AD NO. ____________________________
DTC PROJECT NO. 8-CO-160-UXO-016
REPORT NO. ATC-9266

SHALLOW WATER

UXO TECHNOLOGY DEMONSTRATION SITE

SCORING RECORD NO. 1

SITE LOCATION:
U.S. ARMY ABERDEEN PROVING GROUND

DEMONSTRATOR:
GEOPHEX, LTD.
605 MERCURY STREET
RALEIGH, NC 27603-2343

TECHNOLOGY TYPE/PLATFORM:
GEM-3 ARRAY

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

APRIL 2008

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

DISTRIBUTION UNLIMITED, SEPTEMBER 2008.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
February 2008	Final	27 September through 5 October 2005

4. TITLE AND SUBTITLE
STANDARDIZED UXO TECHNOLOGY DEMONSTRATION SITE SHALLOW WATER SCORING RECORD NO. 1

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Rowe, Gary W.

5d. PROJECT NUMBER
8-CO-160-UXO-016

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Aberdeen Test Center
ATTN: CSTE-DTC-AT-SL-E
Aberdeen Proving Ground, MD 21005-5059

8. PERFORMING ORGANIZATION REPORT NUMBER
ATC-9266

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Evaluation Center
ATTN: SFIM-AEC-ATT
Aberdeen Proving Ground, MD 21005-5401

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
This document corrects discrepancies between the data published in Scoring Record No.1 dated September 2006 and the actual test results. Affected sections are the open water detection and discrimination tables and associated ROC curves.

14. ABSTRACT
This report documents the efforts of Geophex, Ltd. to detect and discriminate inert unexploded ordnance (UXO) using a GEM-3 electric magnetic sensor array modified for use in a shallow water environment. Testing was conducted at ATC, Standardized Shallow Water UXO Technology Demonstration Site. A description of the tested system and an estimate of survey costs along with the analysis of the system performance are provided.

15. SUBJECT TERMS
Geophex, Ltd., UXO Standardized Technology Demonstration Site Program, Shallow Water, GEM-3 Array, MEC

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

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

`

TABLE OF CONTENTS

PAGE

SECTION 1. GENERAL INFORMATION

SECTION 2. SYSTEM UNDER TEST

SECTION 3. SURVEY COST ANALYSIS

SECTION 4. TECHNICAL PERFORMANCE RESULTS

PAGE

SECTION 5. APPENDIXES

<u>SECTION 1. GENERAL INFORMATION</u>

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 so their performance can be characterized. Technologies under development for the detection and discrimination of UXO require independent testing so their performance can be characterized. To that end, the U.S. Army Aberdeen Test Center (ATC) located at Aberdeen Proving Ground (APG), Maryland, has developed a Standardized Shallow Water Test Site. This site provides a controlled environment containing varying water depths, multiple types of ordnance and clutter items, as well as navigational and detection challenges. Testing at this site is independently administered and analyzed by the government for the purposes of characterizing technologies, tracking performance during system development, and comparing the performance and costs of different systems.

The Standardized UXO Technology Demonstration Site Program is a multi-agency program spearheaded by the U.S. Army Environmental Center (USAEC). ATC and the U.S. Army Corps of Engineers Engineering, Research and Development Center (ERDC) provide programmatic support. The Environmental Security Technology Certification Program (ESTCP), the Strategic Environmental Research and Development Program (SERDP), and the Army Environmental Quality Technology Program (EQT) provided funding and support for this program.

1.2 OBJECTIVE

The objective of the Shallow Water Standardized UXO Technology Demonstration Site is to evaluate the detection and discrimination capabilities of existing and emerging technologies and systems in a shallow water environment. Specifically:

a. To determine the demonstrator’s ability to survey a shallow water area, analyze the survey data, and provide a prioritized “Target List” with associated confidence levels in a timely manner.

b. To determine both the detection and discrimination effectiveness under realistic scenarios that varies ordnance, clutter, and bathymetric conditions.

c. To determine cost, time, and manpower requirements needed to operate the technology.

1.3 CRITERIA

The scoring criteria specified in the Environmental Quality Technology - Operational Requirements Document (EQT-ORD) (app D, ref 1) for: A(1.6.a): UXO Screening, Detection and Discrimination document is presented in Table 1-1. Very little information was available on the capabilities of shallow water detection systems when these criteria were developed. However, they were used in the design of the test site, and the five metrics were used to measure system performance in this report.

TABLE 1-1. SCORING CRITERIA

Metric	Threshold	Objective
Detection	80% ordnance items buried to 1 foot and under 8 feet (2.4 m) of water at a standardized site detected	95% ordnance items buried to 4 feet and under 8 feet (2.4 m) of water at a standardized site detected
Discrimination	Rejection rate of 50% of emplaced non-UXO clutter at a standardized site with a maximum false negative rate of 10%	Rejection rate of 90% of emplaced non-UXO clutter at a standardized site with a maximum false negative rate of 0.5%
Reacquisition	Reacquire within 1 meter	Reacquire within 0.5 meters
Cost Rate	$4,000 per acre	$2,000 per acre
Production Rate	5 acres per day	50 acres per day

The ATC shallow water site was designed to evaluate the threshold-detection level of a range of ordnance at the 1-foot + 8-foot requirement. Limited information is available at the objective-detection level. All other measured results will be evaluated against both criteria levels.

1.4 APG SHALLOW WATER SITE INFORMATION

1.4.1 Location

The Aberdeen Area of APG is located in the northeast portion of Maryland on the western shore of the Chesapeake Bay in Harford County. The Shallow Water Test Site is located within a controlled range area of APG.

1.4.2 Soil Type

The area chosen for the shallow water test site was known as Cell No. 3 in a dredge-spoil field. The cell bottom is primarily composed of sediment removed from the Bush River. This is a freshwater site.

1.4.3 Test Areas

a. The test site contains five areas: calibration grid, blind test grid, littoral, open water, and deeper water area. Additional detail on each area is presented in Table 1-2. A schematic of the calibration lanes is shown in Figure 1.

TABLE 1-2. TEST AREAS

Area	Description
Calibration Grid	The calibration area contains 15 projectiles, 3 each 40, 60, 81, 105 and 155 mm. One of each projectile type is buried at the projectile diameter to depth ratio shown in Figure 1. This area is designed to provide the user with a sensor library of detection responses for the emplaced targets an understanding of their resistivity prior to entering the blind test fields. Two "clutter-cloud" target scenarios have been constructed adjacent to this area (fig. 1).
Blind Grid	The blind grid contains 644 detection opportunities. Each grid cell is 2 by 2 m^2. At the center of each cell is either an ordnance item, clutter, or nothing. Surrounding the blind grid on three sides are 3.6-kg (8-lb) shot puts, buried 0.3 meters deep in the sediment. The shot puts can be used as a navigational/Global Positioning System (GPS) check. The GPS coordinates for the center of each grid and the shot put locations are provided to the vendor prior to testing.
Littoral	This is a sloping area on one side of the pond with vegetation growing into the water line. Water depth ranges from 0.3 to 1.8 meters. It contains a variety of navigational and detection challenges.
Open Water	The open water scenario contains a variety of navigational, detection, and discrimination challenges. Water depth varies from 1.8 to 3.4 meters.
Deeper Water	The water depth in this area varies between 3.4 and 4.3 meters.

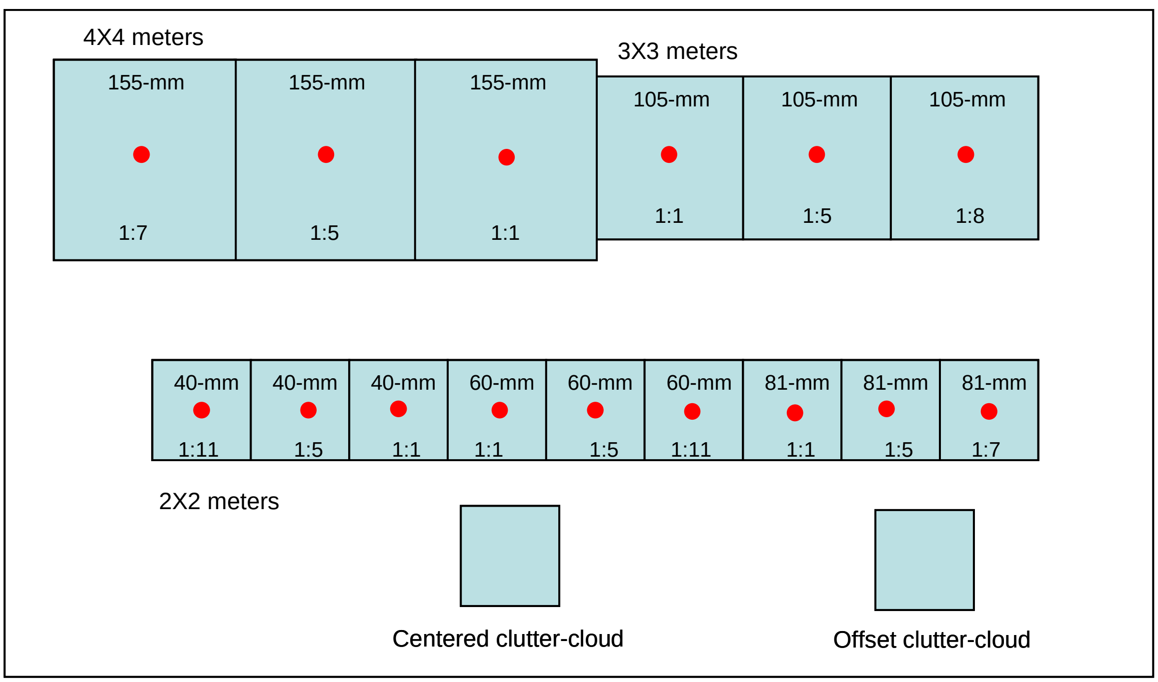

Figure 1. Schematic of the calibration grid.

b. The water depth at this facility during testing is maintained such that the calibration and blind grid areas meet the 2.4-meter (8-ft) detection criterion specified in paragraph 1.3. The test site is approximately 2.8 hectares (6.9 acres) in size.

1.5 GROUND-TRUTH TARGETS

The ground-truth is comprised of both inert ordnance and clutter items. The inert ordnance items are listed in Table 1-3. All items were located in storage sites at APG. The items have not been fired or degaussed.

Clutter items fit into one of three categories: ferrous, nonferrous, and mixed metals. The ferrous and nonferrous items have been further divided into three weight zones as shown in Table 1-4 and distributed throughout all test areas. Most of this clutter is comprised of ordnance components; however, there are also industrial scrap metal and cultural items as well. The mixed-metals clutter is comprised of scrap ordnance items or fragments that have both a ferrous and nonferrous component and could reasonably be encountered in a range area. The mixed-metals clutter was placed in the open water area only.

TABLE 1-3. INERT ORDNANCE TARGETS

Description	Length, mm	Diameter, mm	Aspect Ratio, W/L	Weight, gm
40-mm L70 Projectile	208	40	0.1923	965
60-mm Mortar M49A2	185	60	0.3243	975
81-mm Mortar M374	528	81	0.1534	3969
81-mm Mortar M821	510	81	0.1588	3338
105-mm Projectile M1	445	105	0.2360	13834
155-mm M107 Projectile	684	155	0.2266	41731
8-inch M104/106	856	203	0.2371	89811

TABLE 1-4. CLUTTER WEIGHT RANGES

Clutter Type	Weight Range in Grams		
	Small	Medium	Large
Ferrous	10 to 510	511 to 2200	< 2201
Nonferrous	10 to 270	275 to 800	> 801

SECTION 2. SYSTEM UNDER TEST

2.1 DEMONSTRATOR INFORMATION

Geophex, as part of their Broad Agency Announcement (BAA) submittal (app D, ref 2), provided the information in paragraphs 2.2 through 2.7. ATC's comments on the demonstrated system are provided in paragraph 2.8.

2.2 SYSTEM DESCRIPTION

The GEM-3 array sensor proposed is a system Geophex designed and built in 2003 for a survey of about 52.6 hectares (130 acres) of an underwater area adjacent to Mare Island located within the San Francisco Bay. This site is the first large-scale underwater UXO survey in history. We fabricated an array of three GEM-3s, which were mounted on a long fiberglass rod and pulled behind a powerboat as depicted in Figure 2. The sensor passed rigorous validation test processes over an underwater seeded area at Mare Island in terms of detection rate and a location accuracy of a few decimeters.

Figure 2. Towed sled and the powerboat used to pull it (the supplied Mare Island photographs have been replaced with ATC site-specific ones).

The system is a continuous-wave frequency-domain electromagnetic induction (EMI) sensor that uses a hybrid current waveform to provide simultaneous multi-frequency (typically 10 log spaced) energy in the 90 Hz to 90 kHz band, with each Rx dsp performing digital Fourier transforms at the selected frequencies. A reference coil provides primary field reference (amplitude and phase) for Rx output normalization (units of parts per million (ppm) of the primary field generated electromagnetic force (EMF)).

The GEM-3 sensors employ a primary-field nulling scheme using a secondary concentric transmitter coil in series with the primary transmitter coil but current flowing in the opposite sense. Each sensor fires in sequence so that they do not interfere with each other. The coils can be seen in the photo of the sensor with waterproof housing opened (fig. 3), which are embedded in a fiberglass-laminated foam board.

Figure 3. Integrated array being installed in the waterproof housing.

2.3 DEMONSTRATOR'S SITE SURVEYING METHOD

Two platforms will be employed to operate in different depths. The main platform is similar to the one used in Mare Island, in which the sensor is mounted on a sled with cement filled PVC runners for bottom drag operation, towed using the fiberglass shaft attached to the solid-hull powerboat on a swiveled pivot. This configuration will be used wherever the water is greater than a few feet deep and the boat sensor has room to maneuver. If the water becomes too shallow and the shoreline precludes required turning, a platform mounted on a frame that is rigidly fixed to a small inflatable pontoon boat will be used. The frame fits across the bow of the boat so that the sensor is underneath the front section of the boat at an adjustable depth. The sensor for this configuration is embedded in a fiberglass shell; the sensors are identical to the other configuration. A third configuration may be used along the shoreline in very shallow water, in which the (fiberglass housed) sensor package is floated behind the pontoon boat and pulled with the fiberglass pole.

2.4 DEMONSTRATOR'S QC AND QA

QC **-** daily sensor calibration check with ferrite target; calibration area test.
QA - daily review of data.

2.5 DATA PROCESSING DESCRIPTION

Target detection is achieved by combining multi-frequency data into a single detection channel designed to respond particularly to metal targets and not to geologic anomalies. Several were used including the sum of all quadrature channels, the difference between high frequency and low frequency in-phase channels, the sum of the absolute differences of quadrature channels between all frequency pairs, and the inverse log (frequency) weighted total apparent conductivity. The selected detection channel forms the response stage. The Differential Global Positioning System (DGPS) georeferenced detection channel data are processed with an automatic anomaly picker that identifies target anomalies above a specified threshold, excluding single-point anomalies and overlapping secondary anomalies.

Georeferencing uses the time stamps to interpolate 15 or 30 Hz GEM-5 position between 1-Hz DGPS fixes, and the position for each sensor from spatial interpolation between two DGPS antennas. The raw DGPS latitude/longitude fixes are transformed to Universal Transverse Mercator (UTM) during the post-processing.

Target identification and classification (clutter discrimination) uses a normalized matching of the multi-frequency spectra to a library of known UXO spectral responses. The matching scheme fits an unknown target to the best-fit linear combination of the longitudinal (sensor axis along target long axis) and transverse (sensor axis perpendicular to target long axis) response spectra, allowing for a frequency independent background in-phase response for magnetic soils. The goodness-of-fit to the best fitting item is mapped into a confidence ranking from 0 (definite clutter) to 10 (definite UXO) with 5 corresponding to the clutter misfit threshold. The confidence ranking forms the discrimination stage.

2.6 ATC SURVEY COMMENTS

Only the first two of the proposed three platforms were used during the survey. The platform demonstrated at Mare Island was used in the blind grid, open and deep water, and part of the littoral areas. The second platform was designed and built on-site. This modified platform used the same sensors and instrumentation used in the deeper portions of the test site. Both of the demonstrated platforms experienced technical and/or structural problems.

The wooden mounting platform that connects the towed GEM-3 sensor array to the towing vessel is shown in Figure 4.

Figure 4. Wooden framework connecting the towed sensors to the boat.

The GEM-3 sensor mounted in a wooden frame, installed on the inflated pontoon boat and surveying part of the littoral area is shown in Figures 5 through 7.

Figure 5. GEM-3 sensor array reconfigured for use in the littoral survey.

Figure 6. GEM-3 platform used to survey the littoral region.

Figure 7. Littoral survey.

The Mare Island system experienced two problems: the wooden swiveled-pivot-mount used to attach the towing shaft to the boat had to be reinforced to handle the additional stresses encountered making tight turns at the site. Two DGPS antennas were attached to the towing shaft. The antenna positioned closest to the water experienced frequent signal dropout, invalidating the entire first day of testing. The problem was the boat was blocking the positional signal; the solution was to move that antenna higher on the shaft. The modified configuration is shown in Figure 4.

The second platform had, according to Geophex, been constructed and tested off-site before this survey. Pre-survey quality control checks showed a large amount of EMI caused by the outboard motor. On-site experimentation resulted in placing the GEM-3 array approximately 1 meter below the water's surface to reduce this interference to an acceptable level. While this increased the size of the littoral area that could be surveyed, as compared to their towed platform, it was limited to water depths greater than 1 meter.

Geophex was unable to survey the littoral area completely using the two system platforms. The third platform, a positive buoyancy, pole-guided package was not demonstrated.

SECTION 3. SURVEY COST ANALYSIS

3.1 DATES OF SURVEY

The Geophex, Ltd., GEM-3 array sensor surveyed the site from 27 September through 5 October 2005.

3.2 SITE CONDITIONS

3.2.1 Atmospheric Conditions

An ATC weather station located adjacent to the test site recorded the average temperature and precipitation on an hourly basis for each day of operation. The temperatures listed in Table 3-1 represent the average temperature from 0700 through 1700. The hourly weather logs used to generate this summary are provided in Appendix B.

3.2.2 Water Conditions

Water conditions were monitored using a TIDALITE IV Portable Tide Gauge System©. Data recorded include: water depth and temperature, significant wave height based on the average 1/3 wave height seen over the test period using the Draper/Tucker analysis method, and the full-wave frequency calculated by full-wave mean crossing detection. The values displayed in Table 3-1 are averaged from 0700 through 1700. Detailed information is provided in Appendix B.

TABLE 3-1. SITE CONDITION SUMMARY

Date, 05	Average Air Temperature, °C	Wind, km/h	Average Water Temperature, °C	Water Depth, m	Significant. Wave Height, m	Wave Frequency, Hz
Sep 27	21.9	10.6	20.1	8.44	0.04	0.22
Sep 28	21.9	5.2	19.8	8.22	0.04	0.26
Sep 30	15.8	5.3	18.7	8.11	0.04	0.24
Oct 1	18.9	5.5	17.9	8.12	0.04	0.41
Oct 5	21.9	5.2	18.8	8.19	0.04	0.25

3.3 SURVEY ACTIVITIES

The information contained in this section provides an estimate of the time needed and costs associated with surveying an area with this demonstrator's system. This includes data on equipment setup and calibration, site survey and any resurvey time, and downtime due to system malfunctions and maintenance requirements.

3.3.1 Survey Times

a. A government representative monitored and recorded all on-site activities. These activities are grouped into one of eleven categories. The first eight categories are chargeable to the system while the last three are not. Categorizing these activities provides insight into the technical and logistical aspects of the system. The times recorded in each category are then matched with the number of demonstrator personnel, assigned skill levels and a consistent (across vendor) salary to produce an estimate of the survey costs.

(1) Initial setup/mobilization. Starts at the time when the demonstrator's equipment arrives at the survey site and stops when the system is ready to acquire data.

(2) Daily setup/close-up. Monitors time spent mounting and dismounting the equipment each day.

(3) Instrument calibration. Records the amount of time used for daily quality assurance checks, i.e., sensors, GPS data, survey data quality, etc.

(4) Collecting data. Time spent surveying the test area.

(5) Downtime (non-survey time) due to equipment/data checks. Covers time spent trouble shooting equipment or verifying survey tracks.

(6) Downtime (non-survey time) due to equipment failure. Examples are replacing damaged cables, lost communication with base station, or any other failure that prevents surveying. Some weather related failures would fall into this category for example light-emitting diode (LED) displays darken by the sun, wind creating waves too high to survey in, etc.

(7) Downtime (non-survey time) due to maintenance. Battery replacement and memory downloads are typical examples.

(8) Demobilization. Commencement action once the demonstrator has completed the survey and concluded the final on-site check of the test data and ends when the equipment and personnel are ready to leave the site.

(9) Non-chargeable downtime for breaks and lunch. The demonstrator's company policy sets this standard.

(10) Non-chargeable downtime for weather related causes (i.e., lighting, high wet-bulb heat index, and similar events).

(11) Non-chargeable downtime due to ATC range operating requirements. Danger zone conflicts, lack of support personnel, equipment or other ATC caused delays.

b. Appendix C contains the daily log sheets. Table 3-2 summarizes that information to provide insight into the operational, maintenance, and logistical aspects of the system. Task times have been rounded to 5-minute increments.

TABLE 3-2. TIME ON SITE

Date, 05	**Sep 27**	**Sep 28**	**Sep 29**	**Sep 30**	**Oct 1**	**Oct 3**	**Oct 4**	**Oct 5**	**Activity Totals, hrs**
Activity (daily times recorded in minutes)									
Initial setup	525	-	330	-	-	-	265	-	18.7
Daily setup/close-up	155	125	30	120	110	120	60	35	12.6
Instrumentation calibration	-	30	100	60	-	-	-	-	3.2
Data collection	-	380	-	410	403	325	155	180	30.9
Equipment/data checks	-	60	-	10	-	-	-	-	1.2
Equipment failure	-	50	-	-	-	-	-	70	2.0
Maintenance	-	20	-	-	7	-	-	-	0.5
Demobilization	-	-	-	-	-	-	-	220	3.7
Breaks and lunch	-	-	-	-	-	-	-	-	0.0
Weather related	-	-	-	-	-	-	-	-	0.0
ATC downtime	-	-	-	-	-	-	-	-	0.0
Daily Total, hr	11.3	10.6	7.7	10.0	8.7	7.4	7.4	8.0	
Note: The daily times are rounded to 5-minute intervals.									

3.3.2 <u>On-Site Data Collection Costs</u>

The times associated with the 11 activities have been reduced into the three basic components of the evaluation: initial setup, site survey, and pack-up (demobilization). Note that 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. This combines the actual survey cost with the demonstrator's associated on-site overhead costs

A standardized estimate for labor costs associated with this effort was then calculated using the following job categories: "supervisor" ($95.00/hr), "data analyst" ($57.00/hr), and "site support" ($28.50/hr). The estimated costs are shown in Table 3-3.

TABLE 3-3. CALCULATED SURVEY COSTS

	No. People	Hourly Wage	Hours	Cost
Initial Setup				
Supervisor	1	$95.00	18.7	$1,776.50
Data Analyst	1	$57.00	18.7	$1,065.90
Site Support	1	$28.50	18.7	$532.95
Subtotal:				$3,375.35
Site Survey				
Supervisor	1	$95.00	40.4	$3,838.00
Data Analyst	1	$57.00	40.4	$2,302.80
Site Support	1	$28.50	40.4	$1,151.40
Subtotal:				$7,292.20
Demobilization				
Supervisor	1	$95.00	3.7	$3351.50
Data Analyst	1	$57.00	3.7	$210.90
Site Support	1	$28.50	3.7	$105.45
Subtotal:				$667.85
Total On-site Costs:				$11,335.40

3.4 COST ANALYSIS

The data collection process described above provides an on-site cost guide to compare the performance of this vendor with any other that has demonstrated at the shallow water site. It is not a true indicator of survey costs. Many other expenses have not been included: travel costs, per Diem, off-site data processing and analysis, company overhead, profit, etc.

Calculating the area surveyed is done by plotting the raw GPS coordinates then combining the sensor swath (line spacing and associated overlap).

To determine the number of acres surveyed per day, the total number of hours spent at the test site (table 3-2) is divided by 8 (converts to 8-hour days). The number of acres is then divided by the number of 8-hour days. The cost per acre is determined by dividing the total survey costs (table 3-3) by the same number of acres. This information is summarized in Table 3-4.

TABLE 3-4. GEOPHEX SURVEY COSTS

Area Surveyed (Acres[a])	5.35
Time on-site (8-hour days)	8.9
Calculated survey cost (U.S. dollars)	$11,335
Acres per day	0.29
Cost per acre	$5248

[a]One acre equals 4047 m^2.

Geophex's survey costs are compared with the EQT-ORD criteria in Table 3-5.

TABLE 3-5. TEST RESULTS - CRITERIA COMPARISION

Metric	Threshold	Objective	Geophex
Cost Rate	$4,000 per acre	$2,000 per acre	$5,248 per acre
Production Rate	5 acres per day	50 acres per day	0.29 acres per day

SECTION 4. TECHNICAL PERFORMANCE RESULTS

4.1 AREA SURVEYED

4.1.1 Calculated Area

a. Both the test and scoring methodologies require the demonstrator to survey 100 percent of each of the four test areas (blind grid, open water, littoral, and deeper water). Scoring a partially surveyed area alters the ordnance and clutter sample sizes and test area boundaries and decreases the statistical confidence in the performance statements made for that area. Allowing partial scoring decreases the validity of performance comparisons made between multiple test areas for a single demonstrator and comparisons made between multiple demonstrators for a single test area.

b. Realizing that some systems may not be able to survey 100 percent of a given test area, a ranking system was established. The percent coverage for a given test area is determined by plotting the raw GPS coordinates combined with the sensor swath (line spacing and associated overlap), calculating the area surveyed, and then comparing that surveyed area to the total test area.

$$\frac{\text{Section Surveyed}}{\text{Test Area Size}} \times 100 = \%\ \text{surveyed}$$

c. The demonstrator's system is always scored against the complete ground truth for a given test area regardless of the percentage covered.

4.1.2 Area Assessment

The ranking system and survey results are presented in Table 4-1.

TABLE 4-1. SURVEY RANKING SYSTEM AND RESULTS

Ranking System		**Survey Results**		
% Area Covered	**Ranking**	**Test Area**	**% Area Covered**	**Data Usage**
95 to 100	Met	Blind Grid	100	Direct comparison between systems and areas.
		Open Water	99	
		Deeper Water	98	
90 to 94	Generally met			Comparison between systems and areas. A small negative bias is contained in the reported numbers (bias not quantified in this report).
50 to 89	Partially met	Littoral	82	Reported, not compared between systems or areas. A large negative bias is contained in the reported numbers (bias not quantified in this report).
0 to 49	Not met			Not scored/not reported.

The lower percentage of area covered in the littoral area is a function of the sensor-mounting configuration (fig. 5 and 6). Locations that had a water depth less than approximately 1 meter could not be surveyed with the systems demonstrated.

4.2 SYSTEM SCORING PROCEDURES

a. The scoring entities used in this program are predicated on knowing the composition and location of every detectable item in an area. The deeper water area is the one exception. Ground-truth targets were placed in this area without a pre-survey and clearing operation. Therefore, only the system's probability of detection (P_d) is evaluated in this area.

b. The best indicator of survey performance is the blind grid. This area provides a statistically valid, controlled environment in which the demonstrator must provide a response (ordnance, clutter, or blank) at each of the 644 locations. Comparison of the response and discrimination lists to the ground truth in this area both determines the range of ordnance the system can reliably detect and establishes the baseline to which system performance in all other test areas is measured.

c. The scoring terms and definitions along with an explanation of the ROC curve development and the Chi-square analysis used in this report are in Appendix C.

d. Demonstrator performance is scored in two stages: response and discrimination.

e. The response stage scoring evaluates the ability of the demonstrator's system to detect emplaced ground-truth targets without regard to discriminating ordnance from clutter. In this stage, the GPS locations and signal strengths of all anomalies that the demonstrator has deemed sufficient for further investigation and/or processing are reported. This list is generated with minimal processing, i.e., associating signal strength with GPS location, and includes only signals that are above the system noise level.

f. The discrimination stage evaluates the demonstrator's ability to segregate ordnance from clutter. The same GPS locations reported in the response stage anomaly list are evaluated based on the demonstrator's discrimination process (para 2.1.5). A discrimination stage list is generated and prioritized based on the demonstrator's determination that an anomaly is more likely to be ordnance rather than clutter. Typically, higher output values indicate a higher confidence that an ordnance item is present at a specified location. The demonstrator then specifies the threshold value for the prioritized ranking that provides optimum system performance. This value is the discrimination stage threshold.

g. Both the response and discrimination lists contain an identical number of potential target locations. They differ only in the priority ranking of the declarations.

h. Within both of these stages, the following entities are measured:

(1) Probability of detection (P_d).

(2) Probability of false positive (P_{fp}).

(3) Probability of background alarm (P_{ba})/background alarm rate (BAR).

4.2.1 ROC curves

Based on the entire range of ground-truth targets used at this site, ROC curves were generated for both the response and discrimination stages. In both stages, the probability of detection verses false alarm rates is plotted. False alarms are divided into two groups: those anomalies that correspond to emplaced clutter items, thereby measuring the P_{fp}, and anomalies that do not correspond to any known item, termed background alarms (P_{ba}) in the blind grid area and BAR in all other areas.

The ROC curves for the response and discrimination stages for all areas surveyed are shown in Figures 8 through 13. Horizontal lines illustrate the system performance at the demonstrator's recommended noise level during the response stage, or discrimination threshold level in the discrimination stage. The point where the curve crosses the horizontal line defines the subset of targets the demonstrator recommends digging.

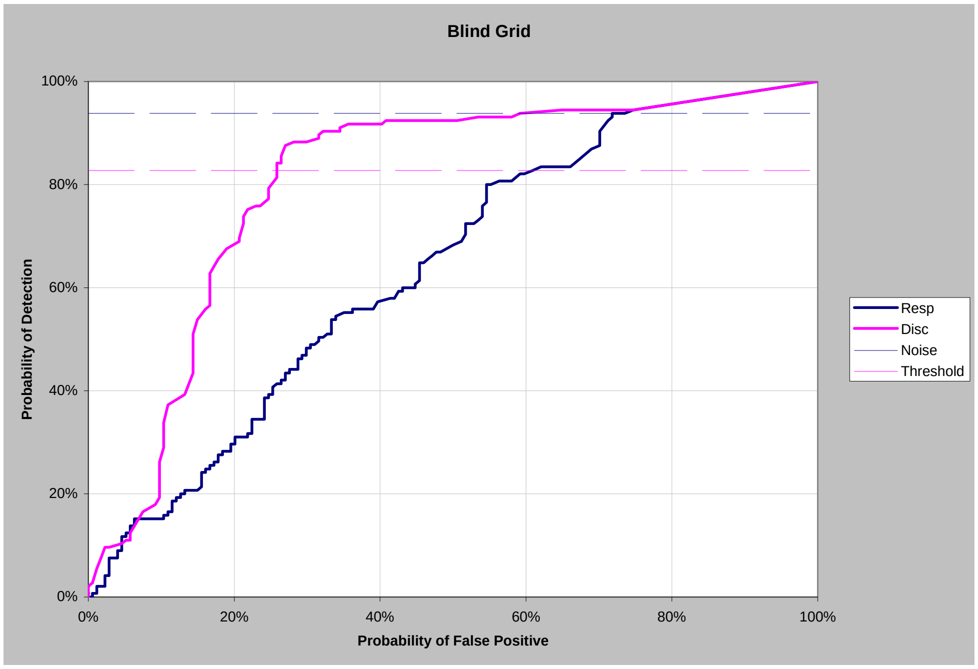

Figure 8. Blind grid P_d versus P_{fp}.

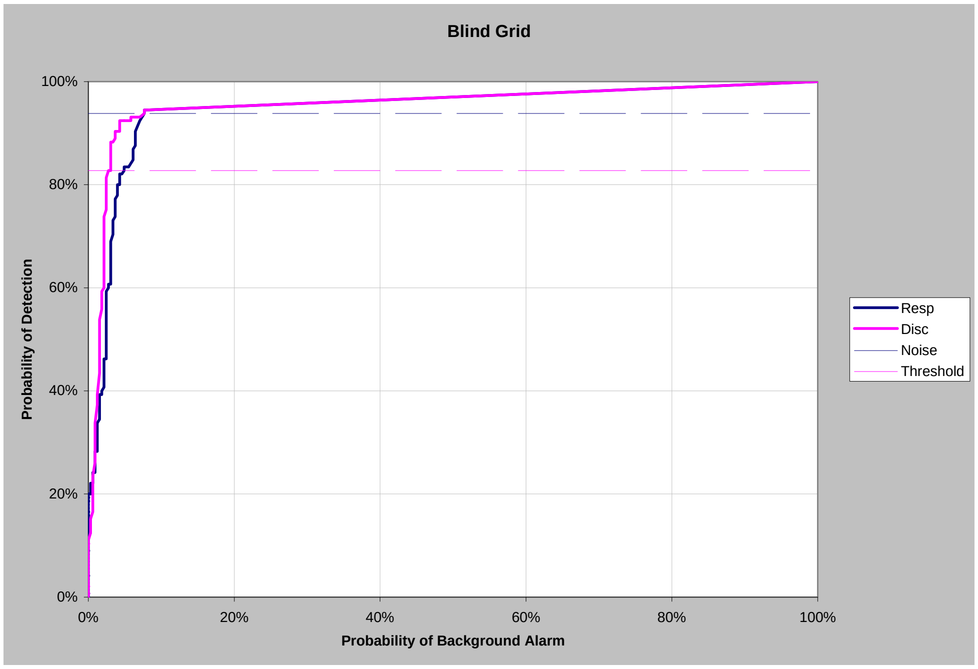

Figure 9. Blind grid P_d versus P_{ba}.

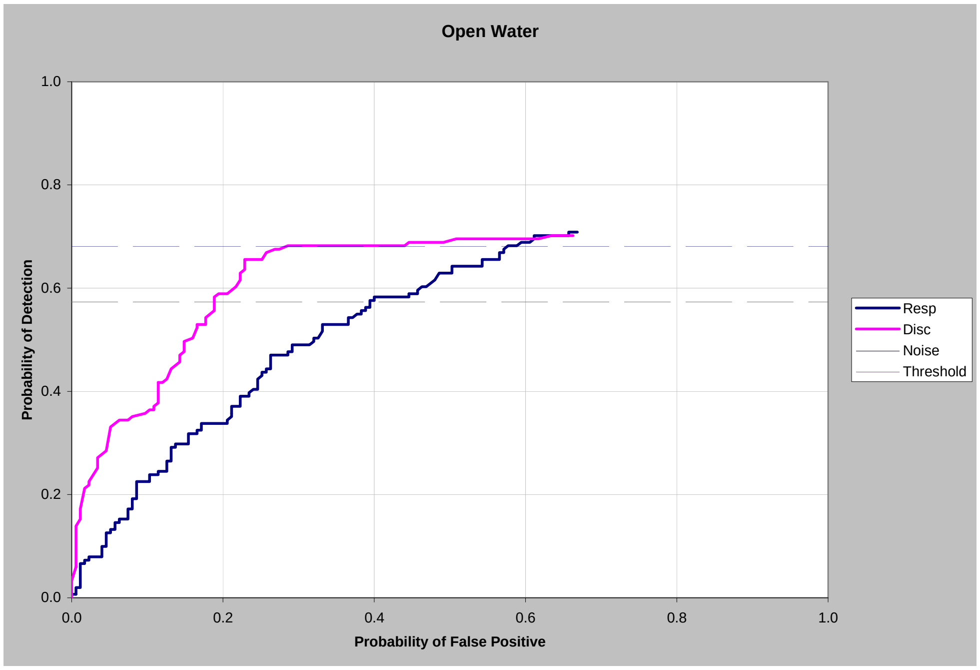

Figure 10. Open water P_d versus P_{fp}.

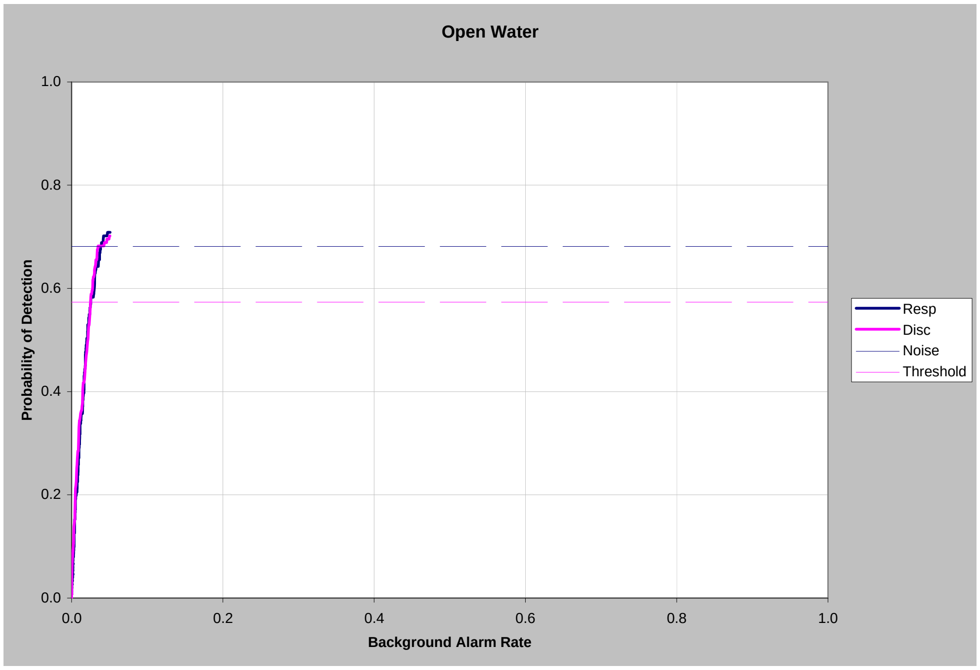

Figure 11. Open water P_d versus BAR.

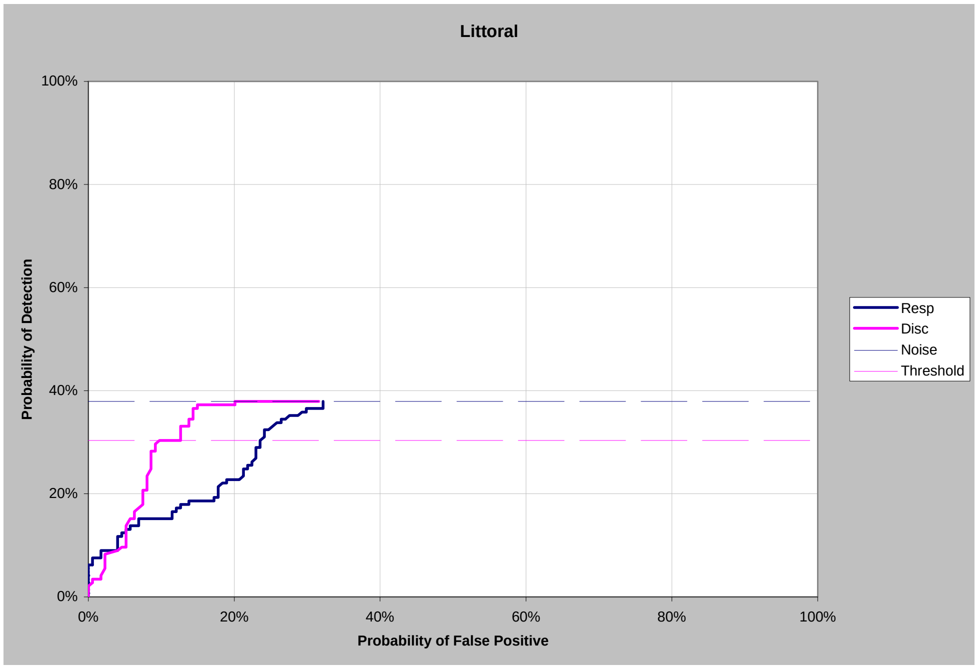

Figure 12. Littoral P_d versus P_{fp}.

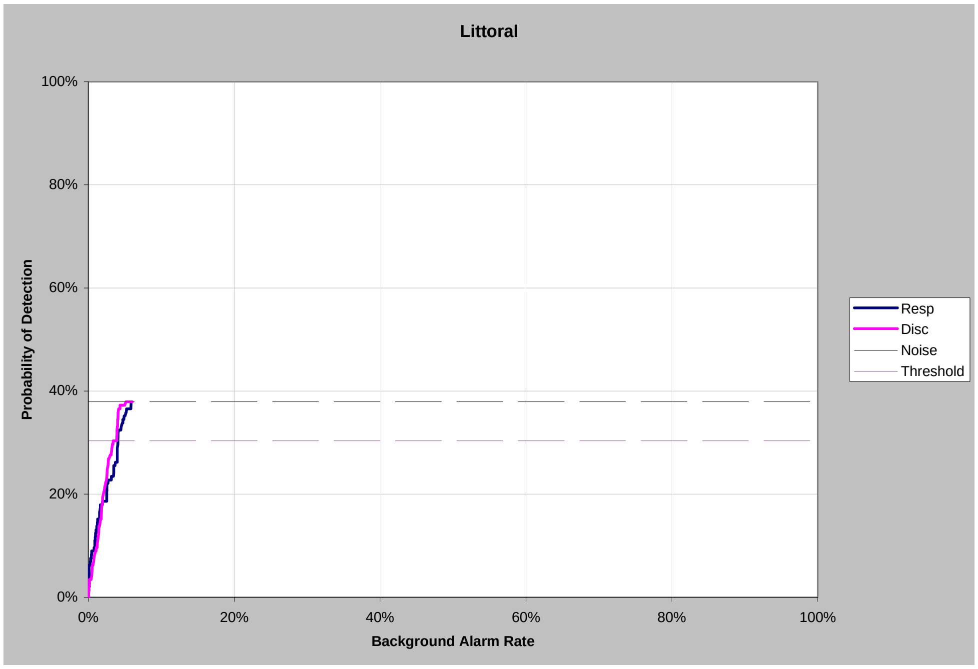

Figure 13. Littoral P_d versus BAR.

4.2.2 Detection Results

Detection results, broken out by stage, area surveyed, and ordnance size are presented in Table 4-2. The results by size indicate how well the demonstrator did at detecting/discriminating ordnance of a given caliber. Overall results summarize ordnance detection over a given area. All values were calculated assuming the number of detections is a binomially distributed random variable. These results are reported at the 90-percent reliability/95-percent confidence levels unless otherwise noted.

TABLE 4-2. SYSTEM DETECTION SUMMARY

Metric	Overall	By Projectile Caliber					
		40 mm	60 mm	81 mm	105 mm	155 mm	8 inch
Blind Grid							
Response Stage							
P_d	93.8%	79.3%	93.1%	96.6%	100.0%	100.0%	
P_d Lower 90% Confidence	90.4%	66.5%	82.7%	87.2%	92.4%	92.4%	
P_{fp}	73.6%						
P_{fp} Lower 90% Confidence	68.8%						
P_{ba}	7.7%						
Discrimination Stage							
P_d	82.8%	58.6%	79.3%	93.1%	89.7%	93.1%	
P_d Lower 90% Confidence	78.1%	45.0%	66.5%	82.7%	78.4%	82.7%	
P_{fp}	25.9%						
P_{fp} Lower 90% Confidence	21.5%						
P_{ba}	2.8%						
Open Water							
Response Stage							
P_d	68.2%	69.0%	65.5%	75.9%	79.3%	60.0%	33.3%
P_d Lower 90% Confidence	62.9%	55.5%	51.9%	62.8%	66.5%	47.8%	9.3%
P_{fp}	57.6%						
P_{fp} Lower 90% Confidence	52.9%						
BAR m^{-2}	0.051						
Discrimination Stage							
P_d	57.3%	69.0%	62.1%	62.1%	58.6%	45.7%	16.7%
P_d Lower 90% Confidence	51.9%	55.5%	48.5%	48.5%	45.0%	34.0%	1.7%
P_{fp}	18.2%						
P_{fp} Lower 90% Confidence	14.8%						
BAR m^{-2}	0.027						
Littoral Region							
Response Stage							
P_d	37.9%	27.6%	34.5%	41.4%	34.5%	51.7%	
P_d Lower 90% Confidence	32.6%	16.8%	22.6%	28.8%	22.6%	38.4%	
P_{fp}	32.2%						
P_{fp} Lower 90% Confidence	27.5%						
BAR m^{-2}	0.061						
Discrimination Stage							
P_d	30.3%	27.6%	27.6%	31.0%	24.1%	41.4%	
P_d Lower 90% Confidence	25.3%	16.8%	16.8%	19.7%	14.0%	28.8%	
P_{fp}	9.8%						
P_{fp} Lower 90% Confidence	7.0%						
BAR m^{-2}	0.034						
Deeper Water							
Response Stage							
P_d	75.9%					75.9%	
P_d Lower 90% Confidence	62.8%					62.8%	
Discrimination Stage							
P_d	75.9%					75.9%	
P_d Lower 90% Confidence	62.8%					62.8%	
Response Stage Noise Level: 20							
Recommended Discrimination Threshold: 5							

4.2.3 System Discrimination

Using the demonstrator's recommended setting, the items that were detected and correctly classified as ordnance were further evaluated as to whether the demonstrator could correctly identify the ordnance type (fig. 14). The list of ground-truth ordnance items was provided to the demonstrator prior to testing.

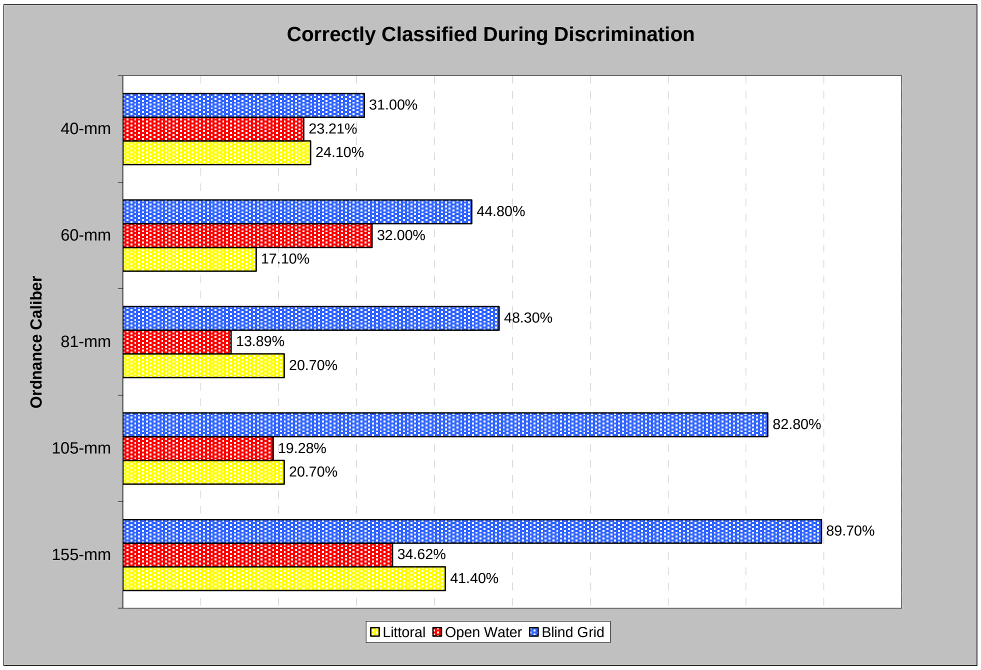

Figure 14. Percent of ordnance correctly classified during the discrimination stage.

4.2.4 System Effectiveness

Efficiency and rejection rates were calculated to quantify the discrimination ability at two specific points of interest on the ROC curve: at the point where no decrease in P_d is suffered (i.e., the efficiency is by definition equal to one) and at the operator-selected threshold. These values are presented in Table 4-3.

TABLE 4-3. EFFICIENCY AND REJECTION RATES

	Efficiency	False Positive Rejection Rate	Background Alarm Rejection Rate
Blind Grid			
At operating point	0.88	0.65	0.64
With no loss of P_d	1.00	0.21	0.00
Open Water			
At operating point	0.84	0.68	0.47
With no loss of P_d	1.00	0.01	0.01
Littoral			
At operating point	0.80	0.70	0.44
With no loss of P_d	1.00	0.38	0.18

4.2.5 Chi-Square Analysis

Chi-square 2 by 2 Contingency Test for comparison between ratios was used to compare performance across test areas with regard to P_d^{res}, P_d^{disc}, P_{fp}^{res}, P_{fp}^{disc}, efficiency, and false alarm rejection rate (R_{fp}). A one-sided Chi-square significance test at the 0.05 significance level was used to compare the survey results from the blind grid to the open water area, when using the same survey system. The intent of the comparison is to determine if the features introduced in each test site had a degrading effect on the performance of the sensor system.

Geophex modified the propulsion and data collection platforms used in the deeper water, blind grid, and open water areas to survey in the shallower littoral region. This change was necessary based on the design of the primary surveying system and the environments to be surveyed. The same operators, electromagnetic (EM) coils, signal processing, and GPS equipment, etc. were used in both areas. Performance comparisons were not made between the different systems used in different areas, i.e., blind grid-littoral and littoral-open water.

The results of the Chi-square performance comparison are presented in Table 4-4.

TABLE 4-4. CHI-SQUARE SIGNIFICANCE TEST RESULTS

Metric	Overall	By Projectile Caliber				
		40 mm	**60 mm**	**81 mm**	**105 mm**	**155 mm**
Blind Grid - Open Water Comparison						
P_d^{res}	Sig	Not	Sig	Sig	Sig	Sig
P_d^{disc}	Sig	[b]	Not	Sig	Sig	Sig
P_{fp}^{res}	Sig					
P_{fp}^{disc}	Sig					
Efficiency	Not					
R_{fp}	Not					
Sig = Significant Not = Not Significant						
[b] No test – Discrimination in the Open Water area is better than in the Blind Grid.						

4.2.6 Location Accuracy

The data points in the following scatter-graphs represent the coordinates of ordnance items in the open water and littoral test areas that were first detected in the response stage within a 0.5-meter radius of their true positions, then correctly identified as ordnance in the discrimination stage. The maximum error represents the 0.5-meter detection limit. The mean error represents the statistical mean of the sample considered.

Each scatter-graph represents both a test area and a data collection platform (para 2.1.3). A visual assessment of both graphs indicates that the location error is a randomly distributed as opposed to a systematic error.

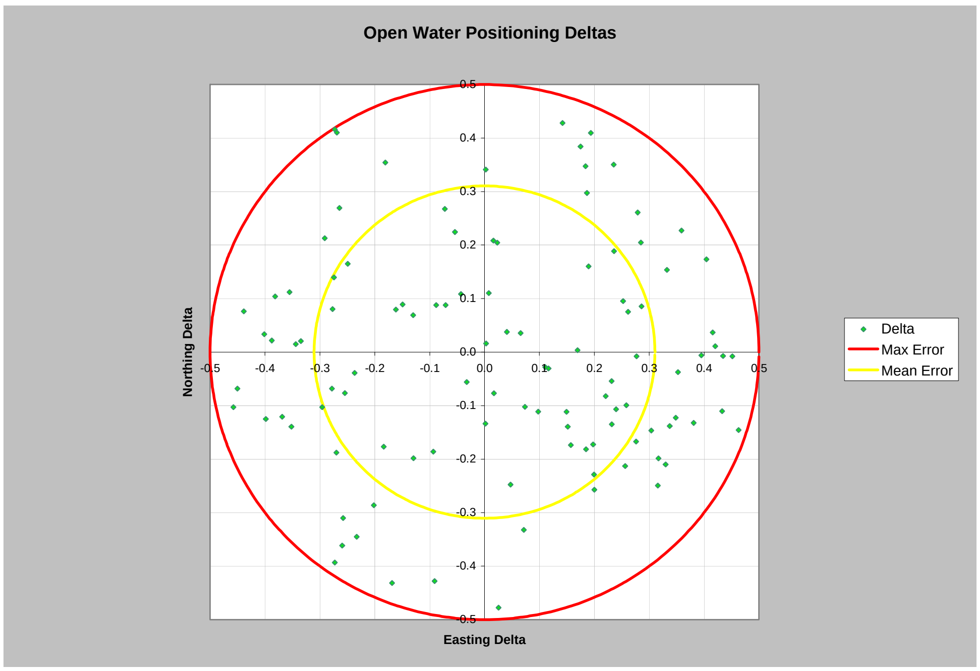

Figure 15. Open water positioning error scatter-graph.

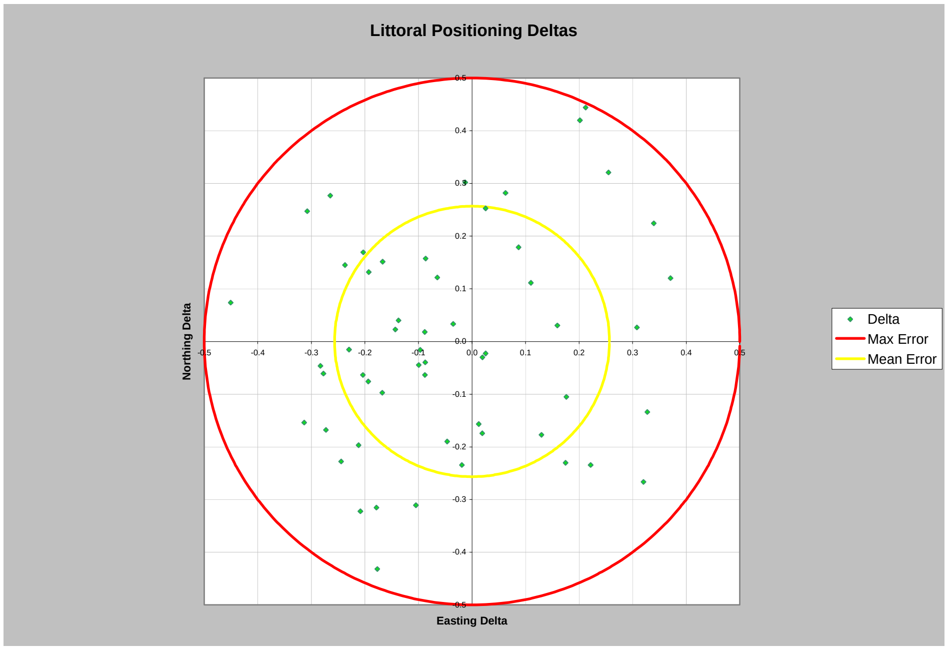

Figure 16. Littoral positioning error scatter-graph.

The comparison between the Geophex test results and the EQT-ORD criteria are presented in Table 4-5.

TABLE 4-5. TEST RESULTS - CRITERIA COMPARISION

Metric	**Threshold**	**Objective**	**Geophex by Area**	
Detection	80% ordnance items buried to 1 foot and under 8 feet (2.4 m) of water.	95% ordnance items buried to 4 feet and under 8 feet (2.4 m) of water.	Blind Grid	93.8%
			Open Water	68.2%
			Littoral	37.9%
Discrimination	Rejection rate of 50% of emplaced non-UXO clutter.	Rejection rate of 90% of emplaced non-UXO clutter.	Blind Grid	65%
			Open Water	68%
			Littoral	70%
	Maximum false negative rate of 10%.	Maximum false negative rate of 0.5%.	Not assessed. An analytical procedure is not available to address this criterion.	
Reacquisition	Reacquire within 1 meter.	Reacquire within 0.5 meters.	The reported detection values are based on ordnance items identified within 0.5 meters of the georeferenced ground-truth targets.	
Note: The blind grid and open water areas are in general accordance with the threshold requirements.				

<u>SECTION 5. APPENDIXES</u>

APPENDIX A. TEST CONDITIONS LOG

ATMOSPHERIC CONDITIONS

Date mm/dd/yy	Time, LST	Average Wind Direction, deg	Average Wind Speed, km/h	Wind Direction Average Standard Deviation, deg	Peak Wind Speed, km/h	Average Temperature, ^{o}C
9/27/05	0700	292	6.1	22	6.1	18.2
	0800	295	5.5	24	5.5	19.4
	0900	356	10.9	21	10.9	20.2
	1000	353	16.3	19	16.3	20.6
	1100	349	14.0	20	14.0	21.2
	1200	343	11.9	23	11.9	22.3
	1300	344	12.4	23	12.4	23.3
	1400	354	10.5	25	10.5	23.9
	1500	360	11.3	23	11.3	24.3
	1600	359	9.7	23	9.7	24.5
	1700	4	9.0	22	9.0	24.1
9/28/05	0700	10	5.8	9	5.8	12.2
	0800	21	5.3	19	5.3	14.6
	0900	68	4.0	35	4.0	18.5
	1000	109	4.5	43	4.5	20.6
	1100	252	4.8	89	4.8	21.9
	1200	221	11.4	22	11.4	22.9
	1300	227	13.2	20	13.2	23.4
	1400	238	13.4	13	13.4	23.6
	1500	224	13.0	16	13.0	24.0
	1600	217	14.3	15	14.3	24.2
	1700	200	12.4	14	12.4	23.8
9/30/05	0700	87	0.8	41	0.8	5.5
	0800	30	1.6	57	1.6	10.2
	0900	359	5.6	28	5.6	13.9
	1000	5	4.2	62	4.2	16.3
	1100	342	4.5	78	4.5	17.3
	1200	192	9.7	38	9.7	17.5
	1300	189	8.2	38	8.2	18.0
	1400	189	7.9	39	7.9	18.6
	1500	211	4.8	78	4.8	19.4
	1600	232	5.3	35	5.3	19.4
	1700	201	5.6	19	5.6	19.1
10/1/05	0700	33	1.0	9	1.0	6.7
	0800	48	1.4	25	1.4	11.5
	0900	92	2.4	34	2.4	17.3
	1000	25	3.2	72	3.2	19.2
	1100	128	3.7	72	3.7	20.8

(CONT'D)

Date mm/dd/yy	Time, LST	Average Wind Direction, deg	Average Wind Speed, km/h	Wind Direction Average Standard Deviation, deg	Peak Wind Speed, km/h	Average Temperature, °C
10/1/05	1200	212	8.4	40	8.4	21.4
	1300	233	8.9	31	8.9	22.0
	1400	238	9.8	16	9.8	22.3
	1500	239	9.2	13	9.2	22.8
	1600	237	6.6	17	6.6	23.0
	1700	218	5.6	23	5.6	22.7
10/5/05	0700	34	3.9	13	3.9	17.8
	0800	43	4.7	20	4.7	18.3
	0900	46	6.0	23	6.0	19.5
	1000	106	6.1	37	6.1	20.3
	1100	73	5.6	33	5.6	20.1
	1200	85	3.5	47	3.5	20.9
	1300	62	6.4	41	6.4	22.6
	1400	113	5.8	39	5.8	24.7
	1500	75	6.4	30	6.4	25.4
	1600	100	4.7	34	4.7	26.3
	1700	65	3.7	22	3.7	26.1

Water conditions documented from the Geophex team from 27 September through 1 October 2005 are shown in Figures A-1 through A-5.

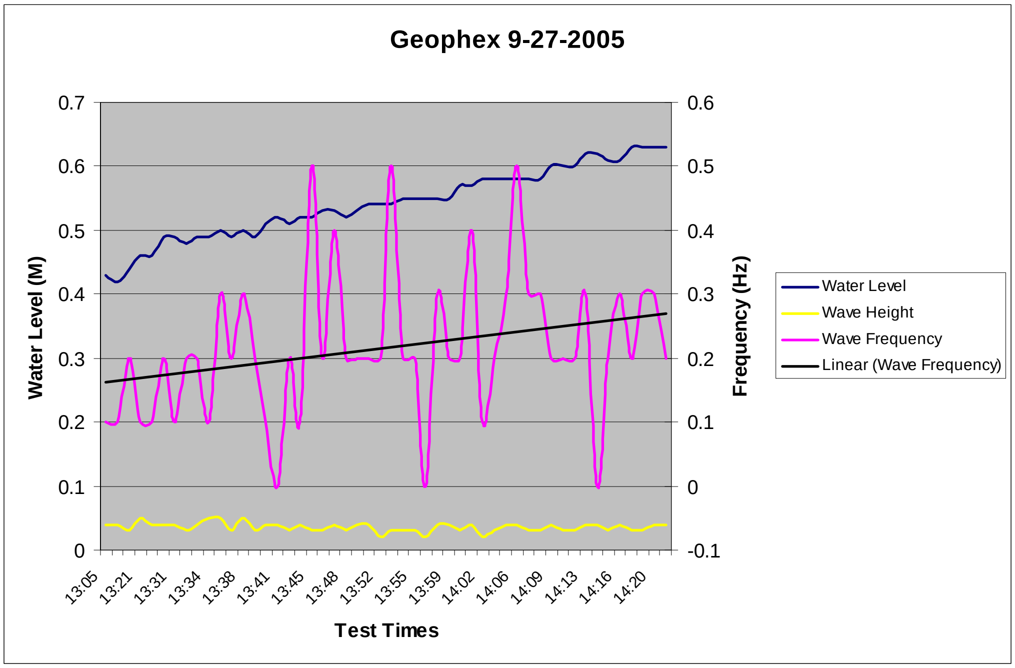

Figure A-1. Water conditions on 27 September 2005.

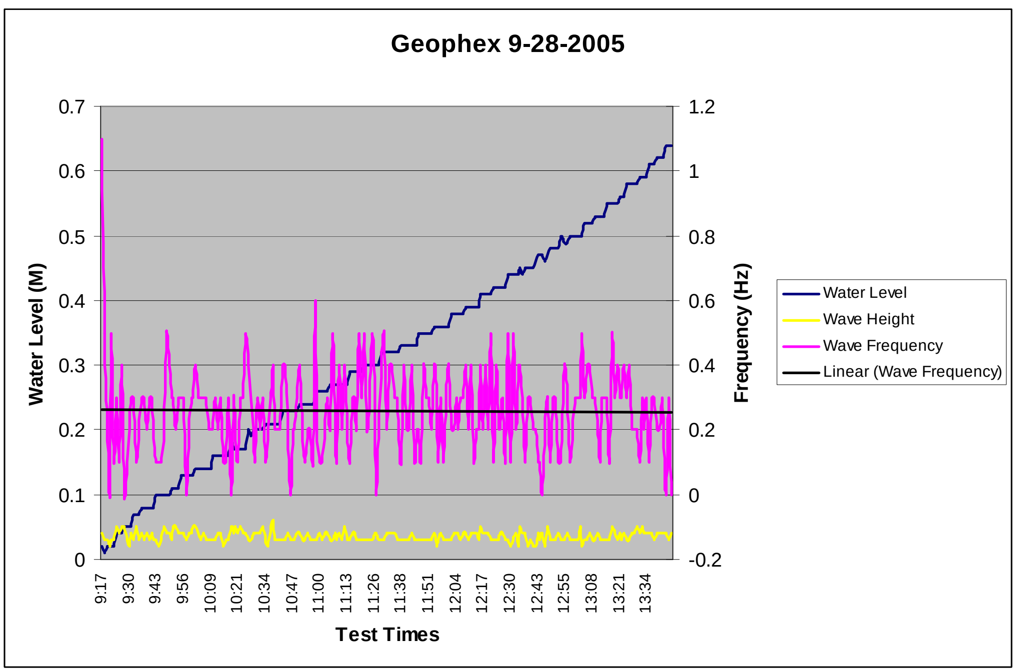

Figure A-2. Water conditions on 28 September 2005.

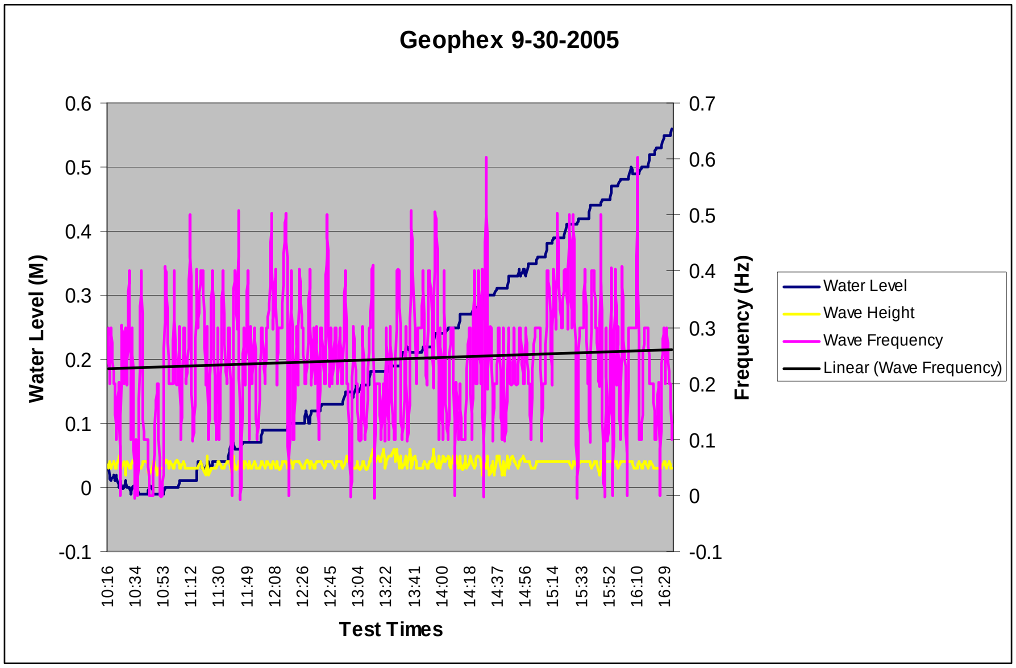

Figure A-3. Water conditions on 30 September 2005.

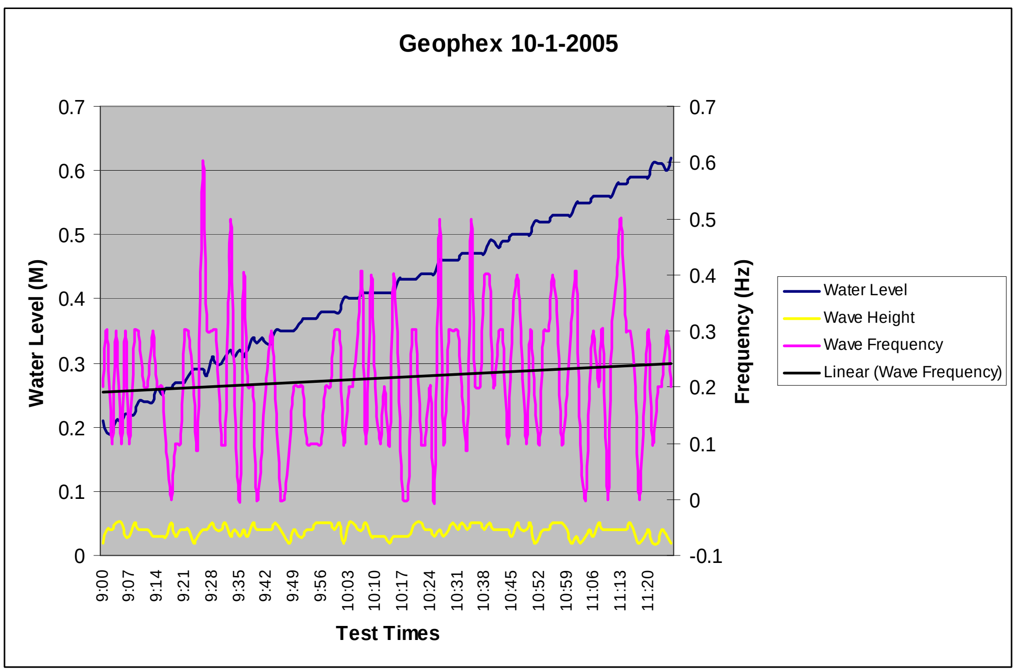

Figure A-4. Water conditions on 1 October 2005.

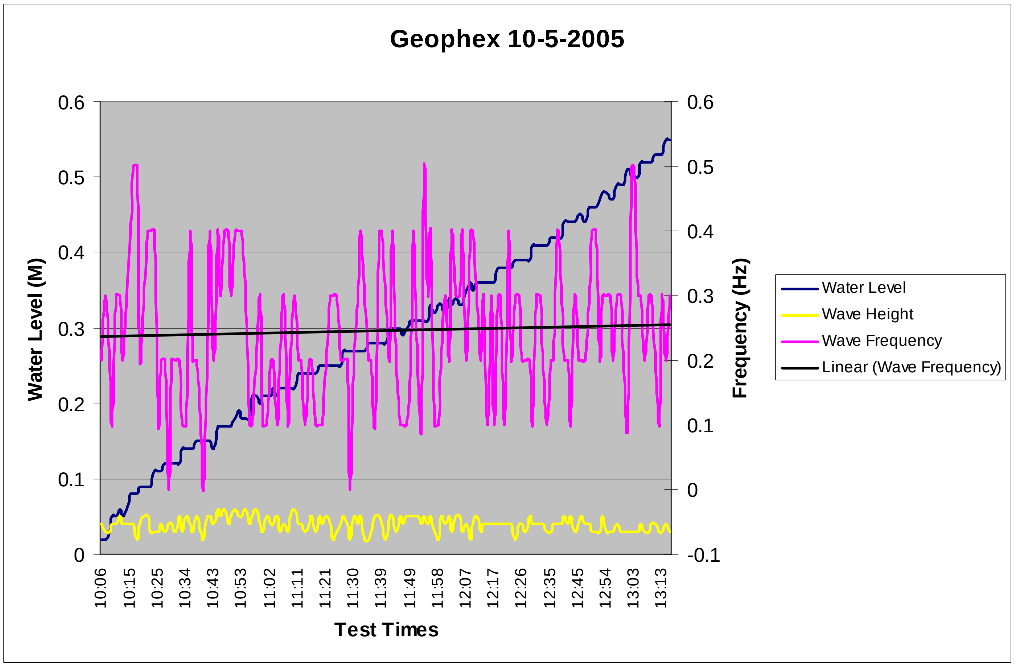

Figure A-5. Water conditions on 5 September 2005.

Date: 9/27/05		*Personnel: Bill SanFilipo, Collin Mellor, and Bradley Carr*		
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0715	0900	Arrived at test site. Launched the boat. Unloaded the truck. Assembled the sled the sensor rests on. Set up the base station. Assembled the wooden frame used to attach the pole and sensor assembly to the boat. Inflated the small pontoon boat for the littoral area.	Initial Setup	105
0900	1530	Problems with the positioning system for the boat. Trimble heads on the towing pole communicating with the base station, but not the head used for boat positioning.	Initial Setup	420
1530	1600	Calibrating sensors with ferrous rods.	Calibration	30
1600	1650	Still no resolution on the GPS problem. Canceled for the day. Perhaps Trimble will provide a solution later this evening. Began packing up for the day. Left test site.	Daily close-up	50

Date: 9/28/05		*Personnel: Bill SanFilipo, Collin Mellor, and Bradley Carr*		
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0715	0830	Arrived at test site. Trimble problem not resolved. Will survey using 'was' function.	Initial Setup	75
0830	1010	Began survey.	Survey	100
1010	1030	Stopped survey to strengthen the wooden frame in the boat that the pole and sensors are attached to. Tight turns produce more stress than anticipated.	Downtime equipment	20
1030	1125	Survey resumes.	Survey	55
1125	1215	Survey stops. One of the two GPS antennas (lower one) experienced signal dropout all morning. The lower antenna was moved toward the top one. The belief is that the boat was blocking reception. All survey data from this morning was lost.	Downtime equipment	50
1215	1600	Starting a new survey.	Survey	225
1600	1700	Post-survey positioning check. Measured target signatures in air.	Calibration	60
1700	1750	Packed up for the day. Left site.	Daily close-up	50

Date: 9/29/05			*Personnel: Collin Mellor and Bradley Carr*	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0800	1200	The weather forecast for today was wind gusts to 40-mph and rain. No survey work was done at the pond. Worked off-site in the morning, on-site in the afternoon.		
0800	0900	Off-site Geophex was working on data reduction. Data problems from Wednesday's survey. Survey paths shown on the navigation screen were wider than they actually were due to a wrong constant in the program (10 vs.1). Cannot mathematically correct due to the varying angle of the pole during turning.	Calibration	60
0800	1130	Off-site Geophex was constructing a frame to mount the sensor under the dingy.	Setup	220
1230	1420	On-site working with dingy assembly.	Setup	110
1420	1500	Positioning checks were made and reference points marked for the pre- and post-survey checks.	Calibration	40
1500	1530	Packed and left site.	Daily close-up	30

Date: 9/30/05			*Personnel: Collin Mellor and Bradley Carr*	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0715	0850	Arrived at test site, set up.	Setup	95
0850	0900	Sensor calibration.	Calibration	10
0900	1410	Began survey.	Survey	310
1410	1420	Changed navigation computer battery.	Downtime maintenance	10
1420	1600	Survey.	Survey	100
1600	1650	Positioning check/sensor checks.	Calibration	50
1650	1715	Packed and left site.	Daily close-up	25

Date: 10/01/05			*Personnel: Collin Mellor and Bradley Carr*	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0740	0840	Arrived at test site, set up.	Setup	60
0840	1415	Collected data with the "big boat."	Survey	335
1415	1422	Changed batteries.	Downtime maintenance	7
1422	1530	Surveyed.	Survey	68
1530	1620	Packed and left site.	Daily close-up	50

Date: 10/03/05			*Personnel: Frank Funak and Bradley Carr*	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0805	0910	Set up, modified boat.	Setup	65
0910	1535	Collected data.	Survey	325
1535	1630	Broke down equipment. Left site.	Daily close-up	55

Date: 10/04/05			*Personnel: Frank Funak and Bradley Carr*	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0805	1240	Set up, mobilized boat.	Setup	265
1240	1515	Collected data around outer edges with small boat.	Survey	155
1515	1615	Broke down equipment. Left site.	Daily close-up	60

Date: 10/05/05			*Personnel: Bill SanFilipo, Frank Funak, and Bradley Carr*	
Start	**Stop**	**Remarks**	**Activity**	**Chargeable, min**
0815	0850	Arrived at test site, set up dingy. Testing yesterday revealed the sensors were picking up interference from the boat motor. (Sensor set ~0.5 m below water surface.)	Setup	35
0850	1000	Base lined sensor. Picked up systematic error. Troubleshot. Lowered sensor 1 meter below boat. Reduced but didn't eliminate noise. One sensor overloaded (same as yesterday). Decided to survey.	Downtime equipment	70
1000	1015	Began survey over the calibration lanes.	Survey	15
1015	1300	Began surveying other areas	Survey	165
1300		End of survey - no QC.		
1300	1640	Began packing equipment. Left site.	Demobilization	220

APPENDIX C. 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.

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

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. 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 less than 0.6 meters in length. When ordnance items are longer than 0.6 meters, the halo becomes an ellipse where the minor axis remains 1 meter and the major axis is equal to the projected length of the ordnance onto the ground plane plus 1 meter.

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 selects the 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 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: An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open water or littoral 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 water only: BAR^{res} = (No. of response-stage background alarms)/(arbitrary constant).

Note that 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 non-ordnance 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: An anomaly in a blind grid cell that contains neither emplaced ordnance nor an emplaced clutter item. An anomaly location in the open water or littoral 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.

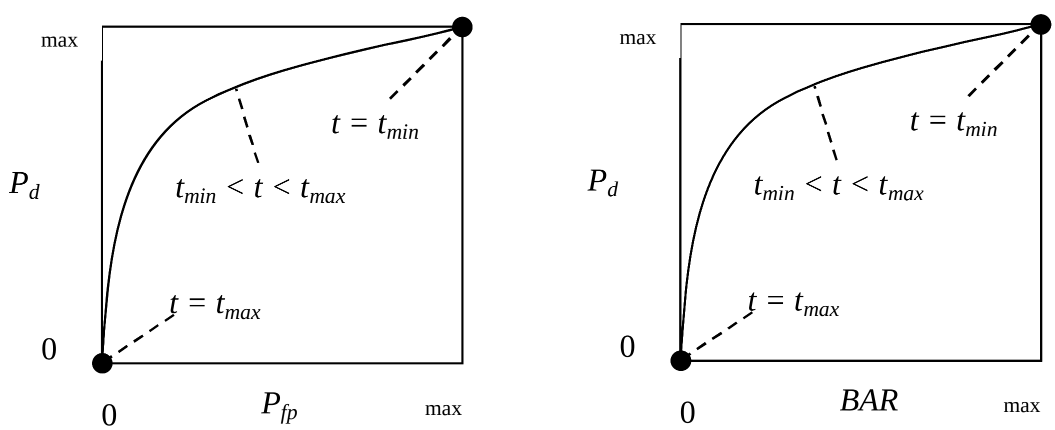

Figure A-1. ROC curves for open-site testing. Each curve applies to both the response and discrimination stages.

---

[1]Strictly 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 ordnance and others are located over clutter or blank spots). In an Open Water scenario, each system suppresses its signal strength reports until some bare-minimum signal response is received by the system. Consequently, the Open Water 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.

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 non-ordnance 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.

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 water: $R_{ba} = 1 - [BAR^{disc}(t^{disc})/BAR^{res}(t_{min}^{res})])$

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 x 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, pages 144 through 151).

A one-sided 2 x 2 contingency table is used in the Shallow Water Site Program to compare each area (Open Water, Littoral, Deep Water) to the Blind Grid since each area introduces a water feature that makes it potentially more difficult to survey than the Blind Grid. The one-sided 2 x 2 contingency table is used 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 feature introduced. A two-sided 2 x 2 contingency table is used to compare performance between any two of the test sites other than the Blind Grid, to determine if there is reason to believe that the proportion of ordnance correctly detected/discriminated by demonstrator X's system is significantly different between those two test sites.

The test statistic of the 2 x 2 contingency table is the Chi-square distribution with one degree of freedom. For the one-sided test, a significance level of 0.05 is chosen which sets a critical decision limit of 3.84 from the Chi-square distribution with one degree of freedom. It 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 Exact Test is used and the critical decision limit is the chosen significance level, which is 0.05 for one-sided tests and 0.10 for two-sided tests. With Fischer's test, if the test statistic (p-value) is less than the critical value, then the null hypothesis of similar performance is rejected in favor of the alternative hypothesis: significantly greater than for the one-sided case or significantly different for the two-sided case.

Shallow-water UXO Detection Test Site examples, where blind grid results are compared to those from the open water and littoral sites and the non-grid sites (open water and littoral) are compared to each other as follows. It should be noted that a significant result does not prove a cause and effect relationship exists between the change in survey area and sensor performance; however, it does serve as a tool to indicate that one data set reflects relatively degraded system performance of a large enough scale 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 areas using the same system (results indicate the number of ordnance detected divided by the number of ordnance emplaced):

	Blind Grid	Open water	Littoral
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 WATER. 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 water. 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 (p-value) 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 water relative to results from the blind grid using the same system.

P_d^{disc}: BLIND GRID versus OPEN WATER. 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 out of 10 emplaced ordnance items were correctly discriminated as such in open water testing. Those four values are used in the Chi-square Contingency Test to calculate a test statistic of 1.12. Since the test statistic is less than the critical value of 3.84, the two discrimination stage detection rates are considered to be not significantly different at the 0.05 level of significance.

P_d^{res}: BLIND GRID versus LITTORAL. Using the example data above to compare probabilities of detection in the response stage, 100 out of 100 and 20 out of 33 are used to calculate a test statistic (< 0.000) 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 (0.61) is considered to be significantly less at the 0.05 level of significance.

P_d^{disc}: BLIND GRID versus LITTORAL. Using the example data above to compare probabilities of detection in the discrimination stage, 80 out of 100 and 8 out of 33 emplaced ordnance items were correctly discriminated as such in open water testing. Those four values are used to calculate a test statistic of 32.01. Since the test statistic is greater than the critical value of 3.84, the smaller discrimination stage detection rate (0.24) is considered to be significantly less at the 0.05 level of significance.

P_d^{res}: OPEN WATER versus LITTORAL. 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 0.10 level of significance.

P_d^{disc}: OPEN WATER versus LITTORAL. 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 two discrimination stage detection rates are considered to be significantly different at the 0.10 level of significance. While a significant result does not prove a cause and effect relationship exists between the change in survey area and change in performance, it does indicate that the ability of Demonstrator X to correctly discriminate seems to have been degraded by features of the littoral area relative to results from the open water using the same system.

APPENDIX D. REFERENCES

1. Environmental Quality Technology - Operational Requirements Document (EQT-ORD) for: A(1.6.a): UXO Screening, Detection and Discrimination.

2. Technical Management Plan, Unexploded Ordnance Detection and Discrimination Demonstration for the APG Standardized UXO Technology Shallow Water Demonstration Site. Submitted in response to the BAA W91ZLK-04-R-0001 by Tetra Tech EC, Inc., 9 August 2005.

3. Practical Nonparametric Statistics, W.J. Conover, John Wiley & Sons, 1980, pages 144 through 151.

APPENDIX E. ABBREVIATIONS

ADST = Aberdeen Data Services Team
APG = Aberdeen Proving Ground
ATC = U.S. Army Aberdeen Test Center
BAA = Broad Agency Announcement
BAR = background alarm rate
DGPS = Differential Global Positioning System
DMM = discarded military munitions
EM = electromagnetic
EMF = electromagnetic force
EMI = electromagnetic interference
EQT = Army Environmental Quality Technology Program
EQT-ORD = Environmental Quality Technology - Operational Requirements Document
ERDC = U.S. Army Corps of Engineers Engineering Research Development Center
ESTCP = Environmental Security Technology Certification Program
GPS = Global Positioning System
LED = light-emitting diode
Log.Sec/Tri-S = Logistics Engineering and Information Technology Company
MEC = munitions and explosives of concern
MEDTC = Military Environmental Technology Demonstration Center
ORD = Operational Requirements Document
P_{ba} = probability of background alarm
P_d = probability of detection
P_d^{res} = probability of detection, response stage
P_d^{disc} = probability of detection, discrimination stage
P_{fp} = probability of false positive
P_{fp}^{disc} = probability of false positive, discrimination stage
P_{fp}^{res} = probability of false positive, response stage
POC = point of contact
PPM = parts per million
PVC = polyvinyl chloride
QA = quality assurance
QC = quality control
Rfp = false alarm rejection rate
ROC = receiver operating characteristics
SERDP = Strategic Environmental Research and Development Program
USAEC = U.S. Army Environmental Center
UTM = Universal Transverse Mercator
UXO = unexploded ordnance

APPENDIX F. DISTRIBUTION LIST

DTC Project No. 8-CO-160-UXO-016

Addressee	No. of Copies
Commander U.S. Army Environmental Center ATTN: SFIM-AEC-ATT (Ms. Bonnie Packer) Aberdeen Proving Ground, MD 21010-5401	1
Commander U.S. Army Aberdeen Test Center ATTN: CSTE-DTC-AT-SL-E (Mr. Gary Rowe)	1
(Library)	1
CSTE-DTC-AT-CS-RM Aberdeen Proving Ground, MD 21005-5055	1
Geophex, Ltd. ATTN: (Mr. Bill SanFilipo) 605 Mercury Street Raleigh, NC 27603-2343	1
Defense Technical Information Center 8725 John J. Kingman Road, Suite 0944 Fort Belvoir, VA 22060-6218	PDF

Secondary "hard-copy" distribution is controlled by Commander, U.S. Army Environmental Center, ATTN: SFIM-AEC-ATT.