



STANDARDIZED
UXO TECHNOLOGY DEMONSTRATION SITE
SCORING RECORD NO. 946

SITE LOCATION:
ABERDEEN PROVING GROUND

DEMONSTRATOR:
DARTMOUTH COLLEGE, THAYER SCHOOL OF ENGINEERING
14 ENGINEERING DRIVE
HANOVER, NH 03755

TECHNOLOGY TYPE/PLATFORM:
HIGH-POWER NRL TEMTADS 2X2
PUSH CART

AREAS COVERED: SMALL MUNITIONS TEST SITE

AD No. _____
ATEC Project No. 2011-DT-ATC-DODSP-FO292
Report No. ATC-12166

Leonard Lombardo
Gene Fabian
Threat Detection and Systems Survivability Branch, Maritime/Threat
Detection and Systems Survivability Division,
Survivability/Lethality Directorate

July 2017

Report Produced by:
U.S. Army Aberdeen Test Center
Aberdeen Proving Ground, MD 21005-5059

Report Produced for:
Strategic Environmental Research and Development Program/Environmental
Security Technology Certification Program, Munitions Management
Alexandria, VA 22350

U.S. Army Test and Evaluation Command
Aberdeen Proving Ground, MD 21005-5001

Distribution Unlimited, July 2017.

The use of a trade name or the name of the manufacturer or a contractor in this report does not constitute an official endorsement or approval of the use of such commercial hardware or software or of service. The report may not be cited for purposes of advertisement.

DISPOSITION INSTRUCTIONS: Destroy this document when no longer needed. Do not return to the originator.

This document contains information EXEMPT FROM MANDATORY DISCLOSURE under the Freedom of Information Act. Exemption 5 (predecisional materials) applies.



REPLY TO
ATTENTION OF

DEPARTMENT OF THE ARMY
US ARMY ABERDEEN TEST CENTER
400 COLLERAN ROAD
ABERDEEN PROVING GROUND, MARYLAND 21005-5059

TEDT-AT-SLM

MEMORANDUM FOR Program Manager - SERDP/ESTCP, Munitions Management,
Dr. Herb Nelson, 4800 Mark Center Drive, Suite 17D08, Alexandria, VA 22350-3600

SUBJECT: Standardized Unexploded Ordnance (UXO) Technology Demonstration Site
Scoring Record No. 946, ATEC Project No. 2011-DT-ATC-DODSP-F0292.

1. Subject report has been approved by this headquarters and is submitted for your information and retention.
2. The point of contact for this office is Mr. Gene L. Fabian, TEDT-AT-SLM, 410-278-7421 or gene.l.fabian.civ@mail.mil.

FOR THE COMMANDER:

Encl

SHEPPARD.TRACY
Y.V.1200534219

Digitally signed by
SHEPPARD.TRACY.V.1200534219
DN: c=US, o=U.S. Government,
ou=DoD, ou=PKI, ou=USA,
cn=SHEPPARD.TRACY.V.1200534219
Date: 2017.08.02 12:35:39 -04'00'

TRACY SHEPPARD
Director, Survivability/Lethality Directorate

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

The 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 the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the 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 (DD-MM-YYYY) July 2017		2. REPORT TYPE Final		3. DATES COVERED (From - To) July 2015	
4. TITLE AND SUBTITLE Standardized UXO Technology Demonstration Site Scoring Record No. 946				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Lombardo, Leonard Fling Jr., Rick Fabian, Gene L.				5d. PROJECT NUMBER 2011-DT-ATC-DODSP-F0292	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Commander U.S. Army Aberdeen Test Center ATTN: TEDT-AT-SLM Aberdeen Proving Ground, MD 21005-5059				8. PERFORMING ORGANIZATION REPORT NUMBER ATC-12166	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Program Manager Strategic Environmental Research and Development Program Environmental Security Technology Certification Program Munitions Management Alexandria, VA 22350				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) Same as Item 8	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution unlimited.					
13. SUPPLEMENTARY NOTES None					
14. ABSTRACT This scoring record documents the efforts of the Dartmouth College, Thayer School of Engineering to detect and discriminate inert unexploded ordnance (UXO) using the APG Standardized UXO Technology Demonstration Site. This scoring record was coordinated by the Standardized UXO Technology Demonstration Site Scoring Committee. Organizations of the committee include the U.S. Army Corps of Engineers, the Environmental Security Technology Certification Program, the Strategic Environmental Research and Development Program, the Institute for Defense Analysis, the U.S. Army Environmental Command, and the U.S. Army Aberdeen Test Center.					
15. SUBJECT TERMS Unexploded Ordnance, UXO, Environmental Security Technology Certification Program, Standardized UXO Technology Demonstration Site, small munitions test site, blind grid, open field, indirect fire, towed array.					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Gene L. Fabian
Unclassified	Unclassified	Unclassified	SAR	41	19b. TELEPHONE NUMBER (Include area code)

ACKNOWLEDGMENTS

Authors:

Leonard Lombardo

Gene L. Fabian

Threat Detection and Systems Survivability (TDSS)
U.S. Army Aberdeen Test Center (ATC)
Aberdeen Proving Ground (APG)

Rick Fling

Aberdeen Test Support Services (ATSS)
Jacobs Technology, Inc.
Aberdeen Proving Ground

TABLE OF CONTENTS

	<u>PAGE</u>
ACKNOWLEDGMENTS	i
 <u>SECTION 1. GENERAL INFORMATION</u> 	
1.1 BACKGROUND	1
1.2 SCORING OBJECTIVES	1
1.2.1 Scoring Methodology	2
1.2.2 Scoring Factors	3
 <u>SECTION 2. DEMONSTRATION</u> 	
2.1 DEMONSTRATOR INFORMATION	4
2.1.1 Demonstrator Point of Contact (POC) and Address	4
2.1.2 System Description	4
2.1.3 Data Processing Description	6
2.1.4 Data Submission Format	10
2.1.5 Demonstrator Quality Assurance (QA) and Quality Control (QC)	10
2.1.6 Additional Records	10
2.2 APG SITE INFORMATION	11
2.2.1 Location	11
2.2.2 Soil Type	11
2.2.3 Test Areas	11
2.3 ATC SURVEY COMMENTS	11
 <u>SECTION 3. FIELD DATA</u> 	
3.1 DATE OF FIELD ACTIVITIES	12
3.2 AREAS TESTED/NUMBER OF HOURS	12
3.3 TEST CONDITIONS	12
3.3.1 Weather Conditions	12
3.3.2 Field Conditions	12
3.3.3 Soil Moisture	13
3.4 FIELD ACTIVITIES	13
3.4.1 Setup/Mobilization	13
3.4.2 Calibration	13
3.4.3 Downtime Occasions	13
3.4.4 Data Collection	14
3.4.5 Demobilization	14
3.5 PROCESSING TIME	14
3.6 DEMONSTRATOR'S FIELD PERSONNEL	14
3.7 DEMONSTRATOR'S FIELD SURVEYING METHOD	14
3.8 SUMMARY OF DAILY LOGS	14

SECTION 4. TECHNICAL PERFORMANCE RESULTS

	<u>PAGE</u>
4.1 ROC CURVES USING ALL MUNITIONS CATEGORIES	15
4.2 ROC CURVES USING ORDNANCE LARGER THAN 20-MM	17
4.3 PERFORMANCE SUMMARIES	20
4.4 EFFICIENCY, REJECTION RATES, AND TYPE CLASSIFICATION	21
4.5 LOCATION ACCURACY	22
4.6 STATISTICAL COMPARISONS	23

SECTION 5. APPENDIXES

A TERMS AND DEFINITIONS	A-1
B DAILY WEATHER LOGS	B-1
C SOIL MOISTURE	C-1
D DAILY ACTIVITY LOGS	D-1
E REFERENCES	E-1
F ABBREVIATIONS	F-1
G DISTRIBUTION LIST	G-1

SECTION 1. GENERAL INFORMATION

1.1 BACKGROUND

Technologies under development for the detection and discrimination of military munitions (i.e., unexploded ordnance {UXO} and discarded military munitions) require testing so that performance can be characterized. To that end, Standardized Test Sites have been developed at Aberdeen Proving Ground (APG), Maryland, and U.S. Army Yuma Proving Ground (YPG), Arizona. These test sites provide a diversity of geology, climate, terrain, and weather as well as diversity in munitions and clutter. Testing at these sites is independently administered and analyzed by the government for the purposes of characterizing technologies, tracking performance with system development, comparing performance of different systems, and comparing performance in different environments (app E, ref 1).

The Standardized UXO Technology Demonstration Site Program is a multiagency program spearheaded by the U.S. Army Environmental Command (USAEC). The U.S. Army Aberdeen Test Center (ATC) and the U.S. Army Corps of Engineers Engineering Research and Development Center (ERDC) provide programmatic support. The program is being funded and supported by the Environmental Security Technology Certification Program (ESTCP), the Strategic Environmental Research and Development Program (SERDP), and the U.S. Army Environmental Quality Technology (EQT) Program.

The Small Munitions Test Site is intended to test the ability of systems to detect anomalies and to discriminate between small ordnance (20-, 37-, and 40-mm projectiles) and clutter when small clutter near the surface partially obstructs the signal. The site is in the form of a grid with a portion of the grid locations having the surface clutter and a portion of the grid locations having no surface clutter to allow for a comparison between conditions.

1.2 SCORING OBJECTIVES

The objective in the Standardized UXO Technology Demonstration Site Program is to evaluate the detection and discrimination capabilities of a given technology under various field and soil conditions. Inert munitions and clutter items are positioned in various orientations and depths in the ground.

The evaluation objectives are as follows:

- a. To determine detection and discrimination effectiveness under realistic scenarios with various targets, geology, clutter, density, topography, and vegetation.
- b. To determine cost, time, and workforce requirements to operate the technology.
- c. To determine the demonstrator's ability to analyze survey data in a timely manner and provide prioritized Target Lists with associated confidence levels.
- d. To provide independent site management to enable the collection of high quality, ground-truth (GT), geo-referenced data for post-demonstration analysis.

1.2.1 Scoring Methodology

a. The scoring of the demonstrator's performance is conducted in two stages: response stage and discrimination stage. For both stages, the probability of detection (P_d) and the false alarms are reported as receiver-operating characteristic (ROC) curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of clutter detection (P_{cd}) or the probability of false positive (P_{fp}). Those that do not correspond to any known item are termed background alarms. The background alarms are addressed as either probability of background alarm (P_{ba}) or background alarm rate (BAR).

b. The response stage scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate munitions from other anomaly sources. For the Blind Test Grid and Small Arms Test Grid response stage, the demonstrator provides a target response from each grid's square along with a threshold below which target responses are deemed insufficient to warrant further investigation. This list is generated with minimal processing and, since a value is provided for every grid square, includes amplitudes both above and below the system noise level. For the open field, the demonstrator provides a list of all anomalies deemed to exceed a demonstrator selected target detection threshold. An item (either munition or clutter) is counted as detected if a demonstrator indicates an anomaly within a specified distance (Halo Radius (R_{halo})) of a ground truth item.

c. The discrimination stage evaluates the demonstrator's ability to correctly identify munitions as such and to reject clutter. For the Blind Test Grid and Small Munitions Test Grid discrimination stage, the demonstrator provides the output of the discrimination stage processing for each grid square. For the open field, the demonstrator provides the output of the discrimination stage processing for anomaly reported in the response stage. The values in these lists are prioritized based on the demonstrator's determination that a location is likely to contain munitions. Thus, higher output values are indicative of higher confidence that a munitions item is present at the specified location. For digital signal processing, priority ranking is based on algorithm output. For other discrimination approaches, priority ranking may be based on rule sets or human judgment. The demonstrator also specifies the threshold in the prioritized ranking that provides optimum performance, (i.e., that is expected to retain all detected munitions and reject the maximum amount of clutter).

d. The demonstrator is also scored on efficiency and rejection ratios, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of munitions detections from the anomaly list, while rejecting the maximum number of anomalies arising from non-munitions items. Efficiency measures the fraction of detected munitions retained after discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the maximum number of munitions detectable by the sensor and its accompanying clutter detection/false positive rate or probability of background alarm.

e. All scoring factors are generated using the Standardized UXO Probability and Plot Program, version 4.

1.2.2 Scoring Factors

Factors to be measured and evaluated as part of this demonstration include:

a. Response stage ROC curves:

- (1) Probability of detection for the response stage (P_d^{res}).
- (2) Probability of clutter detection (P_{cd}).
- (3) Probability of background alarm ($P_{\text{ba}}^{\text{res}}$).

b. Discrimination stage ROC curves:

- (1) Probability of detection for the discrimination stage (P_d^{disc}).
- (2) Probability of false positive (P_{fp}).
- (3) Probability of background alarm ($P_{\text{ba}}^{\text{disc}}$).

c. Metrics:

- (1) Efficiency (E).
- (2) False positive rejection rate (R_{fp}).
- (3) Background alarm rejection rate (R_{ba}).

d. Other:

- (1) Probability of detection by size, depth, and density.
- (2) Classification by type (i.e., 20-, 40-, 105-mm, etc.).
- (3) Location accuracy for single munitions.
- (4) Equipment setup, calibration time, and corresponding worker-hour requirements.
- (5) Survey time and corresponding worker-hour requirements.
- (6) Reacquisition/resurvey time and worker-hour requirements (if any).
- (7) Downtime due to system malfunctions and maintenance requirements.

SECTION 2. DEMONSTRATION

2.1 DEMONSTRATOR INFORMATION

2.1.1 Demonstrator Point of Contact (POC) and Address

POC: Fridon Shubitidze

Address: Dartmouth College, Thayer School of Engineering
14 Engineering Drive
Hanover, NH 03755

2.1.2 System Description (provided by demonstrator)

The standard Naval Research Laboratory (NRL) Time-Domain Electromagnetic Multi-Sensor Towed Array Detection System (TEMTADS) 2x2 is a time-domain electromagnetic induction (EMI) instrument, which has been updated for this demonstration with the intended purpose of improving the detection and classification of small and/or deep UXO. The NRL TEMTADS 2x2's transmitter electronics were updated in the course of SERDP project MR- 2225 to produce higher Tx currents. The instrument has a transmitter (Tx) array with four coplanar square coils, together with four triaxial receivers (Rx) placed at the center of each Tx (fig. 1). Each Rx cube contains three orthogonal coils and thus registers all three vector components of the impinging signals. The Tx coils, with transmitter currents of approximately 13.5 A, illuminate a buried target, and the target responses are collected with a 500-kHz sample rate after turn off of the excitation pulse. The system operates in both static (cued) and dynamic modes. For dynamic mode, the raw decay measurements are grouped into 19 logarithmically-spaced gates whose center times range from 25 μ s to 2.77 ms with 20-percent widths. For cued mode, the raw decay measurements are grouped into 121 logarithmically-spaced gates whose center times range from 25 μ s to 24.35 ms with 5-percent widths. The sensor is placed on a cart (fig. 2), which provides a sensor-to-ground offset of 20 cm or less.

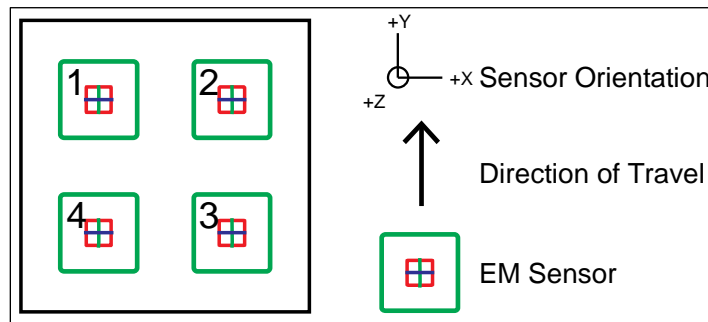


Figure 1. Schematic diagram of the NRL TEMTADS 2x2 EMI sensor array: 2 x 2 sensor array 0.8 x 0.8 m in size. Each unit consists of a 35- x 35-cm Tx loop and 8-cm, triaxial Rx cube.



Figure 2. NRL TEMTADS 2x2 (left) cart alone, and (right) with weather shield removed.

Application of the updated NRL TEMTADS 2x2 is straightforward. In survey mode, the system collects a series of closely-spaced parallel lines of survey data, then the survey data are processed and a list of detected targets positions is developed for cued mode. In this demonstration, where only the Calibration, Blind, and Small Munitions Test Grids are being investigated, the areas are arranged in grids of test cells and the cell center positions are known. Each target position is flagged with a non-metallic pin flag using cm-level Global Positioning System (GPS). The sensor is positioned over each target in turn. With the system positioned over the target, the Tx are activated sequentially and during the Tx, all four Rx record data. The complete set of data for each target is then inverted for target characteristics and classification. In addition, to demonstrate the system's applicability in dynamic mode for targets detection, picking and targets full classifications complete survey data are collected, inverted and classified for the Small Munitions Test Grid.

Support Equipment Required: Overnight storage for the sensor protected from the elements and access to electrical power for battery charging is required. This and workspace for the data quality control analyst located in the building at the test site. The calibration area, blind grid cell centers, and the Small Munitions Test Grid need to be accurately flagged by ATC.

Frequency and Radio Utilization: Dartmouth is licensed for the following frequencies for GPS corrections:

461.0250 MHz	462.1250 MHz	464.5000 MHz	464.6250 MHz	464.7250 MHz
461.0750 MHz	462.3750 MHz	464.5500 MHz	464.6500 MHz	464.7500 MHz
461.1000 MHz	462.4000 MHz	464.6000 MHz	464.7000 MHz	

Dartmouth anticipates using GPS and GPS data radios for this demonstration. Access to one of these frequencies free of interference is required. Dartmouth currently uses 464.6250 MHz, but any of the above frequencies may be used.

Demonstrator's Field Personnel: The personnel on-site were:

Fridon Shubitidze, Dartmouth College.
Dan Steinhurst, Nova Research, Inc.
Glenn Harbaugh, Nova Research, Inc.
Ben Andrews, Nova Research, Inc.

2.1.3 Data Processing Description (Provided by Demonstrator)

Target Selection Criteria:

Targets for this demonstration are those emplaced in the Calibration Lane and the Blind and Small Munitions Test Grids. This will be primarily a cued, classification data collection. This allows for a direct comparison of the demonstration results with the results of earlier demonstrations of other technologies, such as Metal mapper, TEM-8, and OPTEMA. However, time allowing, a dynamic survey will be done over the Small Munitions Test Grid. Since these systems are designed for both cued and dynamic models the comparison would be a useful one. Each data set will be used independently for targets classifications.

- a. What kind of preprocessing (if any) is applied to the raw data (e.g., filtering, etc.)?

The high power NRL TEMTADS 2x2 preprocessing is a batch process of all binary waveform survey data via a EM3D acquire program. The program performs programmable stacking, hardware low-pass filtering, and exports a binary TEM file as well as an ASCII CSV file containing data at each of the time gates along the transmitter current waveform. The raw CSV data are passed to data filtering software, which normalizes data on corresponding Tx currents; subtracts background; forms multi-static response (MSR) data matrix; and estimates its eigenvalues time decays. In addition, the number of extracted eigenvalues above a background threshold are estimated for target detection and picking.

- b. What is the format of the data both pre- and post-processing of the raw data (e.g., ASCII, binary, etc.)?

Both a TEM file and a matching ASCII CSV file are produced by the G&G Science EM3D software. Post-processed data and inversion and classification results are stored as ASCII data files.

- c. What algorithm is used for detection (e.g., peaks of signal surpassing threshold, etc.)?

In the cued mode Calibration, Blind and Small Munition Test Grid surveys, possible target locations are predetermined. However, in dynamic survey mode the Joint Diagonalization (JD) algorithm will be used for target detection and picking. Once targets are detected, then survey data will be inverted using the orthonormalized, volume magnetic source (ONVMS) technique.

- d. Why is this algorithm used and not others?

The JD algorithm has been shown to provide a quick estimate of the number and even types of targets present.

e. On what principles is the algorithm based (e.g., statistical models, heuristic rules, etc.)?

The JD algorithm is based on finding the eigenvalues and eigenvectors of the MSR matrix as a function of time while preserving the order of the eigenvalues which are above the sensor's noise thresholds. This allows a quick determination of the physics-based sources in view of the sensor.

f. What tunable parameters (if any) are used in the detection process (e.g., threshold on signal amplitude, window length, filter coefficients, etc.)?

Site-specific noise threshold.

g. What are the final values of all tunable parameters for the detection algorithm?

The final values for the tunable parameters will be determined for each (Blind and Small Munitions) grid after data are collected. The background threshold values will be determined independently for each area surveyed.

Parameter Estimation:

a. Which characteristics will be extracted from each detected item and input to the discrimination algorithm (e.g., depth, size, polarizability coefficients, fit quality, etc.)?

MSR data, eigenvalues time decays, total ONVMS (effective principal axis polarizabilities), time decays and fit quality.

b. Why have these characteristics been chosen and not others (e.g., empirical evidence of their ability to help discriminate, inclusion in a theoretical tradition, etc.)?

Recent live-site ESTCP discrimination pilot studies indicate that these parameters are the best characteristics for UXO/clutter classification.

c. How are these characteristics estimated (e.g., least-mean-squares fit to a dipole model, etc.)? Include the equations that are used for parameter estimation.

Dartmouth used a least squares approach to recover the parameter vector \mathbf{v} , which in this case contains intrinsic (effective Principal axis polarizabilities) information about the object and its location and orientation. Specifically, if \mathbf{d}^{obs} is the vector of the measured secondary field and $\mathbf{F}(\mathbf{v})$ the forward problem solution, the least squares approach uses as criterion:

$$\underset{\mathbf{v}}{\text{minimize}} \quad \phi(\mathbf{v}) = \frac{1}{2} \|\mathbf{d}^{\text{obs}} - \mathbf{F}(\mathbf{v})\|^2.$$

JD, also referred to as simultaneous matrix diagonalization, is a numerical approach for estimating the common eigenvalues and eigenvectors of a set of related square matrices. Here the JD will be applied to the updated 2x2 3D TEMTAD system's MSR data matrix for detecting and discriminating subsurface. Given the measured MSR matrices $\bar{\bar{H}}(t_q)$ of size $M_T \times M_R$, where M_T and $M_R (= M_T$ by assumption), respectively represent the numbers of Tx and Rx coils, JD finds an orthogonal matrix $\bar{\bar{V}}$ such that the products $\bar{\bar{H}}(t_q) = \bar{\bar{V}} \bar{\bar{D}}(t_q) \bar{\bar{V}}^T$ are as diagonal as possible for

all $q=1,2,\dots,N_t$, where N_t is the number of time gates and $\bar{\bar{v}}^T$ denotes the transpose of $\bar{\bar{v}}$. Studies have shown that the number of nonzero eigenvalues of the MSR matrix (*i.e.*, those above the noise threshold) depend on the number the targets in the snapshot, while their time-decay patterns are intrinsic properties of the individual targets and provide robust targets detection and classification features.

The (ONVMS) model is as a generalized volume dipole model, which in a special limited case coincides with an infinitesimal dipole model approach. The great advantage of the ONVMS technique is that it takes into account mutual couplings between different parts of the target, and avoids matrix singularity problems in case of multiple objects. The technique is applicable for both single and multi-objects cases. Once the magnetic dipole polarizability tensor elements and dipole locations are determined, one could group them by proximity according to the volume distribution, and then for each group calculate the total polarizability tensor. These diagonal elements have been shown to be intrinsic to the object and can be used on its own or combined with other quantities in discrimination processing. Overall, in the ONVMS approach the scattered magnetic field is approximated as a superposition of the fields radiated by these elementary sources, whose amplitudes are normalized with the impinging primary magnetic field. The scattered magnetic field at any point outside the target's volume is represented using orthonormalized functions as:

$$\mathbf{H}(\mathbf{r}) = \sum_{i=1}^{N_t} \bar{\bar{\psi}}_i(\mathbf{R}_i) \cdot \mathbf{b}_i, \quad (6)$$

Where:

$\mathbf{R}_i = \mathbf{r} - \mathbf{r}_i^v$, $\bar{\bar{\psi}}_i(\mathbf{r})$ is an orthonormalized function, and \mathbf{b}_i is orthogonal-function expansion coefficient. First, the orthonormal functions $\bar{\bar{\psi}}_i$ are constructed as linear combinations of the Green's functions using the Gram–Schmidt orthonormalization process, then using the linear independence of Green's Dyadics, the amplitudes of the $\bar{\bar{\mathbf{M}}}_i(p)$ polarizability tensor elements are determined without solving a linear system of equations. Once the $\bar{\bar{\mathbf{M}}}_i(p)$ magnetic dipoles polarizability tensor elements and dipoles locations are determined, one could group them by the volume distribution and for each group calculate the total polarizability tensor. The principal axis effective polarizabilities (total ONVMS) are targets intrinsic properties and used for classification.

d. What tunable parameters, if any, are used in the characterization process (e.g., thresholds on background noise, etc.)?

MSR data, eigenvalue time decays and effective principal-axis polarizabilities for expected ordnance items determined from calibration and previously used data.

Classification:

a. What algorithm is used for discrimination (e.g., multi-layer perception, support vector machine, etc.)?

Generalized fingerprinting for effective polarizabilities and pattern matching for the eigenvalues.

b. Why is this algorithm used and not others?

The algorithm is appropriate for Dartmouth's procedure, which compares fit quality using previously-determined UXO effective polarizabilities with unconstrained fit quality and was proven effective in the ESTCP UXO Classification Study.

c. Which parameters are considered as possible inputs to the algorithm?

Unconstrained fit qualities. After calculating the total ONVMS for both one and multi-targets inversion routines, Dartmouth compares these with standard values obtained from data acquired over the targets in the calibration grid and from Dartmouth's library.

d. What are the outputs of the algorithm (probabilities, confidence levels)?

Closeness of measured response to UXO response.

e. How is the threshold set to decide where the munitions/non-munitions line lies in the discrimination process?

Training data on UXO and clutter acquired in calibration lanes at APG and other demonstration sites.

Training:

a. Which tunable parameters have final values that are optimized over a training set of data and which have values that are set according to geophysical knowledge (i.e., intuition, experience, common sense)?

Quality of fit between effective polarizabilities for test target and library UXO-item is optimized.

(1) For those tunable parameters with final values set according to geophysical knowledge:

(a) What is the reasoning behind choosing these particular values?

Not applicable.

(b) Why were the final values not optimized over a training set of data?

Not applicable.

(2) For those tunable parameters with final values optimized over the training set data:

(a) What training data are used (e.g., all data, a randomly chosen portion of data, etc.)?

All training data on UXO and clutter acquired in calibration lane at APG.

(b) What error metric is minimized during training (e.g., mean squared error, etc.)?

Mean squared error.

(c) What learning rule is used during training (e.g., gradient descent, etc.)?

Not applicable.

(d) What criterion is used to stop training (e.g., number of iterations exceeds threshold, good generalization over validation set of data, etc.)?

Limits of training data.

(e) Are all tunable parameters optimized at once or in sequence (in sequence = parameter 1 is held constant at some common sense values while parameter 2 is optimized, and then parameter 2 is held constant at its optimized value while parameter 1 is optimized)?

All at once.

b. What are the final values of all tunable parameters for the characterization process?

Best threshold setting.

2.1.4 Data Submission Format

Data were submitted for scoring in accordance with data submission protocols outlined on the USAEC Web site www.uxotestsites.org. These submitted data are not included in this report in order to protect GT information.

2.1.5 Demonstrator Quality Assurance (QA) and Quality Control (QC) (Provided by Demonstrator)

General system functionality and individual sensor response are checked daily to ensure adequate system performance. Before beginning survey work each day, one or more standard objects are measured. The resulting signals and inversion results are checked against standard values.

Every 1 to 2 hr, all survey data are transferred to the field data analyst for preliminary data quality checks. The individual sensor files are examined for completeness and consistency. It is at this stage that any sensor malfunctions, etc. are flagged and reported to the field crew for correction.

2.1.6 Additional Records

The following record(s) by this vendor can be accessed via the Internet as Microsoft Word documents at www.uxotestsites.org.

Scoring Record No. No. 945

2.2 APG SITE INFORMATION

2.2.1 Location

The APG Standardized UXO Test Site is located within a secured range area of the Aberdeen Area. The Aberdeen Area of APG is located approximately 30 miles northeast of Baltimore at the northern end of the Chesapeake Bay. The Standardized UXO Test Site encompasses 17 acres of upland and lowland flats, woods, and wetlands.

2.2.2 Soil Type

According to the soils survey conducted for the entire area of APG in 1998, the test site consists primarily of Elkton Series type soil (ref 2). The Elkton Series consist of very deep, slowly permeable, poorly drained soils. These soils formed in silty aeolin sediments and the underlying loamy alluvial and marine sediments. They are on upland and lowland flats and in depressions of the Mid-Atlantic Coastal Plain. Slopes range from 0 to 2 percent.

The U.S. Army Corps of Engineers Engineering Research and Development Center (ERDC) conducted a site-specific analysis in May 2002 (ref 3). The results basically matched the soil survey mentioned above. Seventy percent of the samples taken were classified as silty loam. The majority (77 percent) of the soil samples had a measured water content between 15 and 30 percent with the water content decreasing slightly with depth.

For more details concerning the soil properties at the APG test site, go to www.uxotestsites.org on the Web to view the entire soils description report.

2.2.3 Test Areas

A description of the Small Munitions Test Site at the APG Standardized UXO Test Site is presented in Table 1.

TABLE 1. SMALL MUNITIONS TEST SITE AREAS

Area	Description
Small munitions calibration lanes	Contains three standard munitions items buried in six positions at various angles and depths to allow demonstrators to calibrate their equipment.
Small munitions test grid	Contains 300 grid cells. The center of each grid cell contains either munitions, clutter, or nothing with a portion containing surface clutter.

2.3 ATC SURVEY COMMENTS

None.

SECTION 3. FIELD DATA

3.1 DATE OF FIELD ACTIVITIES (27 through 30 July 2015)

3.2 AREAS TESTED/NUMBER OF HOURS

Areas tested and total numbers of hours operated at each site are presented in Table 2.

TABLE 2. AREAS TESTED AND NUMBER OF HOURS

Area	No. of Hours
Calibration grid	3 hr, 5 min
Small munitions grid	12 hr, 10 min

Note: Table 2 represents the total time spent in each area.

3.3 TEST CONDITIONS

3.3.1 Weather Conditions

An APG weather station located approximately 1 mile west of the test site was used to record average temperature and precipitation on a half-hour basis for each day of operation. The temperatures presented in Table 3 represent the average temperature during field operations from 0700 to 1700 hr, while precipitation data represent a daily total amount of rainfall. Hourly weather logs used to generate this summary are provided in Appendix B.

TABLE 3. TEMPERATURE/PRECIPITATION DATA SUMMARY

Date, 2015	Average Temperature, °F	Total Daily Precipitation, in.
27 July	75.5	0.67
28 July	82.1	0.00
29 July	81.1	0.00
30 July	80.5	0.27

3.3.2 Field Conditions

Dartmouth surveyed the Calibration Grid, Blind Grid and a Small Munitions Test Grid. The field was dry throughout all the Small Munitions Test Grid. The calibration area had a couple of wet spots due to rain early in the morning on the first day of testing.

3.3.3 Soil Moisture

Three soil probes were placed at various locations within the site to capture soil moisture data: blind grid, calibration, open field, and wooded areas. Measurements were collected in percent moisture and were taken twice daily (morning and afternoon) from five different soil depths (1 to 6 in., 6 to 12 in., 12 to 24 in., 24 to 36 in., and 36 to 48 in.) from each probe. Soil moisture logs are provided in Appendix C.

3.4 FIELD ACTIVITIES

3.4.1 Setup/Mobilization

These activities included initial mobilization and daily equipment preparation and breakdown. A four-person crew took 75 min to perform the initial setup and mobilization. A total of 20 min of equipment preparation was accrued, and end of day equipment breakdown totaled 30 min.

3.4.2 Calibration

Dartmouth spent a total of 3 hr 5 min in the calibration lanes, of which 2 hr and 30 min were spent collecting data. One calibration exercise occurred while surveying the small munition grid lasting 5 min.

3.4.3 Downtime Occasions

Occasions of downtime are grouped into five categories: equipment/data checks or equipment maintenance, equipment failure and repair, weather, demonstration site issues, or breaks/lunch. All downtime is included for the purposes of calculating labor requirements (section 5) except for downtime due to demonstration site issues. Demonstration site issues, while noted in the daily log, are considered nonchargeable downtime for the purposes of calculating labor costs and are not discussed. Breaks and lunches are discussed in this section and billed to the total site survey area.

3.4.3.1 Equipment/data checks, maintenance. Equipment data checks and maintenance activities accounted for no site usage time. These activities included changing out batteries and performing routine data checks to ensure the data were being properly recorded/collected. Dartmouth spent 30 min for breaks and lunches.

3.4.3.2 Equipment failure or repair. Two equipment failures occurred during this survey. Battery issues occurred twice. The battery was swapped out each time. The total downtime was 2 hr, 15 min.

3.4.3.3 Weather. No weather delays occurred during the survey.

3.4.4 Data Collection

The total time Dartmouth spent in each test area collecting data is provided in Table 4.

TABLE 4. DATA COLLECTION
TIME PER AREA

Area	Time
Calibration	2 hr, 30 min
Small munition	8 hr, 25 min

3.4.5 Demobilization

The Dartmouth survey crew conducted a demonstration of the calibration and Small Munition Test Grid. Demobilization occurred on 30 July 2015. On that day, it took the crew 45 min to break down and pack up their equipment.

3.5 PROCESSING TIME

The scoring submittal data were provided 6 August 2015.

3.6 DEMONSTRATOR'S FIELD PERSONNEL

Fridon Shubitidze, Dartmouth College.
Dan Steinhurst, Nova Research, Inc.
Glenn Harbaugh, Nova Research, Inc.
Ben Andrews, Nova Research, Inc.

3.7 DEMONSTRATOR'S FIELD SURVEYING METHOD

Dartmouth surveyed the Small Munition Test Grid in a linear fashion. Flags were emplaced to ensure an accurate position of each row while surveying.

3.8 SUMMARY OF DAILY LOGS

Daily logs capture all field activities during this demonstration and are provided in Appendix D.

SECTION 4. TECHNICAL PERFORMANCE RESULTS

4.1 ROC CURVES USING ALL MUNITIONS CATEGORIES

The probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of clutter detection or probability of false positive for capped and uncapped test areas are shown in Figures 3 and 4, respectively. The probabilities plotted against their respective background alarm rate for capped and uncapped conditions are shown in Figures 5 and 6, respectively. All figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator's recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination.

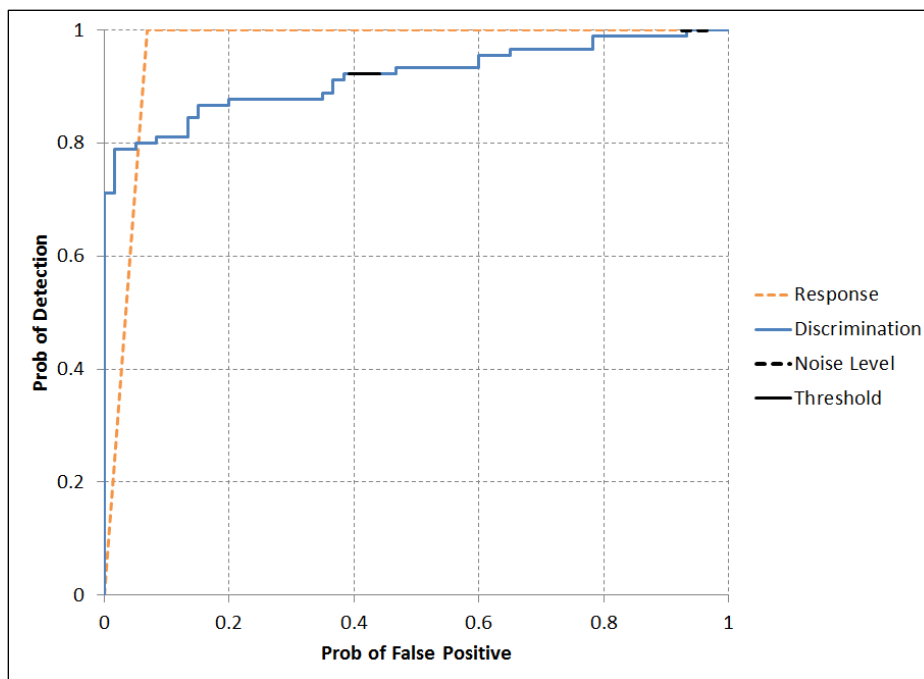


Figure 3. TEMTADS/push cart capped test area probability of detection for response and discrimination stages versus their respective probability of false positive.

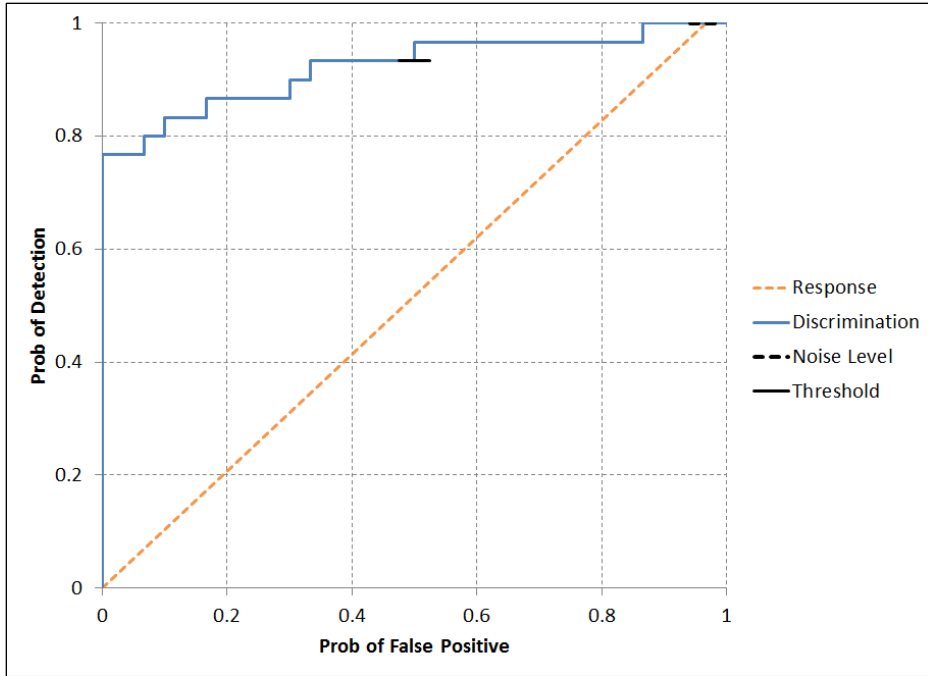


Figure 4. TEMTADS/push cart uncapped test area probability of detection for response and discrimination stages versus their respective probability of false positive.

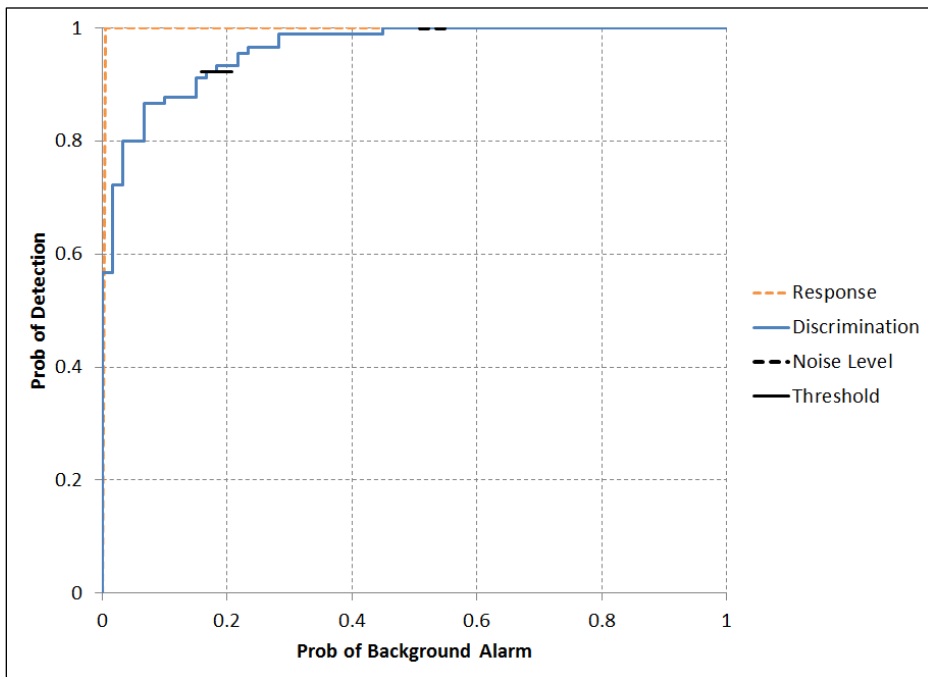


Figure 5. TEMTADS/Pushcart capped test area probability of detection for response and discrimination stages versus their respective probability of background alarm.

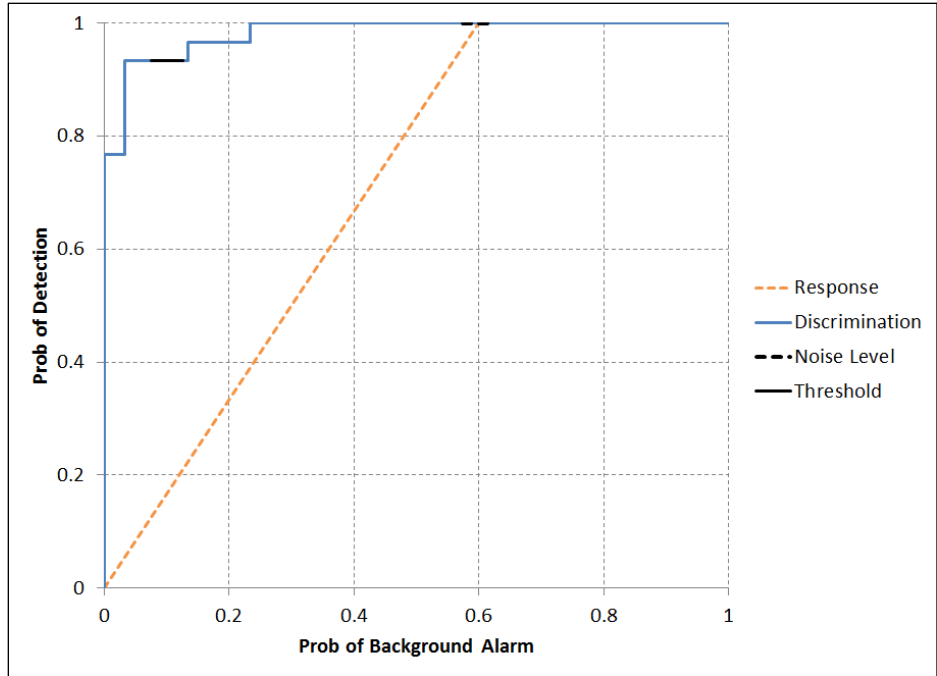


Figure 6. TEMTADS/push cart uncapped test area probability of detection for response and discrimination stages versus their respective probability of background alarm.

4.2 ROC CURVES USING ORDNANCE LARGER THAN 20 MM

The probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of clutter detection or probability of false positive for capped and uncapped test areas when only ordnance larger than 20-mm are scored are shown in Figures 7 and 8, respectively. The probabilities plotted against their respective background alarm rate for capped and uncapped conditions when only ordnance larger than 20 mm are scored are shown in Figures 9 and 10, respectively. All figures use horizontal lines to illustrate the performance of the demonstrator at two demonstrator-specified points: at the system noise level for the response stage, representing the point below which targets are not considered detectable, and at the demonstrator’s recommended threshold level for the discrimination stage, defining the subset of targets the demonstrator would recommend digging based on discrimination.

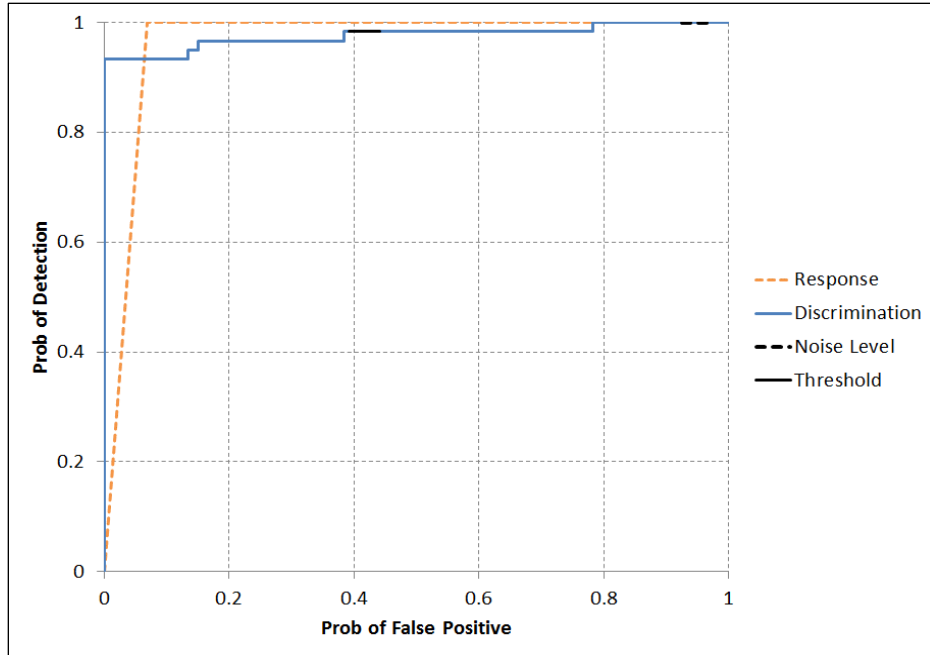


Figure 7. TEMTADS/push cart capped test area probability of detection for response and discrimination stages versus their respective probability of false positive for all ordnance larger than 20 mm.

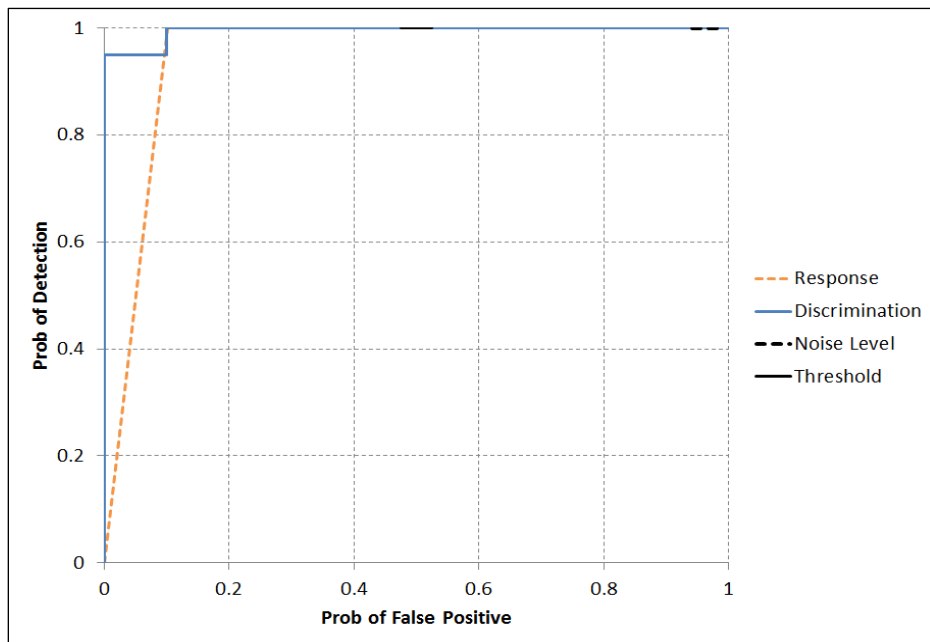


Figure 8. TEMTADS/push cart uncapped test area probability of detection for response and discrimination stages versus their respective probability of false positive for all ordnance larger than 20 mm.

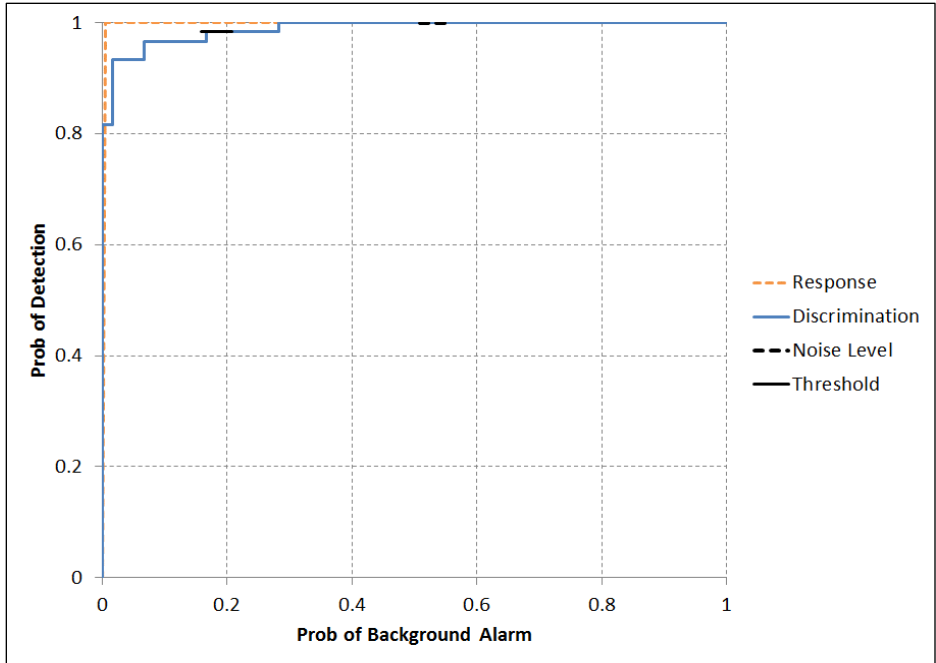


Figure 9. TEMTADS/push cart capped test area probability of detection for response and discrimination stages versus their respective probability of background alarm for all ordnance larger than 20 mm.

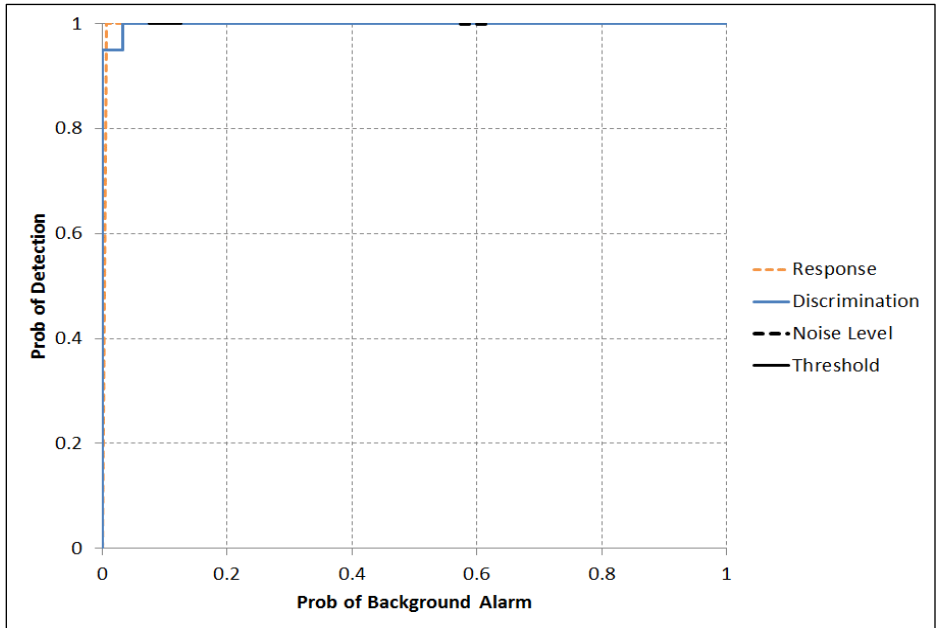


Figure 10. TEMTADS/Pushcart uncapped test area probability of detection for response and discrimination stages versus their respective probability of background alarm for all ordnance larger than 20 mm.

4.3 PERFORMANCE SUMMARIES

Results for capped and uncapped test areas are presented in Table 5a and 5b, respectively. The response stage results are derived from the list of anomalies above the demonstrator-provided noise level. The results for the discrimination stage are derived from the demonstrator's recommended threshold for optimizing munitions related cleanup by minimizing false alarm digs and maximizing munitions recovery. The lower and upper 90-percent confidence limits on P_d , P_{cd} , and P_{fp} were calculated assuming that the number of detections and false positives are binomially distributed random variables.

TABLE 5a. CAPPED TEST AREA RESULTS

Response Stage					Discrimination Stage								
Munitions ^a Scores	P_d^{res} : by type				P_d^{disc} : by type								
	All Types	40-mm	37-mm	20-mm	All Types	40-mm	37-mm	20-mm					
	1.000	1.000	1.000	1.000	0.956	1.000	0.996	0.891					
1.000	1.000	1.000	1.000	0.922	1.000	0.967	0.800						
0.975	0.926	0.926	0.926	0.873	0.926	0.876	0.675						
<i>By Depth^b</i>													
0 to 4D	1.000	1.000	NA	NA	1.000	1.000	NA	NA					
4D to 8D	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000					
8D to 12D	1.000	1.000	1.000	1.000	0.974	1.000	1.000	0.929					
> 12D	1.000	1.000	1.000	1.000	0.600	1.000	0.750	0.286					
Clutter Scores	P_{cd}				P_{fp}								
<i>By Mass</i>													
<i>By Depth^b</i>	All Mass	0 to 0.25 kg	>0.25 to 1 kg	All Mass	0 to 0.25 kg	>0.25 to 1 kg	All Mass	0 to 0.25 kg	>0.25 to 1 kg				
All Depth	0.981	0.966	1.000	0.508	0.428	0.689	0.950	0.909	1.000	0.417	0.303	0.198	0.556
	0.892	0.809	0.918	0.330									
0 to 0.15 m	1.000	NA	1.000	0.667	NA	0.667							
0.15 to 0.3 m	1.000	1.000	1.000	0.548	0.583	0.526							
0.3 to 0.6 m	0.870	0.857	1.000	0.174	0.143	0.500							
Background Alarm Rates													
P_{ba}^{res} : 0.533					P_{ba}^{disc} : 0.183								

^aThe two numbers to the right of the all types munitions result are an upper and lower 90-percent confidence interval for an assumed binomial distribution.

^bAll depths are measured to the center of the object.

TABLE 5b. UNCAPPED TEST AREA RESULTS

Response Stage					Discrimination Stage			
Munitions ^a Scores	P_d^{res} : by type				P_d^{disc} : by type			
	All Types	40-mm	37-mm	20-mm	All Types	40-mm	37-mm	20-mm
	1.000 1.000 0.926	1.000 1.000 0.794	1.000 1.000 0.794	1.000 1.000 0.794	0.982 0.933 0.832	1.000 1.000 0.794	1.000 1.000 0.794	0.945 0.800 0.550
<i>By Depth^b</i>								
0 to 4D	1.000	NA	1.000	1.000	1.000	NA	1.000	1.000
4D to 8D	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
8D to 12D	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
> 12D	1.000	1.000	1.000	1.000	0.600	1.000	1.000	0.000
Clutter Scores	P_{cd}				P_{ip}			
<i>By Mass</i>								
<i>By Depth^b</i>	All Mass	0 to 0.25 kg	>0.25 to 1 kg		All Mass	0 to 0.25 kg	>0.25 to 1 kg	
All Depth	0.996 0.967 0.876	0.994 0.941 0.790	1.000 1.000 0.838		0.630 0.500 0.370	0.650 0.471 0.297	0.736 0.538 0.331	
0 to 0.15 m	1.000	NA	1.000		0.500	NA	0.500	
0.15 to 0.3 m	1.000	1.000	1.000		0.563	0.571	0.556	
0.3 to 0.6 m	0.917	0.900	1.000		0.417	0.400	0.500	
Background Alarm Rates								
P_{ba}^{res} : 0.600					P_{ba}^{disc} : 0.100			

^aThe two numbers to the right of the all types munitions result are an upper and lower 90-percent confidence interval for an assumed binomial distribution.

^bAll depths are measured to the center of the object.

4.4 EFFICIENCY, REJECTION RATES, AND TYPE CLASSIFICATION

Efficiency and rejection rates are calculated to quantify the discrimination ability at specific points of interest on the ROC curve: (1) at the point where no decrease in P_d is suffered (i.e., the efficiency is by definition equal to one) and (2) at the operator selected threshold. These values are presented in Tables 6a and 6b.

TABLE 6a. CAPPED EFFICIENCY AND REJECTION RATES

	Efficiency (E)	False Positive Rejection Rate	Background Alarm Rejection Rate
At operating point	0.92	0.56	0.66
With no loss of P_d	1.00	0.02	0.16

TABLE 6b. UNCAPPED EFFICIENCY AND REJECTION RATES

	Efficiency (E)	False Positive Rejection Rate	Background Alarm Rejection Rate
At operating point	0.93	0.48	0.83
With no loss of P_d	1.00	0.10	0.61

At the demonstrator's recommended setting, the munitions items that were detected and correctly discriminated were further scored on whether their correct type could be identified (tables 7a and 7b). Correct type examples include 20-mm projectile, 37-mm projectile, and 40-mm projectile. A list of the standard type declaration required for each munitions item was provided to demonstrators prior to testing. The standard types for the three example items are 20-mmP, 37-mmP, and 40-mmP.

TABLE 7a. CAPPED CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS MUNITIONS

Size	Percentage Correct
20-mm	0.67
37-mm	0.80
40-mm	0.67
Overall	0.71

TABLE 7b. UNCAPPED CORRECT TYPE CLASSIFICATION OF TARGETS CORRECTLY DISCRIMINATED AS MUNITIONS

Size	Percentage Correct
20-mm	0.70
37-mm	0.80
40-mm	0.70
Overall	0.73

4.5 LOCATION ACCURACY

The mean location error and standard deviations are presented in Tables 8a and 8b. These calculations are based on average missed distance for munitions correctly identified during the response stage. Depths are measured from the center of the munitions to the surface. For the blind grid, only depth errors are calculated because (X, Y) positions are known to be the centers of the grid square.

TABLE 8a. CAPPED MEAN LOCATION ERROR AND STANDARD DEVIATION

	Mean	Standard Deviation
Depth	0.099	0.113

TABLE 8b. UNCAPPED MEAN LOCATION ERROR AND STANDARD DEVIATION

	Mean	Standard Deviation
Depth	0.150	0.161

4.6 STATISTICAL COMPARISONS

Statistical two-sided Fisher’s exact significance tests were used to compare results between the capped and uncapped test areas. The intent of the comparison is to determine if the cap has an effect on the performance of the sensor system. However, any modifications in the UXO sensor system during the test, like changes in the processing or changes in the selection of the operating threshold, will also contribute to performance differences.

The Fisher’s exact test was used to compare capped to uncapped test areas with regard to P_d^{res} , P_d^{disc} , P_{fp}^{res} , P_{fp}^{disc} , P_{ba}^{res} , P_{ba}^{disc} , Efficiency, False Positive Rejection Rate (R_{fp}) and Background Alarm Rejection Rate (R_{ba}). The confidence level used for the test was 0.1. The test results are presented in Table 9. A detailed explanation and example of the Fisher’s exact test application is located in Appendix A.

TABLE 9. FISHER’S EXACT RESULTS - CAPPED VERSUS UNCAPPED

Metric	Ordnance Size			Overall
	20-mm	37-mm	40-mm	
P_d^{res}	Not significant	Not significant	Not significant	Not significant
P_d^{disc}	Not significant	Not significant	Not significant	Not significant
P_{fp}^{res}				Not significant
P_{fp}^{disc}				Not significant
P_{ba}^{res}				Not significant
P_{ba}^{disc}				Not significant
Efficiency				Not significant
R_{fp}				Not significant
R_{ba}				Not significant

SECTION 5. APPENDIXES

APPENDIX A. TERMS AND DEFINITIONS

GENERAL DEFINITIONS

Anomaly: Location of a system response deemed to warrant further investigation by the demonstrator for consideration as an emplaced munitions item.

Detection: An anomaly location that is within R_{halo} of an emplaced munitions item.

Military Munitions: Specific categories of military munitions that may pose unique explosive safety risks, including UXO as defined in 10 USC 101(e)(5), discarded military munitions as defined in 10 USC 2710(e)(2) and/or munitions constituents (e.g., TNT, RDX) as defined in 10 USC 2710(e)(3) that are present in high enough concentrations to pose an explosive hazard.

Emplaced Munitions: A munitions item buried by the government at a specified location in the test site.

Emplaced Clutter: A clutter item (i.e., non-munitions item) buried by the government at a specified location in the test site.

R_{halo} : A predetermined radius about an emplaced item (clutter or munitions) within which an anomaly identified by the demonstrator as being of interest is considered to be a detection of that item. For the purpose of this program, a circular halo 0.5 m in radius is placed around the center of the object for all clutter and munitions items.

Small Munitions: Caliber of munitions less than or equal to 40 mm (includes 20-mm projectile, 25-mm projectile, 37-mm projectile, 40-mm projectile, submunitions BLU-26, BLU-63, and M42).

Medium Munitions: Caliber of munitions greater than 40 mm and less than or equal to 81 mm (includes 57-mm projectile, 60-mm mortar, 2.75-inch rocket, and 81-mm mortar).

Large Munitions: Caliber of munitions greater than 81 mm (includes 105-mm high-explosive antitank (HEAT), 105-mm projectile, and 155-mm projectile).

Group: Two or more adjacent GT items with overlapping halos.

GT: Ground truth

Response Stage Noise Level: The level that represents the signal level below which anomalies are not considered detectable. Demonstrators are required to provide the recommended noise level for the blind grid test area.

Discrimination Stage Threshold: The demonstrator-selected threshold level that is expected to provide optimum performance of the system by retaining all detectable munitions and rejecting the maximum amount of clutter. This level defines the subset of anomalies the demonstrator would recommend digging based on discrimination.

Binomially Distributed Random Variable: A random variable of the type which has only two possible outcomes, say success and failure, is repeated for n independent trials with the probability p of success and the probability $1-p$ of failure being the same for each trial. The number of successes x observed in the n trials is an estimate of p and is considered to be a binomially distributed random variable.

RESPONSE AND DISCRIMINATION STAGE DATA

The scoring of the demonstrator's performance is conducted in two stages: response stage and discrimination stage. For both stages, the probability of detection (P_d) and the false alarms are reported as ROC curves. False alarms are divided into those anomalies that correspond to emplaced clutter items, measuring the probability of clutter detection (P_{cd}) or probability of false positive (P_{fp}). Those that do not correspond to any known item are termed background alarms.

The response stage is a measure of whether the sensor can detect an object of interest. For a channel instrument, this value should be closely related to the amplitude of the signal. The demonstrator must report the response level (threshold) below which target responses are deemed insufficient to warrant further investigation. At this stage, minimal processing may be done. This includes filtering long- and short-scale variations, bias removal, and scaling. This processing should be detailed in the data submission.

For a multichannel instrument, the demonstrator must construct a quantity analogous to amplitude. The demonstrator should consider what combination of channels provides the best test for detecting any object that the sensor can detect. The average amplitude across a set of channels is an example of an acceptable response stage quantity. Other methods may be more appropriate for a given sensor. Again, minimal processing can be done, and the demonstrator should explain how this quantity was constructed in their data submission.

The discrimination stage evaluates the demonstrator's ability to correctly identify munitions as such, and to reject clutter. For the same locations as in the response stage anomaly list, the discrimination stage list contains the output of the algorithms applied in the discrimination-stage processing. This list is prioritized based on the demonstrator's determination that an anomaly location is likely to contain munitions. Thus, higher output values are indicative of higher confidence that a munitions item is present at the specified location. For electronic signal processing, priority ranking is based on algorithm output. For other systems, priority ranking is based on human judgment. The demonstrator also selects the threshold that the demonstrator believes will provide optimum system performance, (i.e., that retains all the detected munitions and rejects the maximum amount of clutter).

Note: The two lists provided by the demonstrator contain identical numbers of potential target locations. They differ only in the priority ranking of the declarations.

RESPONSE STAGE DEFINITIONS

Response Stage Probability of Detection (P_d^{res}): $P_d^{res} = (\text{No. of response-stage detections}) / (\text{No. of emplaced munitions in the test site})$.

Response Stage Clutter Detection (cd^{res}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Response Stage Probability of Clutter Detection (P_{cd}^{res}): $P_{cd}^{res} = (\text{No. of response-stage clutter detections})/(\text{No. of emplaced clutter items})$.

Response Stage Background Alarm (ba^{res}): An anomaly in a blind grid cell that contains neither emplaced munitions nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced munitions or emplaced clutter item.

Response Stage Probability of Background Alarm (P_{ba}^{res}): Blind grid only: $P_{ba}^{res} = (\text{No. of response-stage background alarms})/(\text{No. of empty grid locations})$.

Response Stage Background Alarm Rate (BAR^{res}): Open field any challenge area (including the direct and indirect firing sub areas) only: $BAR^{res} = (\text{No. of response-stage background alarms})/(\text{arbitrary constant})$.

Note that the quantities P_d^{res} , P_{cd}^{res} , P_{ba}^{res} , and BAR^{res} are functions of t^{res} , the threshold applied to the response-stage signal strength. These quantities can therefore be written as $P_d^{res}(t^{res})$, $P_{cd}^{res}(t^{res})$, $P_{ba}^{res}(t^{res})$, and $BAR^{res}(t^{res})$.

DISCRIMINATION STAGE DEFINITIONS

Discrimination: The application of a signal processing algorithm or human judgment to sensor data to discriminate munitions from clutter. Discrimination should identify anomalies that the demonstrator has high confidence correspond to munitions, as well as those that the demonstrator has high confidence correspond to non-munitions or background returns. The former should be ranked with highest priority and the latter with lowest.

Discrimination Stage Probability of Detection (P_d^{disc}): $P_d^{disc} = (\text{No. of discrimination-stage detections})/(\text{No. of emplaced munitions in the test site})$.

Discrimination Stage False Positive (fp^{disc}): An anomaly location that is within R_{halo} of an emplaced clutter item.

Discrimination Stage Probability of False Positive (P_{fp}^{disc}): $P_{fp}^{disc} = (\text{No. of discrimination stage false positives})/(\text{No. of emplaced clutter items})$.

Discrimination Stage Background Alarm (ba^{disc}): An anomaly in a blind grid cell that contains neither emplaced munitions nor an emplaced clutter item. An anomaly location in the open field or scenarios that is outside R_{halo} of any emplaced munitions or emplaced clutter item.

Discrimination Stage Probability of Background Alarm (P_{ba}^{disc}): $P_{ba}^{disc} = (\text{No. of discrimination-stage background alarms})/(\text{No. of empty grid locations})$.

Discrimination Stage Background Alarm Rate (BAR^{disc}): $BAR^{disc} = (\text{No. of discrimination-stage background alarms})/(\text{arbitrary constant})$.

Note that the quantities P_d^{disc} , P_{fp}^{disc} , P_{ba}^{disc} , and BAR^{disc} are functions of t^{disc} , the threshold applied to the discrimination-stage signal strength. These quantities can therefore be written as $P_d^{disc}(t^{disc})$, $P_{fp}^{disc}(t^{disc})$, $P_{ba}^{disc}(t^{disc})$, and $BAR^{disc}(t^{disc})$.

RECEIVER-OPERATING CHARACTERISTIC (ROC) CURVES

ROC curves at both the response and discrimination stages can be constructed based on the above definitions. The ROC curves plot the relationship between P_d versus P_{cd} or P_{fp} and P_d versus BAR or P_{ba} as the threshold applied to the signal strength is varied from its minimum (t_{min}) to its maximum (t_{max}) value.¹ P_d versus P_{fp} and P_d versus BAR being combined into ROC curves are shown in Figure A-1. Note that the “res” and “disc” superscripts have been suppressed from all the variables for clarity.

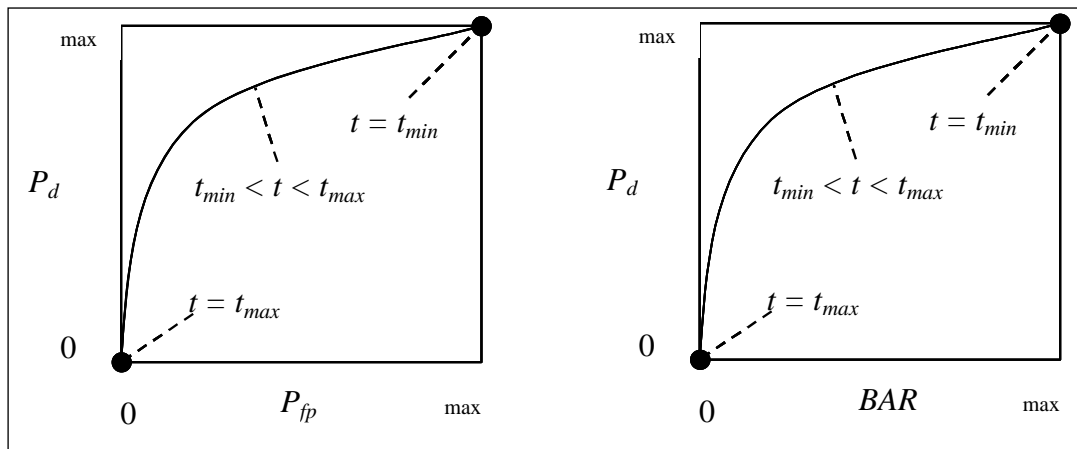


Figure A-1. ROC curves for open field testing. Each curve applies to both the response and discrimination stages.

METRICS TO CHARACTERIZE THE DISCRIMINATION STAGE

The demonstrator is also scored on efficiency and rejection ratio, which measure the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of munitions detections from the anomaly list while rejecting the maximum number of anomalies arising from non-munitions items. The efficiency measures the fraction of detected munitions retained by the discrimination, while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to the entire response list, i.e., the maximum munitions detectable by the sensor and its accompanying clutter detection rate/false positive rate or background alarm rate.

¹Strictly speaking, ROC curves plot the P_d versus P_{ba} over a predetermined and fixed number of detection opportunities (some of the opportunities are located over munitions and others are located over clutter or blank spots). In an open field scenario, each system suppresses its signal strength reports until some bare-minimum signal response is received by the system. Consequently, the open field ROC curves do not have information from low signal-output locations, and, furthermore, different contractors report their signals over a different set of locations on the ground. These ROC curves are thus not true to the strict definition of ROC curves as defined in textbooks on detection theory. Note, however, that the ROC curves obtained in the blind grid test sites are true ROC curves.

Efficiency (E): $E = P_d^{\text{disc}}(t^{\text{disc}})/P_d^{\text{res}}(t_{\text{min}}^{\text{res}})$: Measures (at a threshold of interest) the degree to which the maximum theoretical detection performance of the sensor system (as determined by the response stage t_{min}) is preserved after application of discrimination techniques. Efficiency is a number between 0 and 1. An efficiency of 1 implies that all of the munitions initially detected in the response stage were retained at the specified threshold in the discrimination stage, t^{disc} .

False Positive Rejection Rate (R_{fp}): $R_{\text{fp}} = 1 - [P_{\text{fp}}^{\text{disc}}(t^{\text{disc}})/P_{\text{cd}}^{\text{res}}(t_{\text{min}}^{\text{res}})]$: Measures (at a threshold of interest) the degree to which the sensor system's false positive performance is improved over the maximum false positive performance (as determined by the response stage t_{min}). The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all emplaced clutter initially detected in the response stage were correctly rejected at the specified threshold in the discrimination stage.

Background Alarm Rejection Rate (R_{ba}): $R_{\text{ba}} = 1 - [P_{\text{ba}}^{\text{disc}}(t^{\text{disc}})/P_{\text{ba}}^{\text{res}}(t_{\text{min}}^{\text{res}})]$. Measures the degree to which the discrimination stage correctly rejects background alarms initially detected in the response stage. The rejection rate is a number between 0 and 1. A rejection rate of 1 implies that all background alarms initially detected in the response stage were rejected at the specified threshold in the discrimination stage.

FISHER'S EXACT TEST

Fisher's exact test for differences in probabilities (or 2 by 2 contingency table) is used to analyze two samples drawn from two different populations to see if both populations have the same or different proportions of elements in a certain category. More specifically, two random samples are drawn, one from each population, to test the null hypothesis that the probability of some specified event is the same for both populations.

Two-sided tests are completed to compare the performance factors for the capped and uncapped areas. The confidence level for the tests was 0.1. For the generic 2 by 2 contingency table defined in Table A-1, the probability associated with that exact probability of obtaining that table is given below:

$$p = \frac{\binom{a+b}{a} \binom{c+d}{c}}{\binom{n}{a+c}}$$

The p-value for the two-sided test is obtained by summing the probabilities for each possible table given constant margins whose probability is less than or equal to the probability of the given table.

TABLE A-1. GENERIC 2 BY 2
CONTINGENCY TABLE

	Capped	Uncapped
Detected	a	c
Not detected	b	d

An example follows that illustrates Fisher's exact test using fabricated detection results for capped and uncapped areas. It should be noted that a significant result does not prove a cause-and-effect relationship exists between the two populations of interest; however, it does serve as a tool to indicate a change in system performance at a large enough level than can be accounted for merely by chance or random variation. Note also that a result that is not significant indicates that there is not enough evidence to declare that anything more than chance or random variation within the same population is at work between the two data sets being compared.

Demonstrator X achieves the following overall results for P_d^{res} after surveying the Small Munitions Test Site using the same system (results indicate the number of munitions detected divided by the number of munitions emplaced):

Capped	Uncapped
$3/10 = 0.70$	$9/10 = 0.90$

P_d^{res} : CAPPED versus UNCAPPED. Using the example data above to compare probabilities of detection in the response stage, 3 munitions out of 10 emplaced munitions items were detected in the capped area while 9 munitions out of 10 emplaced were detected in the uncapped area. The resulting contingency table is given in Table A-2.

TABLE A-2. EXAMPLE
CONTINGENCY TABLE

	Capped	Uncapped
Detected	3	9
Not detected	7	1

The probability associated with this exact table is 0.00953. Summing all possible tables given constant margins whose probability is less than or equal to 0.00953 results in a p-value of 0.020. Since 0.020 is less than the specified confidence level 0.1, the result of the test is that there is a statistically significant difference between the capped and uncapped areas with respect to the P_d^{res} factor. While a significant result does not prove a cause-and-effect relationship exists, it does indicate that the difference in detection ability between the capped and uncapped areas of Demonstrator X's system is unlikely to be explained completely by chance or random variation.

APPENDIX B. DAILY WEATHER LOGS

Date, 2015	Time, EST	Average Temperature, °F	Total Precipitation, in.
27 July	0700	70.1	0.00
	0800	70.9	0.00
	0900	71.9	0.00
	1000	73.9	0.00
	1100	75.3	0.00
	1200	76.1	0.00
	1300	76.9	0.00
	1400	76.9	0.00
	1500	77.8	0.00
	1600	79.7	0.00
28 July	1700	80.6	0.00
	0700	71.2	0.00
	0800	73.6	0.00
	0900	77.5	0.00
	1000	80.3	0.00
	1100	82.4	0.00
	1200	83.5	0.00
	1300	85.0	0.00
	1400	85.6	0.00
	1500	87.1	0.00
29 July	1600	88.1	0.00
	1700	88.5	0.00
	0700	68.8	0.00
	0800	72.7	0.00
	0900	77.6	0.00
	1000	80.5	0.00
	1100	82.7	0.00
	1200	83.9	0.00
	1300	85.4	0.00
	1400	86.5	0.00
30 July	1500	86.3	0.00
	1600	84.1	0.00
	1700	83.7	0.00
	0700	77.8	0.00
	0800	78.6	0.00
	0900	80.0	0.00
	1000	81.1	0.00
	1100	82.1	0.00
	1200	82.8	0.00
	1300	84.1	0.00
30 July	1400	85.9	0.00
	1500	83.1	0.00
	1600	74.3	0.27
	1700	76.1	0.00

EST = Eastern Standard Time

APPENDIX C. SOIL MOISTURE

Date: 27 July 2015			
Probe Location	Layer, in.	AM Reading, %	PM Reading, %
Small munitions calibration lanes	0 to 6	27.5	27.4
	6 to 12	33.9	33.9
	12 to 24	37.3	37.1
	24 to 36	41.7	41.6
	36 to 48	58.6	58.5
Small munitions grid	0 to 6		
	6 to 12		
	12 to 24		
	24 to 36		
	36 to 48		

Date: 29 July 2015			
Probe Location	Layer, in.	AM Reading, %	PM Reading, %
Small munitions calibration lanes	0 to 6		
	6 to 12		
	12 to 24		
	24 to 36		
	36 to 48		
Small munitions grid	0 to 6		29.5
	6 to 12		39.7
	12 to 24		45.6
	24 to 36		44.8
	36 to 48		53.7

Date: 30 July 2015			
Probe Location	Layer, in.	AM Reading, %	PM Reading, %
Small munitions calibration lanes	0 to 6		
	6 to 12		
	12 to 24		
	24 to 36		
	36 to 48		
Small munitions grid	0 to 6	29.4	29.4
	6 to 12	39.5	39.6
	12 to 24	45.4	45.5
	24 to 36	44.8	44.7
	36 to 48	53.6	53.6

Date, 2015	No. of People	Area Tested	Status Start Time, hr	Status Stop Time, hr	Duration, min.	Operational Status	Operational Status Comments	Track Method	Pattern	Field Conditions	
27 July	4	Calibration Lanes	0830	0945	75	Initial setup	Initial mobilization	GPS	Linear	Hot dry	Cloudy
			0945	1015	30	Collecting data	Collect data	GPS	Linear	Hot dry	Cloudy
			1015	1215	120	Collecting data	Collect data	GPS	Linear	Hot dry	Cloudy
			1215	1225	10	Downtime due to equipment maintenance/check	Download data	GPS	Linear	Hot dry	Cloudy
			1225	1250	25	Break/lunch	Break/lunch	GPS	Linear	Hot dry	Cloudy
		Blind Test Grid	1250	1540	170	Collecting data	Collect data	GPS	Linear	Hot dry	Cloudy
			1540	1610	30	Daily start, stop	Equipment breakdown	GPS	Linear	Hot dry	Cloudy
28 July	4	Blind Test Grid	0730	755	25	Daily start, stop	Set up equipment	GPS	Linear	Hot dry	Sunny
			0755	805	20	Calibration	Calibration	GPS	Linear	Hot dry	Sunny
			0805	955	110	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
			0955	1005	10	Downtime due to equipment maintenance/check	Change batteries	GPS	Linear	Hot dry	Sunny
			1005	1200	115	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
			1200	1235	35	Break/lunch	Break/lunch	GPS	Linear	Hot dry	Sunny
			1235	1520	165	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
			1520	1530	10	Downtime due to equipment maintenance/check	Download data, battery issue	GPS	Linear	Hot dry	Sunny
29 July	4	Blind Test Grid	0745	0815	30	Daily start, stop	Set up equipment	GPS	Linear	Hot dry	Sunny
			0815	0930	75	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
		Small Munition	0930	1150	160	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
			1150	1230	40	Break/lunch	Break/lunch	GPS	Linear	Hot dry	Sunny
			1230	1235	5	Calibration	Calibration	GPS	Linear	Hot dry	Sunny
			1235	1540	185	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
			1540	1610	30	Daily start, stop	Equipment breakdown	GPS	Linear	Hot dry	Sunny
30 July	4	Small Munition	0745	0805	20	Daily start, stop	Set up equipment	GPS	Linear	Hot dry	Sunny
			0805	0855	50	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
			0855	0920	25	Downtime due to equipment failure	Battery swapped	GPS	Linear	Hot dry	Sunny
			0920	1025	65	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
			1025	1215	110	Downtime due to equipment failure	Battery swapped	GPS	Linear	Hot dry	Sunny
			1215	1320	65	Collecting data	Collect data	GPS	Linear	Hot dry	Sunny
1320	1405	45	Demobilization	Demobilization	GPS	Linear	Hot dry	Sunny			

APPENDIX E. REFERENCES

1. Standardized UXO Technology Demonstration Site Handbook, DTC Project No. 8-CO-160-000-473, Report No. ATC-8349, March 2002.
2. Aberdeen Proving Ground Soil Survey Report, October 1998.
3. Data Summary, UXO Standardized Test Site: APG Soils Description, May 2002.

APPENDIX F. ABBREVIATIONS

APG	= Aberdeen Proving Ground
ATC	= U.S. Army Aberdeen Test Center
ATSS	= Aberdeen Test Support Services
BAR	= background alarm rate
EMI	= electromagnetic induction
EQT	= Environmental Quality Technology
ERDC	= U.S. Army Corps of Engineers Engineering Research and Development Center
EST	= Eastern Standard Time
ESTCP	= Environmental Security Technology Certification Program
GPS	= Global Positioning System
GT	= ground truth
HEAT	= high-explosive antitank
JD	= Joint Diagonalization
MSR	= multi-static response
NRL	= Naval Research Laboratory
ONVMS	= orthonormalized volume magnetic source
P_{ba}	= probability of background alarm
P_{cd}	= probability of clutter detection
P_d	= probability of detection
P_d^{res}	= probability of detection for the response stage
P_d^{disc}	= probability of detection for the discrimination stage
P_{fp}	= probability of false positive
POC	= point of contact
QA	= quality assurance
QC	= quality control
R_{ba}	= background alarm rejection
R_{fp}	= false positive rejection
R_{halo}	= Halo Radius
ROC	= receiver-operating characteristic
SERDP	= Strategic Environmental Research and Development Program
TDSS	= Threat Detection and Systems Survivability
TEMTADS	= Time-Domain Electromagnetic Multi-Sensor Towed Array Detection System
USAEC	= U.S. Army Environmental Command
UXO	= unexploded ordnance
YPG	= U.S. Army Yuma Proving Ground

APPENDIX G. DISTRIBUTION LIST

ATEC Project No. 2011-DT-ATC-DODSP-F0292

Note: A copy of this test report has been posted to the Versatile Information Systems Integrated On-Line (VISION) Digital Library (VDL), <https://vdl.s.atc.army.mil/>.

<u>Addressee</u>	<u>No. of Copies</u>
Program Manager Strategic Environmental Research and Development Program/ Environmental Security Technology Certification Program Munitions Management ATTN: Mr. Herb Nelson 4800 Mark Center Drive, Suite 17D08 Alexandria, VA 22350	1
Dartmouth College, Thayer School of Engineering ATTN: Fridon Shubitidze 14 Engineering Drive Hanover, NH 03755	1
Commander U.S. Army Aberdeen Test Center ATTN: TEDT-AT-SLM (Mr. Gene L. Fabian) TEDT-AT-CSM 400 Collieran Road Aberdeen Proving Ground, MD 21005-5059	1 1
Defense Technical Information Center 8725 John J. Kingman Road, Suite 0944 Fort Belvoir, VA 22060-6218	1

Secondary distribution is controlled by Program Manager, SERDP/ESTCP, Munitions Management, ATTN: Mr. Herb Nelson.