! The Task Force assessment is that much of this wasteful digging can be eliminated by the use of more advanced technology instruments that exploit modern digital processing and advanced multi-mode sensors to achieve an improved level of discrimination of scrap from UXO."[1] The FY06 Defense Appropriation contains funding for the "development of advanced, sophisticated discrimination technologies for UXO cleanup" in the Environmental Security Technology Certification Program (ESTCP).

Significant progress has been made in discrimination technology. To date, these technologies have primarily been tested at constructed test sites, with only limited application at live sites. The routine implementation of discrimination technologies will require demonstrations at real UXO sites under real world conditions.

## 2.2 OBJECTIVE OF THE DEMONSTRATION

Our objective was to advance and improve munitions and explosives of concern (MEC) discrimination performance by validating a decision process that (1) combines statistical analyses of DGM products and LGP methods to enable classification and (2) provides iterative quantitative residual risk assessments that may be used during the excavation phase to determine a stop-digging cutoff. In addition, we sought to test an iterative UXO discrimination and risk analysis process by intelligently sampling selected ground truth for Iteration 2, using the results from Iteration 1.

## 2.3 REGULATORY DRIVERS

Senate Report 106-50, pages 291–293, accompanying the *National Defense Authorization Act for Fiscal Year 2000* (Public Law 106-65),[2] included a provision entitled "Research and development to support UXO clearance, active range UXO clearance, and explosive ordnance disposal." This provision requires the Secretary of Defense to submit to the Congressional defense committees a report that gives a complete estimate of the current and projected costs, to include funding shortfalls, for UXO response at active facilities, installations subject to base realignment and closure (BRAC), and formerly used defense sites (FUDS).

In 2001, the Department of Defense (DoD) reported to Congress: "Decades of military training, exercises, and testing of weapons systems has required that we begin to focus our response on the challenges of UXO . . . . This report provides a UXO response estimate in a range between $106.9 billion and $391 billion in current year [2001] dollars . . . . Technology discovery, development, and commercialization offer some hope that the cost range can be decreased . . . ."[3]

*This page left blank intentionally.*

# 3.0 TECHNOLOGY

## 3.1 TECHNOLOGY DEVELOPMENT

This technology has not been previously developed under grant from ESTCP. Before ESTCP's involvement, the technology was in development since approximately 2002, when Science Applications International Corporation (SAIC) applied RML Technologies, Inc.'s (RML) LGP software to the publicly available data from the Jefferson Proving Grounds IV UXO demonstration test bed.[4] Our UXO discrimination results were by far superior to the best reported results from the demonstrators on these data.[5] Accordingly, using internal financing, RML and SAIC developed and applied an early version of the LGP Discrimination Process to the Jefferson Proving Grounds V[6] EM61 MK2 test bed data. We reported those results in 2004.[7] In addition, in 2004, we developed and reported a technique for iteration through successive rounds of classification using information theoretic methods to select targets at each iteration for improving UXO discrimination performance in subsequent iterations.[8] Then, in 2006, in support of a remedial investigation performed by URS Corporation for F.E. Warren Air Force Base, we applied this technology to production-grade data from an EM61 MK2 to approximately 30,000 TOIs. The result was successful discrimination of all 75 mm and 37 mm projectiles from clutter and a stop-digging threshold that correctly identified a large proportion of all targets as high-confidence Not-UXO.[9]

## 3.2 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY

Key differences between LGP and other learning algorithms are: (1) LGP does not just derive parameters for a specified functional form—it derives the functional form itself and optimizes the parameters of the derived functional form, in one pass; (2) Because LGP software operates directly on populations consisting of Intel machine code functions, it is approximately two orders of magnitude faster than comparable inductive-learning technologies;[10] (3) LGP software has been subjected to extensive in-house and third-party testing on a wide variety of data sets over a 9-year period. Results have been published by RML and SAIC[11] and by third-parties[12]; (4) LGP was designed to prevent, insofar as possible, building models of the training-set noise rather than the signal sought to be modeled. LGP's resistance to fitting noise has been noted in the literature; and (5) The version of Discipulus used in this project uses as its fitness function, the area under the curve (AUC) of the ROC curve defined by the evolved program ranking. In other words, the evolution process is geared toward creating a good ranking. Most other inductive learning algorithms perform some kind of classification and then convert that into a ranking.

A disadvantage of LGP is that it requires experienced data modelers for its operation. It is a very powerful modeling tool because of the breadth of the search it can conduct over a very large solution space—both because of its speed and because it evolves functional form, not just parameterization of a preexisting functional form. If used improperly, it can produce wonderful-looking results on known data and very poor results when applied to new data.

*This page left blank intentionally.*

## 4.0 PERFORMANCE OBJECTIVES

The relevant objectives for Camp Sibert included: (1) TOI retention rate, (2) non-TOI reduction rate, and (3) analysis time (Table 1).

**Table 1. Performance objectives summary for Camp Sibert.**

| Performance Objective | Metric | Data Required | Success Criteria | Result |
|---|---|---|---|---|
| TOI retention rate | Percent TOI correctly classified as TOI at demonstrator stop-digging recommendation | 1. Prioritized dig list<br>2. Excavation results or scoring report | >0.95 | Success |
| Non-TOI reduction rate | Number of false targets eliminated at demonstrator stop-digging recommendation | 3. Prioritized dig list<br>4. Excavation results or scoring report | >40% | Success |
| Analysis time | Person-days in production until stop-digging recommendation | 5. Log of data analysis time | < 60 person-days | Success on two of the three tracks |

The relevant objectives for Camp SLO included: (1) maximize TOI retention rate, (2) maximize non-TOI reduction rate, (3) specification of stop-digging threshold; (iv) minimize number of targets that cannot be analyzed; and (5) minimize the number of blind targets sampled (Table 2).

**Table 2. Performance objectives summary for Camp SLO.**

| Performance Objective | Metric | Data Required | Success Criteria | Result |
|---|---|---|---|---|
| Maximize correct classification of munitions | Number of TOIs retained | Prioritized anomaly lists and scoring reports from the Institute of Defense Analyses (IDA) | Approach correctly classifies 100% of TOIs | Correctly classified 98.6% of TOIs |
| Maximize correct classification of non-munitions | Number of false alarms ($N_{fa}$) eliminated | Prioritized anomaly lists and scoring reports from IDA | Reduction of false alarms by >30% while retaining all TOIs | False alarm rate reduced by 28.4% while retaining all TOIs |
| Specification of no-dig threshold | Probability of correct classification ($P_{class}$) and $N_{fa}$ at demonstrator operating point | Demonstrator specified threshold and scoring reports from IDA | Threshold specified by demonstrator to achieve criteria above | 98.6% of TOIs correctly classified—False alarm rate reduced by 35.9% |
| Minimize number of anomalies that cannot be analyzed | Number of anomalies that must be classified as "unable to analyze" | Demonstrator target parameters | Reliable target parameters can be estimated for >90% of anomalies | Reliable target attributes estimated for 82% of targets |
| Minimize the number of blind targets sampled | Number of targets sampled in the second and subsequent iterations | Requests for ground truth on second and subsequent iterations initial blind data list | Requested ground truth for sampling does not exceed 20% of initial blind targets in the aggregate | 20% of blind targets sampled |

The main failure is misclassifying a TOI as an item that can be left in the ground. Items that may be safely left in the ground included high explosive (HE) fragments, single fins, cultural debris and geology.

# 5.0 SITE DESCRIPTION

The former Camp Sibert consists of mainly sparsely inhabited farmland and woodland and encompasses approximately 37,035 acres near Gadsden, AL. The site is located approximately 50 miles northwest of the Birmingham Regional Airport and 86 miles southeast of the Huntsville International Airport.

The former Camp SLO is approximately 2101 acres situated along Highway 1, approximately 5 miles northwest of SLO, CA. Most of the area consists of mountains and canyons. The site for this demonstration is a mortar target on a hilltop.

## 5.1 SITE SELECTION

These two sites were selected by ESTCP as a progression of increasingly more complex sites for demonstration of the classification process. The first site in the series, Camp Sibert, had only one TOI, the 4.2-inch mortar. Camp SLO was the second site chosen and contained four TOIs: 60 mm, 81 mm, 4.2-inch mortars, and 2.36-inch rockets.

## 5.2 SITE HISTORY

Camp Sibert was acquired in July 1942 by the U.S. Army as a replacement training center for the Chemical Warfare Service (CWS). At Camp Sibert the CWS conducted various training exercises such as smoke screen defense, chemical decontamination, chemical depot maintenance, and chemical impregnation of clothing. Chemical troops equipped the camp with chemical field filling stations, a toxic gas yard, and decontamination areas. The camp was closed at the end of the war in 1945, and the chemical school transferred to Fort McClellan, AL. The Army declared the property excess and transferred it to the War Assets Administration on November 18, 1946, and then to the Farm Mortgage Corporation. The government terminated the leases on the area on December 13, 1946. After decontamination of the various ranges and toxic areas in 1948, the land was transferred back to private ownership. The airfield, however, was transferred to the City of Gadsden.

Camp SLO was established in 1928 by California as a National Guard Camp. Identified at that time as Camp Merriam, it originally consisted of 5800 acres. Additional lands were added in the early 1940s until the acreage totaled 14,959. During World War II, Camp SLO was used by the U.S. Army from 1943 to 1946 for infantry division training that included artillery, small arms ranges, mortar, rocket, and grenade ranges. According to the Preliminary Historical Records Review (HRR), a total of 27 ranges and thirteen training areas were located on Camp SLO during World War II. The U.S. Army used the former camp during the Korean War from 1951 through 1953 where the Southwest Signal Center was established for Signal Corps training. The HRR identified 18 ranges and 16 training areas present at Camp SLO during the Korean War. A limited number of these ranges and training areas were used previously during World War II. Following the Korean War, the camp was maintained in inactive status until it was relinquished by the Army in the 1960s and 1970s.

## 5.3 MUNITIONS CONTAMINATION

The munitions-of-concern at Camp Sibert were 4.2-inch mortars.

At the former Camp SLO study site, 60 mm mortars, 81 mm mortars, 2.36-inch rockets, and 4.2-inch mortars and mortar fragments had been observed before the demonstration.

# 6.0 TEST DESIGN

## 6.1 CONCEPTUAL EXPERIMENTAL DESIGN

The principal objective was to demonstrate an iterative methodology for the use of classification and risk analysis in the munitions response process. The focus was to identify items that may be safely left in the ground.

The ESTCP Program Office coordinated data collection and validation digging activities. All anomalies on the master dig list were investigated. The identities of a small number of the recovered items plus the DGM were provided to the demonstrator for use as "training" data. The identities of the remainder of the targets were retained by the Program Office as "blind" data to validate demonstrator's results.

The demonstrator received and processed the DGM data extract attributes for each Program Office designated target. The project was designed to proceed iteratively. Demonstrator would produce a prioritized dig list for all then "blind" targets, a stop-digging threshold and a probability that any UXO remained on the site, given the then known ground truth and the stop-digging threshold. Demonstrator would then request further ground truth for some of the currently "blind" targets, produce a new dig list and stop-digging threshold, given the then known ground truth. Demonstrator expected and performed two such iterations.

## 6.2 SITE PREPARATION

Before the start of the surveys, each site was seeded with examples of the items of interest under the guidance of the Program Office Seeding Plan. A Calibration Strip containing two of each item of interest and a selection of canonical objects (e.g., metal spheres) was installed near the demonstration site and the site logistics location.

## 6.3 SYSTEM SPECIFICATIONS

This data were acquired using the Naval Research Laboratories' Multisensor Towed Array Detection System (MTADS) the magnetometer MTADS array (MAGMTADS), and EM61 arrays[13] (EM61MTADS). The MTADS hardware consists of a low-magnetic-signature vehicle that measures position, roll, pitch and yaw with great accuracy and that is used to tow different sensor arrays over large areas (10-25 acres/day) to detect buried UXO. The EM61MTADS array is MTADS hardware configured to contain a specially modified EM61 MkII sensor, configured with an overlapping array of three pulsed-induction sensors consisting of 1 m$\times$1 m coils. These data were collected with the EM61MTADS in four-channel mode using delay-time configuration for the four channels of 307, 508, 738, and 1000 µs, respectively. MAGMTADS consists of MTADS hardware configured to contain a linear array of eight geometrics Cs-vapor magnetometer sensors (Geometrics, Inc., G-822ROV/A).

## 6.4 DATA COLLECTION PROCEDURES

EM61MTADS data were collected with nominal down-track spacing of 15 cm and cross track spacing of 50 cm. Because the three transmitters in the EM61MTADS array are synchronized,

data are collected in two orthogonal directions to increase the number of “looks” or directions of illumination of each anomaly by the array.

Magnetometer data were collected with nominal down-track spacing of 10 cm and cross-track spacing of 25 cm. Location of the sensor was measured by real-time kinematic (RTK) Global Positioning System (GPS) receivers.

## 6.5 VALIDATION

After data collection activities, all anomalies (targets) on the master anomaly list assembled by the Program Office were excavated. Each item encountered was identified, photographed, its depth measured, its location determined using cm-level GPS, and the item removed if possible. All nonhazardous items were saved for later in-air measurements as appropriate.

# 7.0 DATA ANALYSIS AND PRODUCTS

## 7.1 DESCRIPTION OF DATA

We received as input fully processed spatially registered EM61 and magnetometer data. We also received target locations and IDs from the ESTCP Program Office and ground truth labels (for training purposes).

## 7.2 OVERVIEW OF PROCEDURES

We took the following steps in this project in this order:

1. Applied data quality assurance (QA)/quality control (QC) and preprocessing.
2. Identified cannot-analyze targets.
3. Characterized each target with a parameterized ellipse.
4. Extracted attributes that characterize each target from the ellipses.
5. Performed modeling and risk analysis iteration.
    a. Built a simple prediscriminator.
        i. Attributed reduction.
        ii. Performed residual risk analysis for prediscriminator.
        iii. Assigned low risk targets to do-not-dig.
    b. Built LGP discriminator on remaining targets.
    c. Performed residual risk analysis on LGP rankings.
    d. Produced Iteration 1 prioritized dig list.
6. Requested and received ground truth for selected blind targets.
7. Performed second modeling and risk analysis iteration (same steps as Iteration 1).

These steps are described in more detail below.

## 7.3 DEFINE TARGET POLYGONS AND ELLIPSES FOR EACH TARGET

We first defined a polygon for each program office target. Figure 1 is an example of such a polygon. We then converted the polygons into ellipses, which defined the spatial region occupied by the target for the remainder of the project.

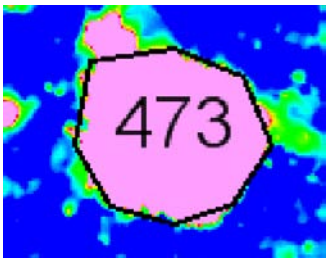



**Figure 1. A target polygon.**

## 7.4 REMOVE CANNOT-ANALYZE TARGETS

We identified targets for which good discrimination was not possible using several criteria: (1) overlapping targets, (2) targets with missing sections of DGM, (3) targets with local data inconsistency, and (4) targets with insufficient DGM density to support a conclusion (not enough data points in the ellipse or one of the measured regions of the ellipse).

Figure 2 is a picture of nine targets that were labeled “cannot-analyze” targets because of target overlap. The red polygons show our attempt to separate them from each other, in our judgment, unsuccessfully.

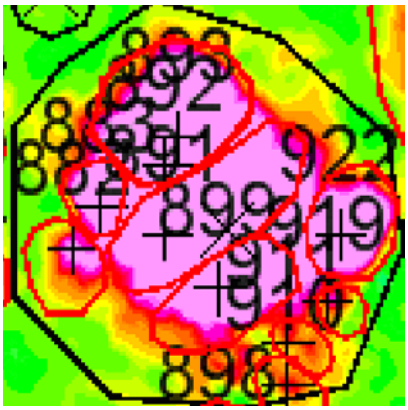


**Figure 2. Example of a cannot-analyze one blob.**

## 7.5 ATTRIBUTE EXTRACTION

Attribute extraction is the process of converting the DGM in the vicinity of a picked target into meaningful statistics about the target. For this project, we extracted and used three types of attributes:

- Attributes that measure a statistic of the amplitude of the signal value of a single channel (Amplitude Statistics)
- Attributes that measure the ratio as between two different channels of Amplitude Statistics (Ratio Statistics)
- Attributes that measure the ratio of adjacent Ratio Statistics (Rate of Change Statistics).

Attributes were calculated on the DGM data points within different regions around the target. Figure 3 illustrates those regions. The ellipse in that figure is the entire ellipse as defined above around the target. The red and blue regions are sub regions in the ellipse from which features are extracted from the DGM data points contained therein.

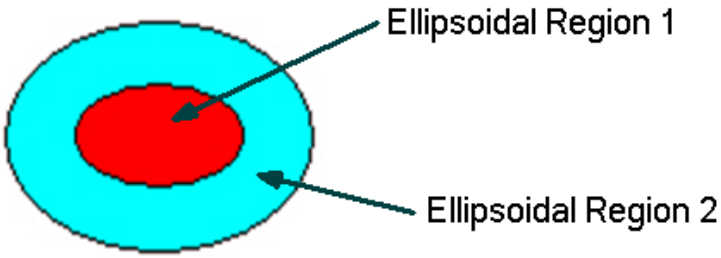



**Figure 3. A simple illustration of ellipsoidal rings for attribute extraction.**

The attributes calculated for each target consisted of the first three moments calculated for each of the different regions around the target, including the entire ellipse and the two subregions as follows:

1. For Amplitude Attributes: The value for channels 1, 2, 3, 4, and sum
2. For Ratio Attributes: The values for all possible ratios between the DGM value for channels 1,2,3, and 4
3. For Rate of Change Attributes: The value of all ratio attributes, respecting the decay order of the channels (e.g., ratio of Channel 1 to Channel 2/ratio of Channel 2 to Channel 3).
4. The result of this process is hundreds of attributes for each target. They are inserted into a control database and used for subsequent analysis.

## 7.6 DESCRIPTION OF A MODELING AND RISK ANALYSIS ITERATION

Each iteration of modeling and risk analysis proceeds in the following steps: (1) Filter out easy-to-find high-probability Not-UXO with a simple prediscriminator; (2) Rank all remaining targets with an LGP ensemble predictor; (3) Set a stop-digging threshold for the ranked targets using residual risk analysis.

We begin our filtering out easy-to-find, high-probability Not-UXO by surveying the existing attributes for the training targets. Using Mutual Information and Chi-Square Binning, we reduce those attributes to a single attribute that ranks the training targets in order of likelihood that the target is UXO. This ranking is the prediscriminator.

At that point, residual risk analysis is performed on the rankings using kernel regression on the training data, regressing probability of UXO as a function of rank. The blue line in Figure 4 shows the modeled probability of UXO in a simple prediscriminator step for Camp SLO.

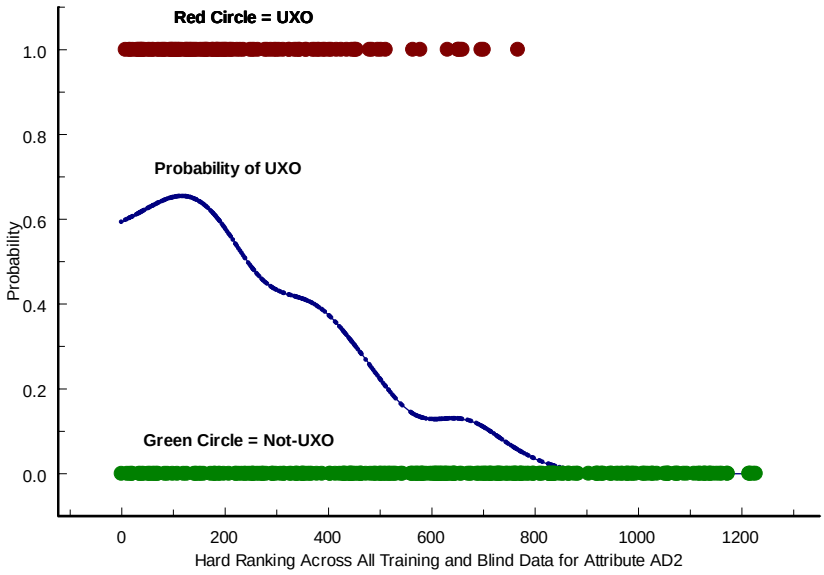



**Figure 4. Prediscriminator model of falling probability of UXO as a function of rank for a simple prediscriminator.**
Red circles show rankings of known UXO in training data. Green circles show rankings of known Not-UXO.

The resulting kernel regression function is then applied to the blind data and we then assess the cumulative probability that UXO remains on site were we to stop digging at each ranked blind target. The ranking at which that probability falls below 0.05 for the entire project is selected as the stop-digging threshold for that step. All targets below that rank may be assigned as high-probability Not-UXO. Figure 5 shows the application of the kernel regression model to the blind data at Camp SLO. The red line shows the cumulative probability at each rank that UXO remains on site. So at the 95% confidence level, we would set the stop-digging threshold between rank 900 and rank 1000. Targets above that rank would be assigned as high-probability Not-UXO.

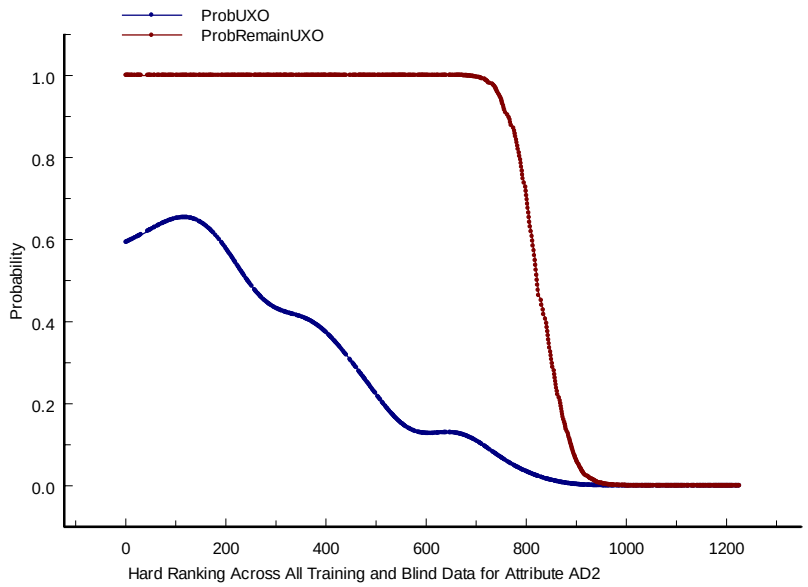



**Figure 5. Prediscriminator model of falling probability of UXO applied to blind data.**

Remaining targets are then the subject of LGP discrimination. To apply LGP, we first reduce our attribute set for the remaining targets to a handful of highly predictive attributes using a collection of tools to reduce attributes. The tools include (1) numeric input binning, (2) maximum relevance minimum redundancy (MRMR), (3) correlation-based feature selection (CFS), (4) decision trees, and (5) Discipulus™ input impacts analysis. These are all well-understood machine-learning and data-mining techniques.

The selected attribute set is then modeled using LGP. To protect against overfitting, we added noise to the training data, used cross-validation to set key LGP parameters, and then generated our discrimination model using bagging techniques.

At the end of this process, we had constructed an LGP ensemble predictor, consisting of 30-50 evolved programs from LGP, each of which had been trained on a different bagged sample from the training data set. The outputs from those thirty programs was reduced to a single predictor for the training and blind targets.

At this point, the prediction for each target is used as a ranking for a residual risk analysis step using kernel regression. A stop-digging threshold is set using the cumulative probability of remaining UXO discussed above. Figure 6 shows the probability models after one of our LGP modeling steps on the blind data. In this figure, the stop-digging threshold would be set at about 600 at the 95% confidence level, and all targets below that would be assigned as high-probability Not-UXO.

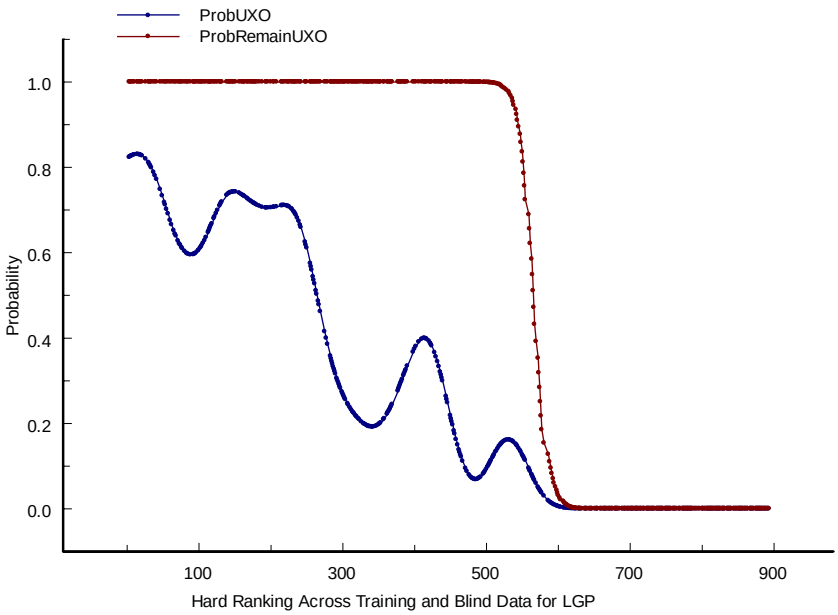



**Figure 6. LGP model of falling probability of UXO as a function of rank after LGP modeling.**

## 7.7 SAMPLING OF ADDITIONAL GROUND TRUTH

When we have finished an iteration of discrimination modeling, the results let us intelligently select specific targets for sampling to help us build better models in the next iteration. We use the probabilities from the risk analysis from the previous iteration (the blue line in Figure 6 would be an example of those probabilities) to make that intelligent selection. This would be the equivalent on an actual site cleanup of requesting that additional targets be dug and then including those targets in additional discrimination steps and risk analysis. As more well-selected targets come in, the models and risk analysis should improve.

Sampling additional ground truth between iterations was performed based on four criteria:

1. *Entropy*. Entropy is a measure of the uncertainty of a target for which ground truth is unknown.
2. *Entropy per Unit of Expected Cost of Sample*. Entropy per unit of expected cost is a criterion designed to get looks at likely UXO at the lowest possible cost. In other words, entropy measures expected information content, and expected cost measures the likelihood that we are digging Not-UXO. Thus entropy per unit of expected cost looks for the targets that provide “cheap” information.
3. *Visual Picks around Training Outliers*. In this project, during the Iteration 1 training, three training UXO targets consistently stood out as more difficult to discriminate than the remainder. We picked blind targets manually in the immediate vicinity of these targets, in attribute space to sample. Figure 7 shows the three outliers in red. The brown ellipse designates the region around those outliers from which we sampled.

4. *Random Sample from Tail of Risk Analysis Probability*. The rankings on our dig list between the last training UXO and the dig threshold comprise a region in which we wish to acquire more information so that the tail of the declining probability is better defined.

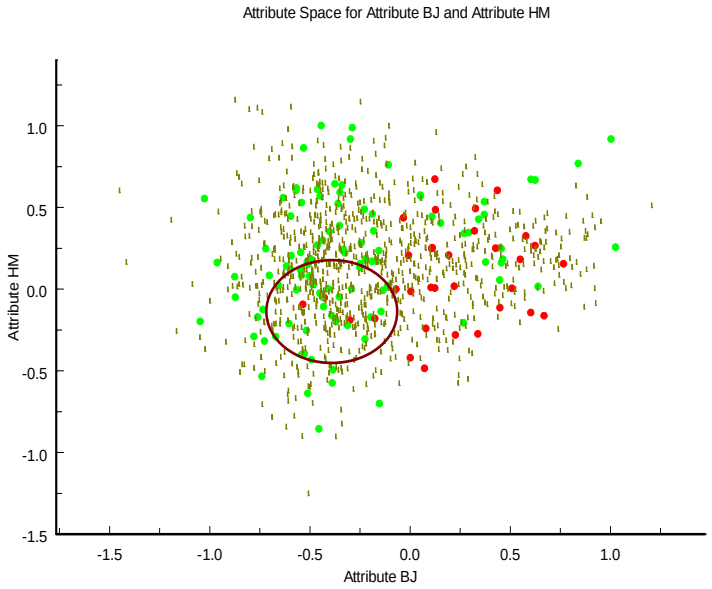



**Figure 7. Region of selection of blind targets for sampling around an outlier UXO.**

*This page left blank intentionally.*

# 8.0 PERFORMANCE ASSESSMENT

## 8.1 CAMP SIBERT

We submitted three dig lists for scoring our analysis of the Camp Sibert data. One dig list was based on our analysis of the EM61 data alone, a second on EM61 and magnetometer data, and a third based on intrinsic magnetic polarizabilities derived from the EM61 data. The following sections show the ROC curves on the blind data for Camp Sibert.

### 8.1.1 EM ONLY

Figure 8 shows that all TOIs were retained above our stop-digging threshold. In other words, we found and dug all UXO. Therefore, this track was a success on this metric.

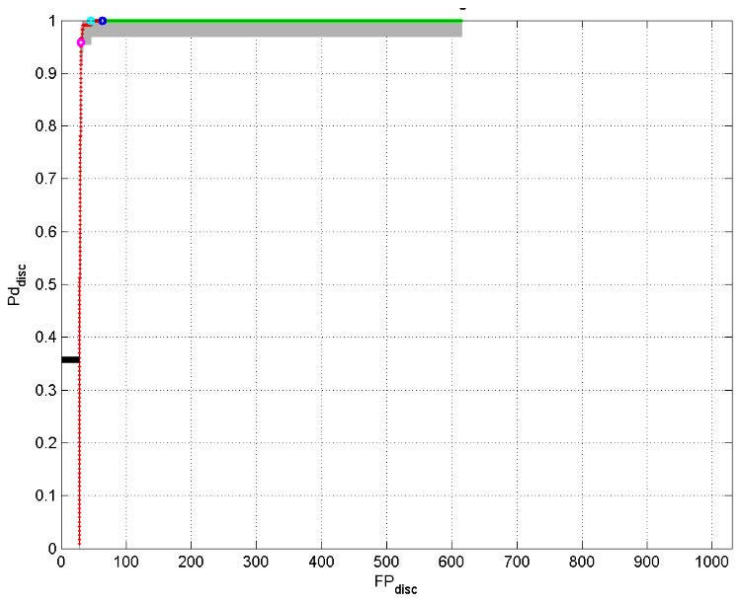



**Figure 8. ROC chart showing blind scoring for EM-only track.**

As noted above, the black line on the left of Figure 8 highlights the cannot-analyze targets. Approximately 4% of the blind targets (29 targets) were classified as cannot-analyze.

Once we started classifying targets (the near-vertical red line that starts at about FP=29), we generated a near-perfect ROC chart—that is, almost all UXO were ranked above all non-UXO; 89.6% of the non-UXO were correctly classified.

### 8.1.2 MAG AND EM

As noted above, the black line on the left of Figure 9 highlights the cannot-analyze targets. Approximately 7% of all blind targets (86 targets) were classified as cannot-analyze.

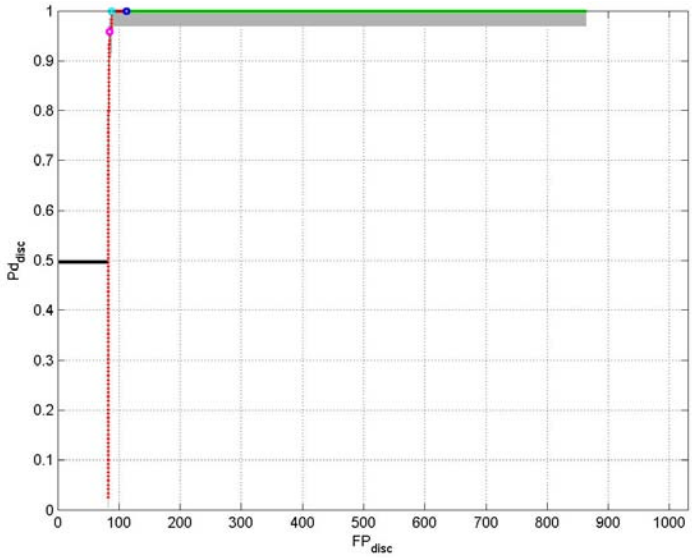



**Figure 9. ROC chart showing blind scoring for combined track.**

Once we started classifying targets (the near-vertical red line that starts at about FP=86), we generated a near-perfect ROC chart—that is, almost all UXO were ranked above all Not-UXO.

The light blue circle shows the final UXO item prioritized on our inversion track dig list. The dark blue circle shows our stop-digging threshold. The key point to draw from these two data is that all UXO were above the stop-digging threshold. That is, no UXO were left in the ground; 86.8% of the non-UXO were correctly classified.

### 8.1.3 INVERSION FEATURES

The black line on the left of Figure 10 highlights the cannot-analyze targets for this track. Approximately 26% of all blind targets (260 targets) were classified as cannot-analyze.

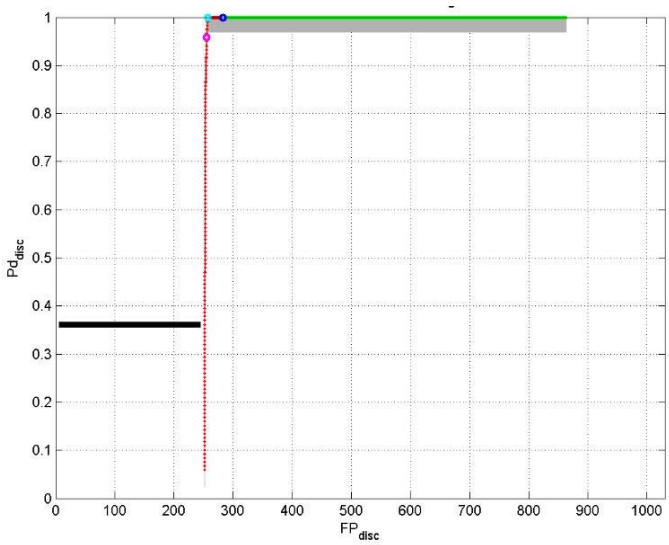



**Figure 10. ROC chart showing blind scoring for inversion track.**

Once we started classifying targets (the near-vertical red line that starts at about FP=260), we generated a near-perfect ROC chart—that is, almost all UXO were ranked above all non-UXO.

The light blue circle shows the final UXO item prioritized on our inversion-track dig list. The dark blue circle shows our stop-digging threshold. The key point to draw from these two data is that all UXO were above the stop-digging threshold. That is, no UXO were left in the ground.

Therefore, this track was a success on this objective, which was 100% retention of TOIs (UXO); 67.1% of the non-TOIs were correctly classified.

## 8.2 CAMP SLO

We submitted two prioritized dig lists—one for each of two iterations—for Camp SLO, based on our analysis of EM61 data. The following sections show the ROC curves generated on the Camp SLO blind targets in both Iteration 1 and Iteration 2. Note that the target set gets smaller from Iteration 1 to Iteration 2. The reason for this is that, after Iteration 1, about 200 blind targets were sampled for ground truth to improve the classification (that is, we learned the ground truth for the targets). Thus, for Iteration 2, those targets had to be and were treated as training targets, not as blind targets any longer.

### 8.2.1 ITERATION 1

Figure 11 shows the ROC curve generated by our prioritized dig list on the blind targets for Iteration 1 at SLO.

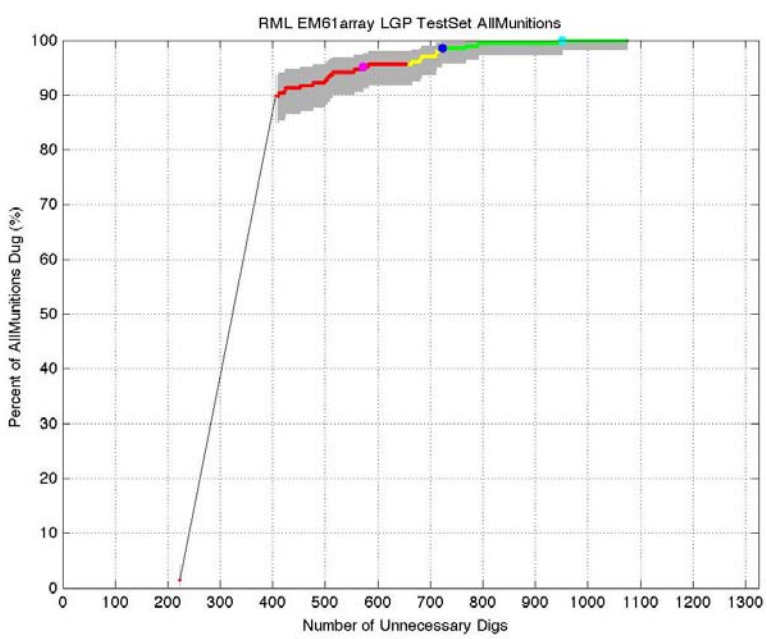



**Figure 11. ROC curve on blind data for Iteration 1 prioritized dig list.**

In this figure, the gray line starts at approximately 220 on the x-axis. That represents all cannot-analyze targets for this iteration. The gray line represents the top-ranked targets on our dig list. They were tied for "first-place." What the gray line indicates is that in the first 180 targets on our dig list, we located 90% of the UXO. The dark blue circle is the point at which we set the stop-digging threshold, and the green line is all targets below the stop-digging threshold. The final UXO was located at the light blue circle at about ranking 950 on the x-axis. Altogether, 98.6% of UXO were ranked above the stop-digging threshold and 1.4% were ranked below the stop-digging threshold.

The areas under the curve for this ROC chart may be measured in two ways. A perfect (or vertical) ROC curve has an AUC of 1.0.

1. Including the cannot-analyze targets, the AUC is 0.683.
2. Including only targets we ranked with our discriminators, the AUC is 0.858.

### 8.2.2 ITERATION 2

Figure 12 is the ROC chart showing the performance of our process on the reduced blind-data set for Iteration 2.

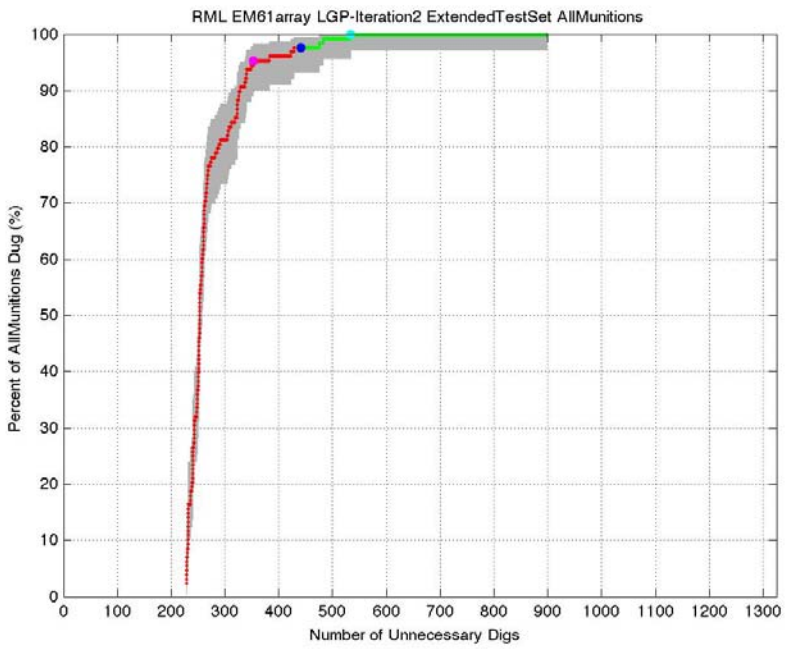



**Figure 12. ROC curve on blind data for Iteration 2 prioritized dig list.**

In this figure, each red dot represents a UXO located on our dig list. The first one is shown at approximately 220 on the x-axis. That gap before 220 represents all cannot-analyze targets for this iteration. This chart shows that we located 90% of the UXO in the first 100 targets ranked by our LGP ensemble predictor or the amplitude discriminator. The dark blue circle in this figure is the point at which we set the stop-digging threshold, and the green line represents all targets below the stop-digging threshold. The final UXO was located at the light blue circle at about

ranking 540 on the x-axis. Altogether, 98.6% of UXO were ranked above the stop-digging threshold and 1.4% were ranked below the stop-digging threshold.

The AUCs for this ROC chart may be measured in two ways.

1. Including the cannot-analyze targets, the AUC is 0.703.
2. Including only targets we ranked with our discriminators, the AUC is 0.936.

### 8.2.3 SAMPLING OF GROUND TRUTH BETWEEN ITERATIONS

Our iterative modeling approach in this project is, as far as we know, unique. The results were quite dramatic. The ROC charts for these two iterations are Figure 11 and Figure 12, respectively. The nearly vertical ROC chart for Iteration 2 is clearly greatly superior to the ROC chart from Iteration 1.

In every respect, the Iteration 2 using the larger training set was superior to or equal to Iteration 1. Table 3 shows that comparison.

**Table 3. Comparison of Iteration 1 and Iteration 2 results.**

| Criterion | Iteration 1 | Iteration 2 |
|---|---|---|
| AUC | 0.858 | 0.936 |
| Count of not-UXO left in ground after last UXO | 124 | 364 |
| Percent not-UXO left in ground after stop-digging | 27.59% | 35.88% |
| False negatives | 3 | 3 |
| False negatives other than mistaken cannot-analyze | 3 | 2 |

The count of Not-UXO ranked below the final UXO approximately tripled while the amount of Not-UXO ranked lower than the stop-digging threshold increased by about 30%.

In addition, the increase in the AUC from Iteration 1 to Iteration 2 is very substantial. The error implied by the AUC is more than halved.

In short, the intelligent sampling of new ground truth between modeling iterations improved the UXO classification significantly by several metrics.

The 200-target request for ground truth between iterations was expected to yield 157.3 Not-UXO and 98.7 UXO. These were straightforward predictions from our Iteration 1 probabilistic risk analysis models. When we received the ground truth from the Program Office, the actual distribution was 162 Not-UXO and 94 UXO. This close match between predicted and actual is a strong validation of the usefulness of the residual risk analysis approach to analyzing a prioritized dig list.

*This page left blank intentionally.*

# 9.0 COST ANALYSIS

The cost reductions in a typical large cleanup project, given these results, would have been quite substantial. Figure 12 shows that, based on the blind data at Camp SLO, about 30% of all targets could have been safely left in the ground, depending on the track. Thus, were there 100,000 targets on a project with a similar ratio of TOI to non-TOI and a similar environmental setting, 30,000 targets would fall below the stop-digging threshold. If the hypothetical project had a ratio of TOI to non-TOI and discrimination results similar to Camp Sibert, almost 90% of the non-TOI could have remained unearthed. In other words, 90,000 out of the 100,000 targets could have been left in the ground.

## 9.1 COST MODEL

A cost decision to use this technology would need to balance the added costs to the project of performing discrimination against the cost-savings occasioned by the targets that may be left in the ground as high-probability Not-UXO. The three main elements are:

1. Data collection costs, since data required for classification may cost more to collect than does data used solely for detecting the presence of anomalies
2. Data analysis costs, since analysis requirements for classification are greater than that required for detection
3. Excavation costs, by identifying some percentage as high confidence clutter, we anticipate savings either from digging fewer holes or changing the safety protocols.

**Table 4. Cost model for a detection/discrimination survey technology.**

| Cost Element | Data Tracked During Demonstration | Estimated Costs |
|---|---|---|
| **Discrimination data processing** | Unit: $ cost per anomaly<br>• Average cost per anomaly over four tracks and 2 years<br>• Time required (hours) per anomaly<br>• Personnel required | • $19.15 per anomaly<br>• 0.19 hours per anomaly<br>• Two to three data analysts |

As a practical matter, these measured costs from the project are, in our opinion, much higher than would occur in actual implementation on a real munitions cleanup site. The main difference arises from the following facts: (1) An actual cleanup project might involve 100,000 targets as opposed to the approximately 1000 to 1500 targets at Sibert and SLO and (2) Many of the cost drivers for discrimination would not increase linearly with the number of anomalies (see Section 9.2).

As an example of the expected economies of scale for larger sites, the discrimination and risk analysis technologies reported here were applied in 2006 to data from an actual remedial investigation at F.E. Warren Air Force Base. There were about 30,000 targets on the portion of that site analyzed. The cost per anomaly at Warren was less than $5. The difference between the $19.15 per-anomaly cost reported above and the $5 Warren cost per-anomaly provides some

measure of the economies of scale that accrue in applying these technologies to the larger anomaly lists involved in actual site cleanups.

For prime contractors, the decision criterion for using these technologies in this regard will recognize that they may be economically applied to sites that meet a minimum threshold for anomalies to be analyzed, depending on what portion of anomalies may remain unexcavated because of the use of discrimination technology. That minimum threshold is almost certainly considerably greater than the number of anomalies involved in either Camp Sibert or SLO.

## 9.2 COST DRIVERS

***Data Collection:*** Generally speaking, data collection costs will be greater for classification than for detection only. The electromagnetic induction (EMI) classification process utilizes sometimes subtle changes in the anomaly shape. Care must be taken during data collection to not only sample the anomaly fine enough, but also to not introduce noise due to inappropriate collection methods. The costs for data collection vary widely, depending on site conditions such as topography, vegetation, geologic background, known munitions types, and weather conditions. We did not gather data in this project, so this cost element was not tracked.

***Data Analysis:*** Data analysis costs will be greater for classification than for detection only. Data analysis costs are affected by the presence of complex geology, which can make filtering and parameter estimation more complicated. The munitions of interest will also have a great effect on complexity and costs of processing, as will anomaly density. In the case considered here, only isolated targets were analyzed and target size proved to be a good attribute, but that will not be the case everywhere. The number of non-munitions that can be removed with high confidence at another site may be much lower. In addition, the job of the processor in determining the important features and training the classifier may be harder.

As noted in Section 9.1, the data analysis costs tracked in this project are probably not reflective of what would occur on an actual site cleanup. Per track, we averaged about 1200 targets per track. A portion of our costs require data analyst judgments about issues such as cannot-analyze target selection and feature selection. These costs would scale approximately linearly with the number of anomalies. Thus, they could be expected to increase at about the same rate as number of anomalies. On the other hand, the other analysis costs consist of processing the data through steps and performing QA/QC on the steps. While computer processing time would increase linearly with number of anomalies, analyst time would not increase nearly so quickly. The F.E. Warren costs reported in Section 9.1 may provide some indication of the economies of scale in performing discrimination on larger projects.

Finally, these were the first projects on which we tracked costs per anomaly. Our observation is that costs per target dropped considerably as our experience running the process increased. Accordingly, the numbers provided in Table 4 are probably high in assessing costs for future research and development projects and are almost certainly high for production projects, which would almost always involve many more anomalies than were found in the Sibert or SLO sites.

***Excavation Cost:*** The costs associated with excavating anomalies vary widely and the goal is to reduce these costs via classification. Safety procedures and nominal burial depth drive

remediation costs. When minimal engineering controls are used, costs as low as $45-$90 per dig have been reported. When safety procedures are far more elaborate due either to the type of munitions or to their proximity to high value objects, the costs per dig are measured in the hundreds of dollars. With regards to burial depth, it is less costly to recover shallow, near-surface items than large deep targets. We did not excavate targets in this project, so we did not track these costs.

## 9.3 COST BENEFIT

The cost benefit of the classification approach relates to savings realized by not excavating items that are not of interest. The ROC curve in Figure 13 shows a three-category classification scheme with a threshold set such that all the items on the right are high confidence non-TOI. Although this is an example ROC only, it is very similar in nature to those presented in Figure 11 and Figure 12. Note that the anomalies to the right of the threshold were correctly classified as high confidence not munitions. Cost savings can be realized, therefore, if we make use of the classification information and remediate accordingly, as illustrated in Figure 13.

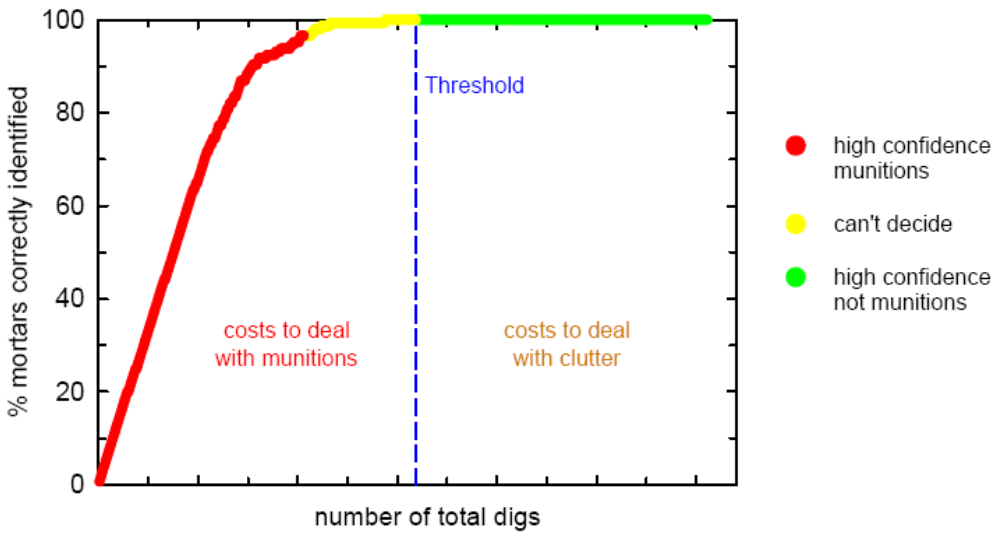



**Figure 13. Example ROC curve that illustrates cost savings due to skillful classification.**

Figure 14 shows how notional costs accumulate through the process of data collection and processing, digging the munitions, and excavation. In the figure, the detection only (solid black line) assumes a lower density data collection for detection only; all anomalies are excavated using intrusive recovery procedures that require trained UXO qualified personnel and safety equipment. The classification 1 (dashed green line) assumes higher density and quality data collection followed by classification processing; all high-confidence clutter items are left unexcavated. Finally, the classification 2 (dotted green line) assumes higher density and quality data collection followed by classification processing, but a less expensive alternative to the current operational methods of intrusive recovery is used on the anomalies determined to be clutter with high confidence.

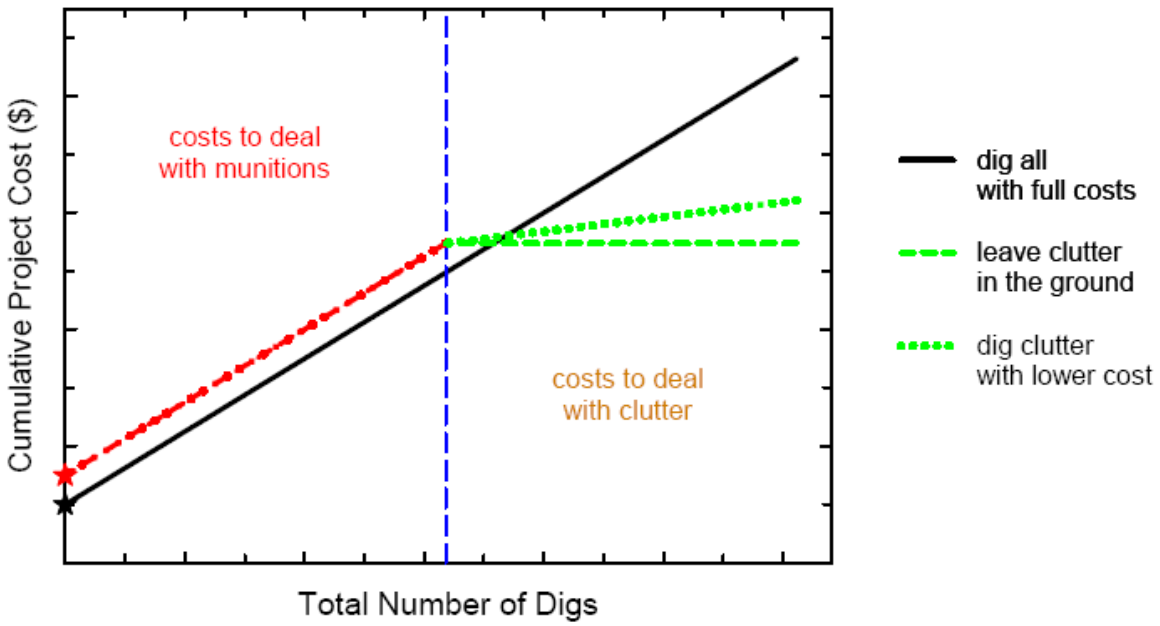



**Figure 14. Conceptual cost model illustrating potential savings from a skillful UXO discrimination project assuming a stop-digging threshold 50% of the way through prioritized dig list.**

The classification examples are tied to the different regions of the ROC curve in Figure 14. There are several important points to note in interpreting this curve:

- The cumulative cost curves start out on the y-axis at different points. This reflects that the initial costs of higher density data collection and processing for classification are higher than the standard methods. The costs of digging the munitions, which must be borne in all cases, are included here.
- The detection-only curve (solid black line) has a constant slope and ends at the total number of anomalies. All detected anomalies are dug using the same procedures at the same costs.
- For both classification examples, all of the items determined to be high confidence munitions or can't-decide must be dug as though they are munitions. Thus, the two classification examples rise at a slope equal to the detection slope until the threshold is reached on the ROC curve where clutter is identified with high.
- In the region where there is high confidence that the remaining anomalies are clutter (green portion of the ROC curve) and it is decided not to dig these anomalies at all, no additional costs are incurred.
- In the region where there is high confidence that the remaining anomalies are clutter and it is decided to dig these anomalies but using alternative dig procedures, additional costs are incurred, but the cost of each of these digs is lower so the slope is more gradual.

- The break point in cost saving will be determined by the true dollars associated with the data collection, processing, and excavation costs—all of which are site-specific. Generally, the more targets on the site, the more cost savings.

The benefits to the participants in the munitions cleanup community are significant.

To begin with, this 2-year project was performed with a technology (LGP), features, and ordering of digs by iteration that are quite different from the standard technologies used for discrimination for munitions response. Its success, along with the success of other demonstrators on the diverse data sets and features sets at Sibert and SLO, represent significant progress toward establishing that information sufficient to solve the UXO discrimination problem exists in the DGM data gathered for cleanup sites and that cost-effective discrimination is possible on real munitions cleanup sites.

These proposals are also the first ESTCP and Strategic Environmental Research and Development Program (SERDP) results using principled entropy-based iteration and residual risk analysis approaches toward discrimination. We believe this is a significant contribution to the community that may improve all existing discrimination technologies. In particular, establishing a solid statistical basis for a stop-digging decision will be a key element in regulatory acceptance of these technologies, and this project is a significant step forward in that regard.

Finally, the demonstrated technologies show significant promise in reducing the number of metallic items that must be excavated to close a site. The number of FUDS that must be cleaned up is quite large and budgets to accomplish that are fixed. The demonstrated technology, if applied to future cleanups, would reduce the excavation costs to close sites substantially and would increase the number of sites that may be closed, given a fixed budget. The end result of this would be that cleanup of our FUDS inventory will take less time and cost less.

*This page left blank intentionally.*

# 10.0 IMPLEMENTATION ISSUES

## 10.1 COST OBSERVATIONS

The discrimination and risk analysis approach demonstrated here utilizes the spatial distribution of the measured EM signatures. As such, it requires high signal-to-noise data with a high degree of spatial precision across the footprint of the anomaly.

The costs to acquire data that will support discrimination decisions are higher than that required if the goal is only to detect the presence of an object. The factors affecting acquisition costs relate to particulars of the sensing system, spatial registration system, the target objectives, and the site environment. Although these costs are not the focus of this demonstration, they are important to the ultimate transferability of this approach.

The analysis costs are also higher if attempts are made to quantitatively classify rather than only to detect. The factors affecting analysis time are significantly affected by (1) the degree to which the anomalies are spatially separated, (2) the number of anomalies, and (3) the amount of geologic related signatures with similar wavelengths as the targeted signatures. The data density is also a factor but only marginally so compared to the factors listed above because it affects computer run time and not analysts' labor.

## 10.2 PERFORMANCE OBSERVATIONS

Discrimination performance is measured by our ability to characterize and classify one object from another. The factors that affect performance, therefore, relate to (1) the similarity (in feature space) between the TOI and non-TOI, (2) our ability to accurately measure the responses, (3) the presence of signatures that spatially interfere or otherwise compete with the UXOs response, and (4) our ability to quantitatively characterize the source objects. Many of these factors are not under our direct control.

The utility of discrimination at a given site is inversely proportionate to the number of can't-analyze targets. The goal is to say something definitive about each anomaly that is selected. In this demonstration, anomalies were selected based on single point amplitudes. The spatial information content was not used during target selection. In an ideal situation, the number of anomalies placed in the can't-analyze category would be zero. The can't-analyze category is necessary in practice, however, because some targets have signal-to-noise ratios that are detectable but not sufficient for data analysis.

## 10.3 SCALE-UP

There are no critical issues with regard to scaling up the demonstration costs reported here to larger, full-scale implementations. The cost categories may not, however, scale linearly. The factors listed in Section 10.1 will determine which, if any, cost categories dominate future technology deployments. Attention should be paid during project planning to the fact that data does not come in all at once on a typical cleanup site. It is often a rolling process where additional data is being constantly acquired over time. This would increase the cost for discrimination and risk analysis by an undetermined amount.

## 10.4 OTHER SIGNIFICANT OBSERVATIONS

There are many technical factors that can affect implementation of the analysis technology discussed in this report. As mentioned earlier, the analysis approach demonstrated here utilizes the spatial distribution of the measured magnetic or EMI signatures. As such, it relies on accurate 3-D spatial measurements as well as on stable geophysical measurements. The measurement of the attitude of the geophysical sensor is also critically important to inverting for meaningful model parameters. If the data going into the inversion routines are noisy or contain systemic problems, the final discrimination decisions will not be acceptable.

## 11.0 REFERENCES


1. Department of Defense. 2003. *Report of the Defense Science Board Task Force on Unexploded Ordnance*. December.

2. Senate Report 106-50, National Defense Authorization Act for Fiscal Year 2000. May 17, 1999. *Research and Development to Support UXO Clearance, Active Range UXO Clearance, and Explosive Ordnance Disposal*. pp. 291–293.

3. Department of Defense. 2001. *Unexploded Ordnance Response: Technology and Cost*, Report to Congress. March.

4. *Advanced UXO Detection/Discrimination Technology Demonstration, U.S. Army Jefferson Proving Ground, Madison, Indiana, Technology Demonstration Plan*. 2000. Washington, DC. Naval Explosive Ordnance Disposal Technology Division (NAVEOD-TECHDIV).

5. Deschaine, L.M., R.A. Hoover, J. N. Skibinski, J. J. Patel, F. D. Francone, P. Nordin, and M. J. Ades. 2002. "Using Machine Learning to Compliment and Extend the Accuracy of UXO Discrimination Beyond the Best Reported Results of the Jefferson Proving Ground Technology Demonstration." In: *Proceedings of the Society for Modeling and Simulation International's Advanced Technology Simulation Conference*, April 2002. San Diego, CA, USA.

6. Cespedes, E. 2001. *Advanced UXO Detection/Discrimination Technology Demonstration—U.S. Army Jefferson Proving Ground, Madison, Indiana*. U.S. Army Corps of Engineers, Engineer Research and Development Center, ERDC/EL TR-01-20.

7. Francone, F.D., L.M. Deschaine, T. Battenhouse, and J.J Warren. 2004. "Discrimination of Unexploded Ordnance from Clutter Using Linear Genetic Programming" In: *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2004)*, June 26–30, 2004, Seattle, WA, USA.

8. Francone, F. D., and L.M. Deschaine. 2004. "Getting It Right at the Very Start—Building Project Models where Data is Expensive by Combining Human Expertise, Machine Learning and Information Theory." In: *Proceedings of the Business and Industry Symposium, 2004 Advanced Simulation Technologies Conference* (Arlington, VA, April 18 - 22, 2004). The Society for Modeling and Simulation International, 162-168.

9. Francone, F., L. Deschaine, and J. Warren. 2007. "Discrimination of Munitions and Explosives of Concern at F.E.Warren AFB using Linear Genetic Programming." In: Proceedings of the Genetic and Evolutionary Computation Conference.

10. Banzhaf, W., P. Nordin, R. Keller, and F. Francone. 1998. *Genetic Programming, an Introduction*, Morgan Kaufman Publishers, Inc., San Francisco, CA at pp 257-264.

Nordin, J.P., F. Francone, and W. Banzhaf. 1999. "Efficient Evolution of Machine Code for CISC Architectures Using Blocks and Homologous Crossover" In: *Advances in Genetic Programming 3*. Chapter 12 (MIT Press, Cambridge, MA).

Fukunaga, A., Stechert, D. Mutz. 1998. "A Genome Compiler for High Performance Genetic Programming" In: *Proceedings of the Third Annual Genetic Programming Conference*, Jet Propulsion Laboratories, California Institute of Technology, Pasadena, CA, Morgan Kaufman Publishers, pp. 86–94.

11. Several years of comparative studies by RML and SAIC are reported in: Francone, F. D., and L.M. Deschaine. 2004. *Extending the Boundaries of Design Optimization by Integrating Fast Optimization Techniques with Machine-Code-Based Linear Genetic Programming, Information Sciences Journal—Informatics and Computer Science*, Elsevier Press, Vol. 161/3-4 pp. 99-120 (see Sections 8.3-8.6 for results of the comparative study) Amsterdam, the Netherlands. In brief summary, RML's LGP software consistently performs as well as the best-tested alternative classification algorithms or better, on blind data. Other learning algorithms sometimes perform as well as RML's LGP algorithm but are not nearly as consistent as RML's LGP in producing high-quality results on unseen, testing data.

12. See: (1) Mukkamala, S., A. Sung, and A. Abraham. 2004. "Modeling Intrusion Detection Systems Using Linear Genetic Programming Approach," in *Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*.

(2) S. Mukkamala, Q. Liu, R. Veeraghattam, and A. H. Sung. 2005. "Computational Intelligent Techniques for Tumor Classification (Using Micro Array Gene Expression Data)." *International Journal of Lateral Computing*, Vol. 2, No. 1, ISSN 0973-208X, pp. 38-45.

(3) Mukkamala, G.D. Tilve, A.H. Sung, B. Ribeiro, and A.S. Vieira. 2006. Computational Intelligent Techniques for Financial Distress Detection. *International Journal of Computational Intelligence Research*.

13. ESTCP MM-0533. MTADS Demonstration at Camp Sibert, Magnetometer/EM61 MkII/GEM-3 Arrays, Technology Demonstration Data Report, G.R. Harbaugh, D.A. Steinhurst, and N. Khadr, September 26, 2007.

# APPENDIX A

# POINTS OF CONTACT

| Point of Contact | Organization | Phone<br>Fax<br>E-Mail | Role |
|---|---|---|---|
| Dr. Frank D. Francone | RML Technologies, Inc.<br>7606 S. Newland Street<br>Littleton, CO 80128 | Phone: (720) 981-8710<br>Fax: (603) 947-9150<br>E-mail: Frank.Francone@RMLTECH.com | Principal Investigator |
| Dr. Dean A. Keiswetter | Science Applications International Corporation<br>120 Quade Drive<br>Cary, NC 275133 | Phone: (919) 677-1560<br>Fax: (919) 678-1508<br>E-mail: Dean.A.Keiswetter@SAIC.com | Principal Investigator |
| Dr. Tom Furuya | Science Applications International Corporation<br>120 Quade Drive<br>Cary, NC 275133 | Phone: (919) 677-1560<br>Fax: (919) 678-1508<br>E-mail: furuyat@SAIC.com | Data Analyst |
| Dr. Herb Nelson | ESTCP<br>901 N. Stuart Street<br>Suite 303<br>Arlington, VA 22203 | Phone: (703) 696-8726<br>E-mail: Herbert.Nelson@osd.mil | Program Manager, Munitions Response (MR) |
| Mr. Peter Knowles | HydroGeoLogic, Inc.<br>11107 Sunset Hills Road<br>Suite 400<br>Reston, VA 20190 | Phone: (703) 696-8726<br>Fax: (703) 478-0526<br>E-mail: pknowles@hgl.com | Program Manager's Assistant, MR |

ESTCP Office
901 North Stuart Street
Suite 303
Arlington, Virginia 22203

(703) 696-2117 (Phone)
(703) 696-2114 (Fax)

E-mail: estcp@estcp.org
www.estcp.org