AD NO. ___________________________
DTC PROJECT NO. 8-CO-160-UXO-020
REPORT NO. ATC-10003

UXO TECHNOLOGY DEMONSTRATION SITE

ACTIVE SITE SCORING RECORD NO. 931

SITE LOCATION:
U.S. ARMY ABERDEEN PROVING GROUND

DEMONSTRATOR:
NAVAL RESEARCH LABORATORY (NRL)
CODE 6110 NAVAL RESEARCH LABORATORY
WASHINGTON, DC 20375-5342

TECHNOLOGY TYPE/PLATFORM:
MAGNETOMETER MTADS/TOWED

PREPARED BY:
U.S. ARMY ABERDEEN TEST CENTER
ABERDEEN PROVING GROUND, MD 21005-5059

JULY 2009

Prepared for:
U.S. ARMY ENVIRONMENTAL CENTER
ABERDEEN PROVING GROUND, MD 21010-5401

U.S. ARMY DEVELOPMENTAL TEST COMMAND
ABERDEEN PROVING GROUND, MD 21005-5055

DISPOSITION INSTRUCTIONS

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
July 2009

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
1 July 2003

4. TITLE AND SUBTITLE
STANDARDIZED UXO TECHNOLOGY DEMONSTRATION SITE SCORING RECORD NO. 931

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
McClung, Stephen J.

5d. PROJECT NUMBER
8-CO-160-UXO-020

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Developmental Test Command
400 Colleran Road
Aberdeen Proving Ground, MD 21005-5059

8. PERFORMING ORGANIZATION REPORT NUMBER
ATC-10003

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Environmental Command
ATTN: IMAE-RTA
314 Longs Corner Road
Warren, MI 48397-5000

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 Naval Research Laboratory (NRL) to detect and discriminate inert unexploded ordnance (UXO) utilizing the APG standardized UXO Technology Demonstration Site Blind Grid, Open Field, and Active Sites. This Scoring Record was coordinated by J. Stephen McClung and the Standardized UXO Technology Demonstration Site Scoring Committee. Organizations on 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
Naval Research Laboratory, APG Standardized UXO Technology Demonstration Site, blind grid, open field, active, magnetometer MTADS/Towed.

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	SAR		19b. TELEPHONE NUMBER *(Include area code)*

TABLE OF CONTENTS

PAGE

SECTION 1. GENERAL INFORMATION

SECTION 2. DEMONSTRATION

SECTION 3. FIELD DATA

SECTION 4. TECHNICAL PERFORMANCE RESULTS

PAGE

SECTION 5. ON-SITE LABOR COSTS

SECTION 6. APPENDIXES

SECTION 1. GENERAL INFORMATION

1.1 BACKGROUND

Technologies under development for the detection and discrimination of munitions and explosives of concern (MEC) - i.e., unexploded ordnance (UXO) and discarded military munitions (DMM) require testing and evaluation in order for their performance to be characterized. It is imperative that this characterization be performed on a realistic test site in order to successfully gauge how well a system may perform at an actual munitions response site. To that end, the Active Response Demonstration Site has been developed at Aberdeen Proving Ground (APG), Maryland. This site provides the ability to test technologies under development on an actual test range that has a large number of UXO, MEC, and DMM that have not been cleared. Realistic characteristics of the Active Response Site include significant quantities of live UXO, range scrap, and excess debris. Testing at this site 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 validating the standardized UXO test sites.

The Active Response 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.

1.2 SCORING OBJECTIVES

The objective in the Active Response Demonstration Site Program is to evaluate the detection and discrimination capabilities of a given technology under realistic conditions. The only UXO that were cleared before vendors were allowed to survey the area are items that pose a safety hazard.

The evaluation objectives are as follows:

a. To determine detection and discrimination effectiveness under a realistic scenario.

b. To determine cost, time, and manpower 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) and geo-referenced data for post-demonstration analysis.

1.2.1 Scoring Methodology

The Active Response Demonstration Site is divided into 20 meter by 20 meter grids. The grids are ranked based upon the density of items that have accumulated in each respective grid cell. After multiple vendors surveyed the area with their UXO detection/discrimination systems, half of the 2 acre site was cleared of all metallic items. This clearing of the metallic anomalies from the 2 acre Active Response Demonstration Site was broken into three phases. In the first phase, the target lists from all of the vendors that have surveyed the site were combined in order to create a master target list that was used in the initial phase of the site clearance. Once Phase 1 was completed, a secondary sweep of the site took place and another recovery operation was performed. After the secondary investigation was completed, the Naval Research Laboratory (NRL) conducted a survey of the site with their Multiple Towed Array Detection System (MTADS). This system is known for its effectiveness and ability to detect metallic items. Once the NRL MTADS surveyed the site, ATC collected their data and conducted another intrusive operation in order to remove any additional anomalies. During each clearance operation, the exact placement of all the metallic items was carefully measured in order to create a GT for each grid cell. Once the GT for each cell was compiled, each item in the GT was classified as being either ordnance or clutter. Clutter items are defined as metallic items that do not have enough explosives to be considered safety hazards. Fuzes that no longer have their boosters, fins, fragmented items, and items that were never part of any ordnance item, for example, were classified as clutter. The remaining objects that pose a safety risk were classified as ordnance. This GT will be used to score all of the vendors that had previously surveyed the site, prior to clearance.

a. The scoring of the demonstrator's performance is conducted in two stages. These two stages are termed the 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 clutter items, measuring the probability of false positive (P_{fp}), and those that do not correspond to any known item, termed background alarms.

b. The response stage scoring evaluates the ability of the system to detect targets without regard to ability to discriminate ordnance from other anomalies. This list is generated with minimal processing.

c. The discrimination stage evaluates the demonstrator's ability to correctly identify ordnance as such and to reject clutter. For the discrimination stage, the demonstrator provides the scoring committee with the output of the algorithms applied in the discrimination-stage processing. The values in this list are prioritized based on the demonstrator's determination that an item is ordnance. Thus, higher output values are indicative of higher confidence that an ordnance item is present at the specified location. For digital signal processing, priority ranking is based on algorithm output. For other discrimination approaches, priority ranking is based on human (subjective) judgment. The demonstrator also specifies the threshold in the prioritized ranking that provides optimum performance, (i.e., that is expected to retain all detected ordnance and rejects the maximum amount of clutter).

d. The demonstrator is also scored on efficiency and rejection ratio, which measures the effectiveness of the discrimination stage processing. The goal of discrimination is to retain the greatest number of ordnance detections from the anomaly list, while rejecting the maximum number of anomalies arising from nonordnance items. Efficiency measures the fraction of detected ordnance retained after discrimination (give ratio), while the rejection ratio measures the fraction of false alarms rejected. Both measures are defined relative to performance at the demonstrator-supplied level below which all responses are considered noise (i.e., the maximum ordnance detectable by the sensor and its accompanying false positive rate or background alarm rate).

e. Depending on the density of items that are in a given grid, there exists the possibility of having anomalies within overlapping halos (halo = 1-m diameter) and/or multiple anomalies within halos. In these cases, the following scoring logic is implemented:

(1) For each anomaly supplied by the vendor, the vendor can be only given credit for finding, at most, one ordnance item. In other words, if a vendor gives only one anomaly that is within 0.5 meters from six grenades, he will only be given credit for finding one of those six grenades.

(2) In situations where multiple anomalies exist within a single R_{halo}, the anomaly with the strongest response or highest ranking will be assigned to that particular GT item. For example, if a vendor supplies two anomalies that are within 0.5 meters from a given ordnance item, and one of the anomalies has a signal level (response level if we are calculating the response stage value, or the discrimination ranking if we are calculating the discrimination stage value) of 0 while another anomaly has a signal level 1, then the anomaly with a signal level of 1 will be given credit for finding that particular GT item. The anomaly with a signal level of 0 will then be free to be possibly attached to another GT item if there is another GT item that is within 0.5 meters from that anomaly.

(3) For overlapping R_{halo} situations, ordnance has precedence over clutter. The anomaly with the strongest response or highest ranking that is closest to the center of a particular GT item gets assigned to that item. Remaining anomalies are retained until all matching is complete. In other words, if a vendor supplies only one anomaly that is within 0.5 meters of both an ordnance and clutter item, the vendor will be given credit for finding the ordnance item. On the other hand, if a vendor supplies only one anomaly that is within 0.5 meters of two ordnance items, then the vendor will be given credit for finding whichever ordnance item is closest to the vendor's anomaly.

(4) Anomalies located within any R_{halo} that do not get associated with a particular GT item are thrown out and are not considered in the analysis. As an example, if a vendor supplies two anomalies that are within 0.5 meters from a GT item, and this is not an overlapping halo situation, then one of the anomalies will be used so that the vendor gets credit for finding this GT item, but the second anomaly will neither be used to give the vendor credit for finding a GT item nor will this item be counted as a background alarm.

(5) All anomalies that are supplied by the vendor that are either outside of the boundary of the active site or are within 1 meter of the boundary of the active site will be thrown out and will not be counted as background alarms nor will they contribute to the vendors P_d or P_{fp}. Likewise, all GT items that are outside of the boundary of the active area or are within 1 meter of the boundary of the active site will be thrown out and will not contribute to the vendor's P_d or P_{fp}. If a vendor supplies an anomaly that is within the active site and more than 1 meter away from the boundary of the active site, and this anomaly is within the halo of a GT item that is closer than 1 meter to the boundary of the active site, but this anomaly is not within the halo of a GT item that is further than 1 meter away from the boundary of the active site, then this anomaly will neither be counted as a background alarm, nor will it contribute to the vendors P_d or P_{fp}.

f. All scoring factors are generated utilizing the Standardized UXO Probability and Plot Program, version 4.0 using the earlier version 3.11 rules so results can be compared to surveys done in the blind grid and open field area of the Standardized UXO Test Site.

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 (P_d^{res}).

(2) Probability of False Positive (P_{fp}^{res}).

(3) Background Alarm Rate (BAR^{res}).

b. Discrimination Stage ROC curves:

(1) Probability of Detection (P_d^{disc}).

(2) Probability of False Positive (P_{fp}^{disc}).

(3) Background Alarm Rate (BAR^{disc}).

c. Metrics:

(1) Efficiency (E).

(2) False Positive Rejection Rate (R_{fp}).

(3) Background Alarm Rejection Rate (R_{BA}).

d. Other:

(1) Location accuracy.

(2) Equipment setup, calibration time, and corresponding worker-hour requirements.

(3) Survey time and corresponding worker-hour requirements.

(4) Reacquisition/resurvey time and worker-hour requirements (if any).

(5) Downtime due to system malfunctions and maintenance requirements.

SECTION 2. DEMONSTRATION

2.1 DEMONSTRATOR INFORMATION

2.1.1 System Description (provided by demonstrator)

The MTADS hardware consists of a low-magnetic signature vehicle that is used to tow linear arrays of magnetometer sensors to conduct surveys of large areas to detect buried UXO. The MTADS tow vehicle, manufactured by Chenoweth Racing Vehicles, is a custom-built off-road vehicle, specifically modified to have an extremely low magnetic signature. Most ferrous components have been removed from the body, drivetrain, and engine and replaced with nonferrous alloys.

The MTADS magnetometers are Cs-vapor full-field magnetometers (Geometrics Model No. 822ROV) selected for low noise and inter-sensor reproducibility. Eight sensors are deployed as a magnetometer array on an aluminum and composite platform. The sensors are sampled at 50 Hz and typical surveys are conducted at 6 mph; this results in a sampling density of about 6 cm along a track with a horizontal sensor spacing of 25 cm.

The sensor positions are measured in real time (5 Hz) using the latest real-time kinematic (RTK) Global Positioning System (GPS) technology. All navigation and sensor data are time-stamped and recorded by the data acquisition computer in the tow vehicle. The Data Analysis System (DAS) employs routines to convert these sensor and position data streams into anomaly maps for analysis.

Figure 1. Demonstrator system, magnetometer (MAG) MTADS/towed.

2.1.2 <u>Data Processing Description (provided by demonstrator)</u>

The MTADS magnetometer array is pulled by the MTADS tow vehicle over the site at approximately 6 mph. Lane spacing is the width of the MTADS tow vehicle, approximately 1.75 meters. Data are recorded from the array at 50 Hz. This results in a down-track sampling interval of about 6 cm and a cross-track sampling interval of 25 cm.

The magnetometer sensors are arranged in a rigid array with the GPS antenna hard mounted on the array so a single GPS measurement suffices. All sensor readings are referenced to the GPS 1-Precise Positioning System (PPS) output to fully take advantage of the precision of the GPS measurements.

The individual data streams (sensor readings, GPS positions, times, etc.) are collected by the data acquisition computer, running a custom variant of the MagLog NT program, and are each recorded in a separate file. These individual data files, which share a root name, include two (magnetometer array) GPS files (one containing the NMEA GGK sentences corresponding to the position of the master antenna and an AVR sentence giving one of the vectors to the secondary antennas, another containing the second AVR sentence, a third containing the Coordinated Universal Time (UTC) time tag, and the fourth containing the computer-time stamped arrival of the GPS PPS). All files are American Standard Code for Information Interchange (ASCII) format.

All these files are transferred to the DAS. They are then checked for data quality, leveled, and the position information is applied to the sensor files. The result is a sequence of positioned measurements of the measured response. This latter file is referred to as raw data.

2.1.3 <u>Data Submission Format</u>

Data were submitted for scoring in accordance with data submission protocols outlined in the Standardized UXO Technology Demonstration Site Handbook. These submitted data are not included in this report in order to protect GT information.

2.1.4 <u>Demonstrator Quality Assurance (QA) and Quality Control (QC) (provided by demonstrator)</u>

There are two items that need to be checked daily to ensure adequate system performance: individual sensor response and reliability of GPS positions. Before beginning survey work each day, the performance of each of the sensors in the array is measured (after a 10 to 15 min warmup) by presenting a standard target to each sensor in turn. The resulting signals are checked against standard values.

The data acquisition system gives the vehicle operator a continuous reading of the quality of the GPS fix. Standard procedure is to take only data with a GPS fix quality of 3 (RTK fixed). Before arriving at the site each day, standard GPS planning software is used to calculate the number of satellites that will be visible to the receivers and the precision dilution of precision (PDOP) achievable minute-by-minute throughout the day. This allows short breaks during periods of poor satellite availability to be planned and keeps data that will have to be discarded later from inadvertently being taken.

At the end of each one-hour survey session, all survey data is transferred to the field data analyst for preliminary data quality checks. This process involves plotting the actual survey path as logged in the GPS files (color-coded by GPS fix quality) to ensure that GPS data of sufficient quality was obtained during the survey. Following this, the individual sensor files are examined for completeness and consistency. It is at this stage that any sensor malfunctions, drifts, etc. are flagged and reported to the field crew for correction. The final task for the field analyst is to calculate a position for each sensor reading and apply it to the reading. The mapped data files are then ready for analysis either in the field, or at a later time.

2.1.5 <u>Additional Records</u>

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

2.2 APG SITE INFORMATION

2.2.1 Location

The APG Active Response Demonstration 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 Active Response Demonstration Site encompasses 1.98 acres of upland and lowland flats.

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.

ERDC conducted a site-specific analysis in May of 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.

SECTION 3. FIELD DATA

3.1 DATE OF FIELD ACTIVITIES (1 July 2003)

3.2 AREAS TESTED/NUMBER OF HOURS

Areas tested and total number of hours operated at each site are summarized in Table 3.

TABLE 3. AREAS TESTED AND NUMBER OF HOURS

Area	Number of Hours
Calibration Lanes	0.00
Active Site	1.60

3.3 TEST CONDITIONS

3.3.1 Weather Conditions

An APG weather station located approximately one 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 4 represent the average temperature during field operations from 0700 to 1700 hours while precipitation data represents a daily total amount of rainfall. Hourly weather logs used to generate this summary are provided in Appendix B.

TABLE 4. TEMPERATURE/PRECIPITATION DATA SUMMARY

Date, 2003	Average Temperature, ^{o}F	Total Daily Precipitation, in.
1 July	79.8	0.00

3.3.2 Field Conditions

NRL surveyed the active site 1 July 2003. The field was dry and the weather was warm that day.

3.3.3 Soil Moisture

Three soil probes were placed at various locations within the site to capture soil moisture data: blind grid, calibration, mogul, 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 3-person crew took 2 hours and 30 minutes to perform the initial setup and mobilization. There was no daily equipment preparation and no end of the day equipment breakdown.

3.4.2 Calibration

NRL spent no time in the calibration lanes.

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 costs (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 routine data checks to ensure the data was being properly recorded/collected. NRL spent no additional time for breaks and lunches.

3.4.3.2 Equipment failure or repair. No time was needed to resolve equipment failures that occurred while surveying the Active Response area.

3.4.3.3 Weather. No weather delays occurred during the survey.

3.4.4 Data Collection

NRL spent a total time of 1 hour and 36 minutes in the Active Response area, all of which was spent collecting data.

3.4.5 Demobilization

The NRL survey crew only surveyed the active site. Demobilization occurred on 1 July 2003.

3.5 PROCESSING TIME

NRL submitted the raw data from the demonstration activities on the last day of the demonstration, as required. The scoring submittal data was provided at a later date.

3.6 DEMONSTRATOR'S FIELD SURVEYING METHOD

NRL surveyed the active site in a linear manner. NRL used line spacing to the width of the magnetometer array itself.

3.7 SUMMARY OF DAILY LOGS

Daily logs capture all field activities during this demonstration and are provided in Appendix D. Activities pertinent to this specific demonstration are indicated in highlighted text.

SECTION 4. TECHNICAL PERFORMANCE RESULTS

4.1 ROC CURVES USING ALL ORDNANCE CATEGORIES

The probability of detection for the response stage (P_d^{res}) and the discrimination stage (P_d^{disc}) versus their respective probability of false positive (P_{fp}) are shown in Figure 2. Both probabilities plotted against their respective BAR are shown in Figure 3, and both 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.

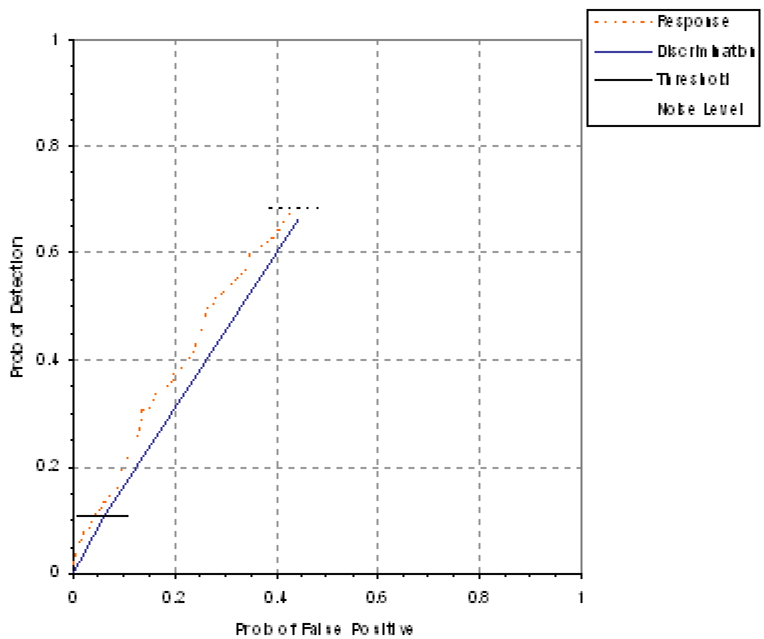

Figure 2. MAGNETOMETER MTADS/TOWED active response P_d^{res} and P_d^{disc} versus their respective P_{fp} over all ordnance categories combined.

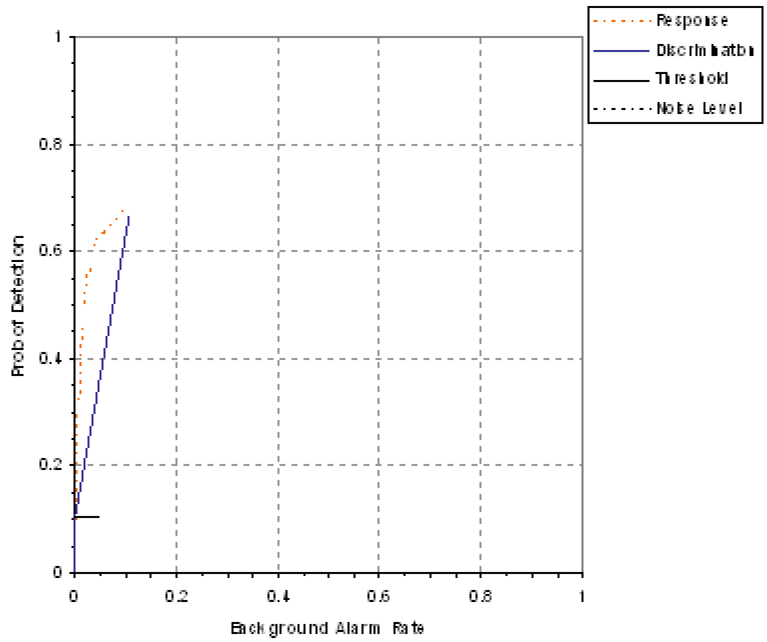

Figure 3. MAGNETOMETER MTADS/TOWED active response P_d^{res} and P_d^{disc} versus their respective BAR over all ordnance categories combined.

4.2 PERFORMANCE SUMMARIES

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 UXO field cleanup by minimizing false digs and maximizing ordnance recovery. The lower 90-percent confidence limit on P_d and P_{fp} was calculated assuming that the number of detections and false positives are binomially distributed random variables.

Results for the active response test are presented in Table 5 (cost results are provided in section 5).

TABLE 5. SUMMARY OF ACTIVE SITE RESULTS FOR MAGNETOMETER MTADS

Metric	**Overall**
RESPONSE STAGE	
P_d	0.68
P_d Low 90% Conf	0.65
P_d Upper 90% Conf	0.72
P_{fp}	0.44
P_{fp} Low 90% Conf	0.41
P_{fp} Upper 90% Conf	0.46
BAR	0.11
DISCRIMINATION STAGE	
P_d	0.11
P_d Low 90% Conf	0.09
P_d Upper 90% Conf	0.13
P_{fp}	0.06
P_{fp} Low 90% Conf	0.05
P_{fp} Upper 90% Conf	0.07
BAR	0.00

A comparison of the P_d, P_{fp}, and P_{ba}/BAR for both the response stage and discrimination stage for the blind grid, the open field, and the active site is presented in Table 6. P_d^{res} versus the respective P_{fp} over all ordnance categories is shown in Figure 6. P_d^{disc} versus their respective P_{fp} over all ordnance categories is shown in Figure 7 by using horizontal lines to illustrate the performance of the demonstrator at the recommended discrimination threshold levels, defining the subset of targets the demonstrator would recommend digging based on discrimination.

TABLE 6. COMPARISON OF BLIND GRID, OPEN FIELD, AND ACTIVE SITE RESULTS FOR MAGNETOMETER MTADS

Blind Grid		**Open Field**		**Active Site**	
Response Stage		***Response Stage***		***Response Stage***	
P_d	0.59	P_d	0.60	P_d	0.68
P_{fp}	0.84	P_{fp}	0.52	P_{fp}	0.44
P_{ba}	0.09	BAR	0.67	BAR	0.11
Discrimination Stage		***Discrimination Stage***		***Discrimination Stage***	
P_d	0.40	P_d	0.56	P_d	0.11
P_{fp}	0.52	P_{fp}	0.32	P_{fp}	0.06
P_{ba}	0.03	BAR	0.62	BAR	0.00

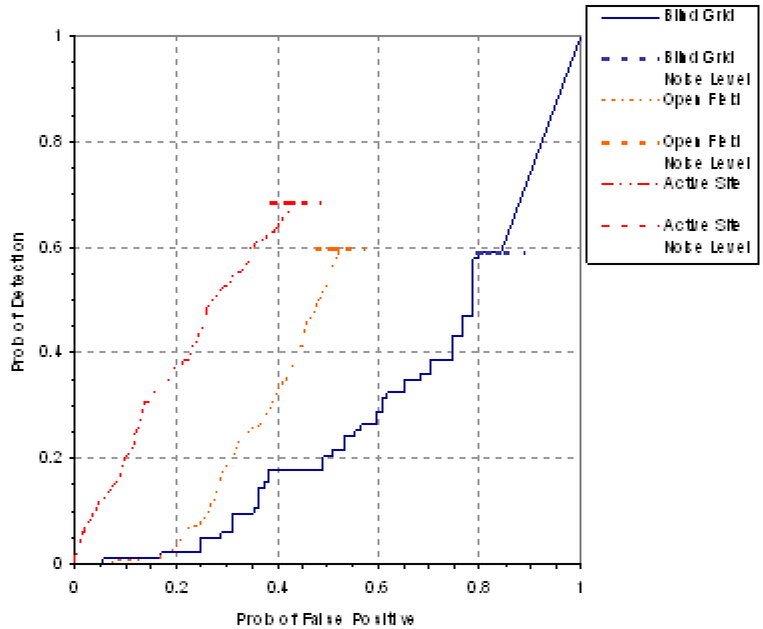

Figure 6. MAGNETOMETER MTADS/TOWED P_d^{res} stages versus the respective P_{fp} over all ordnance categories combined.

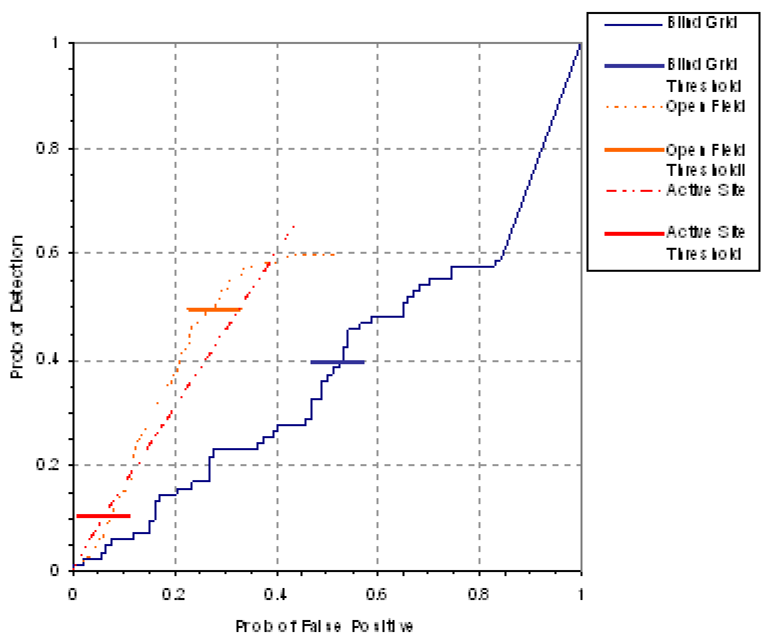

Figure 7. MAGNETOMETER MTADS/TOWED P_d^{disc} versus the respective P_{fp} over all ordnance categories combined.

4.3 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 Table 7.

TABLE 7. EFFICIENCY AND REJECTION RATES

	Efficiency (E)	**False Positive Rejection Rate**	**Background Alarm Rejection Rate**
At Operating Point	0.16	0.86	1.00
With No Loss of P_d	1.00	0.00	0.00

4.4 LOCATION ACCURACY

The mean location error and standard deviations are presented in Table 8. These calculations are based on average missed depth for ordnance correctly identified in the discrimination stage. Depths could not be accurately measured since the discovered ordnance and clutter were discovered and not emplaced. For the active response, no depth errors are calculated and (X, Y) positions are known from the recovery operation.

TABLE 8. MEAN LOCATION ERROR
AND STANDARD DEVIATION (m)

	Mean	**Standard Deviation**
Northing	0.01	0.07
Easting	0.00	0.10

4.5 STATISTICAL COMPARISONS

Statistical chi-square significance tests were used to compare results between the blind grid and active site and the open field and active site scenarios. The intent of the blind grid and active site comparison is to determine if the feature introduced in each scenario has a degrading effect on the performance of the sensor system. The intent of the open field and active site comparison is to determine if the feature introduced in each scenario has any effect, whether a degradation or an improvement, 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 chi-square test for comparison between ratios was used at a significance level of 0.05 to compare blind grid to open field with regard to P_d^{res}, P_d^{disc}, P_{fp}^{res}, and P_{fp}^{disc}, efficiency and rejection rate. These results are presented in Table 9 and Table 10 for the blind grid versus active site and the open field versus active site comparisons, respectively. A detailed explanation and example of the chi-square application is provided in Appendix A.

TABLE 9. CHI-SQUARE RESULTS - BLIND GRID VERSUS ACTIVE SITE

Metric	**Overall**
P_d^{res}	Not significant
P_d^{disc}	Significant
P_{fp}^{res}	Significant
P_{fp}^{disc}	Significant
Efficiency	Significant
Rejection rate	Not significant

TABLE 10. CHI-SQUARE RESULTS - OPEN FIELD VERSUS ACTIVE SITE

Metric	**Overall**
P_d^{res}	Significant
P_d^{disc}	Significant
P_{fp}^{res}	Significant
P_{fp}^{disc}	Significant
Efficiency	Significant
Rejection rate	Significant

SECTION 5. ON-SITE LABOR COSTS

A standardized estimate for labor costs associated with this effort was calculated as follows: the first person at the test site was designated supervisor, the second person was designated data analyst, and the third and following personnel were considered field support. Standardized hourly labor rates were charged by title: supervisor at $95.00/hour, data analyst at $57.00/hour, and field support at $28.50/hour.

Government representatives monitored on-site activity. All on-site activities were grouped into one of ten categories: initial setup/mobilization, daily setup/stop, calibration, collecting data, downtime due to break/lunch, downtime due to equipment failure, downtime due to equipment/data checks or maintenance, downtime due to weather, downtime due to demonstration site issue, or demobilization. The daily activity log is provided in Appendix D. A summary of field activities is provided in Section 3.4.

The standardized cost estimate associated with the labor needed to perform the field activities is presented in Table 11. Note that calibration time includes time spent in the calibration lanes as well as field calibrations. Site survey time includes daily setup/stop time, collecting data, breaks/lunch, downtime due to equipment/data checks or maintenance, downtime due to failure, and downtime due to weather.

TABLE 11. ON-SITE LABOR COSTS

	No. People	Hourly Wage	Hours	Cost
Initial Setup				
Supervisor	1	$95.00	2.5	$237.50
Data analyst	1	57.00	2.5	142.50
Field support	1	28.50	2.5	71.25
Subtotal				**$451.25**
Calibration				
Supervisor	0	$95.00	0.0	0.00
Data analyst	0	57.00	0.0	0.00
Field support	0	28.50	0.0	0.00
Subtotal				**0.00**
Site Survey				
Supervisor	1	$95.00	1.6	$152.00
Data analyst	1	57.00	1.6	91.20
Field support	1	28.50	1.6	45.60
Subtotal				**$288.80**

See notes at end of table.

TABLE 11 (CONT'D)

	No. People	Hourly Wage	Hours	Cost
Demobilization				
Supervisor	0	$95.00	0.0	0.00
Data analyst	0	57.00	0.0	0.00
Field support	0	28.50	0.0	0.00
Subtotal				**0.00**
Total				**$740.05**

Notes: Calibration time includes time spent in the calibration lanes as well as calibration before each data run.
Site survey time includes daily setup/stop time, collecting data, breaks/lunch, downtime due to system maintenance, failure, and weather.

SECTION 6. 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 ordnance item.

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

Munitions and Explosives Of Concern (MEC): Specific categories of military munitions that may pose unique explosive safety risks, including UXO as defined in 10 USC 101(e)(5), DMM 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 Ordnance: An ordnance item buried by the government at a specified location in the test site (for the Active site all 'emplaced' items are items discovered during recovery operations and are not strictly emplaced items).

Emplaced Clutter: A clutter item (i.e., non-ordnance item) buried by the government at a specified location in the test site (for the Active site all 'emplaced' items are items discovered during recovery operations and are not strictly emplaced items).

R_{halo}: A pre-determined radius about the periphery of an emplaced item (clutter or ordnance) within which a location identified by the demonstrator as being of interest is considered to be a response from that item. If multiple declarations lie within R_{halo} of any item (clutter or ordnance), the declaration with the highest signal output within the R_{halo} will be utilized. For the purpose of this program, a circular halo 0.5 meters in radius will be placed around the center of the object for all clutter and ordnance items.

Response Stage Noise Level: The level that represents the point 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 they believe provides optimum performance of the system by retaining all detectable ordnance 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. These two stages are termed the 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 false positive (P_{fp}) and those that do not correspond to any known item, termed background alarms.

The response stage scoring evaluates the ability of the system to detect emplaced targets without regard to ability to discriminate ordnance from other anomalies. For the response stage, the demonstrator provides the scoring committee with the location and signal strength of all anomalies that the demonstrator has deemed sufficient to warrant further investigation and/or processing as potential emplaced ordnance items. This list is generated with minimal processing (e.g., this list will include all signals above the system noise threshold). As such, it represents the most inclusive list of anomalies.

The discrimination stage evaluates the demonstrator's ability to correctly identify ordnance 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 ordnance. Thus, higher output values are indicative of higher confidence that an ordnance 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 ordnance 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} = (No. of response-stage detections)/(No. of emplaced ordnance in the test site).

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

Response Stage Probability of False Positive (P_{fp}^{res}): P_{fp}^{res} = (No. of response-stage false positives)/(No. of emplaced clutter items).

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

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

Response Stage Background Alarm Rate (BAR^{res}): Open field only: BAR^{res} = (No. of response-stage background alarms)/(arbitrary constant).

Note: The quantities P_d^{res}, P_{fp}^{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 $P_d^{res}(t^{res})$, $P_{fp}^{res}(t^{res})$, $P_{ba}^{res}(t^{res})$, and $BAR^{res}(t^{res})$.

DISCRIMINATION STAGE DEFINITIONS

Discrimination: The application of a signal processing algorithm or human judgment to response-stage data that discriminates ordnance from clutter. Discrimination should identify anomalies that the demonstrator has high confidence correspond to ordnance, as well as those that the demonstrator has high confidence correspond to nonordnance 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} = (No. of discrimination-stage detections)/(No. of emplaced ordnance 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} = (No. of discrimination stage false positives)/(No. of emplaced clutter items).

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

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

Discrimination Stage Background Alarm Rate (BAR^{disc}): BAR^{disc} = (No. of discrimination-stage background alarms)/(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 CHARACERISTIC (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_{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] Figure A-1 shows how P_d versus P_{fp} and P_d versus BAR are combined into ROC curves. Note that the “res” and “disc” superscripts have been suppressed from all the variables for clarity.

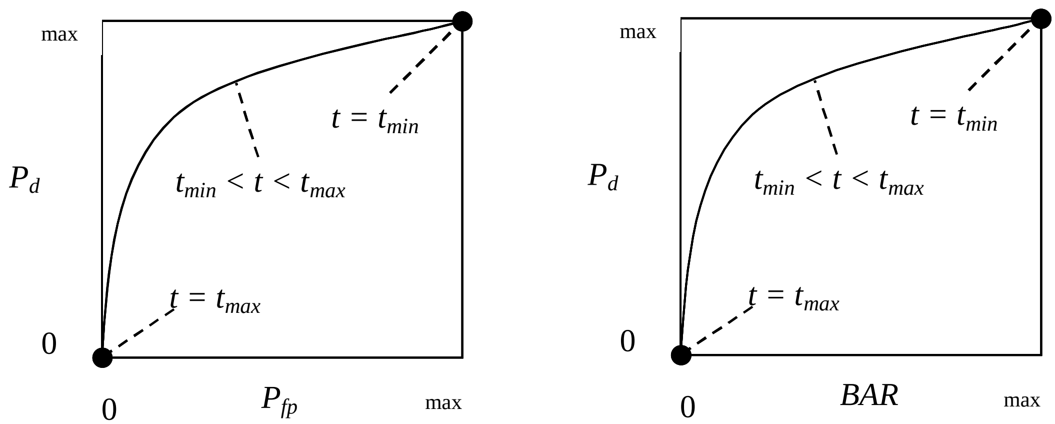

Figure A-1. ROC curves for open field testing. Each curve applies to both the response and discrimination stages.

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 ordnance detections from the anomaly list, while rejecting the maximum number of anomalies arising from nonordnance items. The efficiency measures the amount of detected ordnance 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 ordnance detectable by the sensor and its accompanying false positive rate or background alarm rate.

---

[1]Strictly speaking, ROC curves plot the P_d versus P_{ba} over a pre-determined and fixed number of detection opportunities (some of the opportunities are located over ordnance 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): $E = P_d^{disc}(t^{disc})/P_d^{res}(t_{min}^{res})$; 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 ordnance initially detected in the response stage was retained at the specified threshold in the discrimination stage, t^{disc}.

False Positive Rejection Rate (R_{fp}): $R_{fp} = 1 - [P_{fp}^{disc}(t^{disc})/P_{fp}^{res}(t_{min}^{res})]$; 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}):

Blind grid: $R_{ba} = 1 - [P_{ba}^{disc}(t^{disc})/P_{ba}^{res}(t_{min}^{res})]$.
Open field: $R_{ba} = 1 - [BAR^{disc}(t^{disc})/BAR^{res}(t_{min}^{res})])$.

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.

CHI-SQUARE COMPARISON EXPLANATION:

The chi-square 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 event A (some specified event) is the same for both populations (ref 3).

A 2 by 2 contingency table is used in the Standardized UXO Technology Demonstration Site Program to determine if there is reason to believe that the proportion of ordnance correctly detected/discriminated by demonstrator X's system is significantly degraded by the more challenging terrain feature introduced. The test statistic of the 2 by 2 contingency table is the chi-square distribution with one degree of freedom. Since an association between the more challenging terrain feature and relatively degraded performance is sought for the blind grid versus active site comparison, a one-sided test is performed. A significance level of 0.05 is chosen which sets a critical decision limit of 2.71 from the chi-square distribution with one degree of freedom. For the open field versus active site comparison, there is no assumption of a degraded performance for either site. Therefore, a two-sided test is performed to test for a significant difference in performance in either direction. Using the same significance level of 0.05, the critical decision limit is set to 3.84 from the chi-square distribution with one degree of freedom. For both tests, the value obtained from the chi-square distribution is a critical decision limit because if the test statistic calculated from the data exceeds this value, the two proportions tested will be considered significantly different. If the test statistic calculated from the data is less than this value, the two proportions tested will be considered not significantly different.

An exception must be applied when either a 0 or 100 percent success rate occurs in the sample data. The chi-square test cannot be used in these instances. Instead, Fischer's test is used and the critical decision limit for one-sided tests is the chosen significance level, which in this case is 0.05. With Fischer's test, if the test statistic is less than the critical value, the proportions are considered to be significantly different.

Standardized UXO Technology Demonstration Site examples, where blind grid results are compared to those from the open field and open field results are compared to those from one of the scenarios, follow. 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 that one data set has experienced a degradation 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 after surveying each of the three progressively more difficult areas using the same system (results indicate the number of ordnance detected divided by the number of ordnance emplaced):

	Blind grid	Open field	Moguls
P_d^{res}	100/100 = 1.0	8/10 = .80	20/33 = .61
P_d^{disc}	80/100 = 0.80	6/10 = .60	8/33 = .24

P_d^{res}: blind grid versus open field. Using the example data above to compare probabilities of detection in the response stage, all 100 ordnance out of 100 emplaced ordnance items were detected in the blind grid while 8 ordnance out of 10 emplaced were detected in the open field. Fischer's test must be used since a 100 percent success rate occurs in the data. Fischer's test uses the four input values to calculate a test statistic of 0.0075 that is compared against the critical value of 0.05. Since the test statistic is less than the critical value, the smaller response stage detection rate (0.80) is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and degradation in performance, it does indicate that the detection ability of demonstrator X's system seems to have been degraded in the open field relative to results from the blind grid using the same system.

P_d^{disc}: blind grid versus open field. Using the example data above to compare probabilities of detection in the discrimination stage, 80 out of 100 emplaced ordnance items were correctly discriminated as ordnance in blind grid testing while 6 ordnance out of 10 emplaced were correctly discriminated as such in open field-testing. Those four values are used to calculate a test statistic of 1.12. Since the test statistic is less than the critical value of 2.71, the two discrimination stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{res}: open field versus moguls. Using the example data above to compare probabilities of detection in the response stage, 8 out of 10 and 20 out of 33 are used to calculate a test statistic of 0.56. Since the test statistic is less than the critical value of 2.71, the two response stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{disc}: open field versus moguls. Using the example data above to compare probabilities of detection in the discrimination stage, 6 out of 10 and 8 out of 33 are used to calculate a test statistic of 2.98. Since the test statistic is greater than the critical value of 2.71, the smaller discrimination stage detection rate is considered to be significantly less at the 0.05 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and degradation in performance, it does indicate that the ability of demonstrator X to correctly discriminate seems to have been degraded by the mogul terrain relative to results from the flat open field using the same system.

APPENDIX B. DAILY WEATHER LOGS

Date, 2003	Time, EST	Average Temperature, °F	Total Precipitation, inches
1 Jul	0700	71.6	0.00
	0800	75.0	0.00
	0900	77.4	0.00
	1000	78.6	0.00
	1100	80.4	0.00
	1200	81.7	0.00
	1300	81.7	0.00
	1400	82.0	0.00
	1500	82.8	0.00
	1600	83.3	0.00
	1700	82.9	0.00

(Page B-2 Blank)

APPENDIX C. SOIL MOISTURE

Not available.

APPENDIX D. DAILY ACTIVITY LOGS

Date, 03	No. of People	Area Tested	Status Start Time	Status Stop Time	Duration min.	Operational Status	Operational Status - Comments	Track Method	Pattern	Field Conditions
1 Jul	3	ACTIVE SITE	0730	1000	150	INITIAL SETUP	MOBILIZATION	GPS	LINEAR	SUNNY DRY
1 Jul	3	ACTIVE SITE	1000	1136	96	COLLECTING DATA	COLLECT DATA	GPS	LINEAR	SUNNY DRY

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

ADST	=	Aberdeen Data Services Team
APG	=	Aberdeen Proving Ground
ASCII	=	American Standard Code for Information Interchange
ATC	=	U.S. Army Aberdeen Test Center
ATSS	=	Aberdeen Test and Support Services
BAR	=	Background Alarm Rate
DAS	=	Data Analysis System
DMM	=	discarded military munitions
EQT	=	Environmental Quality Technology
ERDC	=	U.S. Army Corps of Engineers Engineering Research and Development Center
ESTCP	=	Environmental Security Technology Certification Program
GPS	=	Global Positioning System
GT	=	ground truth
HDSD	=	Homeland Defense and Sustainment Division
MAG	=	magnetometer
MEC	=	munitions and explosives of concern
MTADS	=	Multiple Towed Array Detection System
NRL	=	Naval Research Laboratory
PDOP	=	position dilution of precision
POC	=	point of contact
PPS	=	Precise Positioning System
QA	=	quality assurance
QC	=	quality control
ROC	=	receiver-operating characteristic
RTK	=	real-time kinematic
SERDP	=	Strategic Environmental Research and Development Program
USAEC	=	U.S. Army Environmental Command
UTC	=	Coordinated Universal Time
UXO	=	unexploded ordnance

Secondary distribution is controlled by Commander, U.S. Army Environmental Command, ATTN: IMAE-RTA.