SCORING PROTOCOLS

TABLE OF CONTENTS

1	Scoring Objectives	P-2
2	Scoring Methodology	P-2
3	Scoring of the Data	P-4
3.1	Blind Test Grid	P-4
3.1.1	Blind Grid Scoring Example	P-4
3.1.2	Development of ROC Curves - Response Stage	P-6
3.1.3	Development of ROC Curves - Discrimination Stage	P-8
3.1.4	Other Blind Grid Performance Metrics	P-10
3.2	Open Field	P-12
3.2.1	Open Field Scoring Example	P-13
3.2.2	Development of ROC Curves - Response Stage	P-14
3.2.3	Development of ROC Curves - Discrimination Stage	P-16
3.2.4	Other Open Field Performance Metrics	P-18
4	Comparisons Between Test Scenarios	P-19
5	Cost, Time and Manpower Requirements	P-20

1. Scoring Objectives

a. The objective of the Standardized UXO Technology Demonstration Site Program is to evaluate UXO detection and discrimination capabilities under various field and soil conditions. Technologies are evaluated on highly controlled sites with inert munitions and clutter items positioned in various orientations and depths in the ground. After visiting a standardized site, demonstrators will be evaluated based on survey results and field operations.

b. The evaluation criteria are as follows:

(1) Detection and discrimination under realistic scenarios that vary targets, geology, clutter, topography, and vegetation.

(2) Cost, time, and manpower requirements.

(3) Ability to analyze survey data in a timely manner and provide prioritized target lists with associated confidence levels.

(4) Collection of high quality, ground-truth, geo-referenced data for post-demonstration analysis.

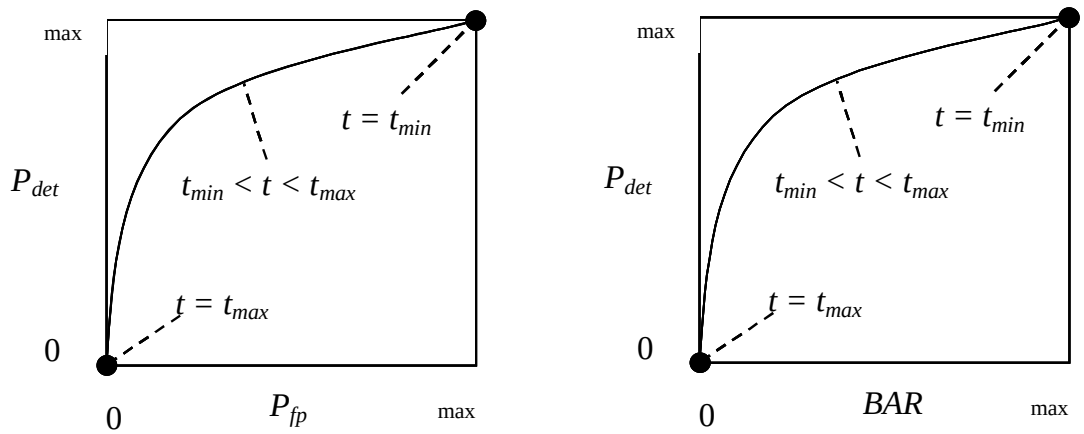
2. Scoring Methodology

a. The scoring of the demonstrator's performance is conducted in two stages. These two stages are termed the RESPONSE STAGE and DISCRIMINATION STAGE.

(1) 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 field 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.

(2) The DISCRIMINATION STAGE evaluates the demonstrator's ability to correctly identify ordnance as such and to reject clutter. For the same field 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 selects the threshold that is believed to provide optimum system performance (i.e., that retains all the detected ordnance and rejects the maximum amount of clutter).

b. For both stages, the probability of detection (P_{det}) 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 (BA). The ROC curves P_{det} versus P_{fp} and P_{det} versus background alarm rate (BAR), as the threshold applied to the signal strength, is varied from its minimum (t_{min}) to its maximum (t_{max}) value. Figure P-1 shows how P_{det}, P_{fp}, and BAR are combined into ROC curves.[1]



Each curve will be calculated for both the response and discrimination stages.

Figure P-1. ROC Curves.

c. 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. EFFICIENCY measures the amount of detected ordnance retained after 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_{det} versus P_{fp} over a predetermined and fixed number of detection opportunities (some of the opportunities are located over targets and others of 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 Test Grid sites are true ROC curves.

3. Scoring of the Data

3.1 Blind Test Grid

a. The Blind Test Grid is designed to test detection systems and algorithms without considering location accuracy or platform effects. The demonstrator is given 400 grid locations (opportunities) to investigate. When the demonstrator investigates opportunities, the following questions must be answered:

(1) What is the signal level in each grid location? (Note: Analog or human judgment requires 1 or 0 indicating either a detection or an empty grid, respectively.)

(2) If the sensor has the ability to discriminate:

(a) What is the numerical result of the discrimination algorithm (electronic, human or some combination) to reflect whether grid location contains ordnance, clutter, or is empty?

(b) If it is ordnance:

1 What is it? (For example: 155-mm projectile, 2.75-inch rocket or BLU-26 submunition.)

2 What is its depth?

3 What is its orientation? (This information is not scored.)

a Azimuth - horizontal rotation relative to magnetic North (0 to 360°).

b Dip - vertical orientation relative to the horizontal plane (0 ± 90°).

b. The data collected by the demonstrator are processed and submitted to the scoring committee in the required format. A computer program compares the demonstrator data to the ground truth database. The program links the grid location in the ground truth database to the demonstrator's data.

3.1.1 Blind Grid Scoring Example.

Table P-1 is an example of the demonstrator's data in the required format. Table P-2 is an example of hypothetical ground truth data. Table P-3 is an example of the computer program linking the grid locations.

TABLE P-1. DEMONSTRATOR DATA FOR BLIND TEST GRID

	Grid Letter	Grid No.	Response Stage	Discrimination Stage/ Ranking	Classification	Type	Depth, m	Azimuth, deg	Dip, deg
1	A	1	0.649	2	C	-	-	-	-
2	A	2	2.852	5	O	20-mm	1.72	142	-69
3	A	3	8.209	33	O	105-mm	0.86	16	-52
4	B	1	1.757	9	C	-	-	-	-
5	B	2	0.491	0	B	-	-	-	-
6	B	3	9.980	52	O	155-mm	0.37	140	40
7	C	1	12.679	8	O	105-mm	0.28	137	70
8	C	2	0.649	7	C	-	-	-	-
9	C	3	2.852	14	O	BLU26	0.72	122	55

O = Ordnance.
C = Clutter.
B = Blank.

TABLE P-2. HYPOTHETICAL GROUND TRUTH DATA

	Grid Letter	Grid No.	Classification	Type	Depth, m	Azimuth, deg	Dip, deg
1	A	1	C	-	-	-	-
2	A	2	O	20-mm	2.00	175	-80
3	A	3	O	105-mm	0.86	23	-40
4	B	1	B	-	-	-	-
5	B	2	B	-	-	-	-
6	B	3	O	105-mm	0.37	145	50
7	C	1	C	-	-	-	-
8	C	2	C	-	-	-	-
9	C	3	O	M42	1.00	135	45

O = Ordnance.
C = Clutter.
B = Blank.

TABLE P-3. COMPUTER PROGRAM LINKING GRID LOCATIONS

	Grid Letter	Grid No.	Response Stage	Discrimination Stage/ Ranking	Classification	Type	Depth, m	Azimuth, deg	Dip, deg
1	A	1	0.649	2	C	-	-	-	-
(gt)	A	1	-	-	C	-	-	-	-
2	A	2	2.852	5	O	20-mm	1.72	142	-69
(gt)	A	2	-	-	O	20-mm	2.00	175	-80
3	A	3	8.209	33	O	105-mm	0.86	16	-52
(gt)	A	3	-	-	O	105-mm	0.86	23	-40
4	B	1	1.757	9	C	-	-	-	-
(gt)	B	1	-	-	B	-	-	-	-
5	B	2	0.491	0	B	-	-	-	-
(gt)	B	2	-	-	B	-	-	-	-
6	B	3	9.980	52	O	155-mm	0.37	140	40
(gt)	B	3	-	-	O	105-mm	0.37	145	50
7	C	1	12.679	8	O	105-mm	0.28	137	70
(gt)	C	1	-	-	C	-	-	-	-
8	C	2	0.649	7	C	-	-	-	-
(gt)	C	2	-	-	C	-	-	-	-
9	C	3	2.852	14	O	BLU26	0.72	122	55
(gt)	C	3	-	-	O	M42	1.00	135	45

C = Clutter.
B = Blank.
Gt = Ground truth.
O = Ordnance.

3.1.2 Development of ROC Curves - Response Stage

a. In the response stage, the computer algorithm scores the demonstrator's data by grid location, comparing the response column to ground truth data. After linking the demonstrator data to the ground truth database, the program thresholds the demonstrator response data. The threshold increment is determined by dividing the range (maximum - minimum) of response stage values into a desired number of increments. Table P-4 shows the derivation of seven threshold increments from the sample demonstrator data of Table P-1.

TABLE P-4. DETERMINATION OF THRESHOLD

Maximum Response Stage Value	12.679
Minimum Response Stage Value	0.649
No. increments (variable)	7.000
Threshold Increment	1.7186

b. At each threshold increment the following are calculated:

<u>Response Stage Probability of Detection</u> (P_d^{res}): P_d^{res} = (# of response-stage detections)/(# of emplaced ordnance).

<u>Response Stage Probability of False Positive</u> (P_{fp}^{res}): P_{fp}^{res} = (# of response-stage false positives)/(# of emplaced clutter items).

<u>Response Stage Probability of Background Alarm</u> (P_{ba}^{res}): P_{ba}^{res} = (# of response-stage background alarms)/(# of empty grid locations).

Note that the quantities P_d^{res}, P_{fp}^{res}, and P_{ba}^{res} are functions of t^{res}, the threshold applied to the response-stage signal strength. Table P-5 continues the example and shows how ROC curve quantities are calculated for the sample demonstrator data of Table P-1 using the threshold increment determined in Table P-4. Sample ROC curves are plotted in Figures P-2 and P-3.

TABLE P-5. EXAMPLE BLIND GRID RESPONSE STAGE ROC CURVE CALCULATIONS

		Increment	1	2	3	4	5	6	7
		Threshold	0.649	1.7186	3.4371	5.1557	6.8743	8.5929	10.311
	Ground		Result = 1 if Response >= Threshold_{inc}; 0 otherwise						
Item	Truth	RESPONSE	RESULT						
1	C	0.649	1	0	0	0	0	0	0
2	O	2.852	1	1	0	0	0	0	0
3	O	8.209	1	1	1	1	1	0	0
4	B	1.757	1	1	0	0	0	0	0
5	B	5.491	1	1	1	1	0	0	0
6	O	9.98	1	1	1	1	1	1	0
7	C	12.679	1	1	1	1	1	1	1
8	C	0.649	1	0	0	0	0	0	0
9	O	2.852	1	1	0	0	0	0	0
Ground Truth Summary:		COUNTS							
# ordnance	4	# detections	4	4	2	2	2	1	0
#fp opportunities	3	# fp	3	1	1	1	1	1	1
#ba opportunities	2	# ba	2	1	1	1	0	0	0
		P_d^{res}	1.0	1.0	0.5	0.5	0.5	0.25	0.0
		P_{fp}^{res}	1.0	0.3333	0.3333	0.3333	0.3333	0.3333	0.3333
		P_{ba}^{res}	1.0	0.5	0.5	0.5	0.0	0.0	0.0

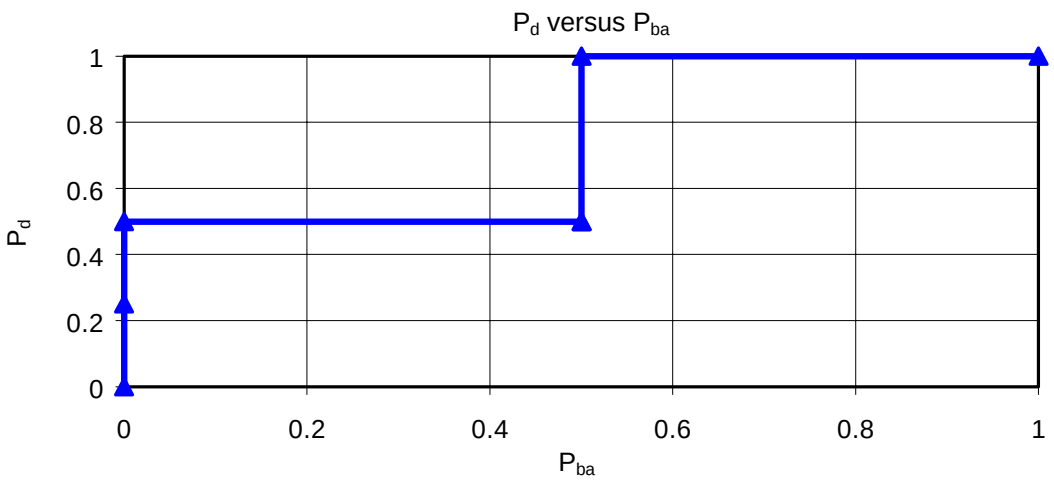


Figure P-2. Blind grid response stage ROC curve.

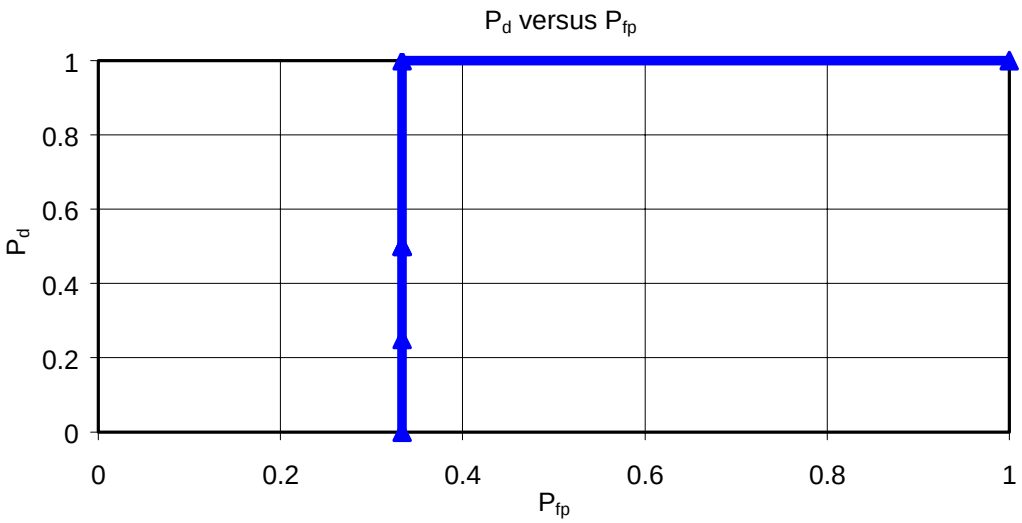


Figure P-3. Blind grid response stage ROC curve.

b. In order to protect the Standardized UXO Test Site ground truth, probabilities and rates will be reported as rounded figures and ROC curves will be smoothed.

3.1.3 Development of ROC Curves - Discrimination Stage

a. In the discrimination stage, the demonstrator's data is already linked by grid location to the ground truth database. Similar to the response stage, the computer program thresholds the demonstrator discrimination data and compares the results to ground truth. The threshold increment is determined by dividing the range (maximum - minimum) of discrimination stage values into a desired number of increments, independent of the number of increments used in the response stage. Table P-6 shows the creation of eight threshold increments, each of size 6.5, from the sample demonstrator data of Table P-1.

TABLE P-6. DETERMINATION OF THRESHOLD

Maximum Discrimination Stage Value	52
Minimum Discrimination Stage Value	0
No. increments (variable)	8
Threshold Increment	6.5

b. At each threshold increment, as well as at the demonstrator's recommended optimum level, the following are calculated:

Discrimination Stage Probability of Detection (P_d^{disc}): P_d^{disc} = (# of discrimination-stage detections)/(# of emplaced ordnance).

Discrimination Stage Probability of False Positive (P_{fp}^{disc}): P_{fp}^{disc} = (# of fp^{disc})/(# of emplaced clutter items).

Discrimination Stage Probability of Background Alarm (P_{ba}^{disc}): P_{ba}^{disc} = (# of BA^{disc})/(# of empty grid locations).

Note that the quantities P_d^{disc}, P_{fp}^{disc}, and P_{ba}^{disc} are functions of t^{disc}, the threshold applied to the discrimination-stage value. Table P-7 shows how discrimination-stage ROC curve quantities are calculated for the sample demonstrator data of Table P-1 using the threshold increment determined in Table P-6. Sample ROC curves are plotted in Figures P-4 and P-5.

TABLE P-7. EXAMPLE BLIND GRID DISCRIMINATION STAGE ROC CURVE CALCULATIONS

		Increment	1	2	3	*	4	5	6	7	8
		Threshold	0	6.5	13	15	19.5	26	32.5	39	45.5
	Ground		Result = 1 if Discrimination >= $Threshold_{inc}$; 0 otherwise								
Item	Truth	DISCRIMINATION	RESULT								
1	C	2	1	0	0	0	0	0	0	0	0
2	O	5	1	0	0	0	0	0	0	0	0
3	O	33	1	1	1	1	1	1	1	0	0
4	B	9	1	1	0	0	0	0	0	0	0
5	B	0	1	0	0	0	0	0	0	0	0
6	O	52	1	1	1	1	1	1	1	1	1
7	C	8	1	1	0	0	0	0	0	0	0
8	C	7	1	1	0	0	0	0	0	0	0
9	O	14	1	1	1	0	0	0	0	0	0
Ground Truth Summary:			COUNTS								
# ordnance	4	# detections	4	3	3	2	2	2	2	1	1
#fp opportunities	3	# fp	3	2	0	0	0	0	0	0	0
#ba opportunities	2	# ba	2	1	0	0	0	0	0	0	0
		P_d^{disc}	1	0.75	0.75	0.5	0.5	0.5	0.5	0.25	0.25
		P_{fp}^{disc}	1	0.6667	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		P_{ba}^{disc}	1	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0

* = Demonstrator's recommended threshold level.

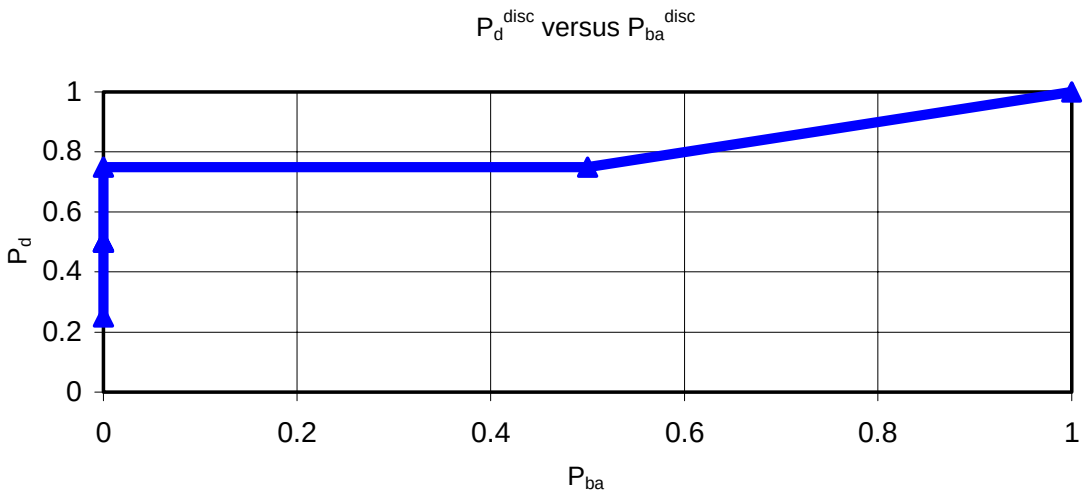


Figure P-4. Blind grid discrimination stage ROC curve.

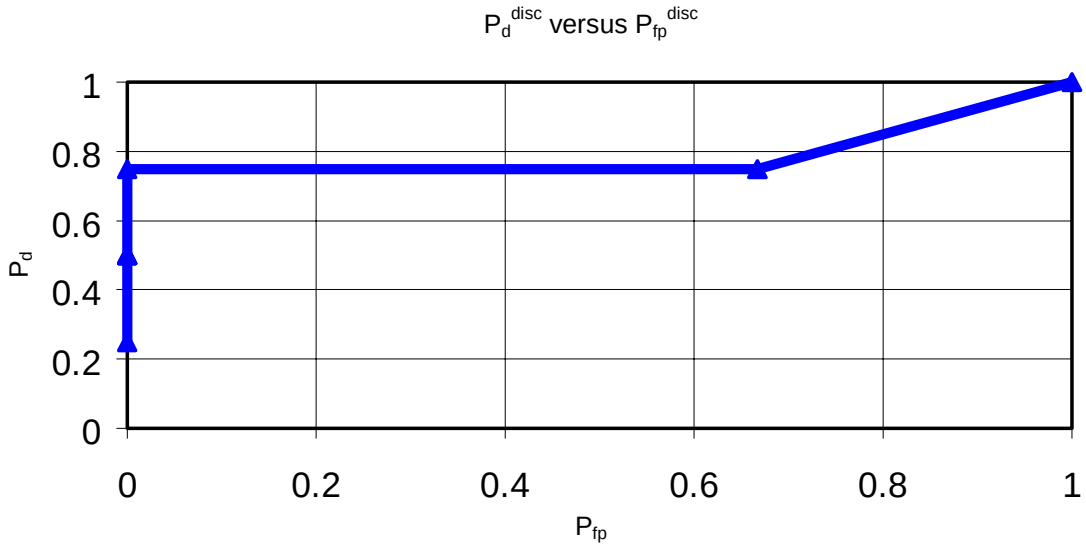


Figure P-5. Blind grid discrimination stage ROC curve.

3.1.4 Other Blind Grid Performance Metrics

- Type Classification: If the demonstrator correctly identifies the item as ordnance, what type is it? The demonstrator will be graded on the ability to correctly identify the type. Table P-8 shows the standardized abbreviations demonstrators must use to classify ordnance. Percent correct will be reported relative to total emplaced ordnance.

TABLE P-8. ABBREVIATIONS FOR TYPE CLASSIFICATIONS

Ordnance	Type	Ordnance	Type
20-mm Projectile	20-mm	40-mm Grenades	40-mm
MK 118 ROCKEYE	MK118	M42 Submunition	M42
BLU-26 Submunition	BLU26	BDU-28 Submunition	BDU28
2.75-in. Rocket	2.75R	57-mm Projectile APC	57-mm
60-mm Mortar	60-mm	81-mm Mortar	81-mm
105-mm Projectile	105-mm P	105-mm Heat Rounds	105-mm
155-mm Projectile	155-mm		

- Location Accuracy: How accurate was the demonstrator in estimating depth? Depth accuracy will be assessed when the demonstrator correctly identifies the detected item as ordnance. Miss distances will be calculated for each ordnance item detected. The average and standard deviation will be reported. To estimate the miss distance for depth the following formula will be used:

  $\text{depth} = \text{depth}_{(\text{demonstrator data})} - \text{depth}_{(\text{ground truth})}$.
  depth = 0.7 5 meters - 0.86 meters = 0.11 meters.

- 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 achieved 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}): $R_{ba} = 1 - [P_{ba}^{disc}(t^{disc})/P_{ba}^{res}(t_{min}^{res})]$); Measures the degree to which the discrimination stage correctly rejects false 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 false alarms initially detected in the response stage were rejected at the specified threshold in the discrimination stage.

- Evaluate P_ds for response and discrimination stage performance by depth (shallow, medium, deep), size (small, medium, large), standard versus non-standard ordnance, and overall at the demonstrator's recommended operating level.

3.2 Open Field.

a. The open field is designed to test entire systems (sensors, algorithms, navigation, and delivery platforms). The demonstrator first must locate potential targets and report the locations in GPS coordinates. The open field scoring is similar to the blind test grid scoring in that the demonstrator's list of response stage and discrimination stage outputs are compared to the ground truth. The data collected by the demonstrator is submitted to the scoring committee in the required format.

b. A computer program compares the demonstrator data to the ground truth database. The program links the ground truth database to the demonstrator's data by matching the northing and easting of the ground truth target items to the demonstrator's data. A halo placed around each target will be used to determine a hit. For all clutter items and ordnance items less than or equal to 0.5 meter in length, a circular halo 0.5 meter in radius will be placed around the center of the object. For ordnance items longer than 0.5 meter, the halo becomes an ellipse where the minor axis remains 0.5 meter and the major axis is equal to the half-length of the ordnance plus 0.5 meter. See Figure P-6 for an example. In order for a demonstrator's position data to be linked with a ground truth target item, both the northing and easting of the point must fall within the halo of the emplaced item. See Figure P-7 for an example.

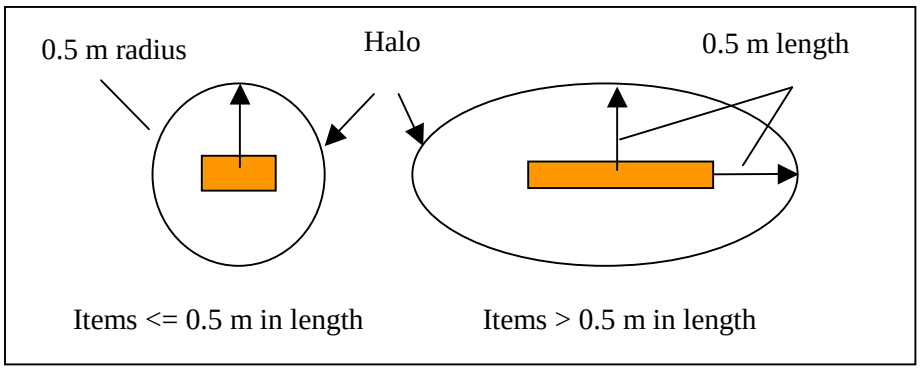


Figure P-6. Target halo.

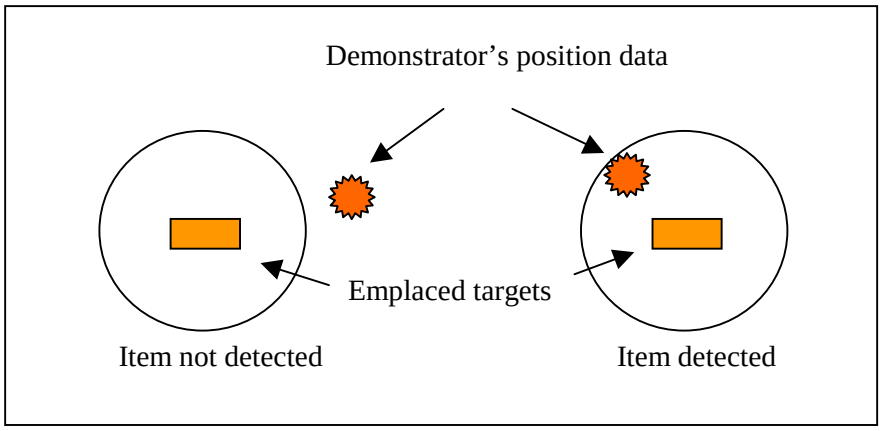


Figure P-7. Position data linked with a target item.

3.2.1 Open Field Scoring Example.

a. Table P-9 is an example of the demonstrator's data in the required format. Table P-10 is an example of hypothetical ground truth data.

TABLE P-9. DEMONSTRATOR DATA FOR OPEN FIELD

	Northing	Easting	Response Stage	Discrimination Stage/ Ranking	Classification	Type	Depth, m	Azimuth, deg	Dip, deg
1	14335954.199	1322776.633	0.649	1	C	-	-	-	-
2	14335810.367	1322622.601	2.852	76	O	20-mm	1.72	142	-69
3	14335769.762	1322340.390	8.209	86	O	105-mm	0.86	16	-52
4	14335723.967	1322372.103	1.757	20	C	-	-	-	-
5	14335667.810	1322381.450	2.491	30	C	-	-	-	-
6	14335714.742	1322349.430	9.980	89	O	105-mm	0.37	140	40
7	14336053.613	1322595.982	12.679	15	O	105-mm	0.28	137	70
8	14335753.177	1322308.994	0.649	40	C	-	-	-	-
9	14335919.549	1322538.777	2.852	52	O	81-mm	1.03	297	- 7

TABLE P-10. HYPOTHETICAL OPEN FIELD GROUND TRUTH DATA

	Northing	Easting	Classification	Type	Depth, m	Azimuth, deg	Dip, deg
101	14335954.399	1322776.533	C	-	-	-	-
102	14335810.267	1322622.501	O	20-mm	1.72	142	-69
103	14335770.862	1322340.290	O	105-mm	0.86	16	-52
104	14335723.867	1322372.153	C	-	-	-	-
105	14335667.910	1322381.300	C	-	-	-	-
106	14335714.842	1322349.533	O	105-mm	0.37	140	40
107	14336053.513	1322595.982	O	105-mm	0.28	137	70
108	14335753.277	1322309.294	C	-	-	-	-
109	14335919.449	1322538.377	O	81-mm	0.96	310	-10

b. Table P-11 is an example of the computer program linking the two data sets for the demonstrator's point 9 only. To establish a link between the two data sets, the distance formula is used to calculate the radial distance, r, between a demonstrator point (dd) and a ground truth location (gt). This equation is:

$$r = ((\text{northing}_{(gt)} - \text{northing}_{(dd)})^2 + (\text{easting}_{(gt)} - \text{easting}_{(dd)})^2)^{1/2}.$$
$$r = ((14335919.449 - 14335919.549)^2 + (1322538.377 - 1322538.777)^2)^{1/2}.$$
$$r = 0.412 \text{ meter.}$$

Since r is less than 0.5 meter, the demonstrator's point 9 will be linked only with this ground truth item. Demonstrator data not linked to items in the ground truth database will be scored as background alarms. Note in the example data sets above that demonstrator's point 3 is

not within 0.5 meter of any item in the ground truth. Demonstrator's point 3 will be scored as a background alarm. In the case of multiple hits within the same halo, the demonstrator data with the strongest signal will be linked to the ground truth item. The remaining hits in that halo will be disregarded.

TABLE P-11. COMPUTER PROGRAM LINKING GPS LOCATIONS

	Northing	Easting	Response Stage	Discrimination Stage/ Ranking	Classification	Type	Depth, m	Azimuth, deg	Dip, deg
Demonstrator's Data (dd)									
9	14335919.549	1322538.777	2.852	52	O	81-mm	1.03	297	- 7
Ground Truth Data (gt)									
109	14335919.449	1322538.377	-	-	O	81-mm	0.96	310	-10

3.2.2 Development of ROC Curves - Response Stage

a. After linking the demonstrator data to the ground truth database the program thresholds the demonstrator response data. The threshold increment is determined in the same way for the open field as for the blind grid (see table P-4 for an example).

b. At each threshold increment the following are calculated:

<u>Response Stage Probability of Detection</u> (P_d^{res}): P_d^{res} = (# of response-stage detections)/(# of emplaced ordnance).

<u>Response Stage Probability of False Positive</u> (P_{fp}^{res}): P_{fp}^{res} = (# of response-stage false positives)/(# of emplaced clutter items).

<u>Response Stage Background Alarm Rate</u> (BAR^{res}): BAR^{res} = (# of response-stage background alarms)/arbitrary constant.

Note that the quantities P_d^{res}, P_{fp}^{res}, and BAR^{res} are functions of t^{res}, the threshold applied to the response-stage signal strength. Table P-12 continues the example and shows how ROC curve quantities are calculated for the sample demonstrator data of Table P-9. (The threshold increment determined in table P-4 applies.) Sample ROC curves are plotted in Figures P-6 and P-7.

TABLE P-12. EXAMPLE OPEN FIELD RESPONSE STAGE ROC CURVE CALCULATIONS

Linked Demo/GT	Ground Truth	Bin No.	1	2	3	4	5	6	7
		Threshold	0.649	2.406	4.812	7.218	9.624	12.03	14.436
			Result = 1 if Response >= Threshold_{bin}; 0 otherwise						
		RESPONSE	RESULT						
1/101	C	0.649	1	0	0	0	0	0	0
2/102	O	2.852	1	1	0	0	0	0	0
4/104	C	1.757	1	0	0	0	0	0	0
5/105	C	2.491	1	1	0	0	0	0	0
6/106	O	9.98	1	1	1	1	1	0	0
7/107	O	12.679	1	1	1	1	1	1	0
8/108	C	0.649	1	0	0	0	0	0	0
9/109	O	2.852	1	1	0	0	0	0	0
NOT Linked									
3/NA		8.209	1	1	1	1	0	0	0
NA/103	O	-	-	-	-	-	-	-	-
Ground Truth Summary:			COUNTS						
# ordnance	5	# detections	4.0	4.0	2.0	2.0	2.0	2.0	1.0
# emplaced		# fp	4.0	1.0	0.0	0.0	0.0	0.0	0.0
Clutter	4	# ba	1.0	1.0	1.0	1.0	0.0	0.0	0.0
		P_d^{res}	0.8	0.8	0.4	0.4	0.4	0.4	0.2
		P_{fp}^{res}	1.0	0.25	0.0	0.0	0.0	0.0	0.0
		BAR^{res}	0.05	0.05	0.05	0.05	0.0	0.0	0.0

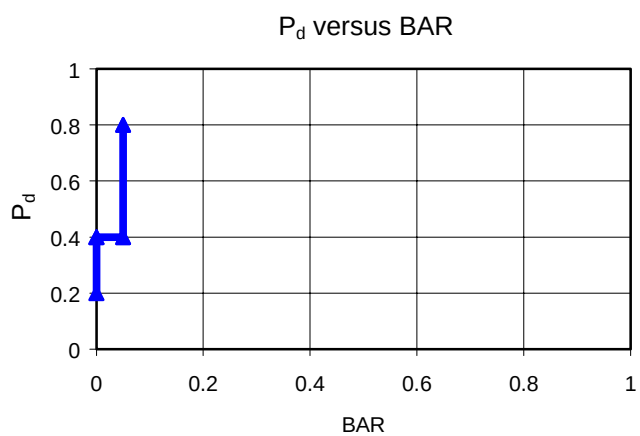


Figure P-6. Open field response stage.

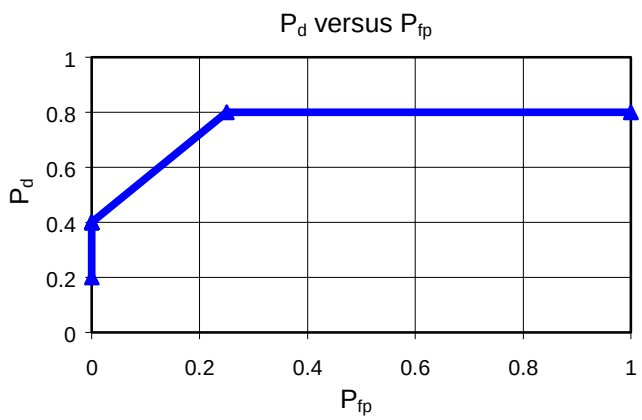


Figure P-7. Open field response stage ROC curve.

c. In order to protect the Standardized UXO Test Site ground truth, probabilities and rates will be reported as rounded figures and ROC curves will be smoothed.

3.2.3 Development of ROC Curves - Discrimination Stage

a. Using the same links between demonstrator and ground truth data sets, the program thresholds the demonstrator's discrimination data. The threshold increment is determined by dividing the range (maximum - minimum) of discrimination stage values into a desired number of increments, independent of the number of increments used in the response stage.

b. At each threshold increment and at the demonstrator's recommended level, the following will be calculated:

<u>Discrimination Stage Probability of Detection</u> (P_d^{disc}): P_d^{disc} = (# of discrimination-stage detections)/(# of emplaced ordnance in the test site).

<u>Discrimination Stage Probability of False Positive</u> (P_{fp}^{disc}): P_{fp}^{disc} = (# of fp^{disc})/(# of emplaced clutter items).

<u>Discrimination Stage Background Alarm Rate</u> (BAR^{disc}): BAR^{disc} = (# of BA^{disc})/(test area).

Note that the quantities P_d^{disc}, P_{fp}^{disc}, and BAR^{disc} are functions of t^{disc}, the threshold applied to the discrimination-stage signal strength. Table P-13 continues the example and shows how ROC curve quantities are calculated for the sample demonstrator data of Table P-9 using ground truth data in Table P-10. Sample ROC curves are plotted in Figures P-8 and P-9.

TABLE 13. EXAMPLE OPEN FIELD RESPONSE STAGE ROC CURVE CALCULATIONS

Linked Demo/GT	Ground Truth	Bin No.	1	2	*	3	4	5	6	7	8
		Threshold	1	14.667	15	29.333	44	58.667	73.333	88	102.67
			Result = 1 if Discrimination >= Threshold_{bin}; 0 otherwise								
		DISCRIMINATION	RESULT								
1/101	C	1	1	0	0	0	0	0	0	0	0
2/102	O	76	1	1	1	1	1	1	1	0	0
4/104	C	20	1	1	1	0	0	0	0	0	0
5/105	C	30	1	1	1	1	0	0	0	0	0
6/106	O	89	1	1	1	1	1	1	1	1	0
7/107	O	15	1	1	1	0	0	0	0	0	0
8/108	C	40	1	1	1	1	0	0	0	0	0
9/109	O	52	1	1	1	1	1	0	0	0	0
NOT Linked											
3/NA		86	1	1	1	1	1	1	1	0	0
NA/103	O	-	-	-	-	-	-	-	-	-	-
Ground Truth Summary:			COUNTS								
# ordnance	5	# detections	4.0	4.0	3.0	2.0	2.0	2.0	2.0	1.0	0.0
# emplaced		# fp	4.0	3.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
clutter	4	# ba	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	0.0
		P_d^{disc}	0.8	0.8	0.6	0.4	0.4	0.4	0.4	0.2	0.0
		P_{fp}^{disc}	1.0	0.75	0.5	0.0	0.0	0.0	0.0	0.0	0.0
		BAR^{disc}	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.0	0.0

* Demonstrator's recommended threshold level.

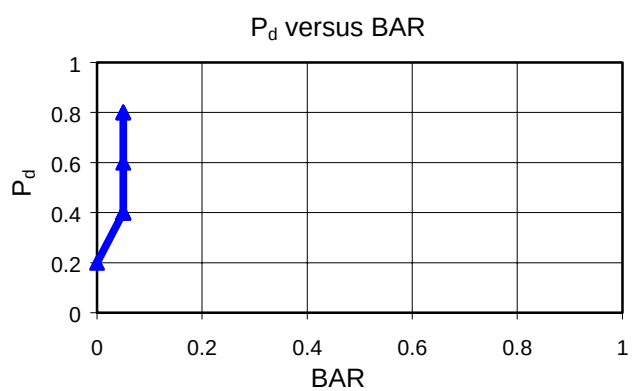


Figure P-8. Open field discrimination stage ROC curve.

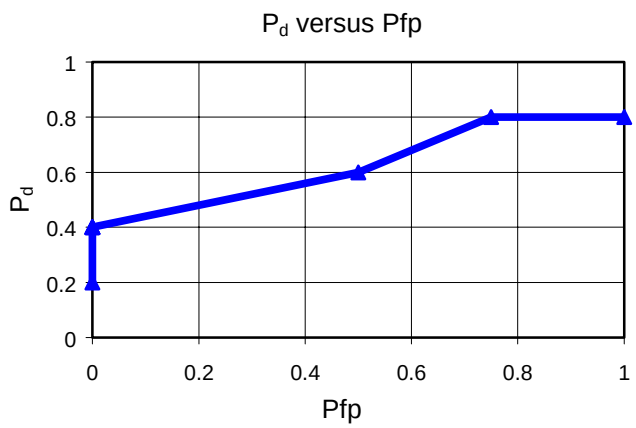


Figure P-9. Open field discrimination stage ROC curve.

3.2.4 Other Open Field Performance Metrics

- Type Classification: If the demonstrator correctly classified the item as ordnance, what type is it? The demonstrator will be graded on the ability to correctly identify the type. Table P-8 shows the abbreviations used to identify ordnance. The demonstrator's point 9 (see Table 11) will be scored as a correct type classification.

- Location Accuracy: How accurate was the demonstrator in estimating depth and ordnance location? Location accuracy will be assessed for those items correctly identified as ordnance. Unidirectional linear miss distances (northing, easting, and depth) will be calculated for each ordnance item detected. The average and standard deviation will be reported for each direction. To calculate miss distance for depth and location accuracy, the following formulas will be used:

  $\text{depth} = \text{depth}_{(dd)} - \text{depth}_{(gt)}$
  depth = 1.03 meter - 0.96 meter = 0.07 meter.

  $\text{northing} = \text{northing}_{(dd)} - \text{northing}_{(gt)}$
  northing = 14335919.449 - 14335919.802 = -0.353 meter.

  $\text{easting} = \text{easting}_{(dd)} - \text{easting}_{(gt)}$
  easting = 1322538.377 + 1322538.602 = -0.232 meter.

- Efficiency (E): $E = P_{det}^{disc}(t^{disc})/P_{det}^{res}(t_{min}^{res})$; 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}.

- <u>False Positive Rejection Rate</u> (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.

- <u>Background Alarm Rejection Rate</u> (R_{BA}): $R_{BA} = 1 - [BAR^{disc}(t^{disc})/BAR^{res}(t_{min}^{res})]$; 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.

- Evaluate P_ds for response and discrimination stage performance by depth (shallow, medium, deep), size (small, medium, large), standard versus nonstandard ordnance, and overall at the demonstrator's recommended operating level. Lower 90 percent binomial confidence limits will be calculated for P_d^{res}, P_{fp}^{res}, P_d^{disc}, and P_{fp}^{disc}.

4. Comparisons Between Test Scenarios

Statistical significance tests are used to compare results between test scenarios (Blind Grid, Open Field, Scenario 1, and Scenario 2) or with past performance. The intent of the comparison is to determine if the feature introduced in each scenario has a degrading effect on the performance of the sensor system. However, any modifications in the UXO sensor system under 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 will be used at a significance level of 0.05 to compare probabilities of detection, probabilities of false positive, efficiency rates, and rejection ratios between test scenarios. Example comparisons between Blind Test Grid and Open Field results are presented in Tables P-14 and P-15.

TABLE P-14. CHI-SQUARE COMPARISON RESULTS, BLIND TEST GRID COMPARED TO OPEN FIELD

	Small	Medium	Large	Overall
P_{det}^{disc}	Sig	Not Sig	Not Sig	Not Sig
P_{fp}^{disc}	Sig	Not Sig	Not Sig	Not Sig

Sig = Significant.

It can be seen from the comparison results that the system's ability to detect small caliber ordnance was significantly reduced in Open Field as compared to its performance in the Blind Test Grid.

TABLE P-15. CHI-SQUARE COMPARISON RESULTS, BLIND TEST GRID COMPARED TO OPEN FIELD

	Small	Medium	Large	Overall
Efficiency	Sig	Not Sig	Not Sig	Not Sig
Rejection Ratio	Sig	Not Sig	Not Sig	Not Sig

The significant difference between efficiencies indicates that detected small caliber items were not correctly identified as ordnance as often in Open Field as they were in the Blind Test Grid.

5. Cost, Time and Manpower Requirements.

Cost, time and manpower requirements will capture field labor costs associated with conducting a field survey. In order for standardized labor costs to be applied among all demonstrators, set labor rates have been established for three labor categories: supervisor, data analyst, and field support. Labor costs will be calculated as follows: The first person at the test site will be designated supervisor; the second person will be designated data analyst; third and following personnel will each be regarded as field support. Labor will be monitored by government on-site representatives on a man-hour basis and will be converted to cost by the application of standardized labor rates as presented in Table P-16. Labor costs include time associated with setup and calibration of equipment, surveying blind test grid, open field and/or scenarios, and demobilization. The labor costs associated with the calibration lanes will not be recorded.

TABLE P-16. LABOR COSTS

Setup and Calibration of Equipment	No. People	Hourly Wage	Hours	Cost
Supervisor	1	$95.00	16	$ 1,520
Data Analyst	1	57.00	16	912
Field Support	4	28.50	16	1,824
Subtotal	6	-	96	4,256
Blind Test Grid				
Supervisor	1	95.00	10	950
Data Analyst	1	57.00	10	570
Field Support	4	28.50	10	1,140
Subtotal	6	-	60	2,660
Open Field				
Supervisor	1	95.00	24	2,280
Data Analyst	1	57.00	24	1,368
Field Support	3	28.50	24	2,052
Subtotal	5	-	120	5,700
Demobilization				
Supervisor	1	95.00	6	570
Data Analyst	0	57.00	0	0
Field Support	3	28.50	6	513
Subtotal	4	-	24	1,083
Total	-	-	300	$13,699