IDA

INSTITUTE FOR DEFENSE ANALYSES

Interpreting Results from the Standardized UXO Test Sites

Michael May
Michael Tuley

January 2007

IDA Document D-3280
Log: H 06-000819

This work was conducted under contracts DASW01 98 C 0067/ W74V8H 05 C 0042, Task AM-2-1528, for the ESTCP/SERDP. The publication of this IDA document does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official position of the Agency.

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-3280

Interpreting Results from the Standardized UXO Test Sites

Michael May
Michael Tuley

PREFACE

This document was prepared for the Environmental Security Technology Program in response to a task titled “ESTCP/SERDP: Assessment of Traditional and Emerging Approaches to the Detection and Identification of Surface and Buried Unexploded Ordnance.”

CONTENTS

EXECUTIVE SUMMARY

INTRODUCTION

The Institute for Defense Analyses (IDA) was tasked by the Strategic Environmental Research and Development Program and the Environmental Security Technology Certification Program to complete a detailed analysis of the results of testing carried out at the Standardized Unexploded Ordnance (UXO) Test Sites. The major purpose of this tasking was to provide data for an Interstate Technology and Regulatory Council report on the status of UXO detection and discrimination technology.* This IDA document provides an overview of the sites, discusses standard data analysis products provided by the U.S. Army Environmental Center, and describes in detail the analysis undertaken by IDA and the results of that analysis.

The focus of the analysis was to provide a data-driven understanding of the performance of sensors currently in widespread use in UXO clearance actions. The method employed was to start with the Army's standard analysis and then apply selective data filters that excluded certain targets to illustrate limitations on performance. As expected, scores improved when munitions that were buried very deeply or clustered with other targets were excluded from the analysis.

The demonstrated sensors could be separated into "good" and "poor" categories. But note that even after filters were applied, the good performers did not achieve probabilities of detection (Pds) of 100% for all munition types. We analyzed data from individual munition items emplaced at the Standardized UXO Test Sites to provide an understanding of why those items were missed. Note that the distribution of clutter, munition types, and their depths at the Standardized Sites is designed to replicate a variety of real-world encounters. The Standardized Sites do not replicate any particular real-world site.

---

* *Survey of Munitions Response Technologies,* SERDP, ESTCP, ITRC, 2006.

PROBABLITIES OF DETECTION AND BACKGOURND ALARM RATES

Figures ES-1 and ES-2 plot the Pds and relative background alarm rates† (rBAR) for the Open Field portion of the two test sites, Aberdeen Proving Ground and Yuma Proving Ground. The good demonstrators occupy the high-Pd and low-rBAR region of the graph. Three Pd scores are reported for each demonstrator. The lowest is the Pd considering all munition targets buried at the site. The middle Pd score applies a filter that ignores targets that were not able to be surveyed or were part of clusters. Generally, using this filter results in a Pd about 5% greater than scoring against all targets. In instances where a large portion of the site could not be surveyed (e.g., it was flooded), this increase was much larger. The highest Pd plotted applies the additional restriction that only targets above (shallower than) 11 times the munition's diameter are considered. This depth, known as the "11× depth," is a rule-of-thumb guideline from the U.S. Army Corps of Engineers that estimates a munition's envelope of detectability.

Although most of the good demonstrators used electromagnetic induction sensors, variations in the way sensors were implemented and the relative number of types being tested make identifying a preferred technology impossible. Variations in performance are seen even among like technologies when implemented and operated differently.

To arrive at meaningful conclusions, we closely examined the good demonstrators, assuming that they implemented and operated sensors in a near-optimum way. Rather than identifying the best technologies, we identified difficult targets and universal factors that enabled optimal performance.

DIFFICULT TARGETS

The targets that were missed by the best demonstrators after the filters were applied come predominantly from two categories: "shadowing" and "halo effect." Shadowing is simply a large target obscuring a smaller nearby signal of interest. The halo effect is a literally a near miss. A hit was defined by declaring an alarm within a fixed distance of a buried munition. When this distance was barely exceeded, a halo-effect miss occurred.

It can be argued that items missed from either shadowing or the halo effect could actually be hits. During a real-world UXO clearance, the excavation team might find the target that was called a miss by the test-site scoring system. Note that this type of find

† The relative rate is reported to protect the ground truth since the test sites are still active.

does not guarantee that the item would be detected if it were isolated. The test sites do not test this issue, so once the filters are applied, some uncertainty remains about what would actually be found in a real clearance action.

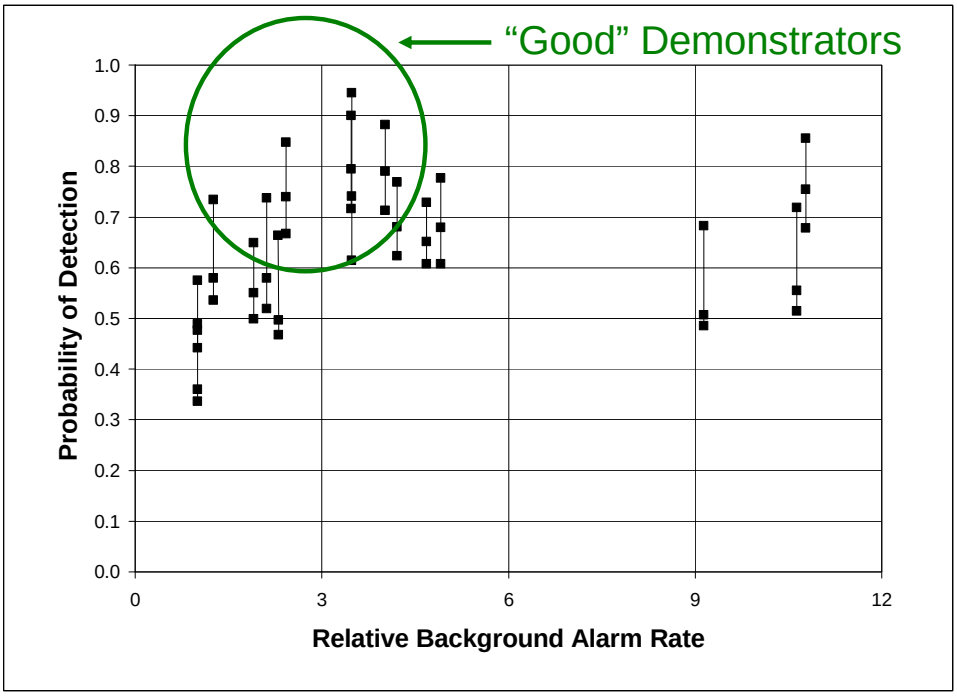

Figure ES-1. Aberdeen Proving Ground Pd vs. rBAR rate.

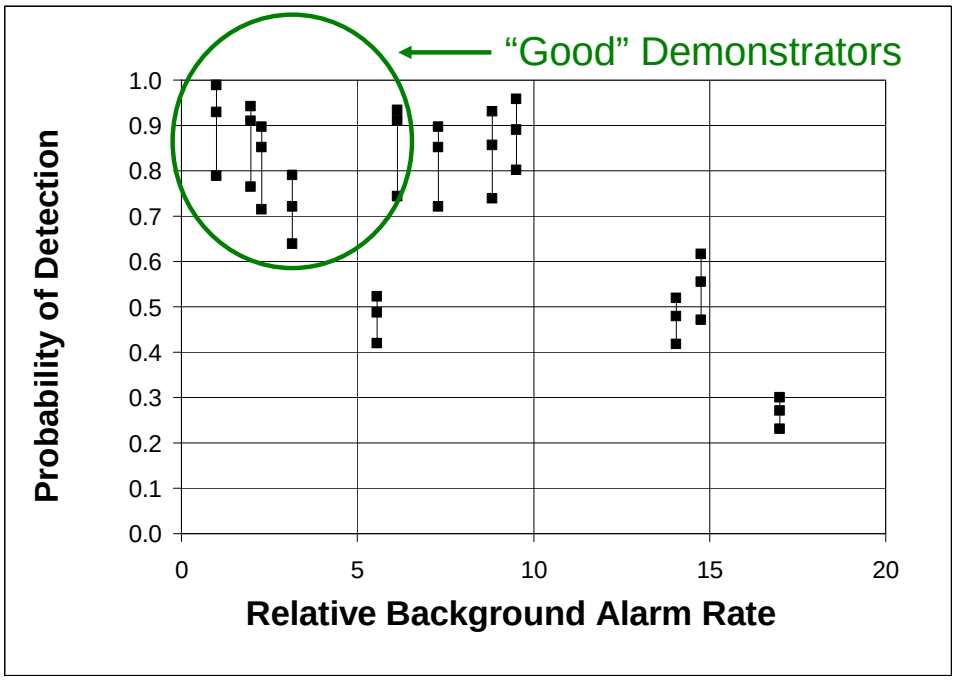

Figure ES-2. Yuma Proving Ground Pd vs. rBAR rate.

FACTORS INFLUENCING OPTIMAL PERFORMANCE

One of the factors that influences performance is good site coverage with sufficient data density to detect and localize targets. This is especially true when trying to find small targets like 20 mm projectiles, where the horizontal extent of the small amplitude signal may be much less than a meter.

Demonstrators that fell into the poor category might have made quality-control errors or signal-processing errors. At least one demonstrator had a sensor with an abnormally high noise level that persisted from Aberdeen Proving Ground to Yuma Proving Ground. At Yuma, the demonstrator declared a few hundred more background alarms than other demonstrators with similar technology (in Figure ES-2 they are off the scale). Other demonstrators had a low number of background alarms, but also a much lower Pd than similarly equipped demonstrators. The most likely reason for these scores is a detection threshold set too high.

CONCLUSIONS

Despite the uncertainty in Pd due to shadowing and the halo effect, the detailed analysis of the Pds by munition type indicates that targets larger than a 60 mm mortar and above the 11× depth should be found greater than 90% of the time. With optimum data analysis and site coverage, this percentage should be nearer to 100% than 90%. For smaller targets, especially as small as a 20 mm projectile, it is not clear that this percentage will exceed 90% without a search designed particularly for finding small targets. While it would be of great value to regulators and stakeholders in UXO cleanup actions to precisely specify the deviation from 100% detection expected in a particular cleanup scenario, the Standardized UXO Test Site results do not provide such precision. Given the number of identically buried, like-type munitions required to make very precise Pd estimates and the number of possible depth and location configurations, it is difficult to envision a practical test site that probes universal variables of the UXO detection problem with great precision.

Note that the characteristics of targets at the Standardized Sites (type, burial depth, angle of emplacement, etc.) do not reflect a particular real-world site. The Standardized Sites attempt to recreate a variety of expected individual encounters and stressing encounters (e.g., deeply buried targets). As a whole, each site is an amalgamation of these encounters, and overall scores at the Standardized Sites are a function of the choices made for parameters such as burial depth, munition types, etc.

I. INTRODUCTION

BACKGROUND AND PURPOSE

The Institute for Defense Analyses (IDA) was tasked by the Environmental Security Technology Certification Program (ESTCP) and the Strategic Environmental Research and Development Program (SERDP) to complete a detailed analysis of the results of testing carried out at the Standardized Unexploded Ordnance (UXO) Test Sites. The major purpose of this tasking was to provide data for an Interstate Technology and Regulatory Council (ITRC), ESTCP, SERDP report on the status of UXO detection and discrimination technology (Reference 1). Much of the description of the Standardized Sites that appears in this document was prepared for the that document, which also contains selected portions of the complete data analysis reported here. This IDA document provides an overview of the sites, discusses standard data-analysis products provided by the U.S. Army Environmental Center (AEC), and describes in detail the analysis undertaken by IDA along with the results of that analysis.

Based on historical data collected in preparation of the ITRC, ESTCP, SERDP report and from the instruments used in demonstrations at the Standardized Sites, two general classes of sensors are predominant in production UXO surveys: magnetometers and electromagnetic induction (EMI) instruments. This analysis evaluates the effectiveness of representative types of the two sensors when operated correctly in the field.

The focus of the analysis was to provide a data-driven understanding of the performance of sensors currently in widespread use in UXO clearance actions. The method employed was to start with the Army's standard analysis and then apply selective data filters that highlighted limitations on performance. After a series of such filters were applied, demonstrators clearly fell into two broad classes, "good" and "poor" performers. Even the good performers did not achieve probabilities of detection (Pds) of unity, so additional analysis was performed on data from individual munition items to provide an understanding of why those items were missed.

Section II of this paper presents a brief overview of the most common technologies studied and provides a description of the Standardized UXO Test Sites. The

scoring methodology used by the AEC and the aspects that are in common with the IDA analysis are discussed in Section III. The overall Pd results are in Section IV, which also describes how the IDA analysis differs from the AEC analysis and the motivation for this approach. Section V examines individual items that were missed and explores the implications of these misses for overall results.

II. OVERVIEW OF SENSORS AND THE TEST SITES

A. CESIUM VAPOR MAGNETOMETERS

Cesium-vapor magnetometers and EMI sensors were the most commonly employed sensors at the test sites. They are also the most common sensors employed in real-world UXO cleanup activities (Reference 1).

Magnetometers are designed to detect static magnetic fields, and they detect the presence of ferrous objects by detecting perturbations in the Earth's magnetic field caused by those objects. Cesium-vapor magnetometers (Figure II-1) are the dominant type of magnetometer demonstrated at the Standardized Sites. Cesium-vapor magnetometers are lightweight, sensitive (fundamental sensitivities of the order of 5 $pT/Hz^{1/2}$), provide a rapid data-collection capability, and can be easily arrayed. These total-field magnetometers are unable to provide vector information.

Cesium-vapor magnetometers make use of the Zeeman effect, in which an ambient magnetic field splits the fine energy levels of the valence electron in a cesium atom. The energy difference between the two levels, where the electron's spin moment is either aligned with the magnetic field or opposes it, is proportional to the strength of the externally applied magnetic field. The cesium-vapor magnetometer measures the RF frequency required to pump the electron from the lower energy level to the higher, which will vary as the magnetometer encounters perturbations in Earth's field. This frequency gives the difference in energy and hence the magnitude of the external field.

B. ELECTROMAGNETIC INDUCTION SENSORS

Unlike magnetometers, EMI sensors are active. They operate by generating a time-varying electromagnetic field, generally using a coil excited by either a pulsed waveform or a sine wave. That transmitted field induces a secondary field in conducting objects that intersects the field lines. The secondary magnetic field can be intersected by a receiving sensor (generally a coil) that provides an indication of the presence of a conducting object. See Figure II-2.

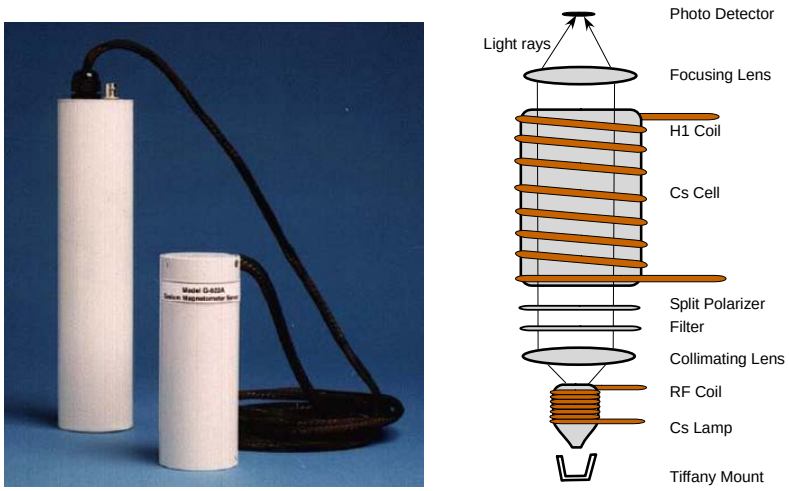

Figure II-1. G-858 Cesium-Vapor Magnetometers.

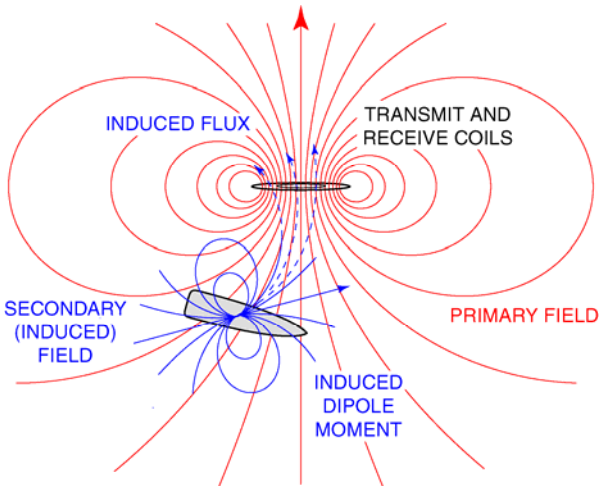

Figure II-2. EMI physics.

Time-domain electromagnetic systems measure the response of the subsurface to a pulsed electromagnetic field as a function of time. Frequency-domain electromagnetic systems measure the response of the subsurface as a function of the frequency of the sensor output. In more advanced instruments, measurements can be made in multiple time gates (time-domain electromagnetic systems) and multiple frequencies (frequency-domain electromagnetic systems), which can increase the information obtained about the physical properties of the targets.

The basic operating principle of time-domain electromagnetic induction involves the use of a wire-loop transmitter carrying a pulsed current that produces a transient magnetic field that propagates into the earth. The magnitude and rate of decay of the fields depend on the electrical properties and geometry of the medium and any subsurface

objects. The time-domain electromagnetic receiver measures the secondary magnetic fields created as a result of the incident magnetic field, which produces eddy currents in the subsurface geology and buried conductive objects. The currents in the conductive earth typically decay at a more rapid rate than the currents in metallic objects. Measurements are made in discrete “time gates,” or time intervals, following the turnoff of the current pulse generated by the transmitter. The early time gates will detect both small and large metallic targets with short and long decay rates, respectively, while the late time gates will detect only larger targets with relatively long response decays. An example of a time-domain electromagnetic sensor examined in this study is the Geonics EM61MKII, which has four time gates spaced between 216 μs and 1.27 ms.

The basic operating principle of the frequency-domain electromagnetic induction method involves a transmitter coil radiating a continuous-wave electromagnetic field at one or more selected frequencies, which induces an electrical current (eddy current) in the earth and subsurface objects. These eddy currents in turn generate a secondary magnetic field. The receiver coil detects and measures this secondary field. The instrument output is obtained by comparing the strength of the secondary field to the strength of the primary field. An example of an frequency-domain electromagnetic instrument analyzed in this effort is the Geophex GEM-3, which can be programmed to transmit up to 10 frequencies, typically covering ~100 Hz to ~24 kHz.

The two domains are capable of producing theoretically equivalent results, but practical implementation issues often result in differences in performance. Time-domain instruments have the advantage of having no transmitted fields present when the response from objects under the earth is being measured; this reduces the dynamic range required of the sensor and in theory improves sensitivity. However, transmit fields do not decay instantaneously after turnoff. This limits the earliest time that received fields may be sampled. The exponential falloff of the received field strength in late time limits how far in time (or equivalently, how low in frequency) useful samples may be obtained. Frequency-domain systems have the advantages of placing all their energy at selected frequencies and continuously exciting a response from targets. However, the reception of a very small receive signal in the presence of a very large transmit signal limits ultimate performance. Bucking coils are often used to attempt to cancel the transmit field component in the receive coil, but this limits receive coil area to a fraction of the transmit coil area and reduces sensitivity. In addition, the useful low-frequency limit on survey instruments is set by motion-induced noise in the receive coil and available integration time.

C. STANDARDIZED UXO DEMONSTRATION SITES

The Standardized Sites were designed as a facility where blind testing could be conducted to provide system performance assessments under realistic conditions. Recognizing the need for release of sufficient target data for demonstrators to understand their system performance, the Army, ESTCP, and SERDP have designed a plan for regular, partial reconfiguration of the sites so that limited amounts of ground truth can be released. IDA received and used the full ground truth in carrying out these analyses, but continuing blind testing limits the details that can be provided in this report about analysis based on portions of the two sites that have not been reconfigured.

It is critical to remember that while the Standardized Sites do contain realistic challenges, the types, relative number, and placement of targets were designed to sample a wide variety of possible real-world sites. Further, the depth distributions were chosen to include challenging targets, and the ratio of clutter to intact munitions is much larger on real-world sites. Aggregate results from the Standardized Sites should not be interpreted as indicative of expected results from a particular (probably very different) real-world site. The value of the Standardized Sites lies in understanding each type of encounter separately.

The Army, in cooperation with ESTCP and SERDP, set up two Standardized Sites for UXO detection and discrimination technology demonstrations. A third, a shallow-water site, has been set up but is not the subject of this analysis. The sites are located at Aberdeen Proving Ground (APG) in Maryland and at Yuma Proving Ground (YPG) in Arizona. To satisfy both the research-and-development community and the technology-demonstration community, the Standardized Sites are made up of three areas, a Calibration Lane, a Blind Test Grid, and a variety of operational challenges. Figure II-3 is an aerial photograph of the APG Standardized Site. Figures II-4 and II-5 show a portion of the Open Field at each site.

The Calibration Lane allows demonstrators to test their equipment, build a site library, document signal strength, and deal with site-specific variables. The calibration portion of the test site contains munitions identical to those buried at other portions of the site and symmetric clutter. These items are buried at known locations, in various orientations, and at three different depths.

The Blind Test Grid allows the demonstrator to operate the sensor system without platform, coordinate system, or operational concerns. The Blind Grid is similar to the Calibration Lane: the demonstrator knows the possible location of targets, but not

whether a target is present, the target depth, or the target type. The operational challenges include a flat, open area (the Open Field), a wooded area (APG only), and an area of rough terrain (both sites). These challenges document the performance of the entire system in conditions similar to actual range operations. The demonstrator does not know the number, type, or location of munitions and clutter that are emplaced. The challenges provide the demonstrator with a variety of realistic scenarios essential for evaluating overall sensor system performance.[1]

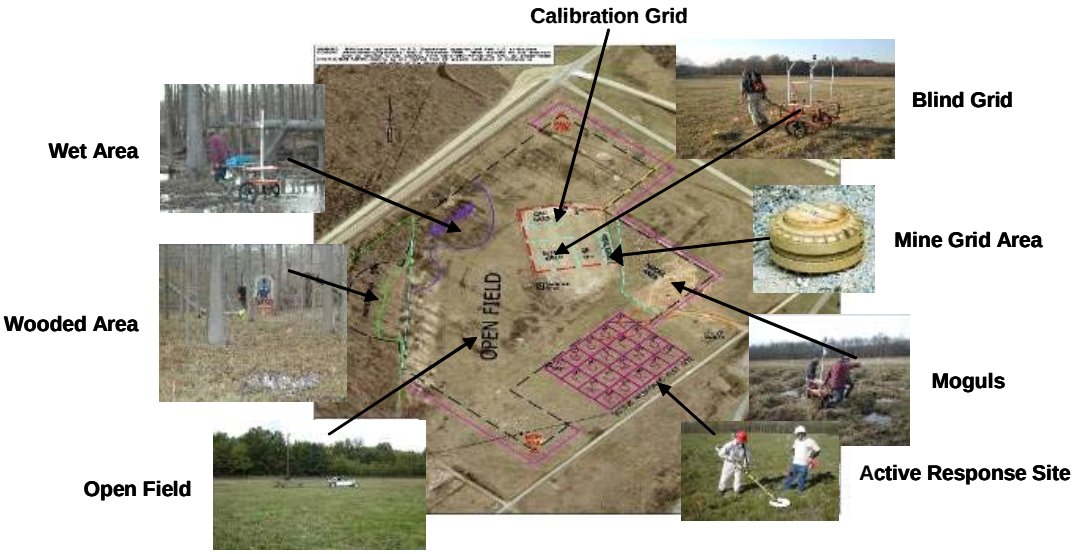

Figure II-3. Aerial photograph of the APG Standardized Site. Total size is about 18 acres.

The distribution of targets at the Standardized Demonstration Sites is designed to replicate a variety of encounters in the field. Burial depths are based on the UXO Recovery Database created by the U.S. Army Corps of Engineers. However, because the Standardized Sites are designed to assess the limits of technology, some targets are deeper than might be expected in normal clearance operations. This depth distribution allowed evaluation of the capabilities of sensors to detect UXO down to the Corps of Engineers rule-of-thumb depth of 11 times the ordnance diameter (the "11× depth")

---

1 For more information, see the Standardized Site Web site, http://aec.army.mil/usaec/technology/uxo03.html.

Figure II-4. A pushcart operating at the APG Open Field.

Figure II-5. A towed array operating at the YPG Open Field.

The targets emplaced at APG and YPG consist of standard targets (Table II-1), nonstandard targets, and clutter. The targets are degaussed before emplacement, although there is anecdotal evidence that this process may not have been 100% effective or that some targets may have reacquired a magnetic moment after degaussing. Nonstandard ordnance items are those that differ from standard ordnance: they may be damaged,

deformed, or from a different subclass of ordnance than the standard set. Emplaced clutter is selected to mimic the types of clutter found on ranges: nails, soda cans, range debris, UXO fragments, etc. The physical properties of each clutter item and nonstandard ordnance are recorded, a photograph taken, and the objects buried. The UXO in the table are grouped according to size as designated by the AEC.

Table II-1. Standardized targets at YPG and APG

Type	Description	Length (mm)	Width (mm)	Aspect Ratio	Weight (lbs)	Size
20 mm	20 mm M55	75	20	3.75	0.25	Small
40 mm	40 mm MK II	179	40	4.48	1.55	Small
40 mm	40 mm M385	80	40	2.00	0.55	Small
M42	Submunition	62	40	1.55	0.35	Small
M75*	Submunition	69	64	0.93	1.19	Small
BLU-26	Submunition	66	66	1.00	0.95	Small
BDU-28	Submunition	97	67	1.45	1.70	Small
57 mm	57 mm M86	170	57	2.98	6.00	Medium
MK118	MK118 ROCKEYE	344	50	6.88	1.35	Medium
60 mm	60 mm M49A3	243	60	4.05	2.90	Medium
81 mm	81 mm M374	480	81	5.93	8.75	Medium
M230	2.75-inch Rocket	328	70	4.69	9.41	Medium
105 mm	M456 HEAT RD	640	105	6.10	19.65	Large
105 mm	105 mm M60	426	105	4.06	28.35	Large
155 mm	155 mm M483A1	803	155	5.18	56.45	Large

* The M75 can also be described as an air-delivered “grenade.” No 37 mm projectiles were present when data was taken for this analysis.

Great effort is made to accurately bury the targets. Holes are dug to a base depth with either a two-man auger or a vehicle-mounted one. With the two-man auger, a positioning template, depth gauge, and dip protractor are used to measure the target’s position, depth (from the local surface to the center of the item), and orientation before covering the target. The significant difference with the vehicle-mounted auger is that diagonal penetration holes can be dug. The emplacement crew calculates the angle and depth necessary to emplace the target according to the range plan. The angle and depth are double-checked with a rod, and then the target is inserted into the hole.

III. TEST SITE SCORING METHODOLOGY

The standard scoring for the test sites is based on a two-stage evaluation of system performance. The first stage, called the "Response Stage," is designed to test only the ability of a system to *detect* anomalies, whether they are ordnance or clutter. The second stage is the Discrimination Stage. It tests whether the system, having detected an item, can differentiate between clutter and UXO. The IDA analysis in this document only uses Response Stage information to calculate probabilities of detection. The metrics reported for this stage depend upon whether the Blind Grid or Open Field areas are being considered. The scoring protocols and the terms used in scoring were the result of a cooperative effort among the Army and ESTCP/SERDP, with input solicited from organizations, such as IDA, that had been heavily involved in scoring UXO and countermine testing. More complete descriptions of the scoring methodologies are found in documents on the Standardized Site Web site.

A. BLIND GRID SCORING

In the Blind Grid, buried items appear at the center of a grid square, so there is no navigation uncertainty. The only uncertainty is in the presence or absence of an object in the square and, if an object is present, whether it is a munition item or clutter. In the Response Stage, three measures of performance are defined for the Blind Grid:

- Response Stage probability of detection (Pd^{res})—the number of Response Stage "detections" (grid squares declared to contain an object that actually do contain a munition) divided by the number of emplaced munitions
- Response Stage probability of false positive (Pfp^{res})—the number of Response Stage "false positives" (grid squares declared to contain an object that actually contain clutter) divided by the number of emplaced clutter items
- Response Stage probability of background alarm (Pba^{res})—the number of Response Stage "background alarms" (grid squares declared to contain an object that are actually empty) divided by the number of empty grid locations

Demonstrators provide AEC with a list of sensor outputs for each of the grid squares, generally ordered in the Response Stage from strongest to weakest. The values of the three metrics depend on the threshold chosen, below which signals are ignored. From this list, a receiver operating characteristics (ROC) curve that plots Pd^{res} vs. Pba^{res} as a

function of threshold can be calculated. Figure III-1 provides a notional ROC curve. The left edge of the curve represents those squares with the highest response, hence highest threshold; the upper right corner is for the lowest threshold. This curve always begins at (0,0) and ends at (1,1). The ideal curve would rise vertically from the origin to the point (0,1) and extend horizontally to (1,1), since in that case, all UXO appears with higher responses than any empty grid squares. A similar curve can be generated with Pfp^{res} as the abscissa. However, it is not particularly meaningful in the Response Stage, except insofar as it indicates the relative strength of response from emplaced ordnance vs. emplaced clutter items. This quantity can be used as a reference in the Discrimination Stage.

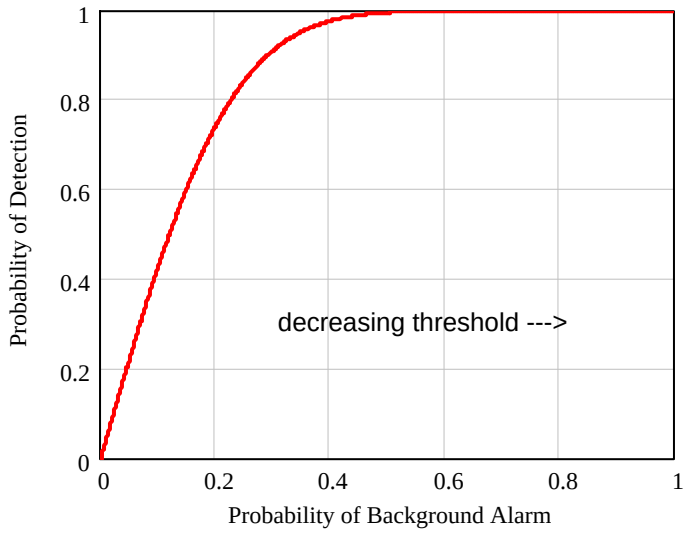

Figure III-1. Example Receiver Operating Characteristics Curve.

For the Discrimination Stage, the relevant metrics are Pd^{disc}, Pfp^{disc} and Pba^{disc}, calculated in the same manner as the Response Stage metrics. For the Discrimination Stage, demonstrators submit a different list, ordered in likelihood of the response being from UXO, from most likely to least likely. ROC curves are generated for the Discrimination Stage, but in this case, the ROC plotting Pd^{disc} vs. Pfp^{disc} is of more interest because the major purpose of discrimination is to separate ordnance that would have to be dug from scrap that can be left in the ground. Additional measures of performance calculated in the discrimination stage include efficiency, false-positive rejection rate, and background-alarm rejection rate. These metrics were not used in the IDA analysis, and they are not discussed further here (see the Standardized Site Web page for more information). In this analysis, "Pd" denotes the response stage value, Pd^{res}.

B. OPEN FIELD SCORING

Open Field scoring employs similar, but not identical, metrics to the Blind Grid scoring. Pd and Pfp are calculated exactly as they are in the Blind Grid, but Pba is no longer a good measure of background alarms because there is not a finite number of locations where detection calls can be made. Therefore, Pba is replaced with background alarm rate (BAR), which is normally defined as the number of background alarms divided by the area surveyed. For Standard Site reporting, BAR is normalized by an arbitrary constant to protect ground truth. In this study, the lowest number of false alarms reported at the site is used to normalize the BAR.

In the Open Field, a set of rules specifies which detection declarations are to be associated with munitions, clutter, or background alarms. Each munition and clutter object is assumed to have a halo around it. If a declaration falls within the halo, that UXO item or clutter item is declared detected. If a declaration is outside the halo, a background alarm is declared. For the Standardized Sites, the nominal halo radius is 0.5 m. For UXO in Table II-1 classified as small or medium, a circle of radius 0.5 m is drawn, with the center at the center of the munition item. For large UXO, defined as those items whose length exceeds 0.5 m, the halo is an ellipse (Figure III-2). The semi-minor axis of the ellipse is 0.5 m, but the semi-major axis is determined by the projected length of the UXO as seen from the surface:

$$\text{Halo semi-major axis} = \frac{L_{\text{UXO}}}{2}\text{Cos}\,\theta + \frac{1}{2},$$

where L_{uxo} is the length of the ordnance item and θ is the dip angle from the horizontal.

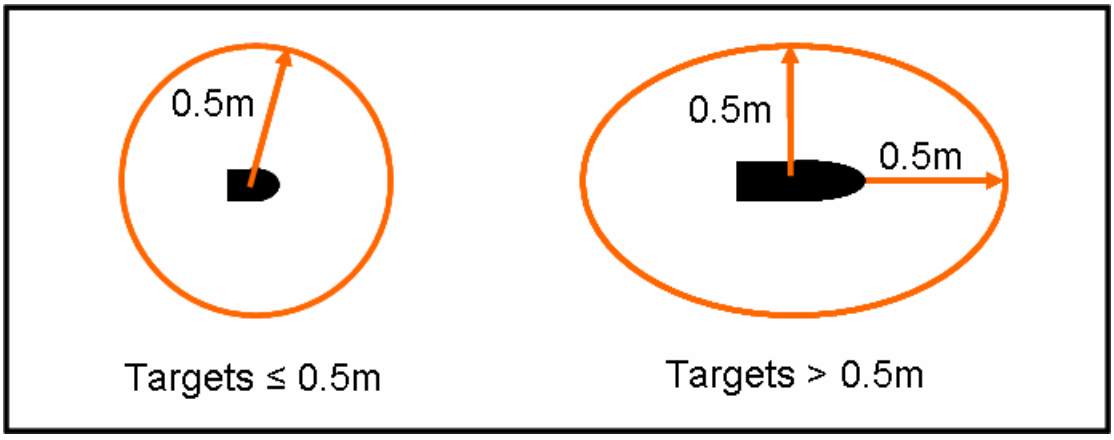

Figure III-2. Illustration of the scoring halo for targets less than and greater than 0.5 m in length.

Determining whether a demonstrator's detection call is associated with an isolated UXO or clutter item depends on whether it lies on or within the halo, as shown in Figure III-3. The Army and IDA have independently developed scoring software that makes the determination, and the results have been compared to ensure that they match.

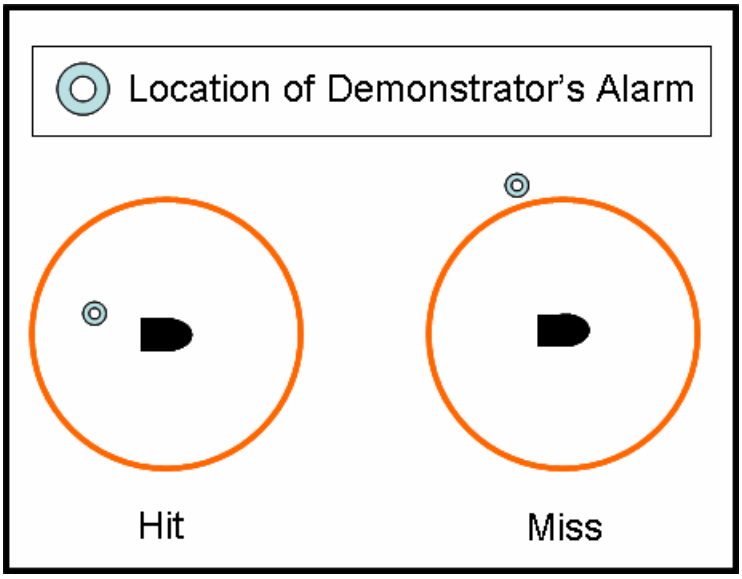

Figure III-3. Illustration of demonstrator detection calls lying inside and outside the scoring halo.

For targets that are not isolated (i.e., where target halos overlap) or for multiple demonstrator detection calls that lie inside a single halo, a procedure to disambiguate calls was agreed to by the Army and ESTCP/SERDP and codified by IDA (see appendix, Section C (p. A-17)). The protocol allows only one target to be associated with a single detection call.

The standardized scoring system is rigid and treats each demonstrator equally; human judgment does not enter into the scoring. While unambiguous and fair, the procedure does not fully assess sensor system performance in certain cases. These cases are discussed in detail in Section V. An example of a shortcoming is a region where target signatures overlap. In fact, there are several large clusters of ordnance at the Standardized Sites where many signatures combine to make, essentially, one large target. It was the desire to understand fundamental sensor capabilities and to analyze the factors on which performance depends that drove the IDA analysis described in this report. One component of that analysis is to evaluate these clusters separately. Of course, human judgment enters since we need to define what a "cluster" is.

IV. IDA ANALYSIS

A. IDA METHODOLOGY AND OVERALL RESULTS

IDA's analysis of data from the Standardized UXO Test Sites differs from AEC's standardized scoring system primarily because the IDA analysis is not *standardized*: the IDA analysis allows for scoring demonstrators, using separate criteria. The intent is to gain a better understanding of each sensor system as it is intended to be used. The IDA analysis excludes clutter and munitions that were inaccessible to a particular demonstrator on a case-by-case basis. It also separately considers the performance of the best scoring demonstrators to learn the capabilities of technologies with "best practices" implementation. The IDA analysis encompasses data from 2002 through early 2005, before the 2005 reconfiguration of the ground truth at APG.

The IDA analysis also considers site-wide factors that are not considered by AEC: the effect of clusters of targets[2] and Pd as a function of the depth-to-diameter ratio.

Aggregate measures such as Pd were calculated by comparing the location of buried munitions in the ground truth to a list of suspected locations of possible UXO provided by each demonstrator. The suspected locations were those reported during the Response Stage of the standardized analysis. Pd's for different target sets were generated by flagging the items in each list as desired. The detailed investigation of individual targets used processed data from the demonstrators' sensor databases visualized in Oasis montaj.[3] The demonstrators' databases included geo-referenced sensor outputs. Geo-referenced databases were not available for all demonstrators. Some were not reported, while others did not record digital data.

The analysis found that the detection of munitions about the size of a 60 mm mortar or larger was not generally limited by the signal strength down to the 11× depth. The Pd's reported in the Standardized Site analysis rarely exceeded 80%, but these

---

2 A "target" is any munition or clutter item buried at a Standardized UXO Test Site for the purpose of testing UXO detection technology. The definition does not include accidentally occurring items or natural background.

3 Oasis montaj is a data analysis and visualzation software suite produced by Geosoft Inc.

standardized results included targets deeper than 11× or that were in clusters. Some of the misses generated at the response stage of the standardized analysis were for mundane reasons (see Section V for details). The standardized results are further prejudiced by the target distributions, which included significant numbers of 20 mm and 40 mm projectiles. These small targets were difficult to find, even for the better demonstrators.

Detectability of targets larger than about 60 mm persisted beyond the 11× depth. In the case of the largest munitions, the deepest seeded targets could be found down to the deepest seeded depth. When only considering the best performers measured against isolated targets larger than 20 mm projectiles, misses of targets shallower than the 11× depth were almost always explained by deficiencies in survey procedures.

The results from the Standardized Sites must be put in context. The fraction of targets found by a sensor system at a site is driven by the relative number of difficult and easy targets. Real-world sites may contain only a few kinds of munitions with similar impact parameters. For example, an antitank range may have been used to fire several types of large munitions that were not expected to deeply penetrate the ground. The probability of detecting unexploded rounds at such a site would be very high. If a real-world site does contain difficult targets, expectations for detection should reflect the limits of the sensor. The Standardized Sites give information about each particular encounter, and general expectations for a real-world site can be built from understanding these encounters.

B. SCORING METRICS FOR THE IDA ANALYSIS

1. Scoring Halo for the Open Field

The fundamental scoring tool is the “scoring halo.” IDA used a C++ based scoring program to compare the surveyed locations of targets and the locations of alarms (suspected targets) reported by demonstrators. Targets within the scoring halo were called “hits” (or “found”). Those outside the scoring halo were called “misses.” The IDA analysis focused only on munition-type targets, and not clutter, by discarding all hits on clutter targets before calculating Pd and BAR. More precisely, alarms that were matched to clutter items did not contribute to the background-alarm rate; background alarms were only those alarms that were not within the halo of *any* target, including clutter.

Adjustments to the background-alarm rate and accidental hits related to the finite size of the scoring halo were not considered.[4]

A critical property of the scoring halo is its size. Like the standardized scoring system, the IDA analysis used a 0.5 m radius halo for most targets. For large targets (105 mm and 155 mm projectiles, 105 mm HEAT rounds, and 500 lb bombs), the scoring halo was an ellipse (Figure III-2). The major axes of the scoring halo and these large munitions were aligned in the same direction.

The estimated geo-location accuracy of demonstrators' alarms and the likelihood of being able to reacquire and excavate a target during an actual remediation affected the choice of halo size. For targets that were not part of a cluster, the standardized scoring halo usually provided an unambiguous method to determine hits and misses, but near misses still occurred. Figure IV-1 shows processed data from an EM61MKII pushcart near a relatively shallow (20 cm deep) 81 mm mortar. This demonstrator was the only one to miss this target. A contributing factor to this miss appears to be the irregularly spaced survey path, shown by the dots in this figure. Other near misses occurred when the horizontal extent of an anomaly was much larger than 1 m. In this instance, the alarm may "mark" the anomaly, but still fall outside the target halo. While these relatively rare situations affected the aggregate Pd measures very little, they did explain some anomalous misses of otherwise obvious targets and illustrated limitations of the standardized halo scoring method. Note that placing the halo around the target or the alarm is a symmetrical operation for both circular and elliptical halos since the azimuth of the major axis of the ellipse is fixed by the target's azimuth .

2. Three-Stage Filters for Aggregate Pd and FAR

The IDA analysis used three successively more restrictive filters to study aggregate performance on three like subsets of the ground truth. Note that some nonferrous items (Mk118 shaped-charge bomblets and 40 mm rifle grenades) were emplaced at the test sites. They are never included in the Pd calculation for *any* result derived solely from magnetometer data. Because they were not considered when matching magnetometer alarms to targets, a background alarm could occur within the halo of these nonferrous items.

---

[4] In the case of Blackhawk's results, a tremendous number of alarms were submitted. The finite halo effects from the resulting large BAR only worsens the aggregate results reported in this study. A complete discussion of finite halo effects can be found in Reference 2.

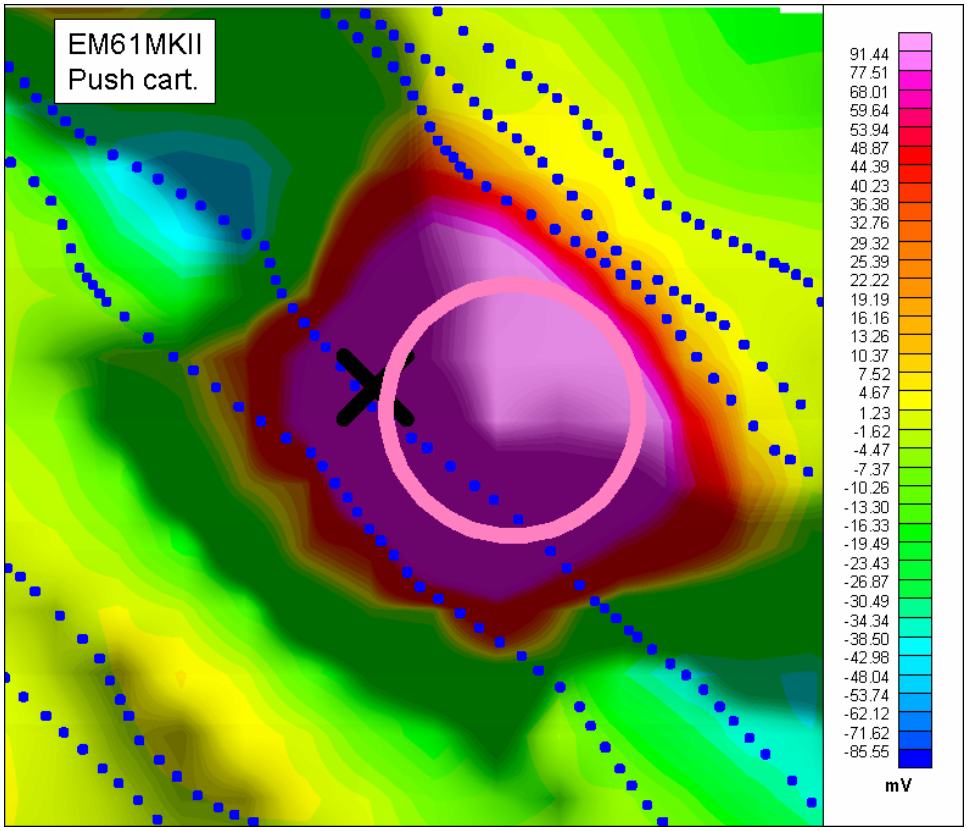

Figure IV-1. Miss of an 81 mm mortar. The black X marks the location of an 81 mm mortar at a depth of 20 cm. The circle is 1 m in diameter and is centered on a reported alarm. The data are from the 366 µs timegate of an EM61MKII mounted on a pushcart. The dots indicate the locations where data was taken. The colored grid was made in Oasis montaj.

a. Probability of Detection vs. *All* Munitions and Background-Alarm Rate

The first filter calculated the fraction of *all* munitions found by each demonstrator and the corresponding background-alarm rate. Knowing the Pd for all munitions buried at a site allowed a direct comparison to the AEC standardized scoring system to ensure that the IDA scoring program was working properly. The IDA scoring at this stage closely matches the protocol used by AEC (see appendix, Section C (p. A-17)), although the scoring programs were developed independently. To compare scores, the BAR was computed by dividing the number of alarms that were not in the scoring halo of any target by the same factor (not publicly released) used by AEC to calculate the BAR reported in the standardized analysis.

In most cases, the IDA scoring system agreed with the AEC report within the AEC rounding error (5%). Table IV-1 shows the differences. IDA generally scored submittals provided by AEC. A few submittals were provided directly by the

demonstrator. The file names of the scoring submittals passed to IDA from AEC contained a shorthand code for a date (test date or processing date), the demonstrator, the test site, the testing area, and the sensor type. Some difficulties involved in tracking demonstrators were that demonstrators tested multiple times, duplicate files were encountered, and file names did not account for minor variances in equipment (EM61 vs. EM61MKII). While there is not a one-to-one match between IDA-scored submittals and AEC-published scoring reports, the IDA-scored submittals are reported here in accordance with the file name provided to IDA. Direct comparisons of hits and misses indicate that these differences reflect the inability to fully reconcile IDA's and AEC's file lists, rather than fundamental differences in scoring standards. The noted differences are reported to explain the lack of a one-to-one correspondence between the AEC-published scoring reports and those scores in this analysis. In any event, we judge the differences to be small enough not to affect any conclusions drawn from the data.

Table IV-1. Differences between IDA and AEC scoring. Both BARs use AEC convention.

Demonstrator	Site	IDA Pd	AEC Pd	IDA BAR	AEC BAR	Comment
GeoCenters Combined EM/Magnetometer	APG	60%	55%	0.3	0.15	IDA has 2002 test for combined result, not 2004
Gtek TM4 Magnetometer	YPG	65%	55%	1	0.75	Multiple files submitted for scoring.
HFA	YPG	50%	45%	0.55	0.5	Scoring or rounding error possible.
NRL GMTADS	APG	60%	70%	0.25	0.20	Multiple files submitted for scoring.

b. "Able to Survey" and "Non-Clustered" Targets Filter

After calculating overall Pd and BAR that matched the similar calculation by the AEC standardized analysis, the IDA analysis considered the change in Pd that results from excluding targets that could not be surveyed because of an obstacle or that were buried in large clusters. Obstacles included areas of the test site that were inaccessible because of flooding, vegetation that blocked survey instruments, and man-made objects like fences. Figure IV-2 describes one of the clusters of large ordnance at APG.[5] Figure IV-3 shows an obstacle at YPG. Figure IV-4 shows the maximum extent of

[5] The color scale for all grids in this report were chosen to highlight changes near the mean value. Extreme values near targets may far exceed the minimum and maximum values listed in the scale.

flooding at APG in June 2004. Figure IV-5 shows the wash area at YPG. Figures IV-6 and IV-7 are photographs of the terrain at APG and YPG. Lack of site coverage signified by a low density of survey tracks is considered to be the result of poor navigation or field technique, but not a reason to filter missed targets from the ground truth before reporting a Pd.

Large clusters were identified by plotting the locations of ground truth items and identifying localized masses of targets by eye. Several of these clusters were intentionally emplaced at each Standardized Site and were easy to identify. The problem of locating targets within these clusters with enough precision to place them in individual scoring halos is twofold. First, it is technically challenging to disentangle multiple overlapping signals. Second, the practice of several demonstrators was to mark the extent of the combined signals and ignore the difficult problem of resolving the internal structure. Such clusters may be encountered during a real-world remediation effort, but it is probably not necessary for the survey sensor to resolve the location of individual targets within the cluster.

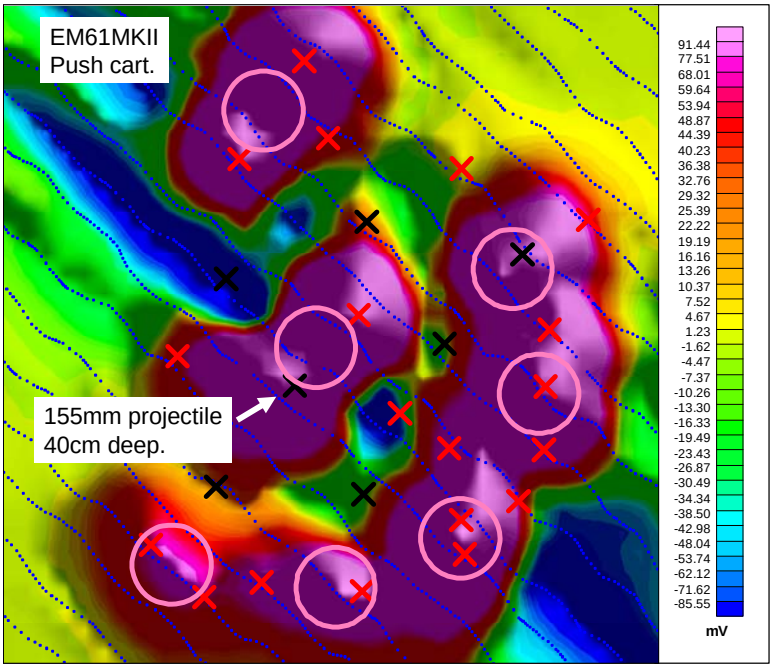

Figure IV-2. Large cluster at APG. The X's are emplaced targets (red for clutter, black for munitions). The circles are 1 m in diameter and centered on the demonstrator's alarms. The white arrow points to a shallow 155 mm projectile. The projectile's long axis is pointing 107 degrees relative to the top of the figure. It is inclined at 36 degrees above the horizontal. The alarm falls outside the elliptical scoring halo, and it was scored as a miss for this demonstrator. The demonstrator has apparently marked the peaks in the hemispherical anomaly and the center of the central anomaly.

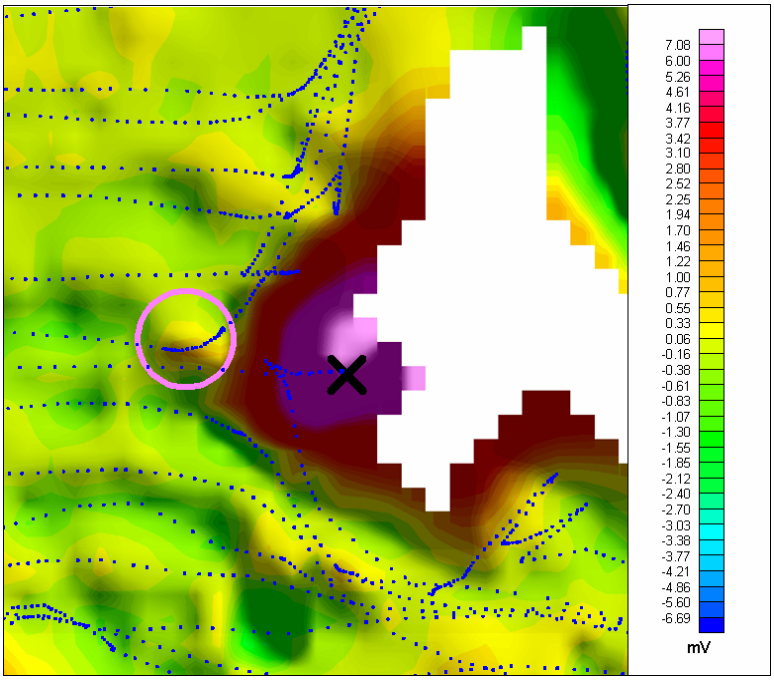

Figure IV-3. Unknown obstacle at YPG. Data from an EM63 pushcart. The black X is an 81 mm mortar target.

In cases where greater than 1% of a site was not surveyed because of flooding or terrain issues, the BAR was calculated by dividing the number of false alarms by an estimate of the area actually surveyed. This estimate was made by breaking the test site into 50 m × 50 m squares and estimating the fraction of each grid that had been surveyed. The surveyed area was not adjusted to reflect the removal of clusters or small obstacles from the ground truth prior to calculating the BAR. Ground conditions and vegetation density at YPG are highly variable, depending on recent weather. Table IV-2 and Table IV-3 (p. IV-11) record the demonstrators who missed a large portion of each site. At both sites these were demonstrators who used a towed array.

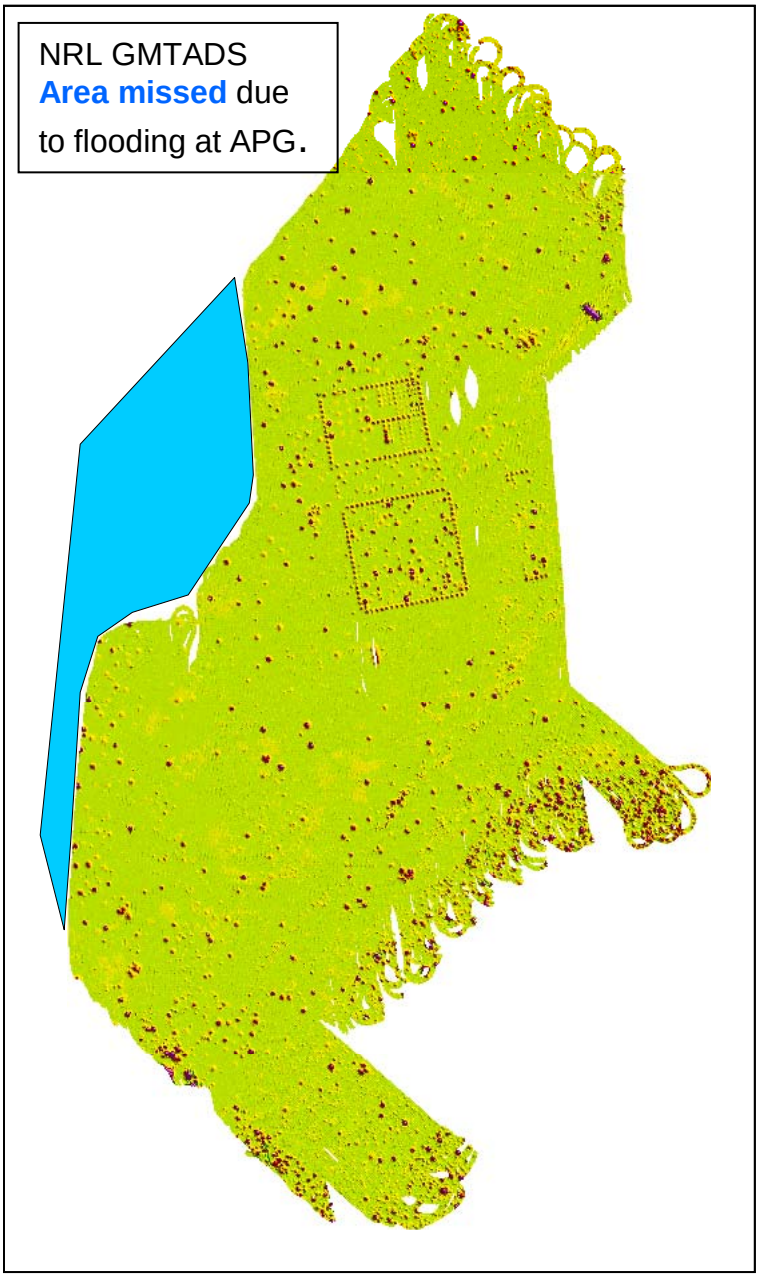

Figure IV-4. The greatest area missed due to flooding in June 2004 at APG is indicated in blue. The data grid is from the NRL GMTADS sensor.

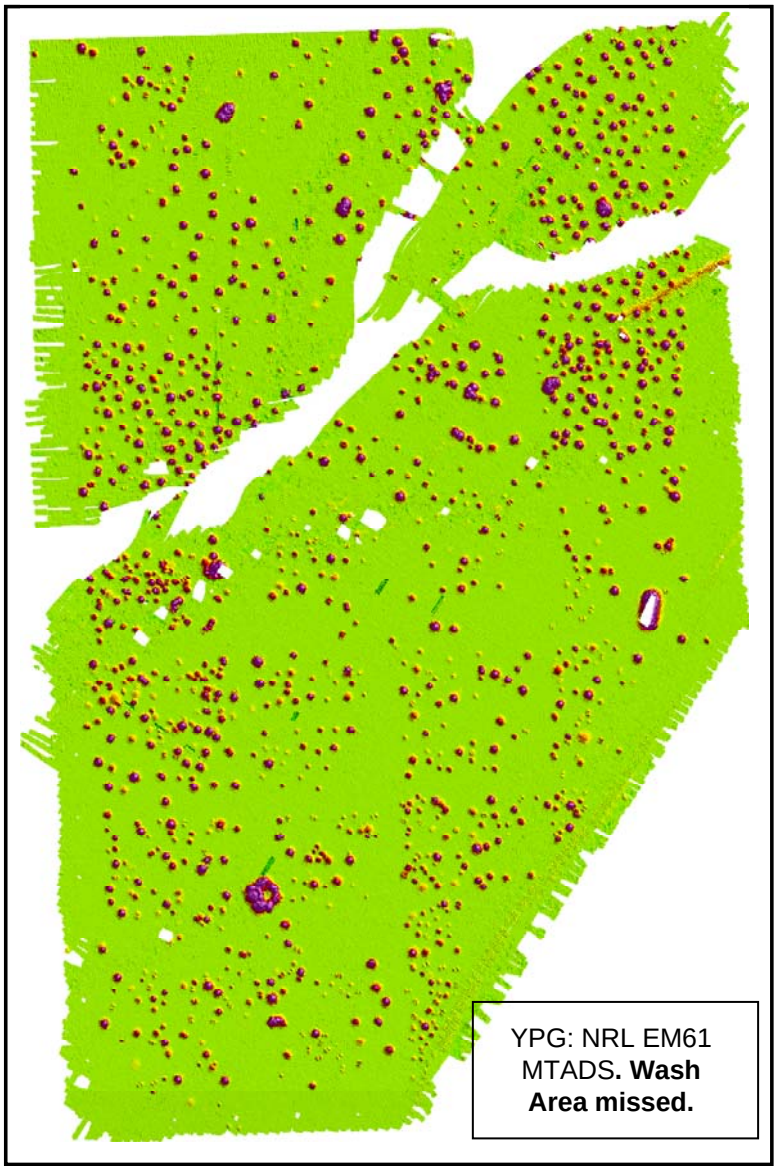

Figure IV-5. Data from YPG showing poor coverage in the wash area.

Figure IV-6. Geophex hand-held GEM3 type sensor at APG. This photo illustrates the tendency of the Open Field edge nearest to the Wooded Challenge Area to flood.

Figure IV-7. NRL GMTADS array navigating the wash at the YPG Open Field.

Table IV-2. Fraction of APG site missed because of flooding.

Open Field Not Surveyed	APG
NRL Mag	5%
NRL EM61	6%
NRL GMTADS	14%
GeoCenters STOLS 2004	6%

Table IV-3. Fraction of YPG site inaccessible because of desert-wash terrain.

Open Field Not Surveyed	YPG
NRL EM61	5%
NRL GMTADS	5%

With all filters of the IDA analysis, the number of background alarms used to make the plots in Section IV.D (p. IV-13) is the same for the all-munitions filter for each demonstrator. Only background alarms not in the scoring halo of *any*[6] target were considered, so excluding targets from the Pd calculation for a particular filter did not change the numerator of the BAR. The area of each cluster is ignored because it is typically of the order 1/1,000th of an entire test site.

c. 11× Filter: Corps of Engineers' Depth Threshold

The next filter applied was the 11× filter. The 11× depth rule-of-thumb is a general approximation that ignores the details of munition composition, shape, and orientation. *It is also static: it does not consider potential advances in detectors that would enlarge the envelope of likely detectability*. Nonetheless, the 11× depth is a convenient way to separate easy and more difficult targets. Section IV.E.2 (p. IV-30) shows that targets can often be detected deeper than the 11× depth.

The results presented in Figures 18–21 apply the 11× filter as an AND filter along with the "Able to Survey" filter described above. The Pd calculated with these two filters only considers munitions that were expected to be easy to find (above the 11× depth), could be surveyed, and were not in large clusters. Those munitions that were still missed after the application of these filters spurred a failure analysis of individual misses. Many

6 There is one exception. The removal of nonferrous items from the ground truth prior to matching alarms and targets allows a magnetometer to have a background alarm in the vicinity of a nonferrous target.

of the interesting results from this study come from understanding why seemingly easy targets are missed.

d. Relative Background Alarm Rate

While the BAR was used as an internal metric to analyze results, to maintain the security of the ground truth, this study does not report the actual BAR. Like the AEC standardized reports, a relative background-alarm rate (rBAR) is reported. The graphs in this analysis depict the rBAR, where unity is assigned to the demonstrator with the lowest actual number of background alarms. AEC's standardized reports used a different scaling factor. To get a sense of the real numbers, the typical BARs at APG were 50–100 per acre. At YPG, typical values were just a few tens per acre. For the poorer performers at YPG, this drives the rBAR to a very high number.

C. ADDITIONAL METRICS

1. 100% Detection Depth and Depth of Deepest Detection

We would like to know at what depth the Pd for a particular munition approaches 100%. Only munitions targets of a particular type that were isolated[7] (no neighboring target within 2 m) and were able to be surveyed were *considered* when measuring this depth. This analysis defines the 100% detection depth as the depth at which all considered munitions that were shallower, or at the same depth, were found. The most striking thing observed about the actual 100% detection depth for a given munition is that demonstrators often miss items much more shallow than the 11× depth. Note that requiring a target to be "isolated" is more strict than "not in a cluster." Isolation is required for metrics that are very sensitive to single misses.

The depth of deepest detection for each munition type, regardless of non-detected shallower munitions, was also studied. It is often deeper than the 11× depth for a given munition. This metric is affected by the depth distributions at the test sites. In most cases, the distribution of depths at the sites probes all depths of interest for a given munition. The exceptions are bomblets, which are only buried at shallow depths, and very large munitions, which are often detected even at their deepest seeded depths.

---

[7] Isolated targets were determined using Oasis montaj. Circles of 1 m radius were plotted around targets, and overlapping circles were flagged as not isolated.

Both the 100% detection depth and the depth of deepest detection are plotted for individual demonstrators and selected munition types to illustrate the variance across like technologies.

2. Pd as a Function of Depth

The Pd for a particular munition as a function of depth shows at what depth, and how rapidly, the signal falls off. As with the 100% detection depth, only isolated targets were considered. Munition items were sorted by burial depth into bins, each with a width that was one-sixth of that munition's 11× depth. The results from multiple demonstrators' using like technologies were *combined* to calculate the Pd for a depth bin to provide a reasonable number of munition items in each bin. To ensure that the results reflect the optimum implementation of a detection technology, only the best performing demonstrators were used.

D. RESULTS FROM ALL DEMONSTRATORS

1. Overview

The study included 19 demonstrators for the APG Open Field and 17 demonstrators in the YPG Open Field. At APG, 20 data sets were plotted; the NRL MTADS magnetometer reported two lists of alarms to IDA, one using a high threshold (suitable for discrimination) and one at a lower threshold (suitable for the response stage only). Tables IV-4 and IV-5 (p. IV-16) list the demonstrators at APG and YPG reported in this study, but do not include all demonstrators who tested sensors. Some demonstrators were excluded because there were questions about the completeness of the alarm list that was passed to IDA. Other demonstrators tested after IDA's study had ended.

2. Blind Grid Results

Figures IV-8 and IV-9 plot the results from the APG and YPG Blind Grids. The Blind Grid results are presented only as a performance baseline under very controlled conditions. Only Blind Grid results from those demonstrators that went on to test in the Open Field at a particular test site are shown.

Two Pd scores are presented for each demonstrator. The lowest is for the Pd calculated considering all UXO targets buried in the Blind Grid. The higher Pd score excludes those UXO buried deeper than the 11× depth. The Pba is calculated considering

all *blanks* in the grid, so it is the same for both Pd scores. Note that while the applied 11× depth filter increases the Pd of each demonstrator, it does not radically change the relative position of each demonstrator. This is in part because the analysis explicitly links the Pba for both Pds.

The Blind Grid results represent the sensors' performances against targets at fixed locations. Important variables such as navigation and the density of survey lines that affect performance in the Open Field scenario are not tested in the Blind Grid scenario. Demonstrators have the option of collecting Blind Grid data in either a survey mode (sensors moving at a constant rate during data collection) or a cued mode (sensor stepped across grid). The mode used was not reported and so is not considered in the IDA analysis.

Under these controlled conditions scores are expected to be better than in the Open Field (Figures IV-11 through IV-14). At the APG Blind Grid, some demonstrators scored over 90% Pd against all targets while scoring a Pba below 20%. Note that Pba for the Blind Grids may be biased somewhat by encroachment into empty squares of signals from large items in adjacent grid squares. At YPG, the NRL GMTADS had *no* background alarms and scored *100%* Pd against targets that were above the 11× depth. Note that the geophysical conditions at the two sites are different. The variance in the electromagnetic background at YPG is much smaller than at APG, but YPG does contain some naturally occurring magnetically active areas.

Despite these better demonstrators, there were scores with Pd's falling below 90%, even for targets above the 11× line. The Geophex GEM3E failed to score even 40% Pd at APG Blind Grid, but the NRL GMTADS (a different design based on the same fundamental technology) was one of the better demonstrators at both sites. EM61MII pushcart type technology had similar tendencies. Shaw did not score above 90% against targets above the 11× line, while TetraTech/Foster Wheeler (TTFW) scored nearly 100% at both YPG and APG. The Pba for Shaw was about 10% at both sites, while TTFW had no background alarms at YPG.

These scores show that the scoring system provides information on the relative performance of sensor systems *as they were used* at a particular site. Very similar technologies may perform very differently if implemented or operated differently. These differences could be caused by factors not measured at the Standardized Test Sites: setting a detection threshold, choosing optimum features from multichannel sensors to make a detection decision, and optimum quality-control strategies. Some implementation

features, such as Pd as a function of survey-track spacing and the performance of like sensors on different types of platforms could potentially be measured using the Standardized Site data, but there are insufficient samples to make precise statements about them.

3. Open Field Pd and BAR Results

In the Open Field scenario, three Pd scores are reported for each demonstrator corresponding to the three filters. The lowest is the Pd considering all munition targets buried at the site. The middle Pd score applies the filter that removes targets that could not be surveyed or were part of clusters. Generally, scoring against the set of targets passed through this filter yields a Pd about 5% higher than when scoring against all targets. In instances where a large portion of the site was not surveyed (e.g., APG NRL GMTADS), this increase was much larger. The highest Pd makes the additional restriction that only targets above the 11× depth are considered. The conditions applied to the lowest and highest Pd scores in the Open Field are roughly equivalent to the two scores reported for the Blind Grid. Figure IV-10 provides a key to the Open Field Pd and rBAR plots, which we call pseudo-ROC plots.[8]

As at the Blind Grid, there are scores with excessively low Pd or high BAR. Pd scores, shown in Figures IV-11 through IV-14, are generally lower than the same demonstrator's score at the Blind Grid. Even though there is added difficulty in the Open Field scenario, after filters are applied some demonstrators attain a Pd greater than 90% without an excessive BAR. Four demonstrators performed consistently well at both sites: NRL MTADS towed array (using both EM61MKII and GEM3 type sensors), TetraTech Foster Wheeler's (TTFW) EM61MKII pushcart, and NAEVA's towed EM61MKII array. All used digital geophysics. The few digital magnetometers had poorer scores than the digital EMI sensors. The best digital magnetometers attained Pd's somewhat lower than the "good" EMI systems, and their BARs were several times higher.

At APG and YPG, the analog magnetometer (Mag & Flag) sensors from Parsons and HFA and the digital magnetometer from the NRL MTADS sensor have roughly the same rBAR, but the digital magnetometer has a 10–20% greater Pd. At APG, the Mag & Flag sensors did poorly against targets below the 11× line.

---

8 Like a ROC plot, Pd vs. rBAR is plotted, and regions of high and low threshold can be discerned. The points represent each demonstrator's chosen threshold, however, not a decreasing threshold for the same sensor.

Ideally, a good demonstrator should have both high Pd and low BAR. The circled demonstrators in the Open Field pseudo-ROC plots are "good" demonstrators that are used in the Pd by depth calculation in Section IV.E.2 (p. IV-30). The good demonstrators are selected to capture the best performance of the sensors. Demonstrators that did not score in the high-Pd, low-BAR regime may have done so for reasons other than their sensor's limitations. In one limit, if a demonstrator set a relatively high detection threshold, it would lead to low overall Pds. This strategy is based on severely limiting background alarms, while the UXO problem suggests achieving high Pds should take precedence over low background alarm rates. In the other limit, some demonstrators have very high BAR scores, but have Pd's comparable to demonstrators with much lower BARs. Here, the demonstrator may be setting a low threshold, but may also have an inadvertently high noise level.

Table IV-4. Key to APG Demonstrators.

GEM3 Type		NRL	3x GEM3	Towed Array
		GeoPhex	GEM3E	Towed Array and Push Cart
EM61 Type		TTFW	EM61MKII	Push Cart
		NRL	3x EM61 Variant	Towed Array
		NAEVA	EM61MKII	Towed Array
		Shaw	EM61MKII	Push Cart
		GeoCenters	EM61MKII	Towed Array
		Black Hawk	EM61MKII	Pull Cart
Other EMI		Gtek	TM5 EMU	Sling
		Zonge	nanoTEM3D	Push Cart
Fused EMI and Mag.		GeoCenters, 2002	STOLS, Fused EM & Mag	Towed Array
		Black Hawk	Fused EM and Mag	Pull Cart
Magnetometer Type		NRL (low threshold)	8x G822 Variant	Towed Array
		NRL (high threshold)	8x G822 Variant	Towed Array
		Gtek	TM4 (G822A)	Sling Array
		Black Hawk	4x G822	Pull Cart
		GeoCenters	5x G822A	Towed Array
Mag/EM and Flag		Parsons	EM61MKII	Analog Push Cart (EM and Flag)
		HFA	Schonstedt	Analog Hand Held (Mag and Flag)
		Parsons	Schonstedt	Analog Hand Held (Mag and Flag)

Table IV-5. Key to YPG demonstrators.

GEM3 Type		NRL	3x GEM3	Towed Array
		ERDC	GEM3	Push Cart
EM61 Type		TTFW	EM61MKII	Push Cart
		NRL	3x EM61 Variant	Towed Array
		Shaw	EM61MKII	Push Cart
		GeoCenters	EM61MKII	Towed Array
		Black Hawk	EM61MKII	Pull Cart
Other EMI		Gtek	TM5 EMU	Sling
		ERDC	EM63	Push Cart
Fused EMI and Mag.		GeoCenters	STOLS, Fused EM and Mag	Towed Array
		Black Hawk	Fused EM and Mag	Pull Cart
Magnetometer Type		NRL	8x G822 Variant	Towed Array
		Gtek	TM4 (G822A)	Sling Array
		Black Hawk	4x G822	Pull Cart
		GeoCenters	5x G822A	Towed Array
		ERDC	TM4	Sling
Mag/EM and Flag		Parsons	EM61MKII	Analog Push Cart (EM and Flag)
		HFA	Schonstedt	Analog Hand Held (Mag and Flag)
		Parsons	Schonstedt	Analog Hand Held (Mag and Flag)

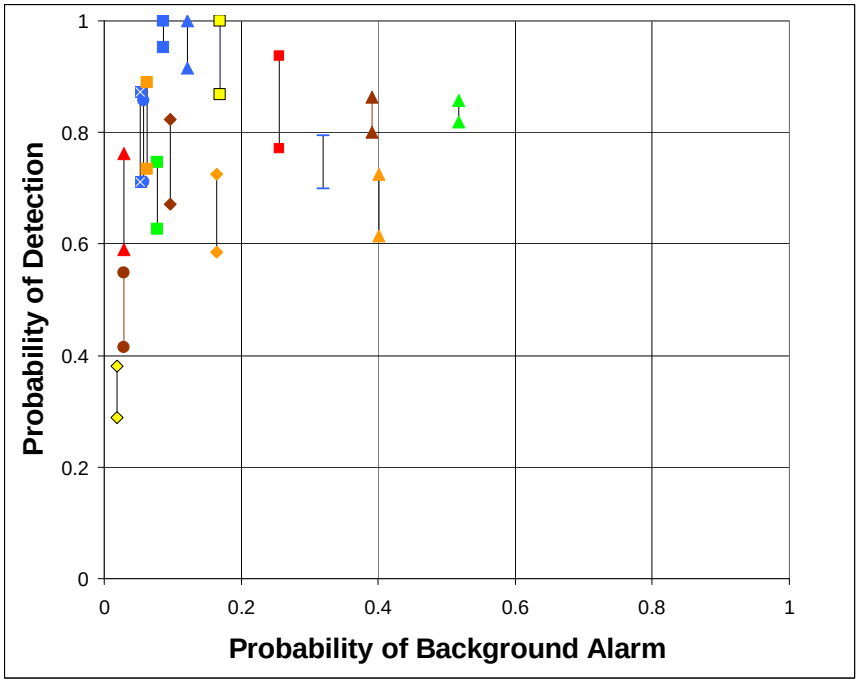

Figure IV-8. APG Blind Grid, Pd vs. Pba. Low Pd: all targets. High Pd: only targets above 11×.

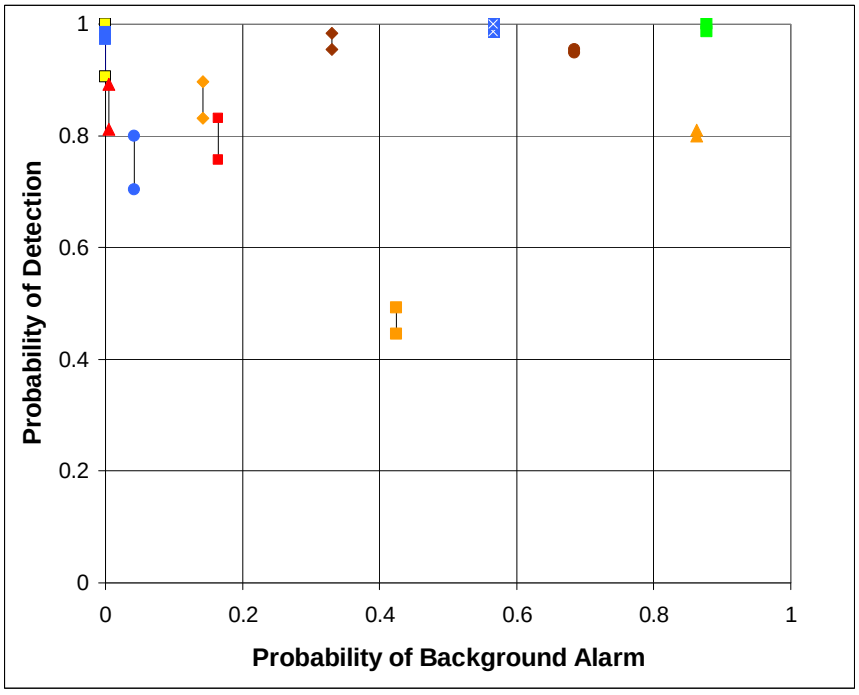

Figure IV-9. YPG Blind Grid, Pd vs. Pba. Low Pd: all targets. High Pd: only targets above 11×.

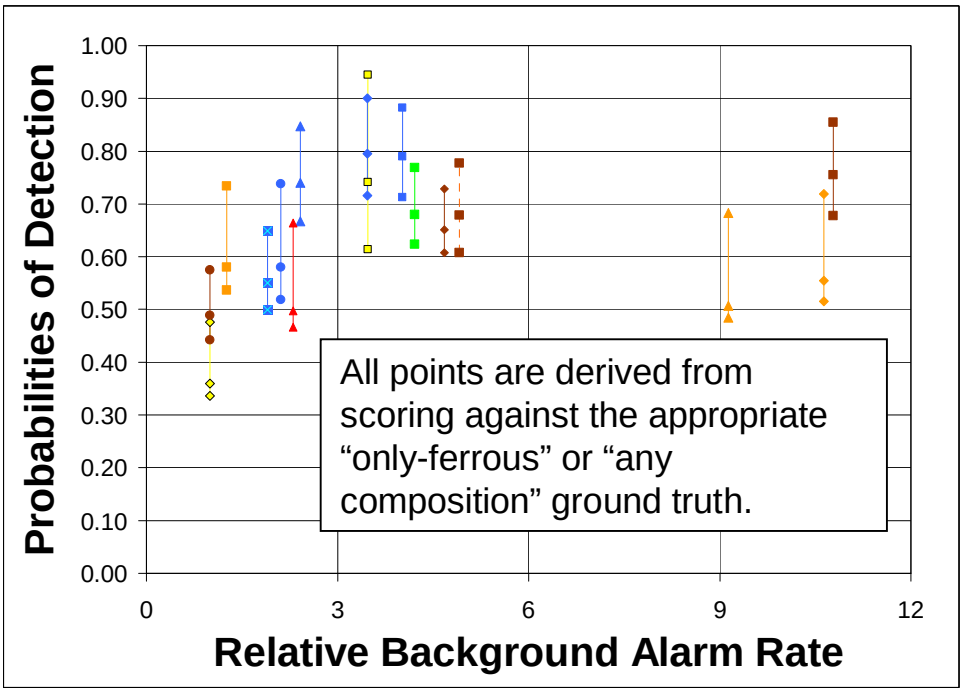

Figure IV-10. Key to Open Field pseudo-ROC plots.

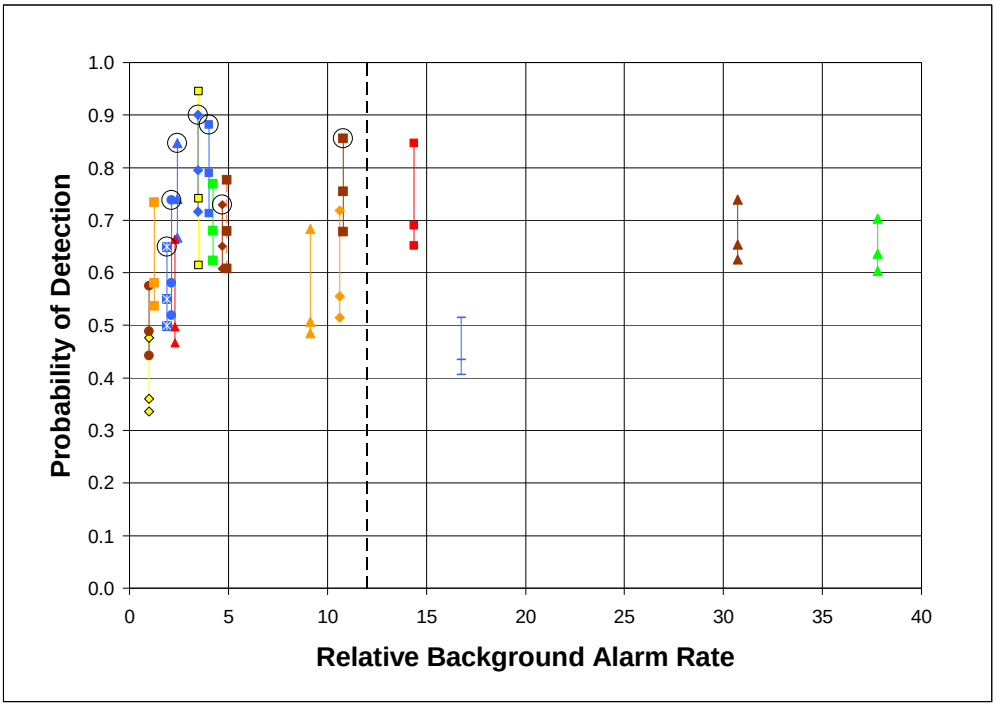

Figure IV-11. APG Open Field, Pd vs. rBAR. The area to the left of the dotted line is expanded in Figure IV-12. The circles denote demonstrators whose results were used to calculate Pd as a function of munition depth in Section IV.E.2 (p. IV-30).

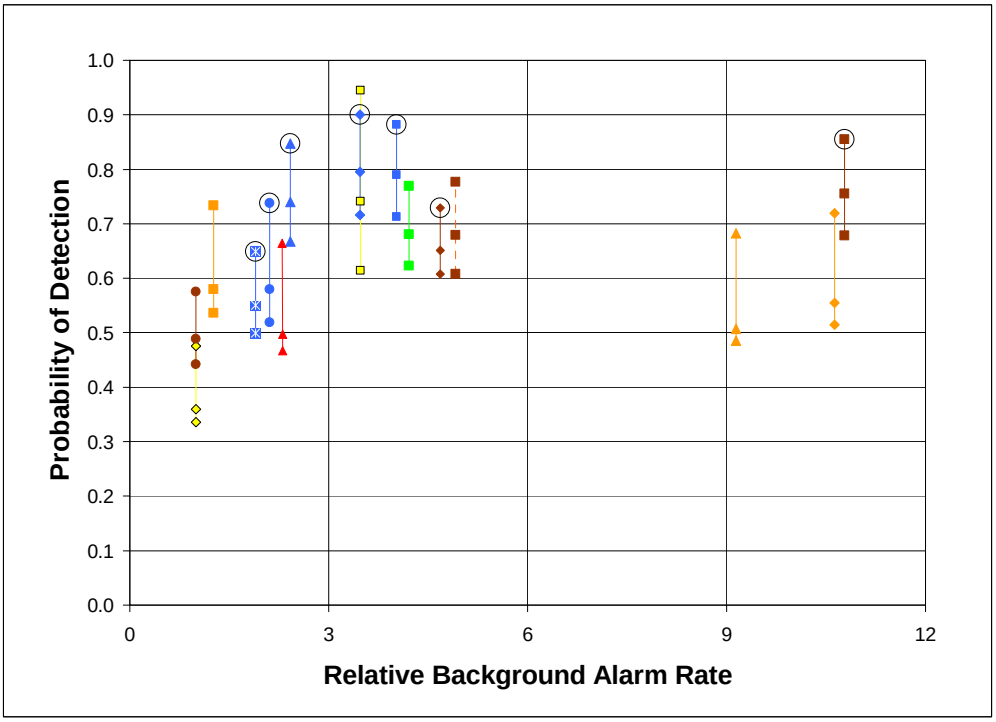

Figure IV-12. APG Open Field, Pd vs. rBAR. Zoom. The circles denote demonstrators whose results were used to calculate Pd as a function of munition depth in Section IV.E.2 (p. IV-30).

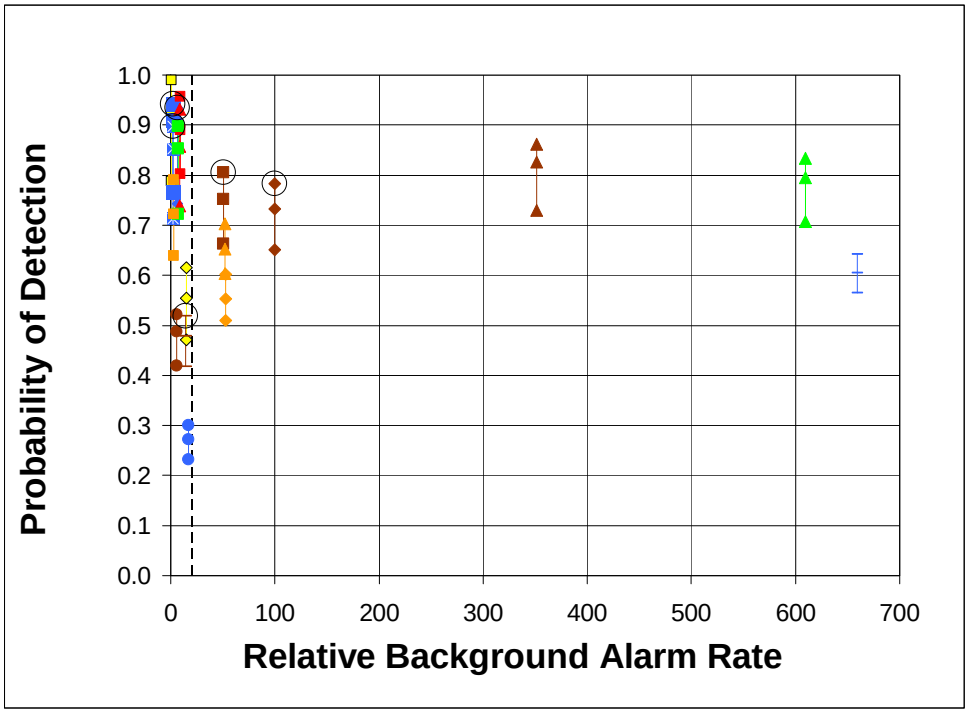

Figure IV-13. YPG Open Field, Pd vs. rBAR. The area to the left of the dotted line is expanded in Figure IV-14. The circles denote demonstrators whose results were used to calculate Pd as a function of munition depth in Section IV.E.2 (p. IV-30).

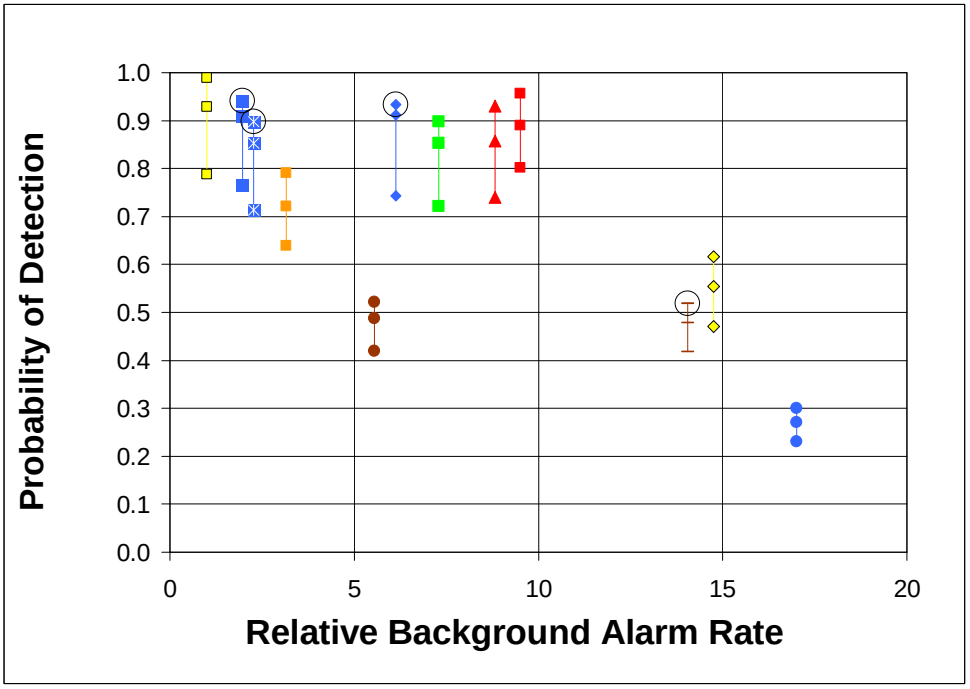

Figure IV-14. YPG Open Field, Pd vs. rBAR. Zoom. The circles denote demonstrators whose results were used to calculate Pd as a function of munition depth in Section IV.E.2 (p. IV-30).

4. 100% Detection and Deepest Detection Depth

In the Open Field, the depths of 100% detection were usually less than the 11× depth. This was true even for the better performing demonstrators. The plots in this section also introduce a filter that considers only munitions that have no neighbor within 2 m. The addition of this filter eliminates the arbitrariness in the definition of "cluster." Because the depth of deepest detection is of interest, the 11× filter is not applied in this section.

Figures IV-15 through IV-18, IV-20, and IV-21 provide bar-and-whisker plots showing 100% detection depths and the depths of deepest detection by munition type and by site. For each plot, the solid bar indicates the 100% detection depth, the whisker indicates the depth of deepest detection, and the solid line marks the 11× depth. Tables IV-6 and IV-7 provide the YPG and APG keys to the figures.

In most cases, the depths of the seeded UXO probed sensor sensitivity to depths greater than the 11× depth. Exceptions are submunitions and bomblets, which are expected to be found near the surface because they are not employed in a way that would enable them to penetrate very deeply. Also shown in this section are the depths of deepest detection. In many cases, this depth is limited by the deepest item seeded at the site.

Inspection of these results, which include all demonstrators, shows the variance in 100% detection depth. We ideally would like to have a precise measurement of the deviation of Pd from 100% for a given set of circumstances (e.g., munition type, site soil, number of clusters, etc.). The variance in 100% detection depth shows, however, that for a given munition type, the depth of certain detectability changes greatly from demonstrator to demonstrator.

The results in this section are broken down by site and munition type. The plots for 20 mm projectiles, 60 mm mortars, and 155 mm projectiles at APG and YPG are shown here; plots for several other selected munitions are shown in the appendix, Section A. Even at this coarse level of analysis, there are *at most* a few tens of munitions in each category (the exact number is not revealed to conceal the ground truth).

Figures IV-15 and IV-16 show the results for 20 mm projectiles. A significant number of demonstrators missed the shallowest 20 mm at each site (no solid blue bar). While none of the demonstrators that missed the shallowest 20 mm found extremely deep 20 mm projectiles, the depth of deepest detection for these demonstrators was near the 11× depth. It is hypothesized that these shallow misses were for reasons other than low

signal. Section V.C (p. V-9) examines the misses from some of the better performing demonstrators to test this hypothesis.

Figures IV-17 and IV-18 show results for 60 mm mortars. The sensitivity of the 100% detection depth to misses unrelated to signal strength can easily be seen in these graphs. Demonstrator 4 at APG and YPG was the NRL EM61 MTADS towed array, which had good overall performance. At both sites, the depth of deepest detection for 60 mm mortars was comparable. However, at YPG the sensor missed the shallowest 60 mm mortar. Figure IV-19 shows raw sensor data from near this miss. Note that although the signal from the 60 mm mortar is strong, it was not "detected" according to the halo used by the scoring system.

Table IV-6. Bar-and-whisker key. APG demonstrators grouped by technology type.

#	Demonstrator	Sensor Type	Transport Mode
1	Naval Research Laboratory (NRL)	GEM3 Type	TA
2	Geophex	GEM3 Type	TA
3	TTFW	EM61 Type	PC
4	NRL	EM61 Type	TA
5	NAEVA Geophysics, Inc.	EM61 Type	TA
6	Shaw Environmental, Inc.	EM61 Type	PC
7	Geocenters, Inc	EM61 Type	TA
8	Blackhawk GeoServices	EM61 Type	PC
9	Gtek	TM5 Sling (dual sensor)	Sling
10	*Zonge Engineering and Research Organization, Inc.*	nanoTEM3D	PC
11	Geocenters, Inc	Fused EM/Mag	TA
12	Blackhawk GeoServices	Fused EM/Mag	PC
13	Naval Research Laboratory (NRL)	Mag (high threshold)	TA
14	Naval Research Laboratory (NRL)	Mag (low threshold)	TA
15	Gtek	Mag	Sling
16	Blackhawk GeoServices	Mag	PC
17	Geocenters, Inc	Mag	TA
18	Parsons EM&F	EMI (analog)	PC
19	Human Factors Applications, Inc. (HFA)	Mag (Analog)	HH
20	Parsons	Mag (Analog)	HH

Key:
TA: towed array.
PC: pushcart.
Sling: man-carry with shoulder-strap type supports.
HH: hand-held, wand-type.
Analog: no digital record of sensor data was made (i.e., "mag and flag").

Table IV-7. Bar-and-whisker key. YPG demonstrators grouped by technology type.

#	Demonstrator	Sensor Type	Transport Mode
1	Naval Research Laboratory (NRL)	GEM3 Type	TA
2	Engineer Research and Development Center (ERDC)	GEM3 Type	PC
3	TTFW	EM61 Type	PC
4	NRL	EM61 Type	TA
5	Shaw Environmental, Inc.	EM61 Type	PC
6	Geocenters, Inc	EM61 Type	TA
7	Blackhawk GeoServices	EM61 Type	PC
8	Engineer Research and Development Center (ERDC)	EM63	PC
9	Gtek	TM5 Sling (dual sensor)	Sling
10	Geocenters, Inc	Fused EM/Mag	TA
11	Blackhawk GeoServices	Fused EM/Mag	PC
12	Naval Research Laboratory (NRL)	Mag	TA
13	Gtek	Mag	Sling
14	Blackhawk GeoServices	Mag	PC
15	Geocenters, Inc	Mag	TA
16	Engineer Research and Development Center (ERDC)	Mag	Sling
17	Parsons EM&F	EMI (analog)	PC
18	Human Factors Applications, Inc. (HFA)	Mag (Analog)	HH
19	Parsons	Mag (Analog)	HH

Key:
TA: towed array.
PC: pushcart.
Sling: man-carry with shoulder-strap type supports.
HH: hand-held, wand-type.
Analog: no digital record of sensor data was made (i.e., “mag and flag”).

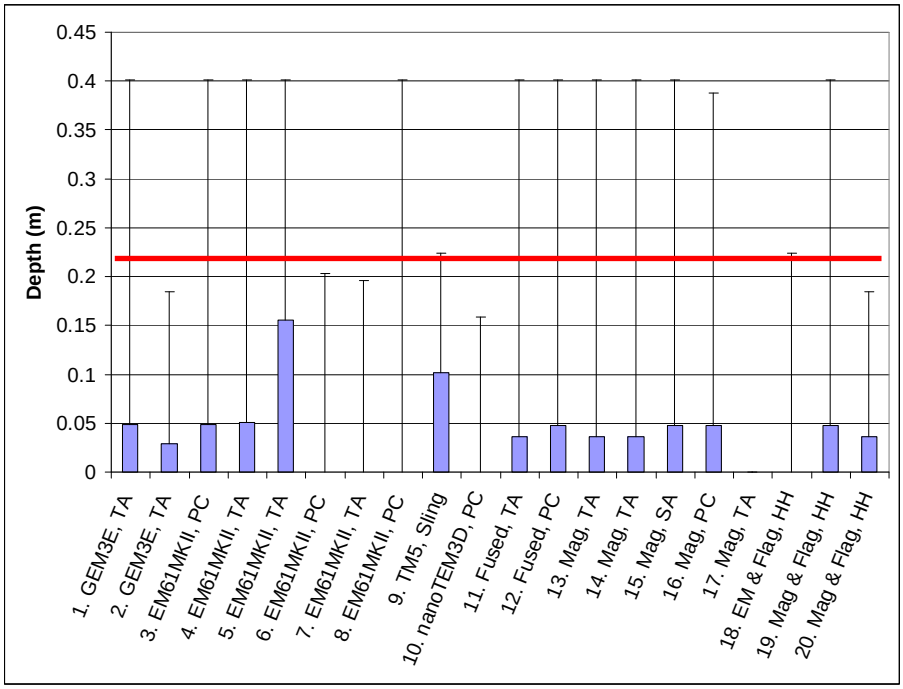

Figure IV-15. APG. 20 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark). The red line is the 11× Corps of Engineers depth.

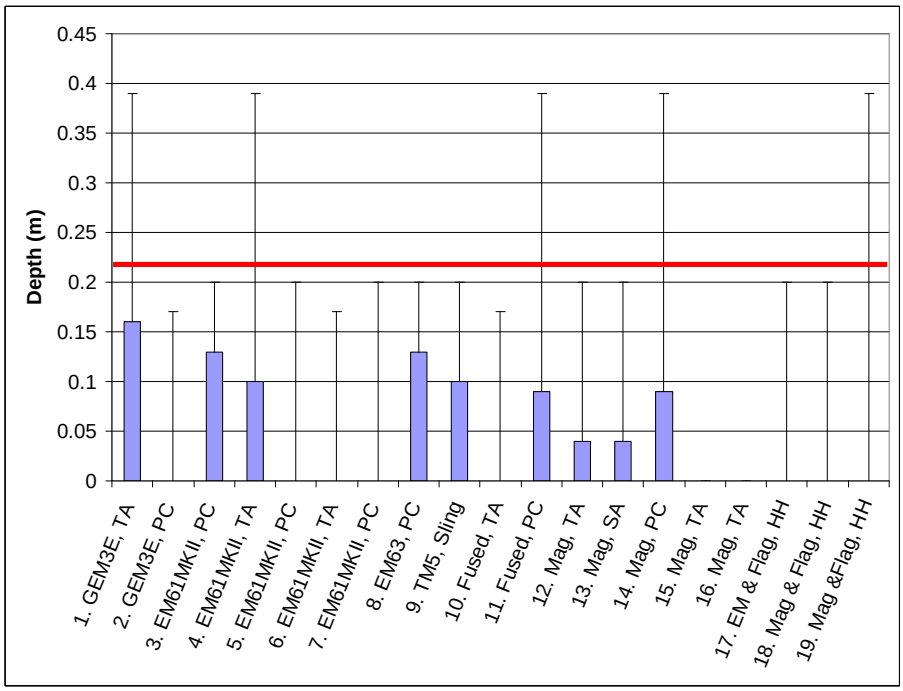

Figure IV-16. YPG. 20 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark). The red line is the 11× Corps of Engineers depth.

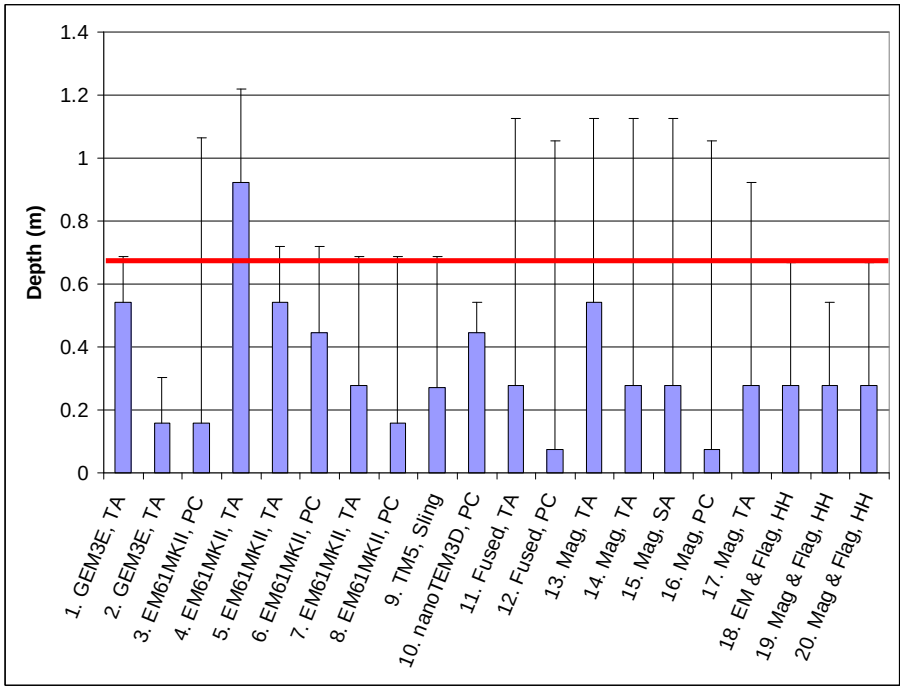

Figure IV-17. APG. 60 mm mortar, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark). The red line is the 11× Corps of Engineers depth.

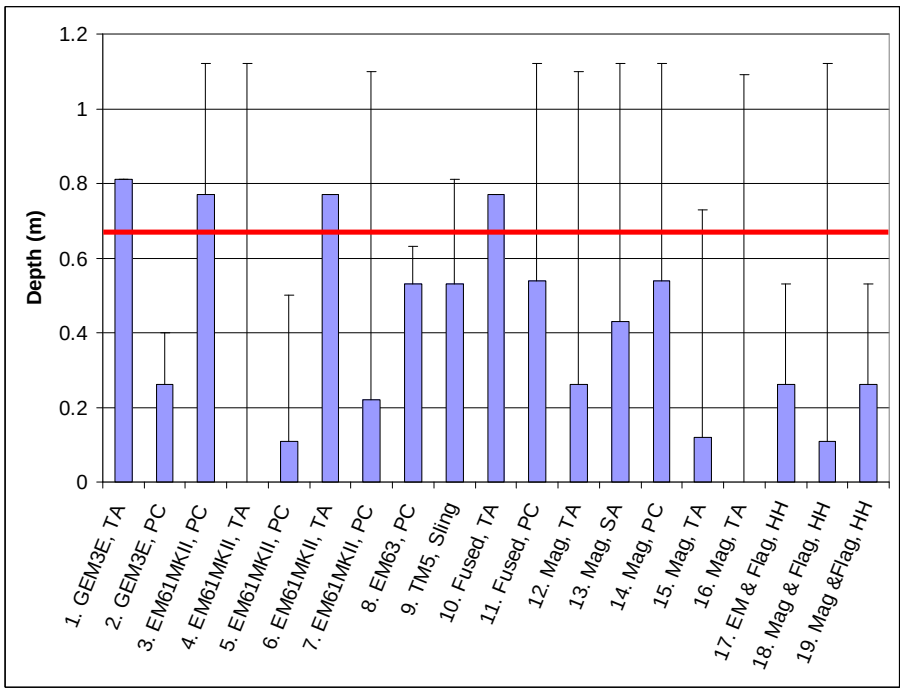

Figure IV-18. YPG. 60 mm mortar, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for YPG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

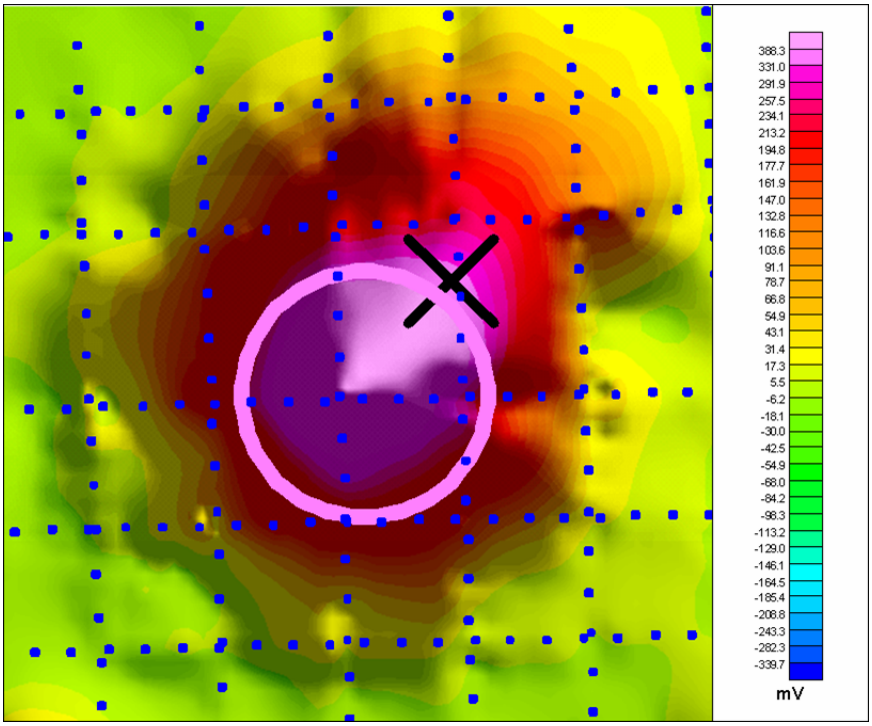

Figure IV-19. Miss of the *most shallow* 60 mm mortar at YPG by an otherwise "good" demonstrator (NRL EM61 MTADS). The mortar is indicated by the black X. The center of the circle is the location of the alarm. Note that the signal is strong, but that its peak is shifted away from the location of the mortar.

Figures IV-20 and IV-21 show data for 155 mm projectiles at APG and YPG. Some demonstrators found 100% of these munitions, but many demonstrators missed these large munitions at depths much shallower than the 11× depth.

Variability across demonstrators can be seen in the 100% detection depth and depth of deepest detection as well. Results from demonstrator 1 (NRL GMTADS) and demonstrator 2 (GEM3E pushcart) are from the same fundamental technologies, but one difference is that the transmit moment (power) for the GMTADS is larger. At APG, the GEM3E was operated by Geophex, its manufacturer. At YPG, the GEM3E was demonstrated by ERDC. As seen in the pseudo-ROC charts in Section IV.D.3 (pp. IV-15ff), the NRL GMTADS towed array scored a greater overall Pd than the other GEM3E-based sensors at APG and YPG. That advantage is retained by the NRL GTADS in every case presented in the 100% detection and depth-of-deepest detection charts, and the performance difference is often great.

TTFW (demonstrator 3) and Shaw (demonstrator 6 at APG, demonstrator 5 at YPG) both used an EM61MKII pushcart. For 20 mm projectiles, Shaw missed the shallowest one at both sites. For 60 mm mortars, TTFW missed a shallow one at APG,

but had a better deepest depth at both sites. For 155 mm projectiles, the situation is similar to the 60 mm mortars.

For all munitions, shallow 100% detection depths may be driven by poor sensor implementation or operation. These misses, such as the 60 mm mortar miss shown in Figure IV-19 by the NRL EM61 MTADS, may also illustrate the limits of the scoring system.

Although this variability suggests that individual hits and misses at each depth be investigated, security of the ground truth precludes showing the detailed depth distribution of munition targets. In addition, too few targets are in each category to assemble a reasonable number of targets in suitably narrow depth bins.

To solve this problem, we sum the hits and misses from several different demonstrators to form a larger pool of munitions. This larger pool provides a sufficient number of munitions in each depth bin to make a gross estimate of Pd as a function of depth. Section IV.E (p. IV-28) describes how these demonstrators were selected. The Pd as a function of depth results are in Section IV.F (p. IV-34).

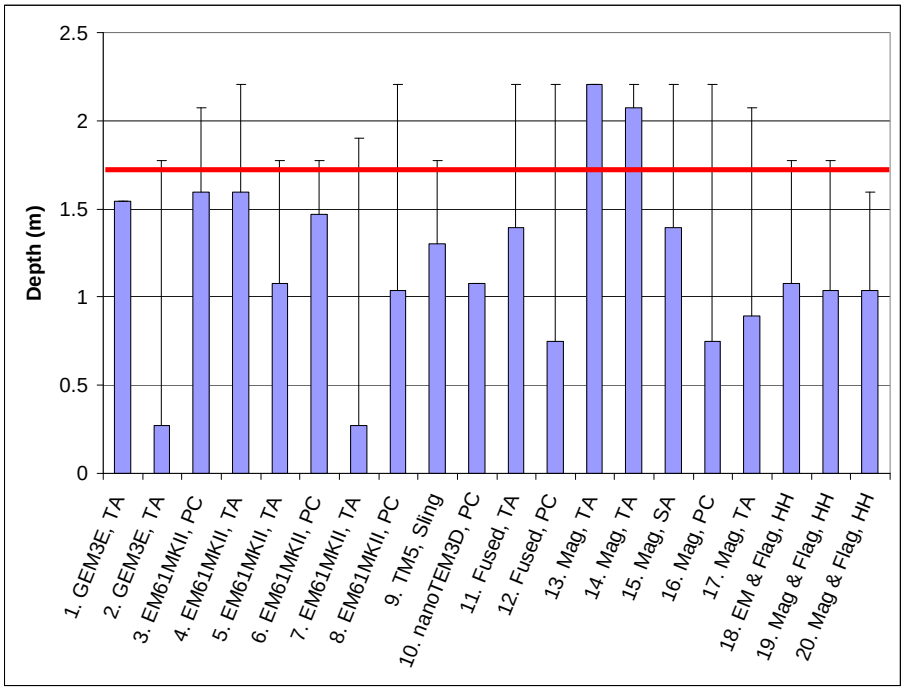

Figure IV-20. APG. 155 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark). The red line is the 11× Corps of Engineers depth.

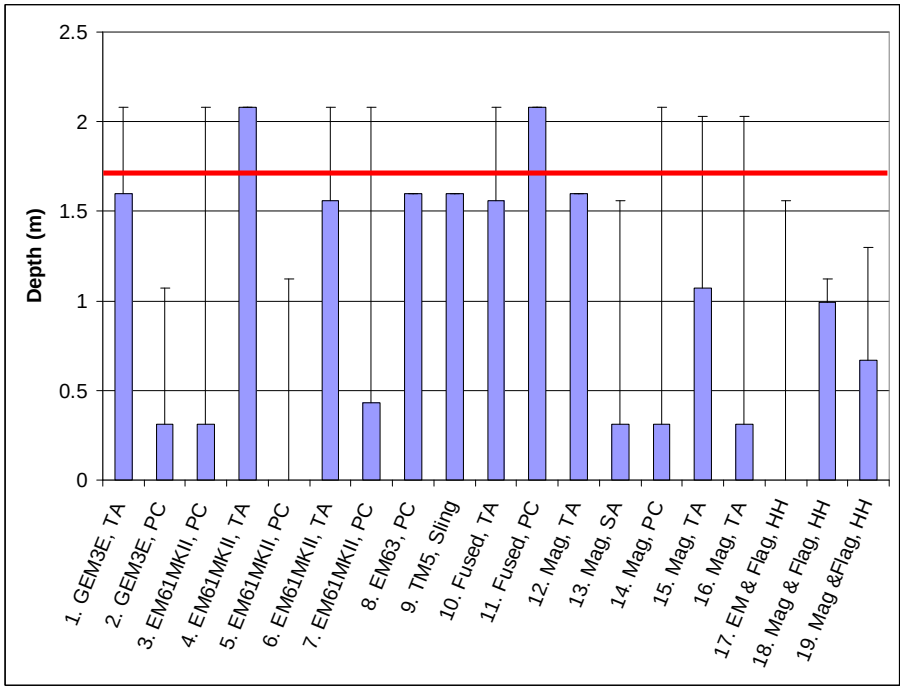

Figure IV-21. YPG. 155 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark). The red line is the 11× Corps of Engineers depth.

E. ANALYSIS OF BETTER PERFORMING DEMONSTRATORS

1. Fraction of Misses Due to Low Signal

Why are isolated munitions above the 11× line missed? We might expect the Pds recorded in the pseudo-ROC curves of Section IV.D.3 (p. IV-15) to be near 100% with all filters in place. Like the example shown in Figure IV-19, these missed munitions were not in large clusters where the detection was obviously ambiguous. The sensor could physically access them, and they were above the 11× depth. To shed light on these misses, they were examined by inspecting sensor data in Oasis montaj (when available) for a set of demonstrators with high Pd that spanned technology types at each site. The Oasis montaj databases contain the output from each sensor channel. Although each type of sensor reported different quantities (e.g., millivolts in the receiver coil for EM61 and nanoteslas for a cesium-vapor magnetometer), these databases mapped the fundamental response of the sensor without significant post-processing or interpretation. The results of

an analysis to determine whether low signal[9] was responsible for most misses are recorded in Table IV-8.

Table IV-8. Better demonstrators spanning technology type. The percentage value indicates the fraction of misses (excluding 20 mm projectiles) that were apparently due to a low signal. Most of the misses accounting for the 30% value at YPG were from two of the demonstrators listed. If those two are not included, the low-signal miss rate at YPG is comparable to APG.

Study of Individual Missed Munitions: Misses of Munitions Above 11×, Not in a Large Cluster, and Able to be Surveyed		
	Demonstrators Included	**Misses Likely Due to Low Signal**
APG	TTFW, NRL (EM61 type, GEM3, Mag)	5%
YPG	TTFW, NRL (EM61 type, GEM3, Mag), GeoCenters EM61 type, ERDC EM63, Gtek TM5	30%

Excluding 20 mm projectiles, at least 50 misses were examined at each site. At APG 5% of the misses and at YPG 30% of these misses had no obvious explanation other than low signal. Further, at YPG, 80% of the apparent low-signal misses were from two of the seven demonstrators whose misses were examined (the ERDC EM63 and NRL MTADS Mag). The MTADS Mag suffered from a heightened geologic background. The overlapping signatures from naturally occurring magnetic anomalies could be identified as the source of several of the MTADS Mag misses. These are included as low-signal misses because the magnetic background is an intended feature of the site. At YPG, less than 10% of the misses by the other five sensors were due to low signal. *Thus, low signal rarely appears to be a limiting factor in detecting UXO above the 11× depth.*

In most cases the reason for the miss was identifiable in the gridded data as either halo effect or shadowing by an overlapping signature. In the case of a halo effect, an alarm was present near the ground-truth item, but not within the halo itself. Shadowing could be detected by the influence of a known target near the missed munition. Misses where there was no obvious signal in the data were classified as likely due to low signal. These types of misses are discussed in detail in Section V.

9 A rigorously defined signal-to-noise ratio was not used because of the multiple channels in some sensors, the difficulty of locally defining "background noise," and incomplete knowledge of how the demonstrators selected detections (i.e., what signal above what threshold required an alarm).

2. Pd by Munition Type

To calculate the Pd as a function of depth, the results from several *good* demonstrators using *like* technologies were aggregated.[10] "Good" refers to demonstrators whose performance lies in the upper left of the Pd vs. rBAR graphs. The particular good demonstrators used in this section are circled in the Pd vs. rBAR pseudo-ROCs (Figures IV-11 and IV-13) and listed in Table IV-9. "Good" *is a loosely defined term at the Standardized Sites, and it in no way judges the performance of demonstrators in any other situation*. For the purposes of this section, *like* means either EM61MKII or cesium-vapor magnetometer sensors.

Table IV-9. List of "good" demonstrators for "like" technologies.

	APG	YPG
EMI (EM61MKII)	NRL MTADS	NRL MTADS
	GeoCenters	GeoCenters
	Shaw	TTFW
	NAEVA	
	TTFW	
Magnetometer*	NRL MTADS	NRL MTADS
	Gtek	Gtek
		ERDC

* Note that from the Open Field Pd-rBAR results, a case could be made to inlude the GeoCenters magnetometer result as good. GeoCenters operated its STOLS sensor as a combined magnetometer/EMI. While the results for each sensor were given separately and as a fused result, the better of the two (EMI) was chosen to capture GeoCenters results to avoid any biasing of the results from knowledge gained by the other sensor.

Note that the pool of good demonstrators listed in Table IV-9 to calculate the Pd as a function of depth differs from those in Table IV-8. In Table IV-8, other demonstrators were included to increase the number of munitions considered for like technologies.

For the good demonstrators, the aggregate Pd's for all munitions approach or exceed 90% after a very restrictive set of filters (above 11×, not in a cluster, and able to survey) is applied. This Pd value is dependent on the relative number of difficult and easy munition types seeded in the site. When the Pd is segregated by munition type, 20 mm

[10] Depth was chosen as the most important variable. Other variations present, but not analyzed separately due to small numbers, are inclination, standard or nonstandard target, and offset of the survey track from the target's center.

projectiles prove to be the most difficult to find. The ERDC magnetometer found *no* 20 mm projectiles.

Table IV-10 shows the Pd by munition type for the good demonstrators at APG, while Table IV-11 provides similar data for YPG. The "11×+ Exclusions" column shows the Pd with the 11× and Able to Survey filters applied. The column to the far right shows the Pd with the additional condition that there is no target within a 2 m horizontal distance of the munition target that is included in the Pd calculation. This column eliminates the effect of shadowing. The most common effect of the 2 m isolate filter is to increase the Pd slightly over the "11×+ Exclusions" column. However, even after this filter is applied, 100% Pd is rarely achieved by any demonstrator. Note that Pd sometimes decreases when only isolated targets are considered. In these cases, the isolation condition removed items that had been credited as hits under the less restrictive filter conditions. For example, a large, shallow item and a nearby small, deep item are *both* removed by the isolation condition.

The Pd for 20 mm projectiles is generally lower than for other munition types. Among the demonstrators listed in Table IV-10 for APG, values from the 2 m isolate column range from 24% (GeoCenters EM61) to 81% (NRL EM61 MTADS). The 20 mm projectiles are discussed more in Section V.C (p. V-9).

The misses shown in Tables IV-10 and IV-11 are not due to ambiguities in the scoring system, deeply buried targets, or overlapping signals. They may be due to halo effect or low signal.[11] Thus, in the Pd-as-a-function-of-depth plots in Section IV.F (p. IV-34), it is expected that deviations from 100% detection are due primarily to low signal and secondarily to halo effects. Note that, except for 20 mm projectiles, the *good* demonstrators almost always find 90% or more of the items above 11×. These cases are highlighted in Tables IV-10 and IV-11. The submunitions are also rarely missed, but they are not buried very deeply in relation to their size (as would be expected for a air-scattered munition).

[11] In rare cases, the reason for a miss is unknown. There is a high signal, no overlap with another signal, and no alarm nearby. It is likely that these misses were due to errors in handling the raw data.

Table IV-10. APG—Pd by munition type for demonstrators included in the Section IV.F (IV-34) plots. The "11×+ Exclusions" column shows the Pd with the 11× and Able-to-Survey filters applied. The column to the far right shows the Pd with the additional condition that there is no target within 2 m of the munition target that is included in the Pd calculation.

NRL MTADS Mag	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.85	0.89
20mm Projectile	0.52	0.52
40mm Projectile	0.89	1.00
60mm Mortar	0.79	0.91
81mm Mortar	0.95	1.00
2.75" Rocket	0.96	0.95
105mm Projectile	1.00	1.00
155mm Projectile	1.00	1.00

NRL EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	1.00	1.00
20mm Projectile	0.80	0.81
40mm Projectile	0.78	1.00
60mm Mortar	0.93	1.00
81mm Mortar	0.90	0.94
2.75" Rocket	0.91	0.90
105mm Projectile	1.00	1.00
155mm Projectile	0.94	0.94

Gtek TM4 Mag	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.69	0.67
20mm Projectile	0.28	0.29
40mm Projectile	0.80	0.88
60mm Mortar	0.79	0.82
81mm Mortar	0.80	0.83
2.75" Rocket	0.81	0.88
105mm Projectile	0.86	0.83
155mm Projectile	0.89	0.89

NAEVA EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.92	0.89
20mm Projectile	0.72	0.76
40mm Projectile	0.90	1.00
60mm Mortar	0.86	0.91
81mm Mortar	0.90	0.94
2.75" Rocket	0.67	0.72
105mm Projectile	0.93	0.92
155mm Projectile	0.89	0.89

GeoCenters EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.62	0.78
20mm Projectile	0.20	0.24
40mm Projectile	0.60	0.75
60mm Mortar	0.71	0.73
81mm Mortar	0.80	0.83
2.75" Rocket	0.65	0.71
105mm Projectile	0.86	0.92
155mm Projectile	0.67	0.67

TTFW EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	1.00	1.00
20mm Projectile	0.68	0.76
40mm Projectile	0.80	0.88
60mm Mortar	0.86	0.91
81mm Mortar	0.90	0.94
2.75" Rocket	0.81	0.80
105mm Projectile	0.93	0.92
155mm Projectile	0.94	0.94

Shaw EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	1.00	1.00
20mm Projectile	0.28	0.33
40mm Projectile	0.80	0.88
60mm Mortar	0.79	0.91
81mm Mortar	0.85	0.89
2.75" Rocket	0.70	0.76
105mm Projectile	0.86	0.83
155mm Projectile	0.89	0.89

Table IV-11. YPG—Pd by munition type for demonstrators included in the Section IV.F (p. IV-34) plots. The "11×+ Exclusions" column shows the Pd with the 11× and Able-to-Survey filters applied. The column to the far right shows the Pd with the additional condition that there is no target within 2 m of the munition target that is included in the Pd calculation.

NRL MTADS Mag	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.96	1.00
20mm Projectile	0.37	0.38
40mm Projectile	0.92	0.91
60mm Mortar	0.84	0.88
81mm Mortar	0.91	0.93
2.75" Rocket	1.00	1.00
105mm Projectile	1.00	1.00
155mm Projectile	1.00	1.00

GeoCenters STOLS EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.96	1.00
20mm Projectile	0.42	0.40
40mm Projectile	0.90	0.89
60mm Mortar	0.94	1.00
81mm Mortar	0.94	0.97
2.75" Rocket	1.00	1.00
105mm Projectile	1.00	1.00
155mm Projectile	0.96	0.96

Gtek TM4 Mag	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.87	0.89
20mm Projectile	0.44	0.42
40mm Projectile	0.58	0.55
60mm Mortar	0.76	0.80
81mm Mortar	0.85	0.87
2.75" Rocket	0.95	0.95
105mm Projectile	0.92	0.96
155mm Projectile	0.89	0.88

NRL EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.96	1.00
20mm Projectile	0.73	0.72
40mm Projectile	0.83	0.91
60mm Mortar	0.91	0.96
81mm Mortar	1.00	1.00
2.75" Rocket	0.82	0.88
105mm Projectile	0.96	0.96
155mm Projectile	0.96	1.00

ERDC Mag	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.61	0.72
20mm Projectile	0.00	0.00
40mm Projectile	0.33	0.36
60mm Mortar	0.45	0.44
81mm Mortar	0.71	0.75
2.75" Rocket	0.62	0.65
105mm Projectile	0.85	0.84
155mm Projectile	0.71	0.80

TTFW EM61 type	11x+ Exclusions	AND 2m-Isolate
Munition	Pd	Pd
BDU28	0.96	1.00
20mm Projectile	0.85	0.85
40mm Projectile	0.92	1.00
60mm Mortar	0.91	1.00
81mm Mortar	0.97	0.97
2.75" Rocket	0.95	1.00
105mm Projectile	1.00	1.00
155mm Projectile	0.89	0.96

F. PROBABILITY OF DETECTION AS A FUNCTION OF DEPTH

In Figures IV-22 through IV-33, the aggregate Pd for good demonstrators is reported for 20 mm projectiles, 60 mm mortars, and 155 mm projectiles. The uncertainties shown represent a 70% confidence level, calculated assuming a true detection probability, P, and true miss probability, 1–P. Given the total number of targets encountered in a particular bin, the observed Pd (fraction detected) is a random sample from a binomial distribution whose most probable value is P. The uncertainty expresses the 70% confidence interval in which P is expected to lie when the observed Pd is indicated by the black dot.

The curve fit to the data is:

$$Pd(d) = \frac{1}{2} - \frac{1}{2}\tanh\left(\frac{a-d}{b}\right),$$

where d is depth, and a and b are parameters determined from a least-squares fit to the observed probabilities in each bin. The fit is not weighted by the uncertainties.

Many factors dictate the precise shape of the curve, including the background noise distribution, the data-analysis method, and field techniques. The *tanh* function was chosen as a fitting function solely because it approximates the global features of the probability-of-detection curve. At low depths, this equation is nearly one, and at great depths, it is nearly zero. Terms a and b describe at what depth the probability passes below 50% and how steeply the probability descends from one to zero. Figures IV-34 and IV-35 summarize the values of the fit parameters. The fit is omitted in cases where there were too few populated bins or the numerical fit did not converge.

The Pd-by-depth plots shown in this section represent a small, medium, and large munition type (various other munitions are plotted in the appendix, Section B (p. A-6)). These plots are also arranged by sensor type: EM61MKII and cesium-vapor magnetometer. The 11× depth for each munition is binned into six increments (each bin is 11/6 of the munition's diameter deep). Note that some depth bins contain no samples.

Three different regimes of detectability are shown. For 20 mm projectiles, the Pd rises at shallow depths, but is generally no greater than 80% for projectiles just below the surface. The YPG magnetometer data (Figure IV-31) scores 100% Pd in the shallowest bin, but the fit is consistent with the Pd being bound by a number somewhat less than 100%. Note the size of the uncertainty bars on this plot. Even after summing demonstrators, there are not a great number of targets in each depth bin.

For medium munitions, the detectability curve is very near 100% at the very shallow depths and has a transition region where the Pd falls. The 60 mm plots in this section show this transition and that the deepest munitions are generally found only about 20% of the time. Other medium munitions, like the 81 mm mortar, show similar curves, although the width of the transition region from 100% Pd to very low Pd varies by site and munition type. Note that the Pd does not always fall to zero for the deepest medium munitions.

For large targets like the 155 mm projectile, the burial depths are not always deep enough to probe the transition region from 100% to significantly lower Pd. Figure IV-27 shows the results for magnetometers at APG. The Pd for the deepest depth bin is 100%. For large targets (especially the 155 mm), the fit parameters (Figures IV-34 and IV-35) have a large variance because there are not always enough deeply buried targets to fit the transition region.

1. APG, EMI, Probability of Detection as a Function of Depth

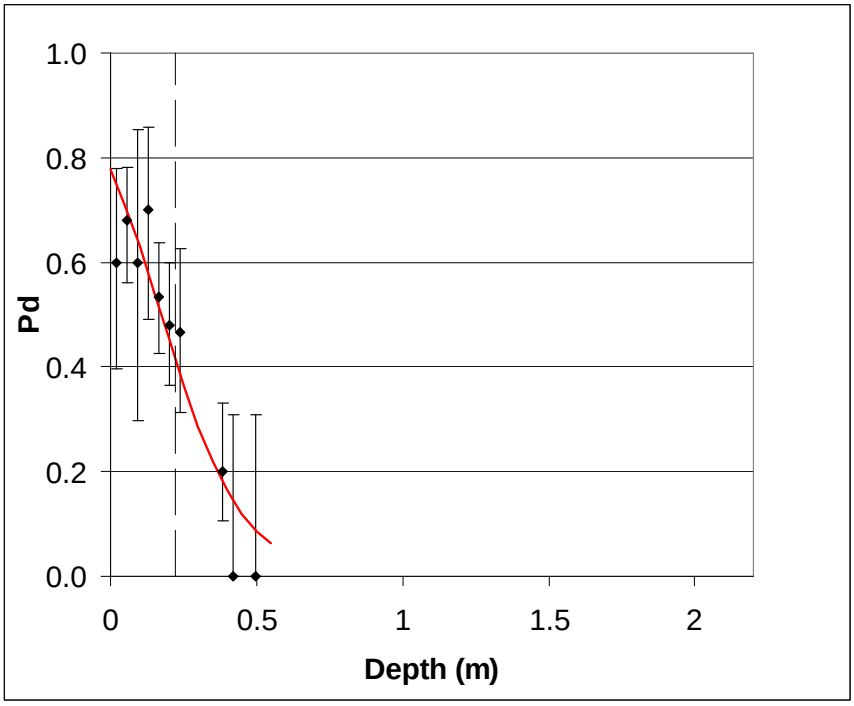

Figure IV-22. 20 mm projectile, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

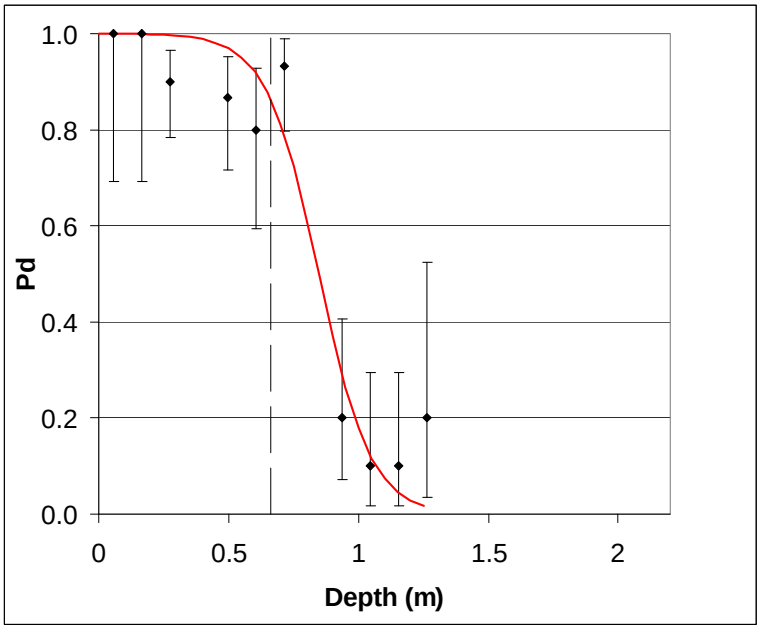

Figure IV-23. 60 mm mortar, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

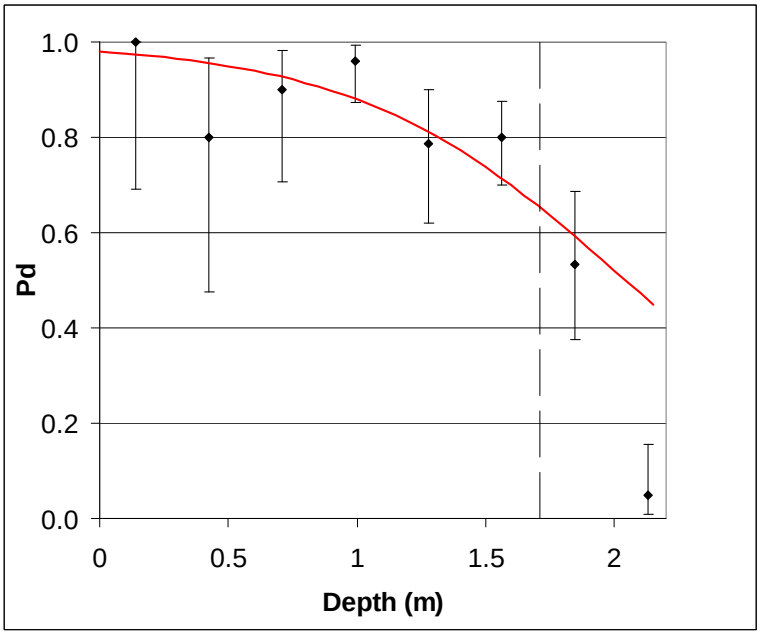

Figure IV-24. 155 mm projectile, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

2. APG, Magnetometer, Probability of Detection as a Function of Depth

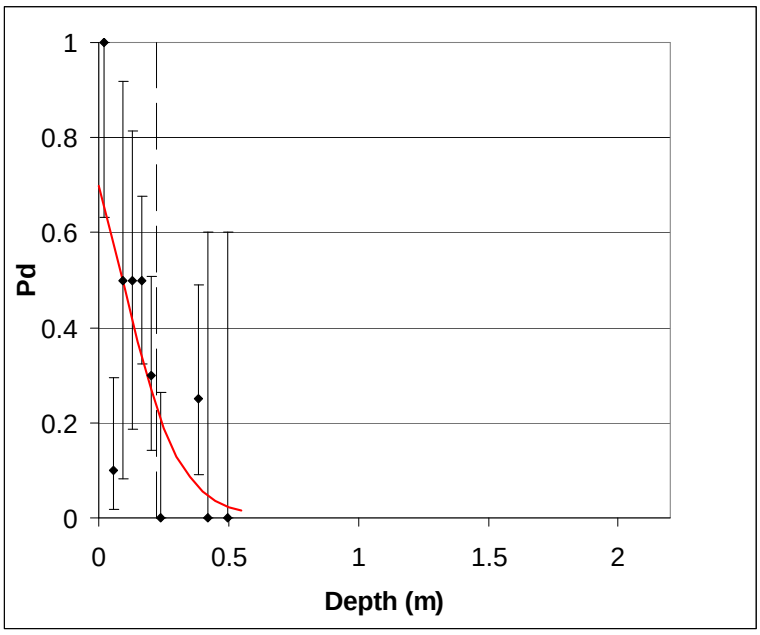

Figure IV-25. 20 mm projectile, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

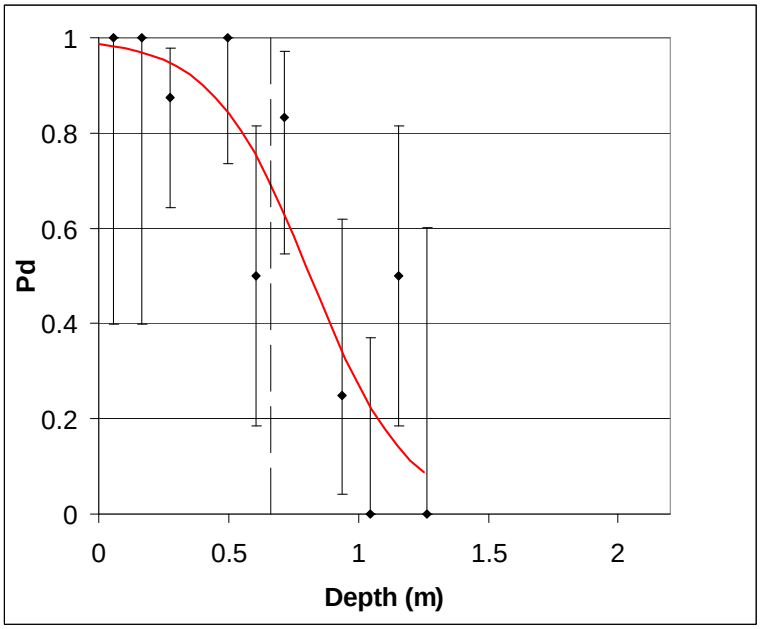

Figure IV-26. 60 mm mortar, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

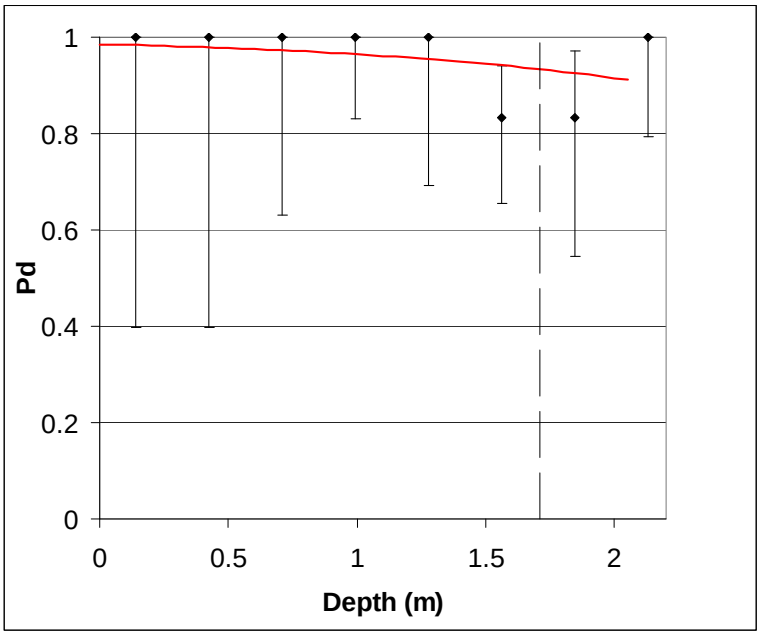

Figure IV-27. 155 mm projectile, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

3. YPG, EMI, Probability of Detection as a Function of Depth

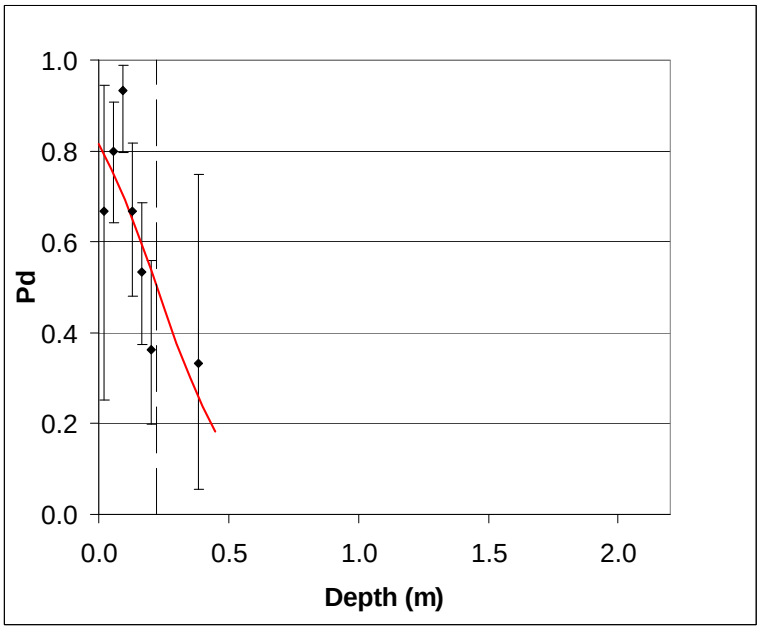

Figure IV-28. 20 mm projectile, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

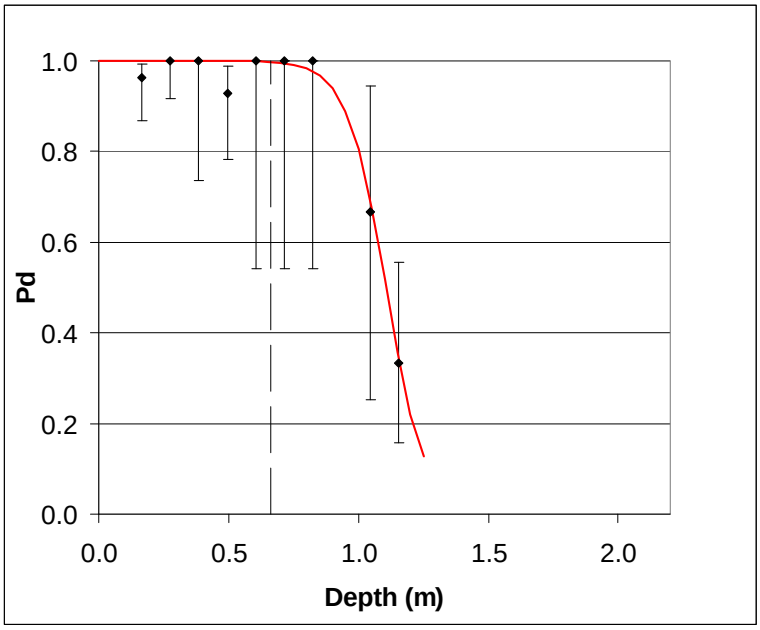

Figure IV-29. 60 mm mortar, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

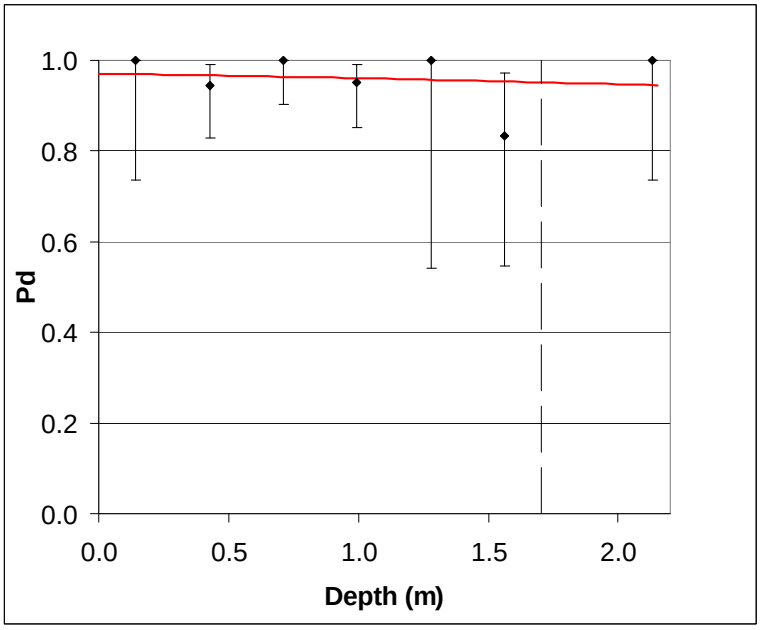

Figure IV-30. 155 mm projectile, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

4. YPG, Magnetometer, Probability of Detection as a Function of Depth

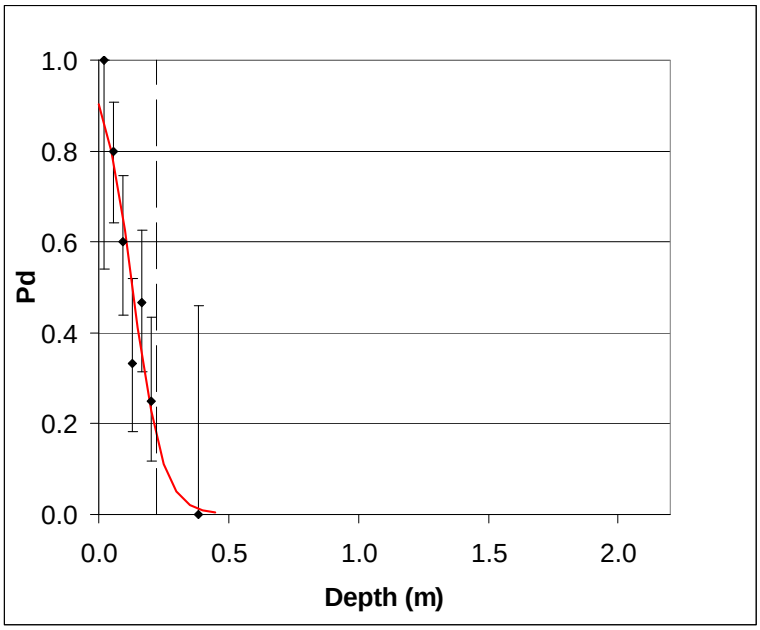

Figure IV-31. 20 mm projectile, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

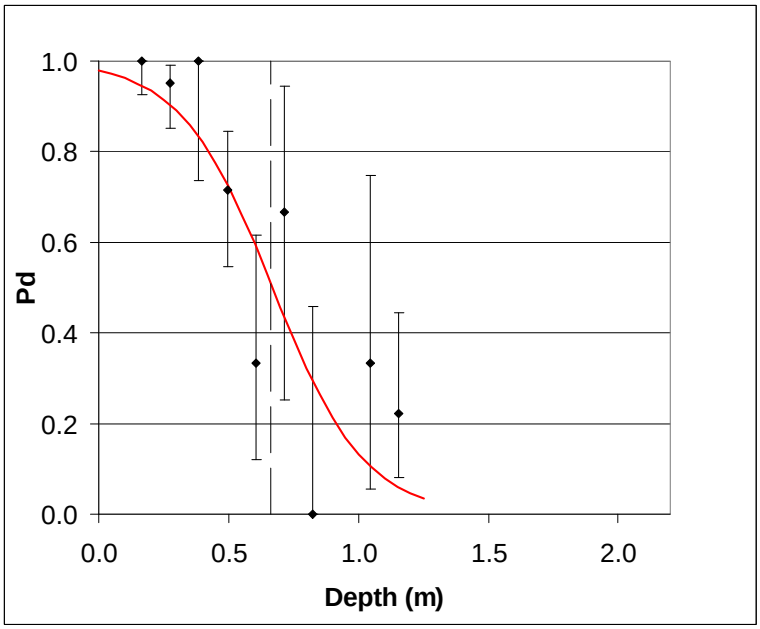

Figure IV-32. 60 mm mortar, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

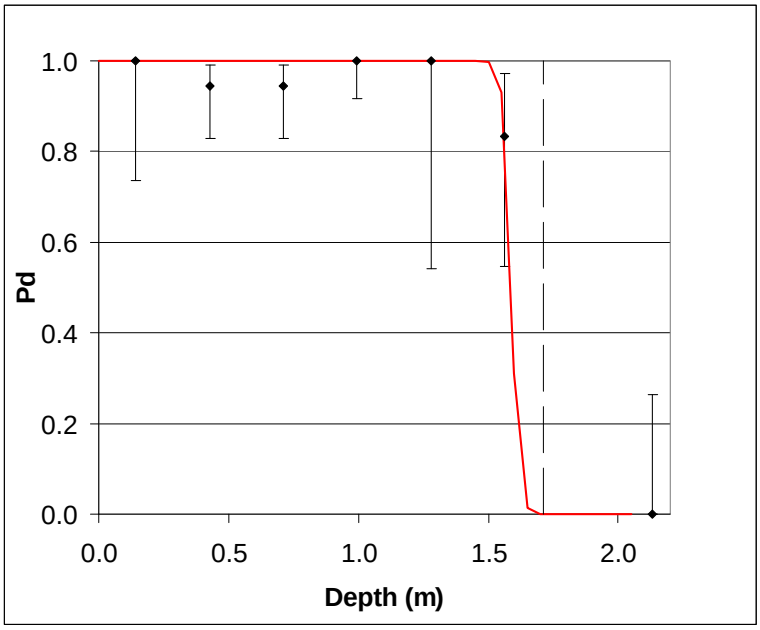

Figure 33. 155 mm projectile, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

Figure IV-34 plots the *a* parameter from each of the fits presented in this section and in the appendix against munition diameter. The *a* parameter describes the depth at which the fitted probability-of-detection curve equals 50%. The 11× depth is also plotted in the figure as a reference. The 11× depth is a rule-of-thumb guide for detectability, but no threshold Pd that should be satisfied for munitions buried at the 11× depth is specified. To compare the detection curves in this section to detectability at the 11× depth, it is assumed that fitted curves whose value is greater than 50% at the 11× depth are consistent with the detectability envisioned by the Corps of Engineers when the rule of thumb was defined.

For 20 mm projectiles the detection curve for the good demonstrators is underperforming the 11× estimate—the detection curve falls to 50% before the 11× depth in all cases. The detectability of larger munitions is more consistent with the 11× rule of thumb. Of course, as shown in the Pd-by-depth plots of this section, the fits do not imply that 100% of the ordnance above 11× is actually detected.

Figure IV-35 shows the scaling factor, *b*, in the fitted probability-of-detection curves vs. munition diameter. This parameter describes the width, in meters, of the depth region where the transition from 100% Pd to 0% Pd occurs. The trend is less well

pronounced than for the *a* parameter, but shows a general tendency to increase with diameter.

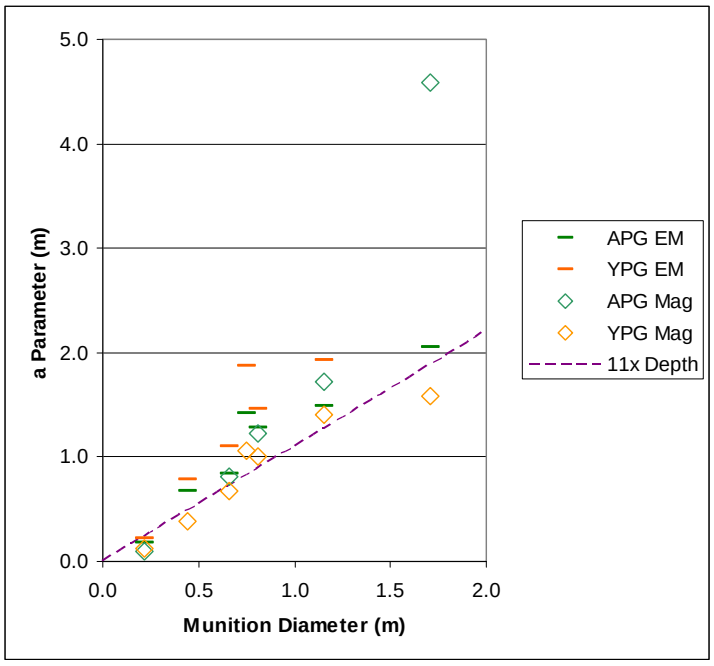

Figure IV-34. Plot of the 50% detection depth (*a* parameter in meters in each fitted probability-of-detection curve) vs. munition diameter. The dotted line represents the 11× depth. The YPG EMI point for the 155 mm projectiles is well off the graph at *a* = 12 m.

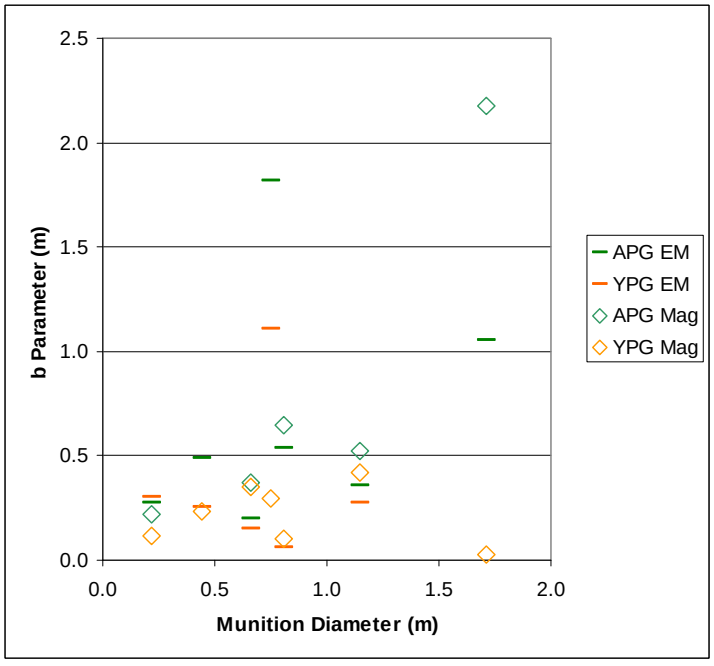

Figure IV-35. *b* parameter in meters. Plot of the transition scale over which the fitted probability-of-detection curve falls from nearly 100% Pd to nearly 0% Pd vs. munition diameter. A smaller value of *b* means a sharper transition from 100% Pd to 0% Pd. The YPG EMI value is off the plot at 6.7 m.

From a theoretical standpoint, EMI technology relies upon a signal that falls off as the sixth power of the sensor-target separation (sensor-induced dipole signal in the target). Magnetometers rely on sensing a signal created by Earth's magnetic field, which is nearly uniform across the regions of interest. The signal sensed by the magnetometers falls off only as the third power of the sensor-target separation. While other factors like sensor and background noise, processing, and transmitter power play a huge role, note that the potentially large signal suppression from three additional powers of the separation distance is not preventing the EMI systems from performing at least as well as the magnetometer systems for this set of target/target depth combinations.

V. INDIVIDUAL MISS AND FAILURE ANALYSIS

Even when shallow, isolated targets are considered, misses still occur. This section examines the reasons for those misses.

Most of the misses at depths above the 11× depth can be attributed to halo effect or to shadowing, rather than a fundamentally low signal. In rare cases, other conditions may be responsible for the miss. There are isolated examples where the signal from the munition was enormous, but the munition was very close to the boundary of the site. We can speculate that the demonstrator processed the data incorrectly and marked the anomaly as out of bounds. Such errors are ignored in this analysis because while they may indeed result in a munition not being marked with an alarm, the source of such errors is difficult to verify without exhaustive records of how the alarm lists were made by the demonstrator. Instead, this part of the study focuses on misses that were apparently due to the site configuration or the demonstrators' quality of coverage, both of which are well documented.

This section also examines some of the poorer performing demonstrators. Analysis of raw data from some of the excessively high rBAR or low Pd demonstrations reveals likely reasons for those results. For example, the Blackhawk data were exceptionally noisy. Their high rBAR[12] at APG and YPG indicates a demonstrator-specific systematic problem. In another example, the GeoCenters APG magnetometer raw data were not leveled properly. Two magnetometers in the array consistently read 20 nT higher than the others. GeoCenters' low Pd in conjunction with a low rBAR is consistent with a threshold set too high to avoid the leveling problem or with statistical noise induced by correcting it after the data were taken.

A. NOISE COMPARISON OF "POOR" AND "GOOD" DEMONSTRATORS

Figures V-1 through V-3 show the same region of the APG Open Field—a zoomed-out region around the large cluster shown in Figure IV-2. These figures demonstrate the variability in performance between demonstrators. The gridded data are

---

12 In fact, at APG roughly 20% of the site was within some alarm halo. Blackhawk's results from both sites are severely prejudiced by lucky hits.

from three sensors operated by BlackHawk, GeoCenters, and NRL. The first two were part of combined EMI and Mag arrays, where only the output from one of the sensors was used in the figures (an EM61 type for Blackhawk and a cesium-vapor magnetometer for GeoCenters). The NRL data are from an array of EM61-type sensors. The NRL data were selected as a good demonstrator and they are included for comparison to the Blackhawk data.

In these figures, a 9 m × 7 m box is shown. There are no emplaced targets in this area. The average and standard deviation of the gridded data are recorded below each figure. While the statistics inside this box are not representative of the entire site's background, the nature of the signal relative to the large nearby cluster is instructive.

The Blackhawk data are exceptionally noisy. A diagonal band of increased noise repeats across the site. Blackhawk's data are excessively noisy at YPG, too. The alarms are indicated in the figure, and the source of the high rBAR in the pseudo-ROC plots of Section IV.D.3 (p. IV-15) is obvious.

GeoCenters data show a DC offset of about 20 nT for two of the magnetometers in the array. GeoCenters has very few alarms compared with Blackhawk. The offset may have limited how low the threshold could be set. The grid shown is *not* necessarily a visualization of the same quantity used by GeoCenters to arrive at the confidence levels they reported. For example, GeoCenters may have attempted to remove the offset, and its relatively low overall Pd scores may reflect uncertainties in that process. The details of the processing are simply not known. The NRL data are much cleaner, with almost no mean offset and a small standard deviation. This likely represents near-ideal operation of the EM61MKII for the environment.

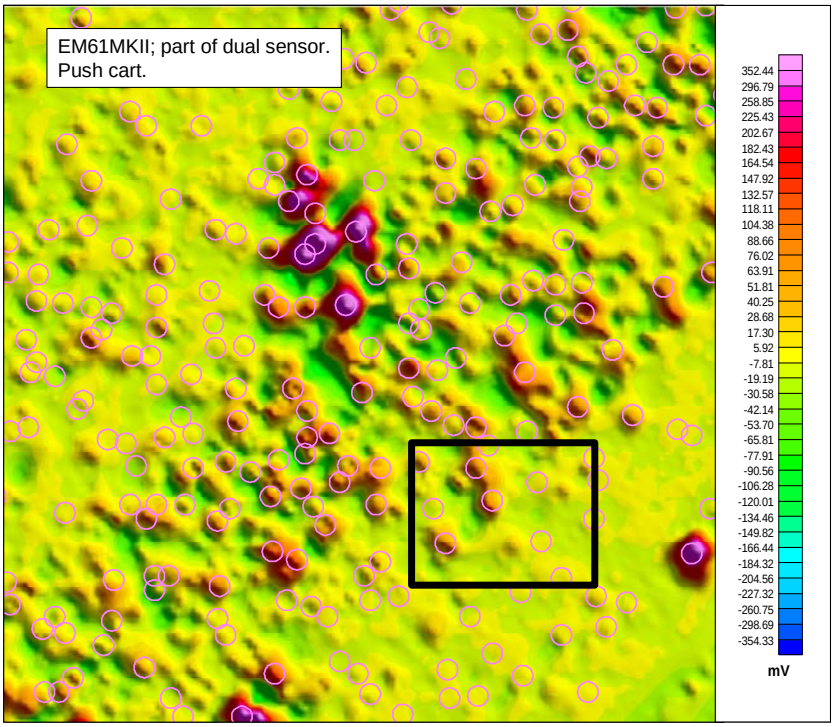

Figure V-1. BlackHawk EMI noise. Within the 9 m × 7 m box, the signal ranged from –40 mV to 80 mV, with a mean and standard deviation of –5 mV and 15 mV, respectively.

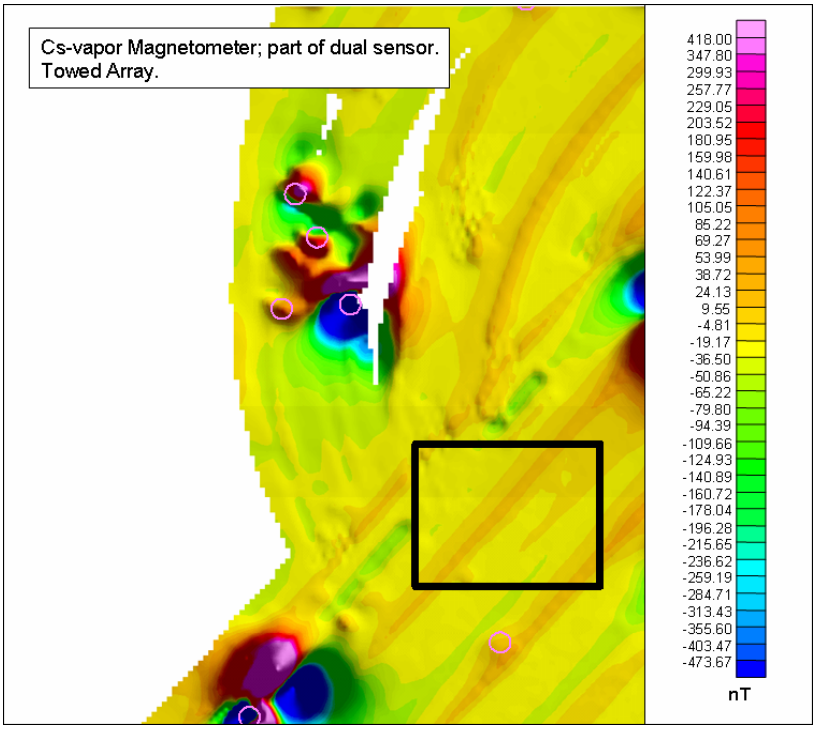

Figure V-2. GeoCenters STOLS magnetometer noise. Within the 9 m × 7 m box, the signal ranged from –40 nT to 10 nT, with a mean and standard deviation of –20 nT and 11 nT, respectively.

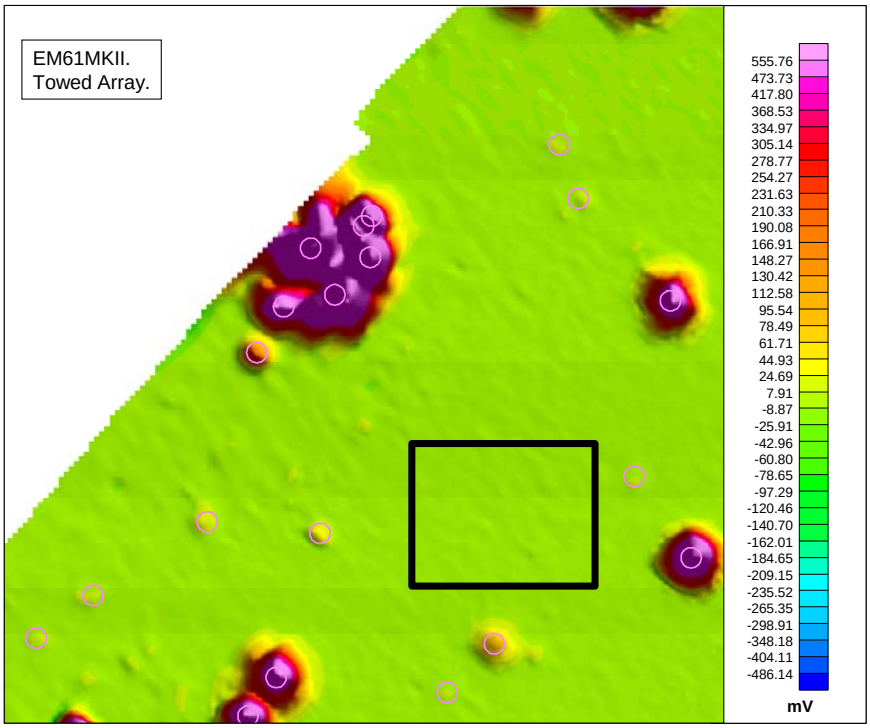

Figure V-3. NRL EM61 Noise. Within the 9 m × 7 m box, the signal ranged from –9 mV to 7 mV, with a mean and standard deviation of –0.5 mV and 1.6 mV, respectively.

Shadowing Misses

Removing overlapping signals in aggregate Pds generally produces a Pd higher than one calculated by including overlaps. The size of the difference depends on the number of overlapping signals emplaced at the Standardized Sites. Overlapping signals are likely to be encountered at real-world sites, though their number depends on the anomaly density at each site.

Overlapping signal misses are the most worrisome type of miss observed at the Standardized Sites because shadowed items may fall in the depth range where near 100% detection is assumed. While the standard operating procedure advocated by the Corps of Engineers requires “clearing the hole” with a hand-held sensor after excavating suspected UXO, shadowed items may be far enough from the larger anomaly that they would not be found. The standardized test sites do not test the efficiency of clearing holes after an excavation, so the existence of shadowing events underscores the importance of clearing the hole without offering quantitative information on clearing technique.

Figures V-4 through V-9 show two cases of shadowing for three different sensors. The first case (Figures V-4 through V-6) is an 81 mm mortar (black X) at 40 cm depth. It is 1.4 m from a 4 to 10 kg clutter item (red X) that is at 10 cm depth. The 81 mm mortar

was found by only three demonstrators at the site. The second case (Figures V-7 through V-9) is a 60 mm mortar that is at its 11× depth (66 cm). It is 1.2 m from a 57 mm projectile that is buried at 25 cm. The 60 mm mortar was found by only two of the demonstrators at the site. The data shown in these two examples are from three sensors: an EM61MKII type time-domain electromagnetic sensor, a GEM3E type frequency-domain electromagnetic sensor, and a cesium-vapor magnetometer. They were used by TTFW and NRL (both the GMTADS and MTADS magnetometer). The EM61 data are from the bottom coil in the first 366 µs timegate. The GEM3E data are an average over the quadrature midrange frequencies and the magnetometer is the magnitude of the total field. While the raw data could be reanalyzed, emphasizing different sensor channels, it is not clear that this would greatly increase the chances of finding shadowed items. These examples were missed by demonstrators who scored very well overall and apparently had a good way to construct confidence levels from multichannel sensors. Exploiting a particular signature to detect multiple targets would be a valuable tool, but is beyond the scope of this analysis.

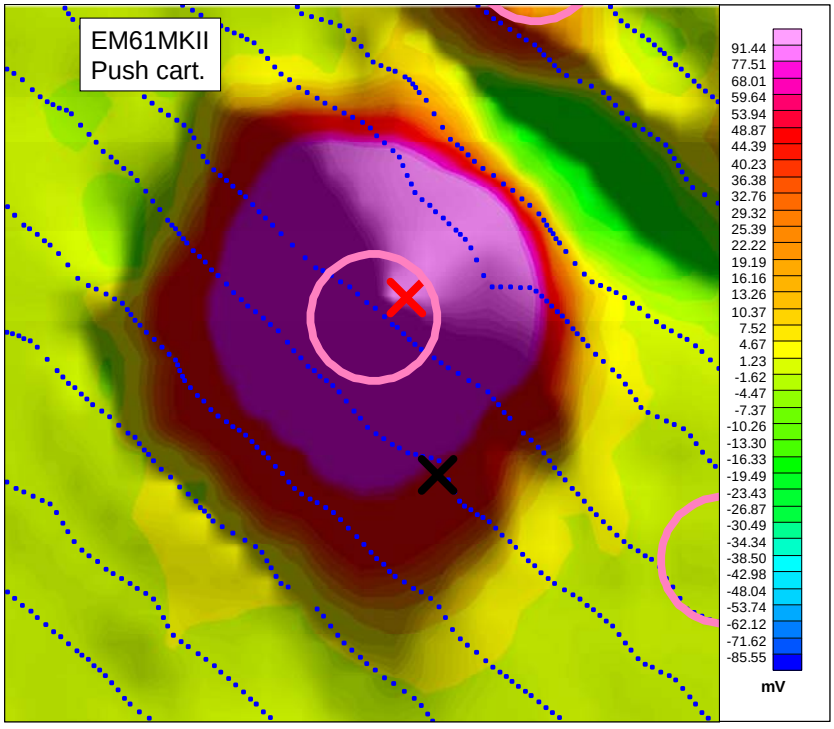

Figure V-4. Shadowed 81 mm mortar at 40 cm. The red X marks a "4–10 kg" clutter object at 10 cm depth. EM61MKII pushcart data.

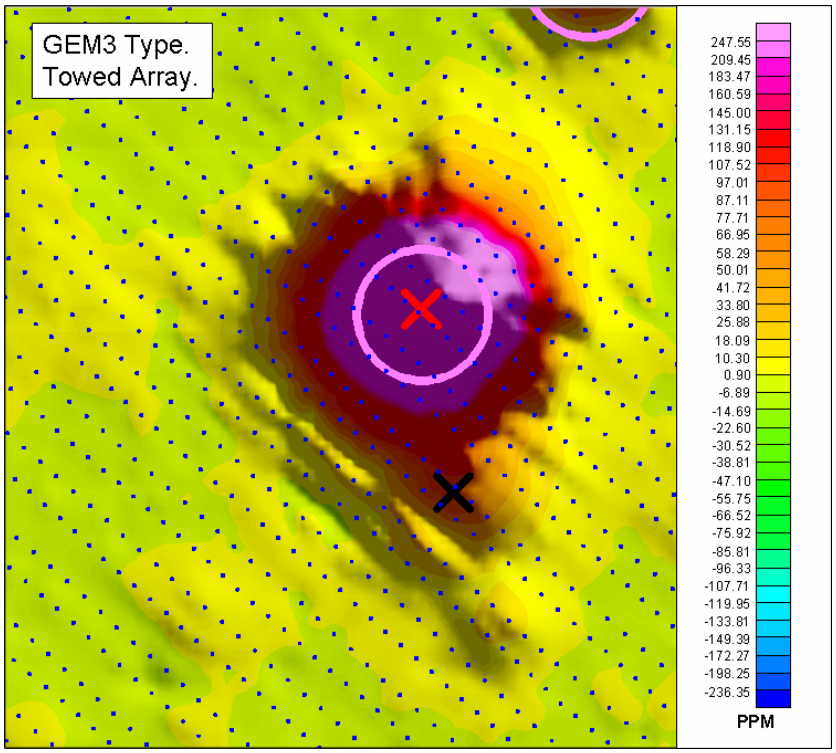

Figure V-5. Shadowed 81 mm mortar at 40 cm. The red X marks a "4–10 kg" clutter object at 10 cm depth. GEM3E towed-array data.

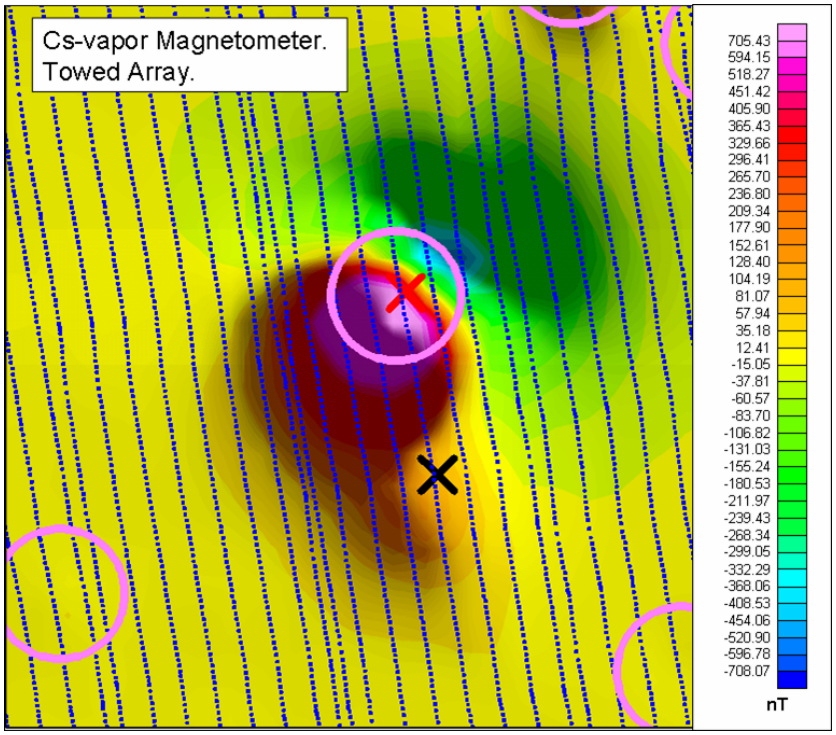

Figure V-6. Shadowed 81 mm mortar at 40 cm. The red X marks a "4–10 kg" clutter object at 10 cm depth. Cesium-vapor magnetometer towed-array data.

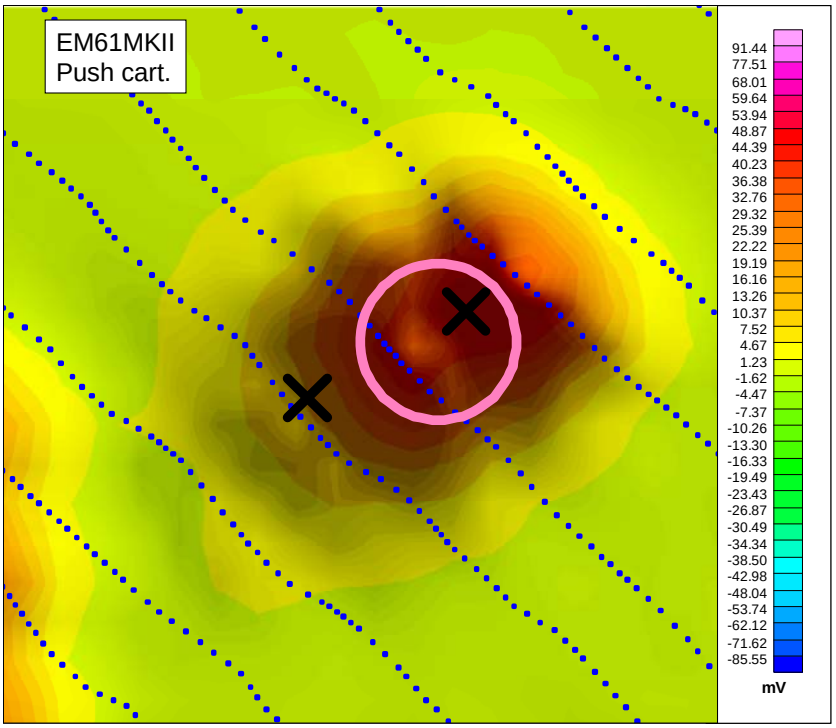

Figure V-7. Shadowed 60 mm mortar (lower left) at 66 cm. The other item is a 57 mm projectile at 25 cm, EM61MKII pushcart data.

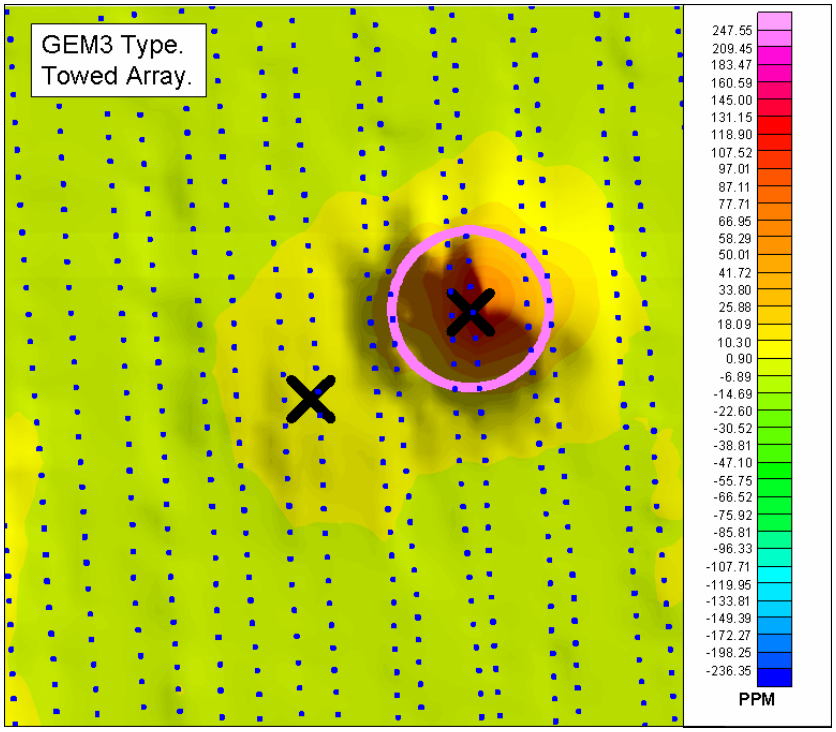

Figure V-8. Shadowed 60 mm mortar (lower left) at 66 cm. The other item is a 57 mm projectile at 25 cm, GEM3E towed-array data.

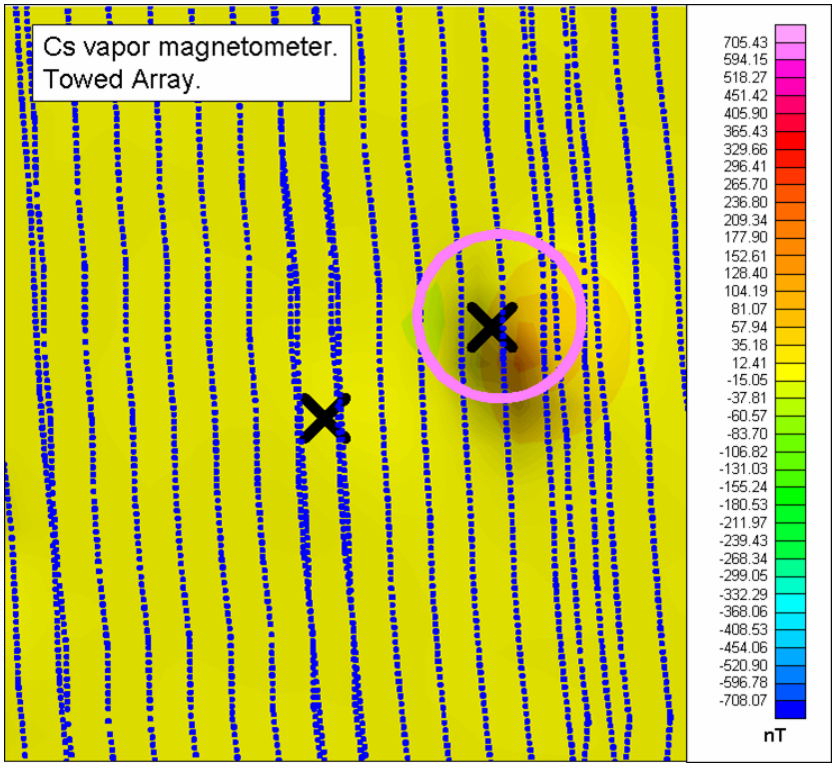

Figure V-9. Shadowed 60 mm mortar (lower left) at 66 cm. The other item is a 57 mm projectile at 25 cm, Cesium-vapor magnetometer towed-array data.

B. HALO EFFECT

After shadowing, the next most common reason for a miss was halo effect—an alarm near the target, but outside the scoring halo. Some of the alarms scored as misses by the software were only a few centimeters outside the scoring halo. In some instances, the horizontal extent of the missed target's signature was larger than the scoring halo, and it is possible that if the dig list was actually excavated, the target would have been found. Note that there is a significant distinction between finding a target and discriminating a munition from clutter. This analysis focuses on the response (or detection) stage. In the discrimination stage, some of the detected targets are declared as clutter. In a real cleanup, they would be left in the ground. A large part of the penalty associated with the halo effect is transferred to the discrimination stage. The location error associated with halo-effect misses may be detrimental to the physical analysis of the signal that determines what items are safe to leave behind.

Although these near misses are called "halo effects," this does not mean that the halo is too small to accurately reflect target finds. Rather, the number of near misses that does occur illustrates a limitation of the scoring system: the selection of a definite size for

the halo. Any halo of definite size will return near misses. An alternative (albeit a complicated one) would be to score the alarms with a smooth function that decreased with distance from the target.

A contributing factor to the halo effect is survey-track spacing. TTFW demonstrated a pushcart using a 0.5 m track spacing. TTFW did well overall; however, it had more halo-effect misses than any of the other demonstrators in the better demonstrator analysis (Section IV.E.2 (p. IV-30)). The actual track spacing was irregular and diverged to 1 m in some places. Arrays typically had track spacing of 20–40 cm. For medium and large munitions targets, most demonstrators' regular track spacings intersected a target's signature in several places. For smaller targets (like the 20 mm projectile), good coverage (navigation) and narrow track spacing were required to ensure several encounters with the target. Low data density near the target will affect the ability of algorithms to invert the anomaly and match it to a buried source at a particular location. Figures V-10 and V-11 show some examples of the halo effect.

C. 20 mm PROJECTILES

The 20 mm projectiles were excluded from the miss analysis because they were missed so frequently. Although shadowing and halo effects were seen for 20 mm projectiles, they were often missed because of a low signal. For 20 mm projectiles, the Pd-by-depth graphs in Section IV.F (p. IV-34) show that even for the shallowest items, the Pd does not approach 100%. For larger items the horizontal distance from the sensor's location to the target is usually irrelevant since many survey tracks intersect the anomaly caused by the target. Intended survey-track separations vary from 20 cm for closely spaced arrays up to 50 cm for some pushcarts. Poor navigation can increase this distance from lane to lane. For 20 mm sized targets, often just one or two survey tracks cross over the anomaly caused by the target. Each sensor model's sensitivity varies across its width, and the target signature is also a function of its orientation. Fewer encounters with the target increase the likelihood that it will be encountered in only a less sensitive way. This compounds the difficulty of the already inherently small signal.

Figures V-12 through V-14 show four 20 mm projectiles that happened to be buried near each other at APG. Two of the four items are above the 11× depth. The data are from the TTFW's EM61MKII type sensor, the GMTADS GEM3E type, and the MTADS magnetometer. The TTFW pushcart found two of the four projectiles, but did a relatively poor job of locating them. The NRL MTADS array variants had narrower track separations and fared better at placing the alarm near the target. We did not do a

comprehensive analysis of signal or Pd as a function of track spacing, but evidence suggests that track spacing becomes important for small targets where the horizontal scale of the anomaly is about the same size or smaller than the track spacing. That is, even with perfect navigation, sufficiently dense survey tracks should be designed when looking for smaller targets.

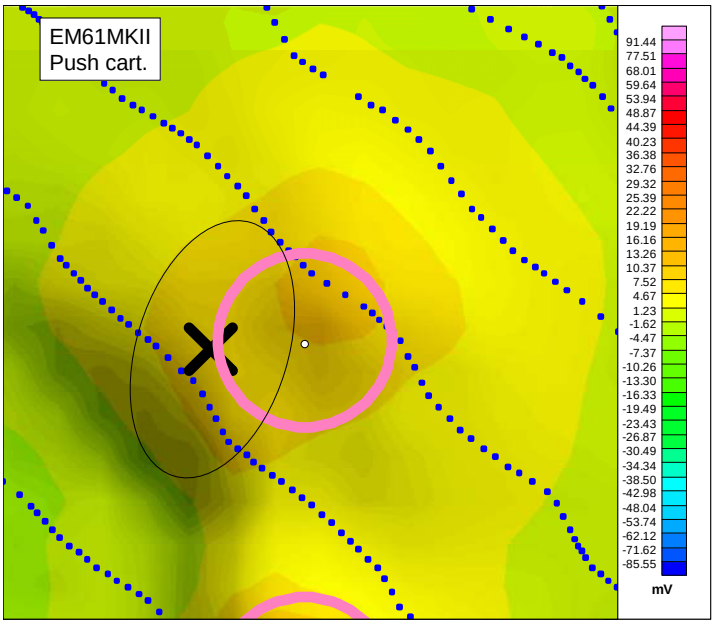

Figure V-10. Halo effect. TTFW EM61MKII type sensor. 155 mm projectile that is 1.6 m deep. Found by 12 of 20 demonstrators at APG. The oval scoring halo around the large 155 mm is explicitly shown. The pink circle is 1 m in diameter and centered on the alarm (white dot). Note that TTFW had the most instances of halo-effect misses. TTFW did well overall, but irregular and wide (0.5 m) track spacing seems to have hurt its performance.

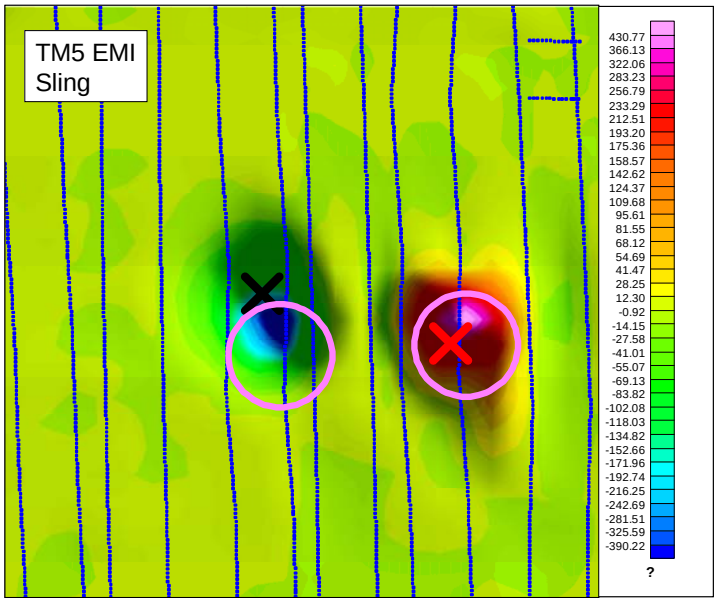

Figure V-11. BDU-28, 15 cm deep at YPG. Found by 15 of 19 demonstrators. The data are from the Gtek TM5 sensor. The question mark below the legend illustrates another ambiguity in analyzing raw data. The Gtek data were smoothed by high and low pass filters although many of the details of this process were not reported. This scale is likely in mV units. While the demonstrators were often helpful in attempting to reconstruct their analysis, they were not required to exhaustively document it.

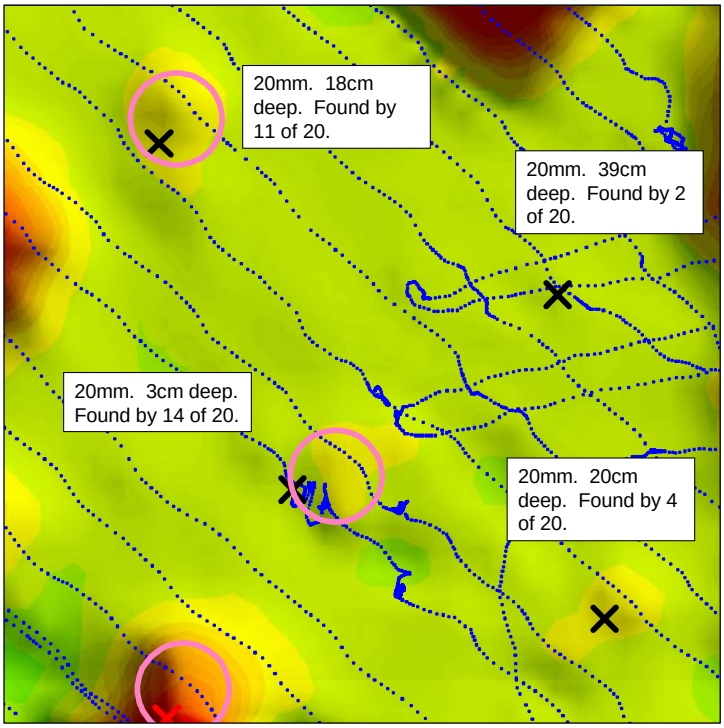

Figure V-12. EM61MKII type sensor operated by TTFW. Four 20 mm projectiles at APG. Note the difficulty locating the 3 cm deep target and the 20 cm deep target without an alarm. While overall results were good from this demonstrator, irregular track spacing is a concern.

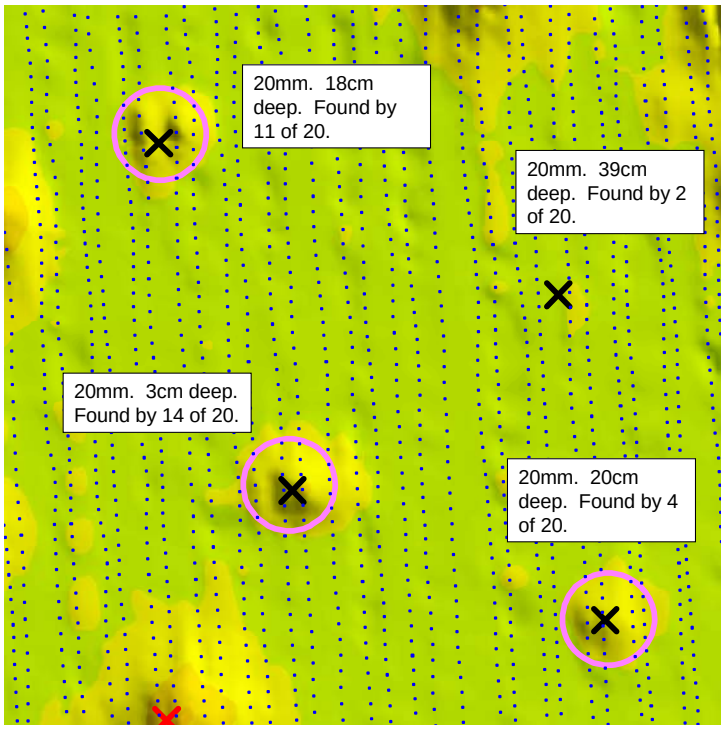

Figure V-13. GMTADS GEM3E type sensor. The GMTADS was used as an array with three sensors on the towed platform. Note the denser and more regular data spacing compared to the TTFW pushcart.

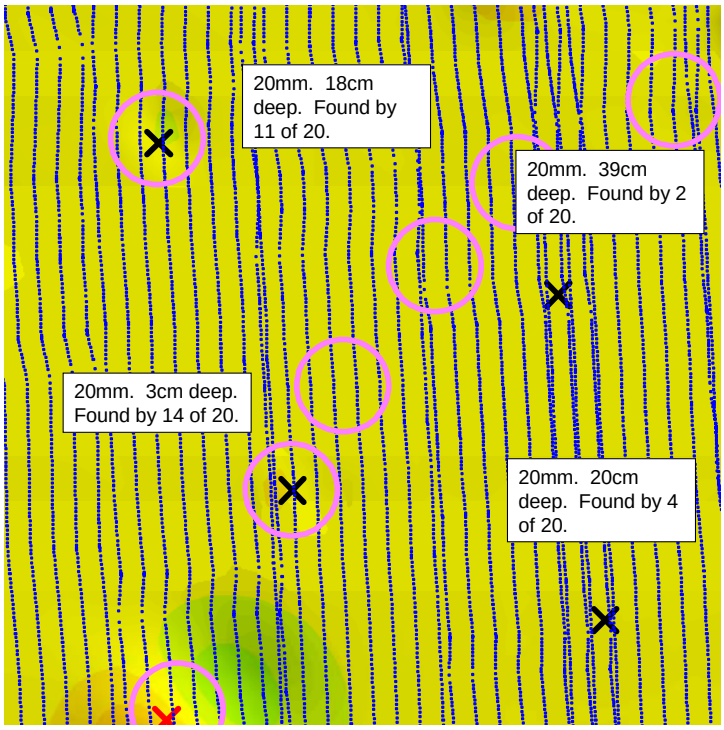

Figure V-14. The MTADS cesium-vapor magnetometer. High track density but less sensitivity to the 20 mm projectiles.

VI. CONCLUSIONS

Results from the Standardized UXO Test Sites are reported in terms of probability of detection and rBAR. Results reported in this way are *site specific* and depend on the scoring system. Range of burial depths, the type of munitions at the site, and a demonstrator's operation of sensor technology at that site will affect detection performance. The two largest issues with the scoring system are the handling of large clusters and the halo effect. The fixed-halo scoring system tends to underestimate the number of munition targets in large clusters that would be excavated in a real-world cleanup action. Isolated targets counted as misses because of the halo effect defined in this study may also be reacquired and successfully excavated during a cleanup if the intent is to excavate *all* the anomalies that were detected. If an attempt is being made to *discriminate* which anomalies do not need to be excavated, the location error associated with halo-effect misses may be detrimental. The number of these types of misses at the Standardized Sites depends on the number of clusters that happened to be emplaced and the number of halo-effect misses. Note that the number of halo-effect misses depends upon how data were collected and analyzed by each demonstrator and on inherent positioning error in survey equipment.

The Standardized Sites have several characteristics that differentiate them from a real-world UXO clearance. First, some of the munitions are buried at challenging depths. A real-world cleanup action would set standards for success that accounted for the capability of the sensor as well as likely UXO penetration depths. Second, the Standardized Sites contain a wide variety of munitions to test a large part of the spectrum of detection capability. Real-world ranges may have had a very limited purpose (and consequently few munition types), so the cleanup plan could be optimized for those munitions. Third, the relative numbers of targets and their depth distribution may not be indicative of a real-world site. Real-world sites have vastly more clutter than UXO.

Several filters were introduced in this analysis to obtain results from the Standardized Sites that were applicable to a more restricted class of target (e.g., above 11×). The objective was to measure the Pd on isolated targets that were actually surveyed and to understand the effect of the difficult depth distribution at the sites. This facilitates a comparison to common practice at geophysical prove-outs.

After adding the filters, the two most common causes for misses by better performing demonstrators, shadowing and the halo effect, affected the overall estimation of real-world performance in unknown ways. Shadowing is simply the limiting case of a small cluster. Limiting the target set to items isolated by an arbitrary distance removes some ambiguity from the scoring method, but it does not reward a demonstrator for any ability to separate nearby targets. Which target groupings should be separable and which should be considered too difficult, given current technology, are subject to debate. Because halo-effect misses were also demonstrator specific, no uniform filter could be applied to the overall Pd scores to account for them. In addition to the common misses, a very few items were missed for no obvious reason, even by the better performing demonstrators. It is hypothesized that errors occurred in the demonstrator's data analysis. The Standardized Site data reporting system was not designed to track these errors.

Despite the ambiguities in the scoring system, the ground-truth filters, the Pd-by-depth analyses of better demonstrators, and the Pds for munition types indicate that *targets larger than a 60 mm mortar that are above the 11× depth should be found greater than 90% of the time*. With optimum data analysis and site coverage this percentage should be nearer to 100%. For smaller targets, especially those as small as a 20 mm projectile, it is not clear that this percentage will approach 90% without a search designed particularly for finding small targets. While it would be of great value to regulators and stakeholders in UXO cleanup actions to precisely specify the deviation from 100% detection expected in a particular scenario, the Standardized UXO Test Site results do not provide such precision. Given the number of identically buried like-type munitions required to make very precise Pd estimates, the number of possible depth and location configurations, and the uncertainty inherent in not physically excavating found targets, it is difficult to envision a practical test site that probes universal variables of the UXO cleanup problem with great precision.

REFERENCES

1. *Survey of Munitions Response Technologies*, SERDP, ESTCP, ITRC, 2006.

2. T. Altshuler et al., *Demonstrator Performance at the Unexploded Ordnance Advanced Technology Demonstration at Jefferson Proving Ground (Phase I) and Implications for UXO Clearance*, IDA Paper P-3114 (Alexandria, Va.: Institute for Defense Analyses, 1995).

APPENDIX

A. 100% DETECTION DEPTHS AND DEPTHS OF DEEPEST DETECTION

This section shows more bar and whisker charts like those of Section IV.D.4 (p. IV-21). A selection of munitions that span type and size are presented. The solid blue bar represents the 100% detection depth. The whisker is the depth of the deepest found target.

1. APG

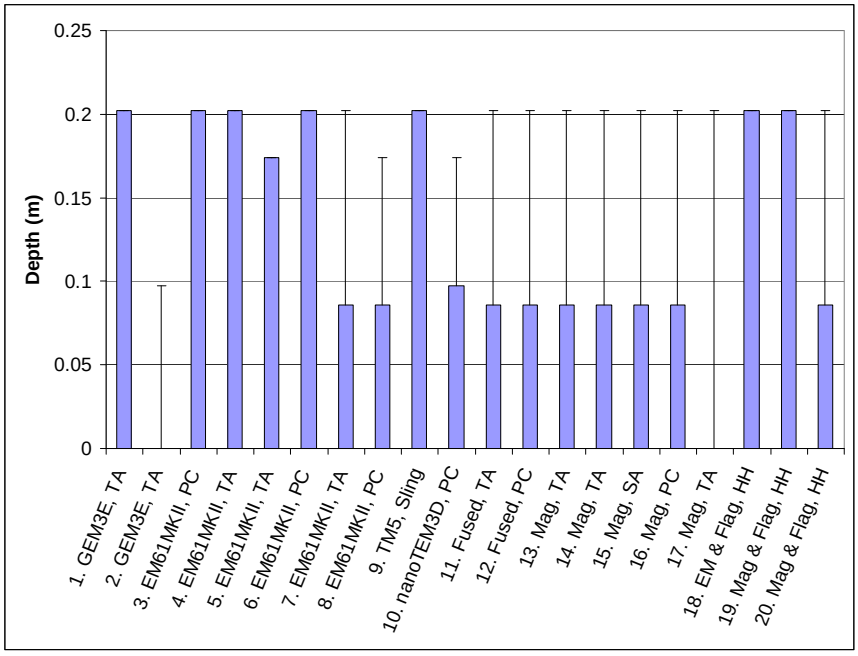

Figure A-1. BDU-28 bomblet, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for APG Open Field demonstrators. The 11× Corps of Engineers depth is much deeper than the deepest seeded target; bomblets are not expected to penetrate very deeply.

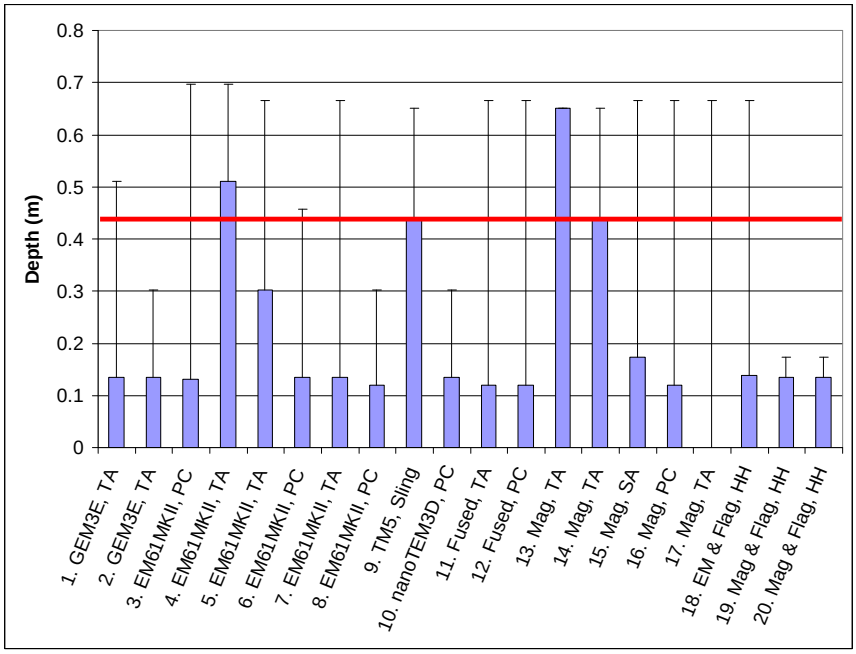

Figure A-2. 40 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for APG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

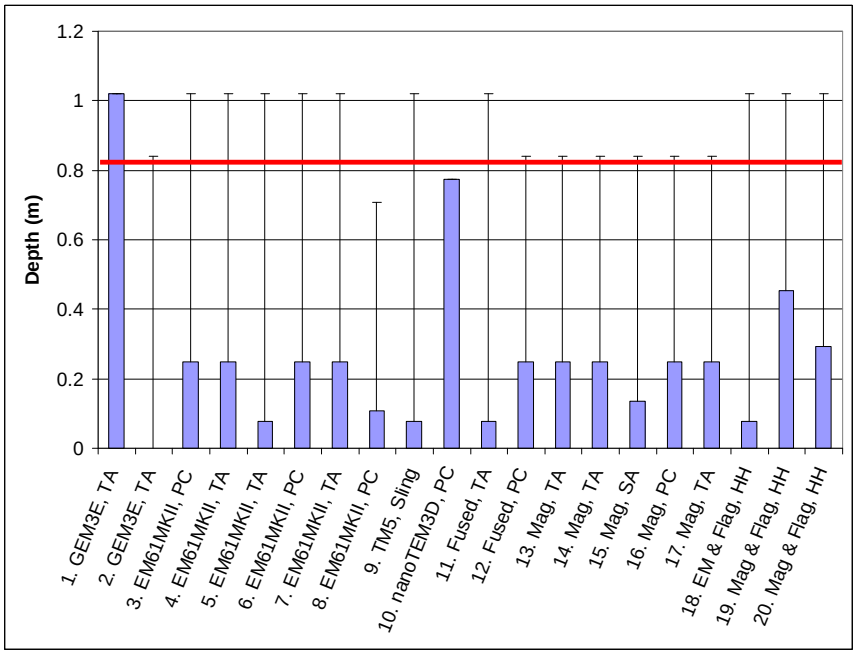

Figure A-3. 2.75-inch rocket, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for APG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

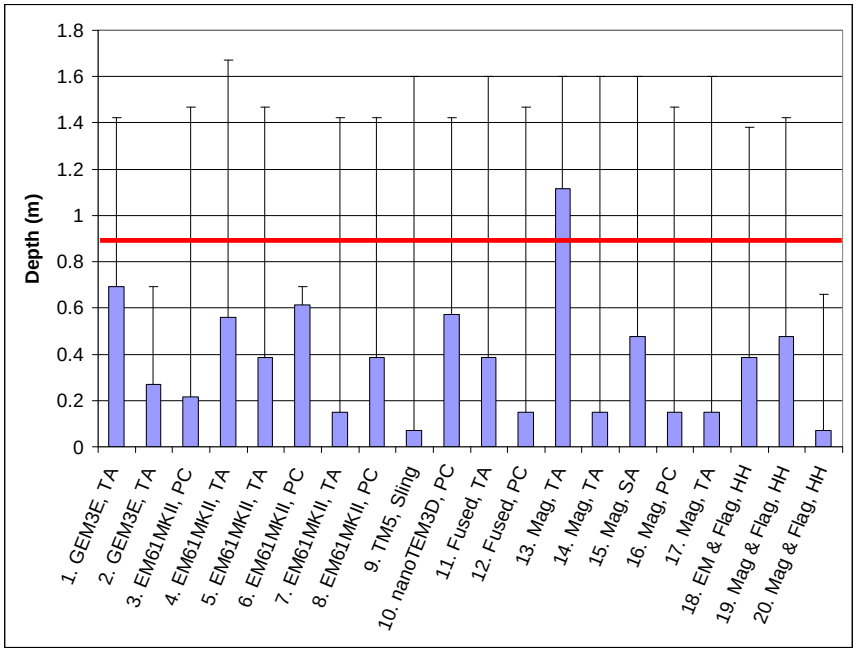

Figure A-4. 81 mm mortar, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for APG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

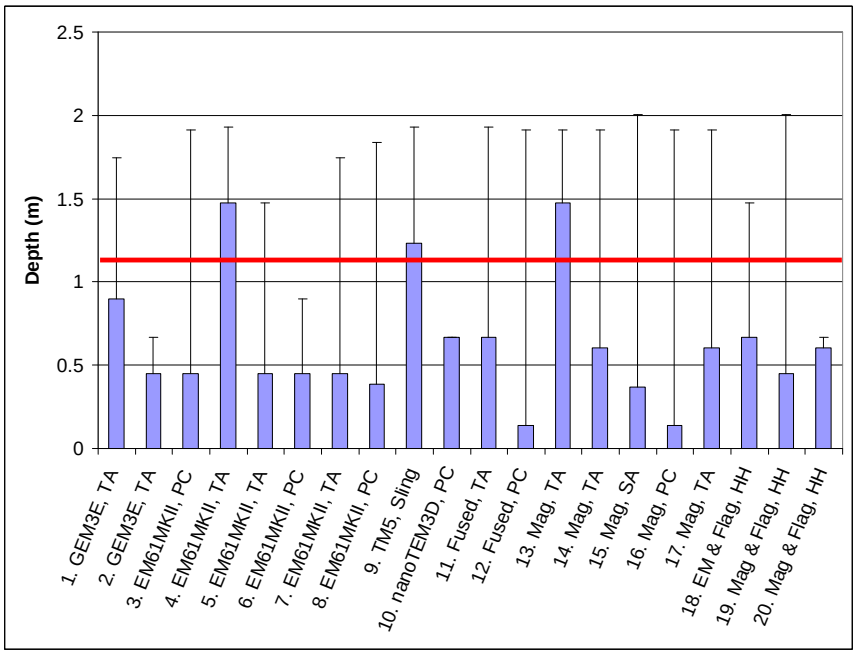

Figure A-5. 105 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for APG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

2. YPG

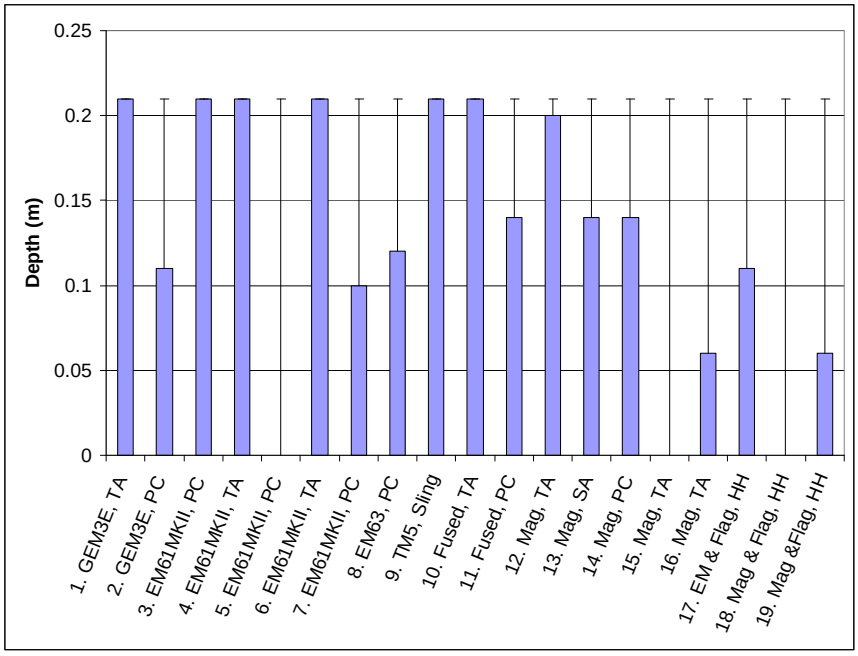

Figure A-6. BDU-28 bomblet, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for YPG Open Field demonstrators. The 11× Corps of Engineers depth is much deeper than the deepest seeded target; bomblets are not expected to penetrate very deeply.

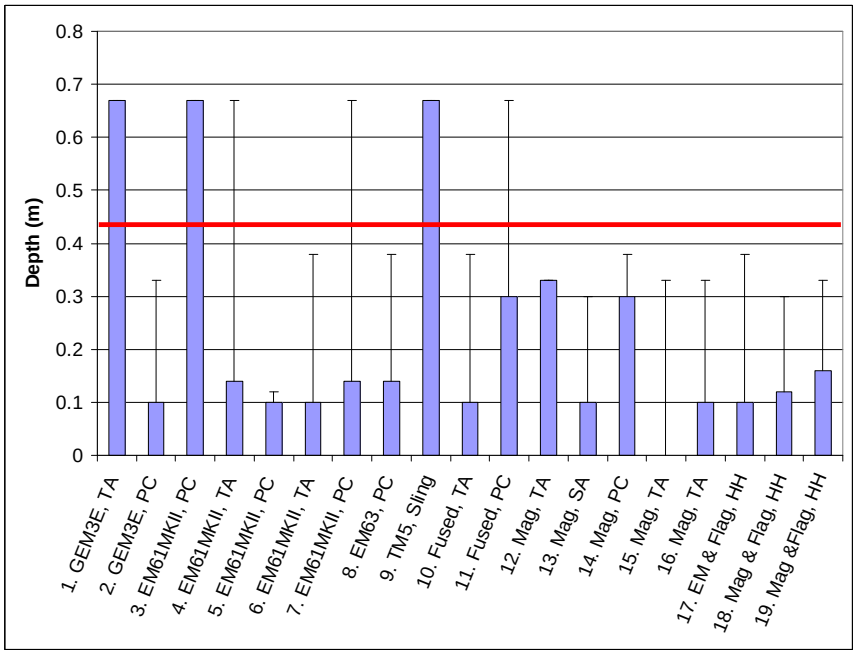

Figure A-7. 40 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for YPG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

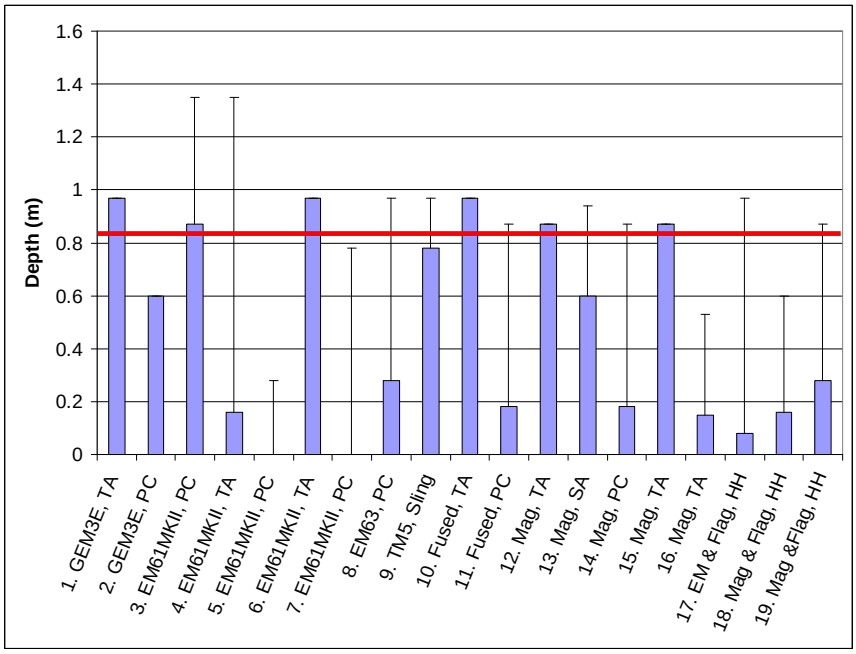

Figure A-8. 2.75-inch rocket, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for YPG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

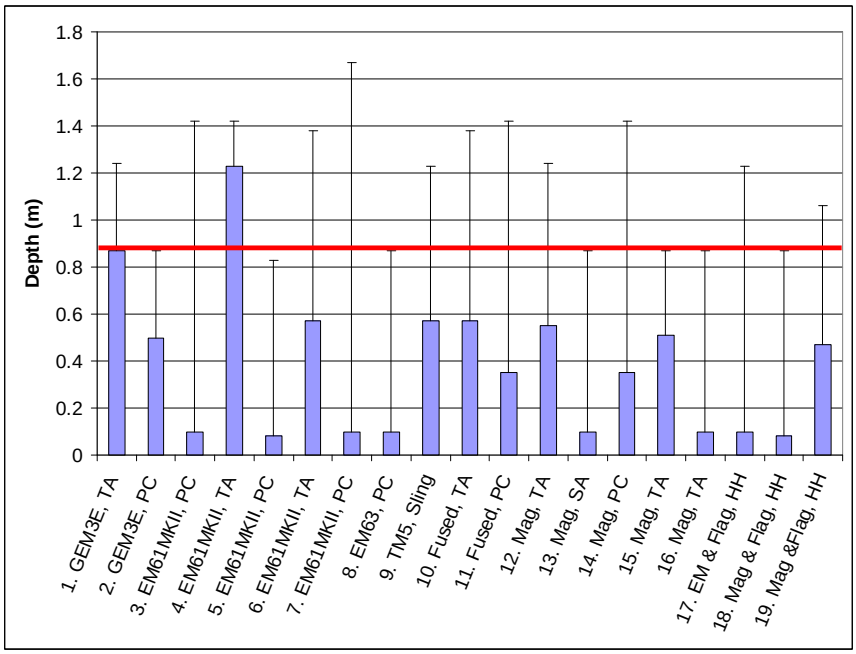

Figure A-9. 81 mm mortar, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for YPG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

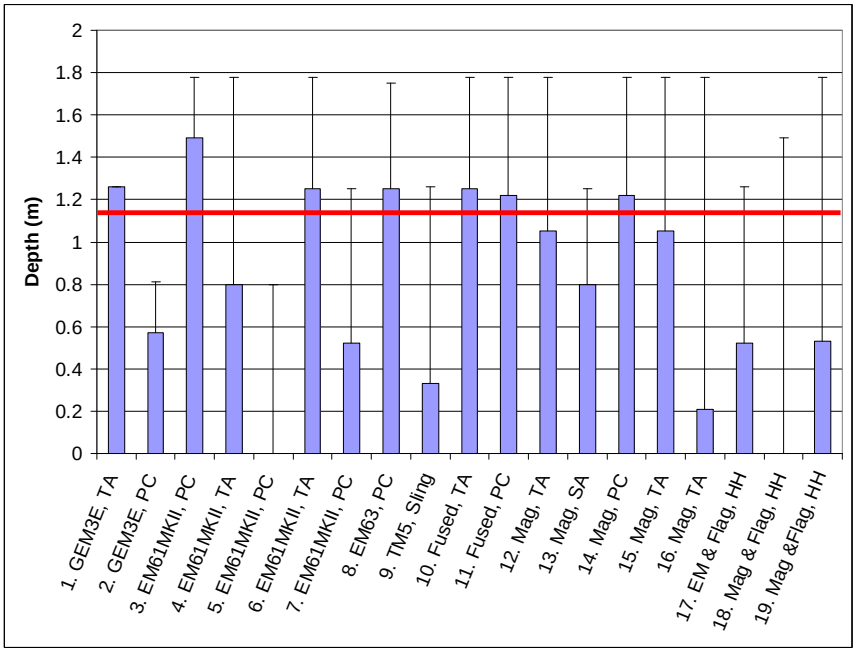

Figure A-10. 105 mm projectile, 100% detection depth (solid bars) and depth of deepest detection (horizontal hash mark) for YPG Open Field demonstrators. The red line is the 11× Corps of Engineers depth.

B. Pd AS A FUNCTION OF DEPTH

This section presents additional Pd as a function of depth graphs for a selection of munitions that spans size and type. A detailed explanation of the graphs is in Section IV.F (p. IV-34). Recall that each graph uses hits and misses that are summed across a set of "good" demonstrators to calculate the Pds that are shown. The graphs are sorted by munition type, sensor technology, and test site.

1. APG

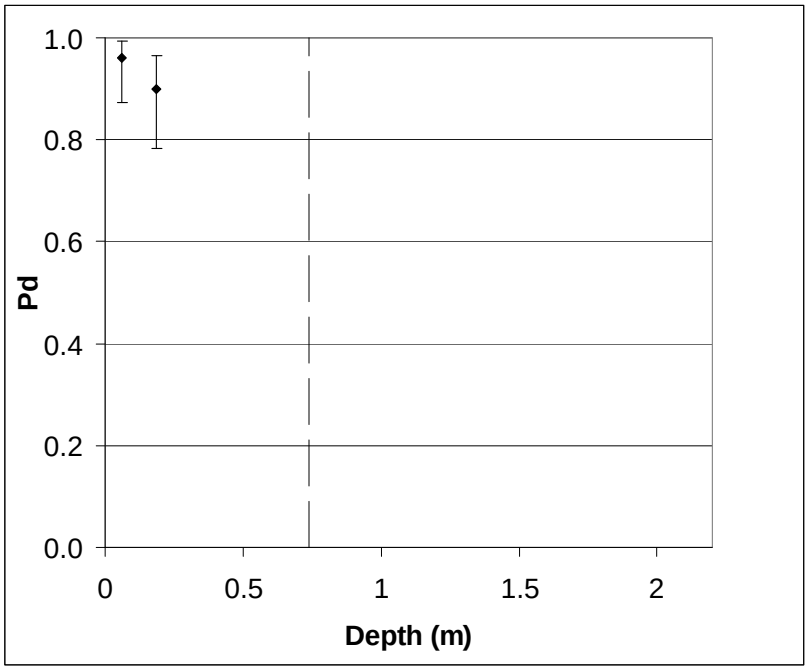

Figure A-11. BDU-28 bomblet, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The dashed vertical line marks the 11× Corps of Engineers depth.

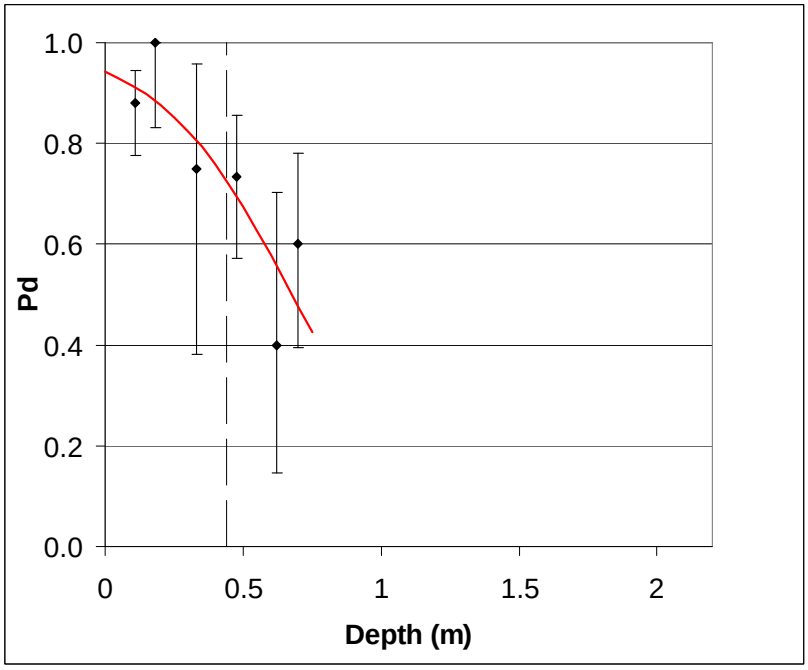

Figure A-12. 40 mm projectile, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

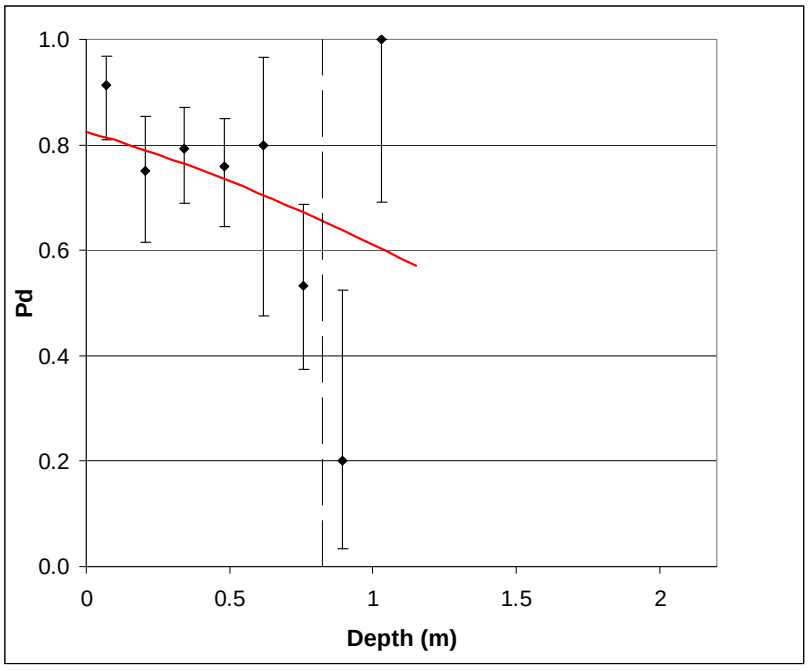

Figure A-13. 2.75-inch rocket, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

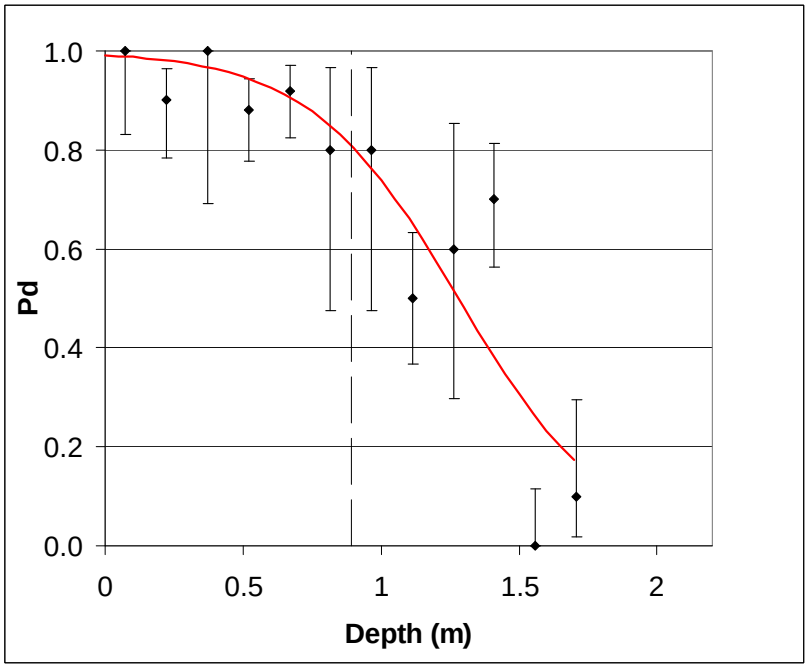

Figure A-14. 81 mm mortar, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

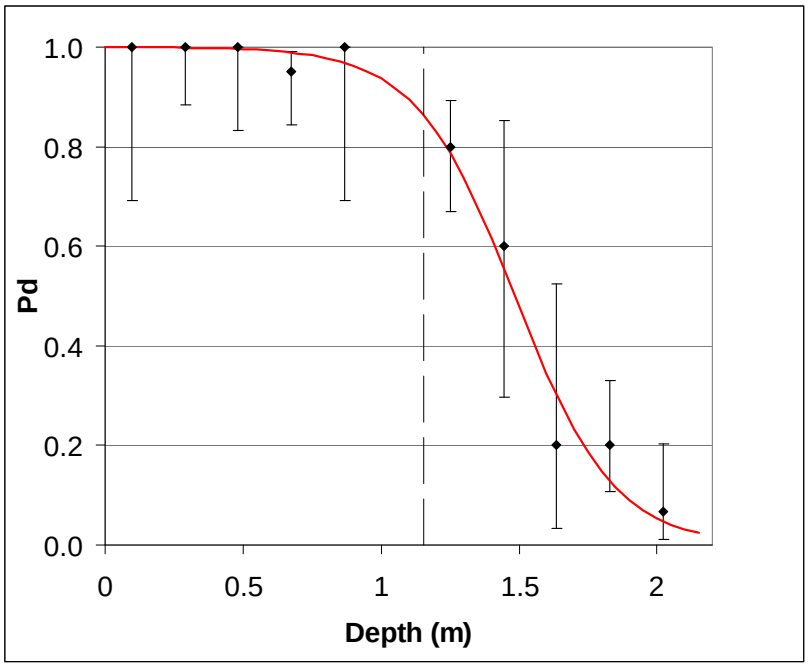

Figure A-15. 105 mm projectile, EMI, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

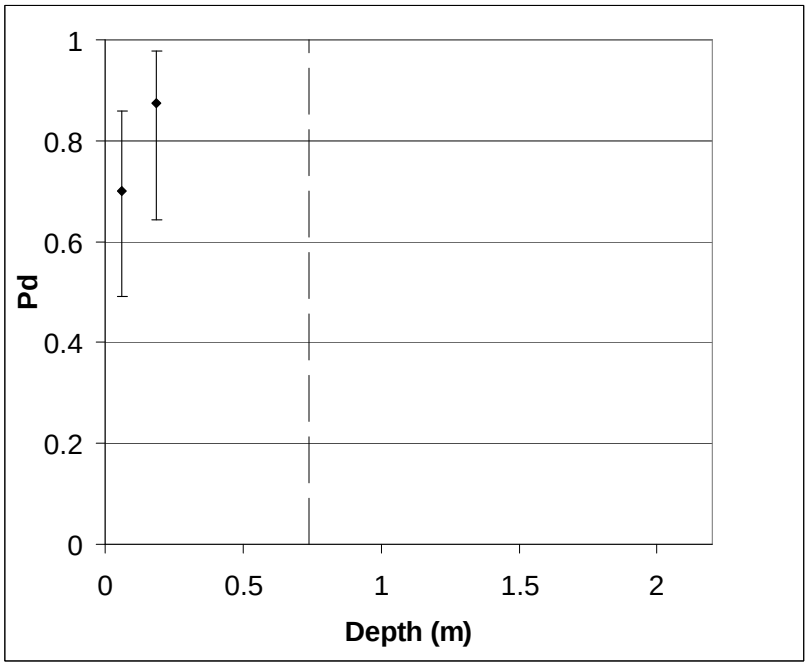

Figure A-16. BDU-28 bomblet, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The dashed vertical line marks the 11× Corps of Engineers depth.

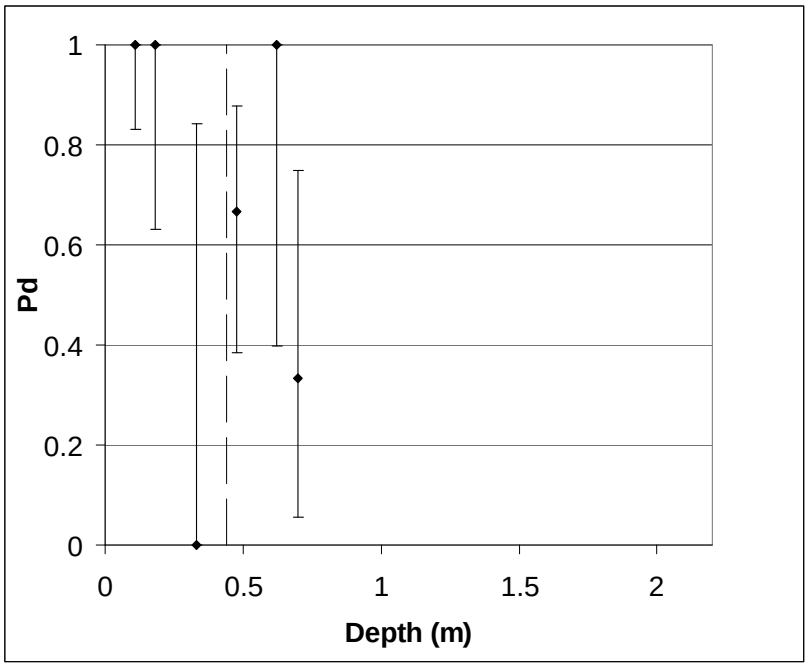

Figure A-17. 40 mm projectile, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The dashed vertical line marks the 11× Corps of Engineers depth.

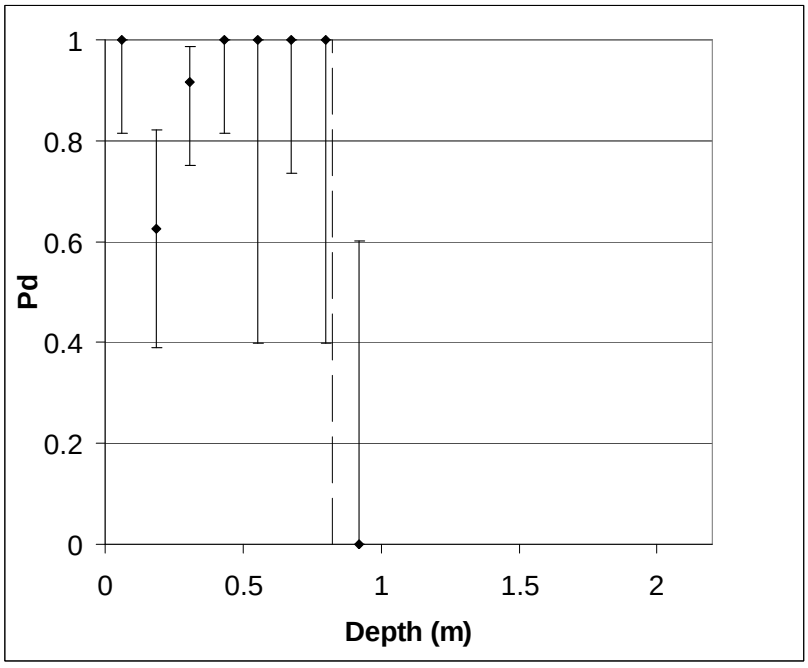

Figure A-18. 2.75-inch rocket, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The dashed vertical line marks the 11× Corps of Engineers depth.

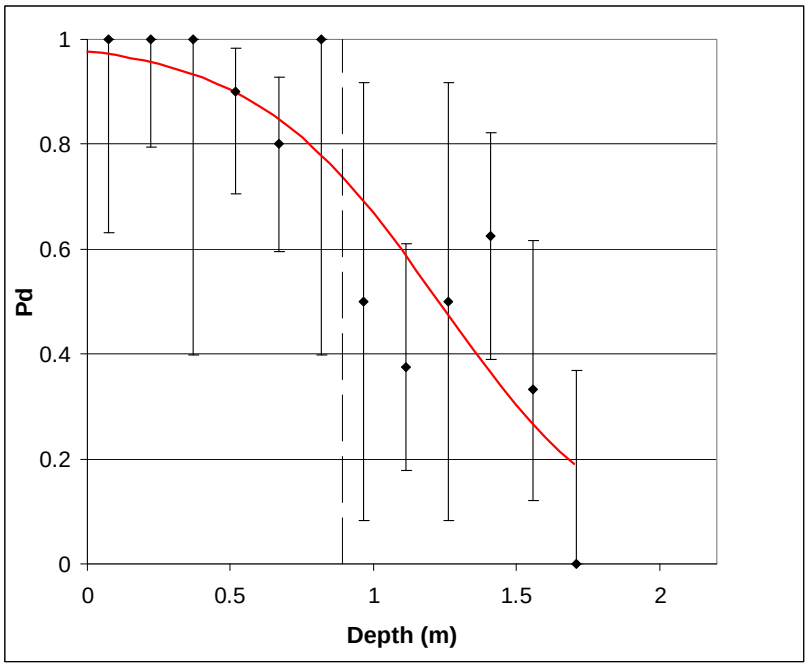

Figure A-19. 81 mm mortar, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

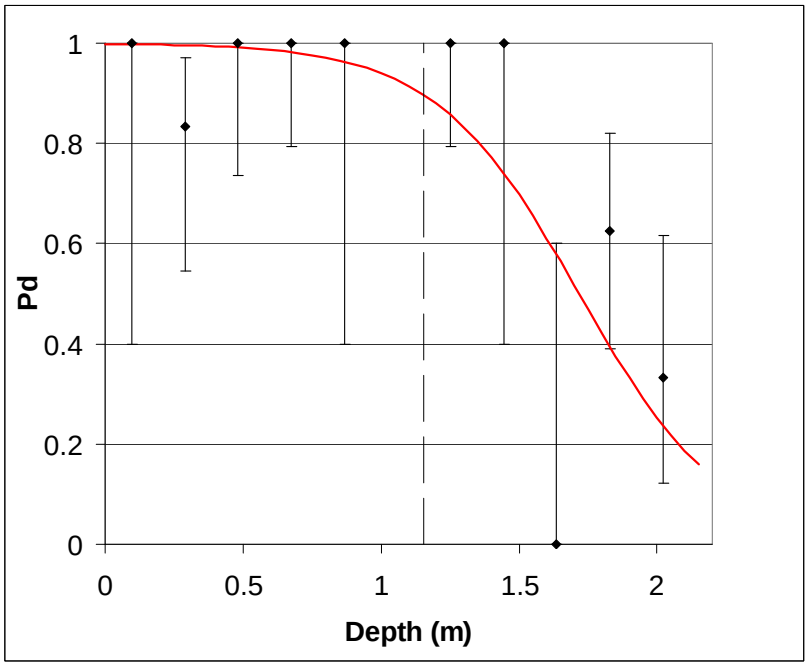

Figure A-20. 105 mm projectile, magnetometer, APG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

2. YPG

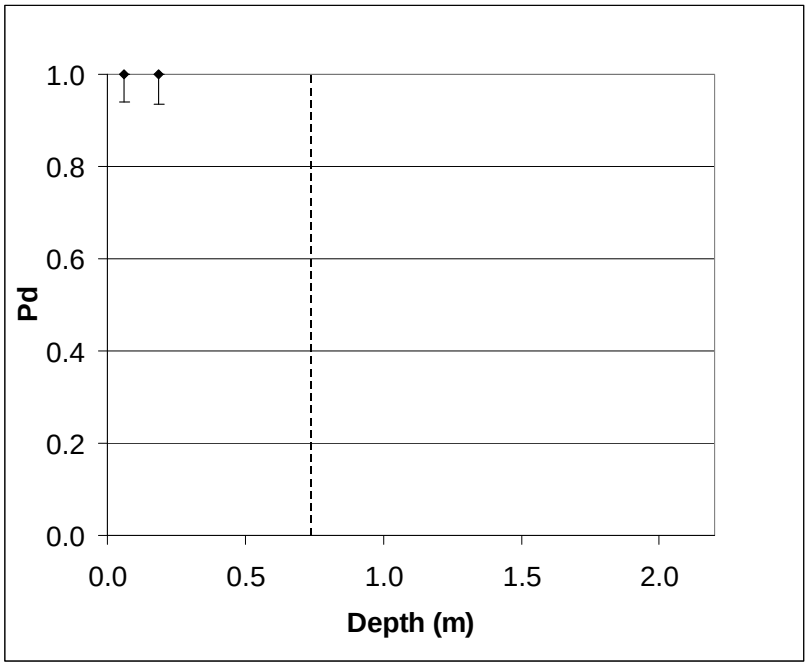

Figure A-21. BDU-28 bomblet, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The dashed vertical line marks the 11× Corps of Engineers depth.

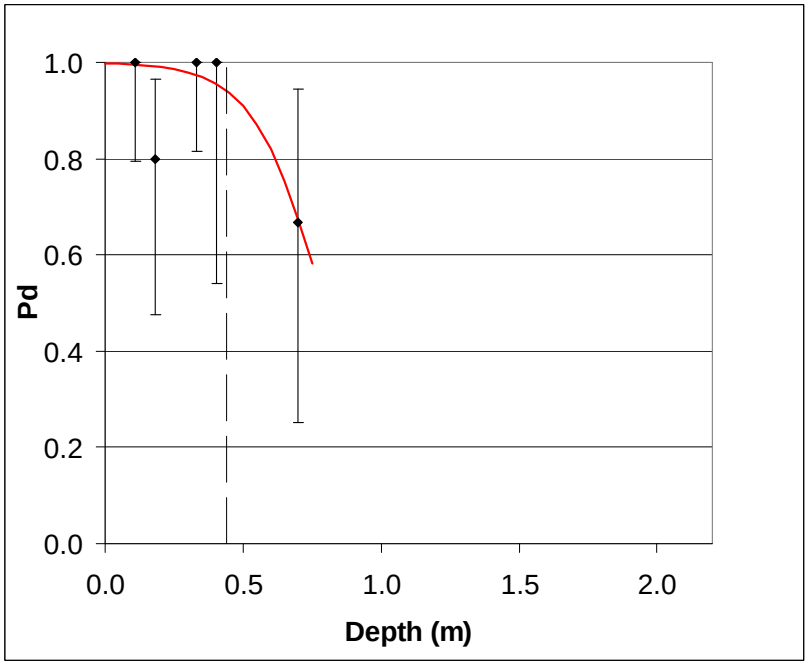

Figure A-22. 40 mm projectile, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

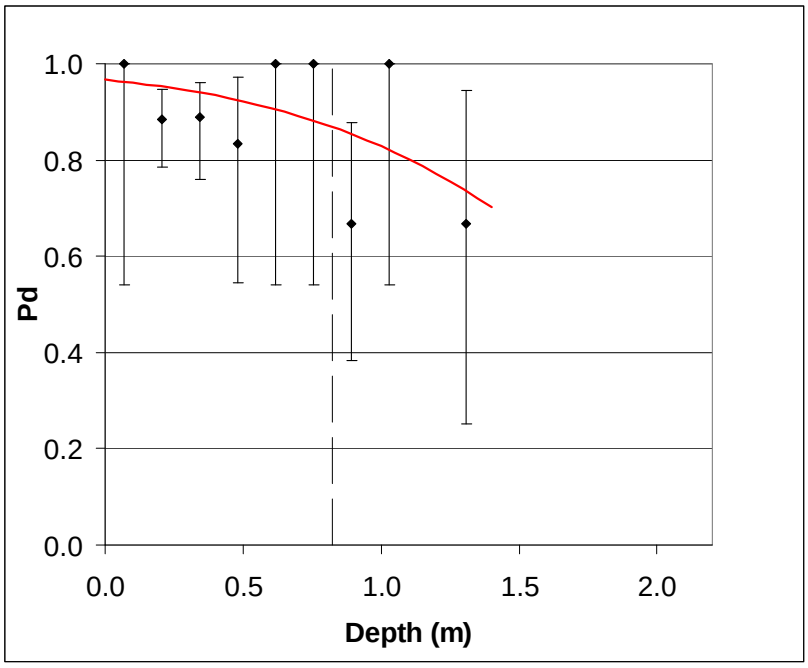

Figure A-23. 2.75-inch rocket, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

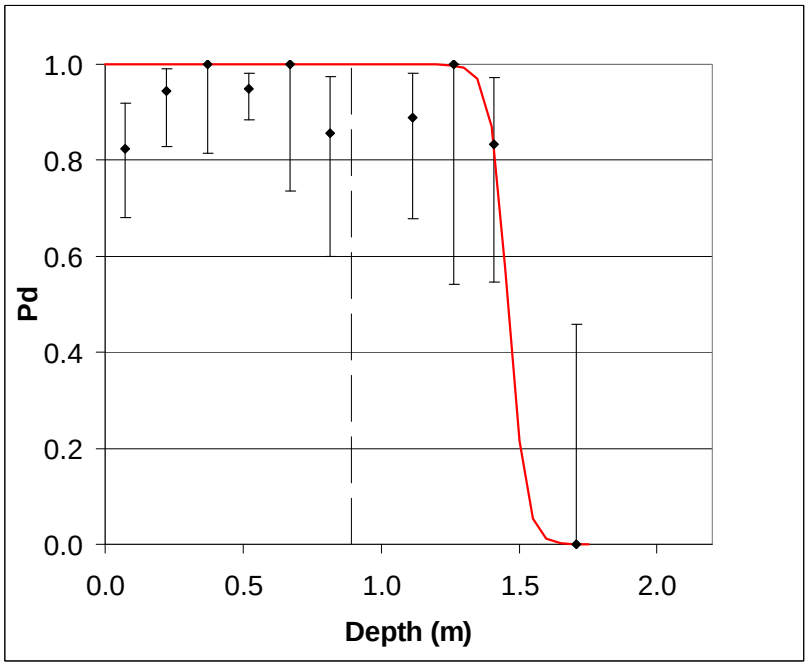

Figure A-24. 81 mm mortar, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

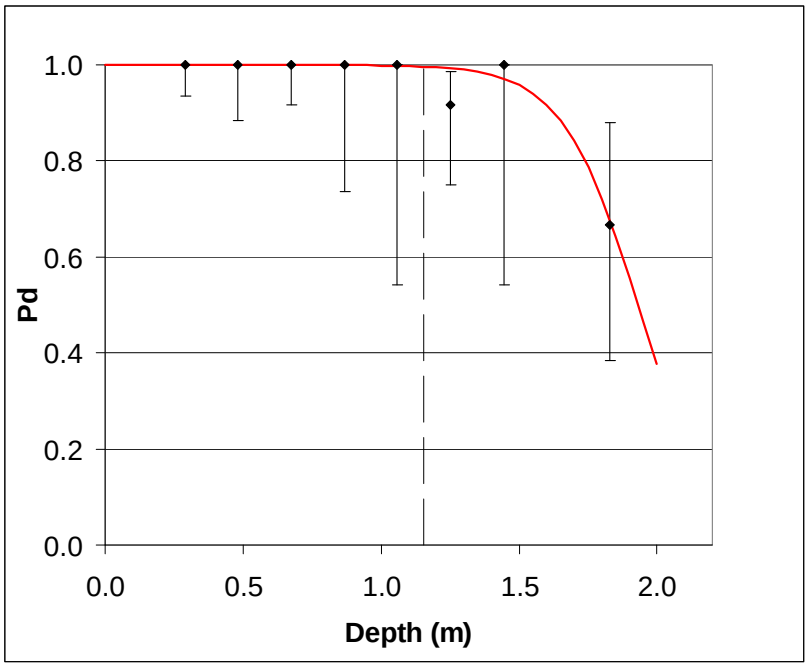

Figure A-25. 105 mm projectile, EMI, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

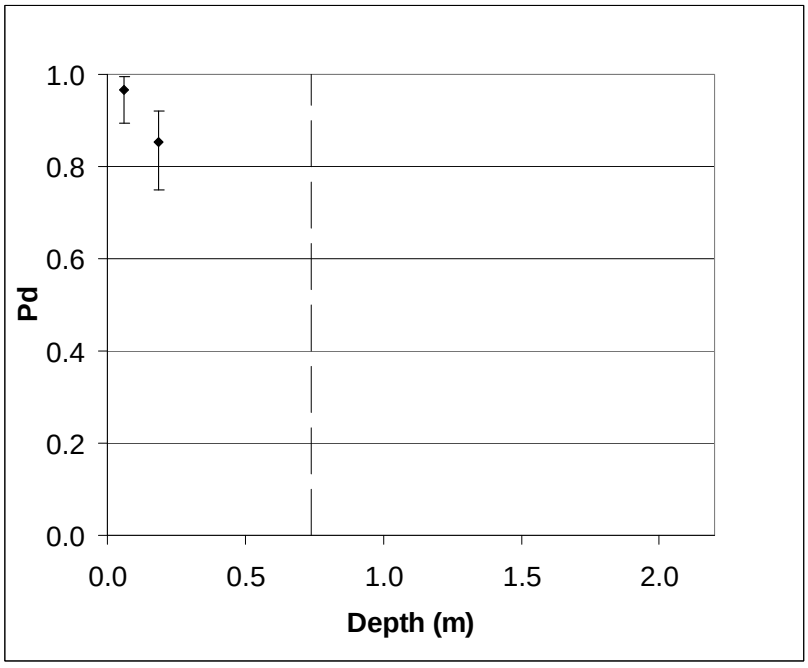

Figure A-26. BDU-28 bomblet, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The dashed vertical line marks the 11× Corps of Engineers depth.

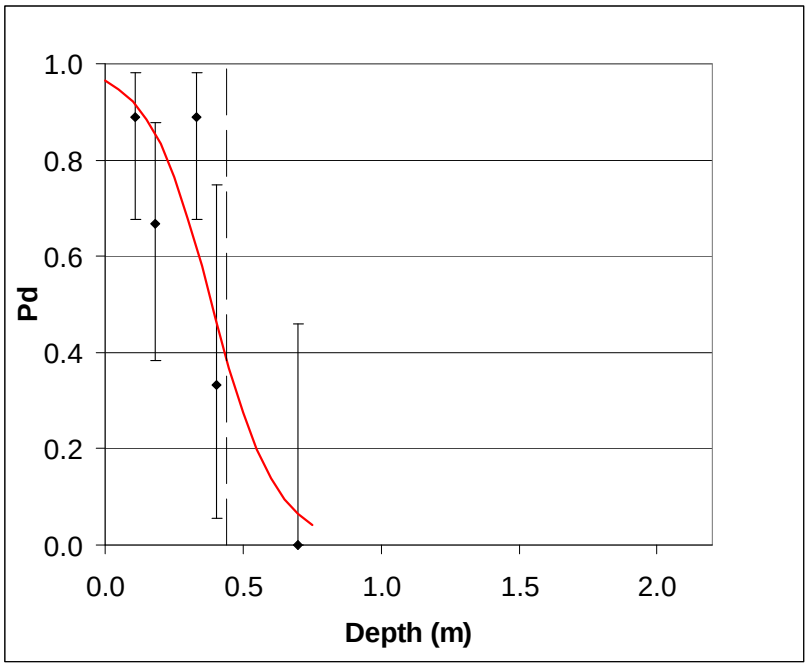

Figure A-27. 40 mm projectile, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

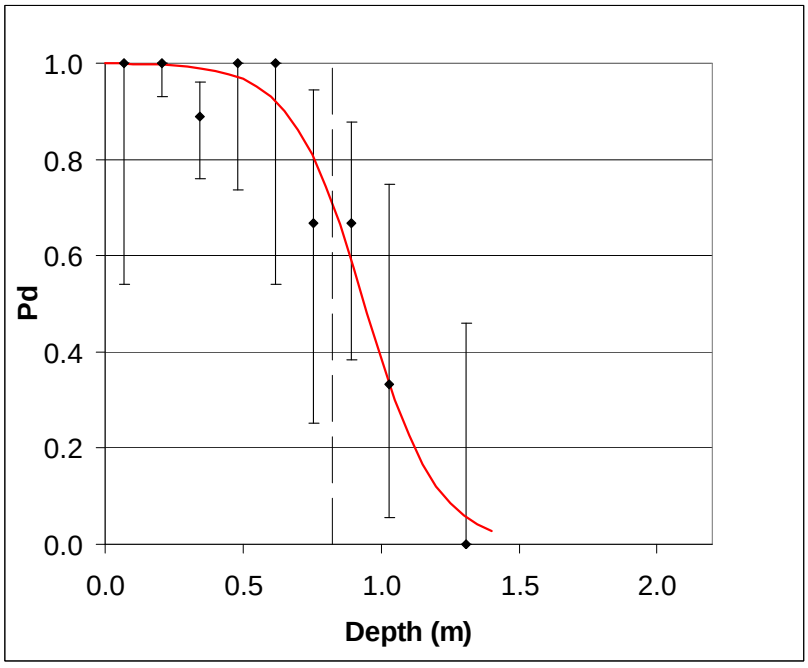

Figure A-28. 2.75-inch rocket, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

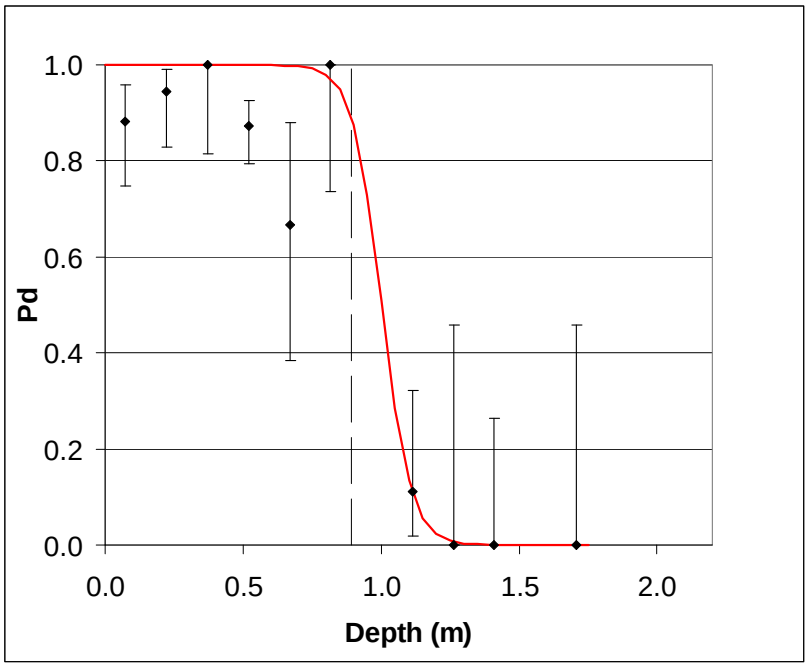

Figure A-29. 81 mm mortar, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

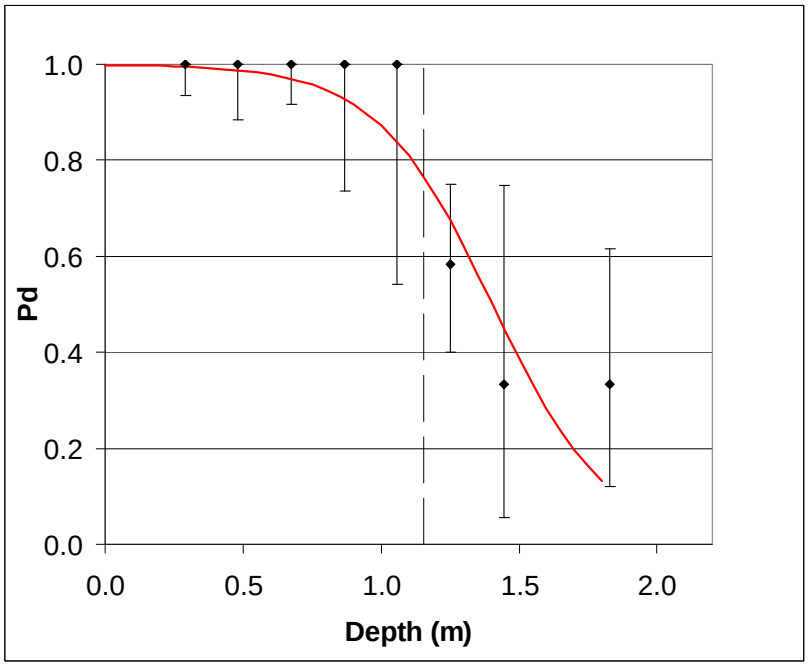

Figure A-30. 105 mm projectile, magnetometer, YPG. Pd as a function of depth. The uncertainty represents a 70% confidence level. Depth bins are one-sixth of the 11× Corps of Engineers depth wide, and the Pd is plotted at the center of the bin. The red line is an empirical fit. The dashed vertical line marks the 11× Corps of Engineers depth.

C. SCORING MEMORANDUM

This memorandum details the scoring protocol agreed on by SERDP/ESTCP, IDA, and AEC.

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive / Alexandria, Virginia 22311-1882 / 703-845-2000

Science and Technology Division

Memo

To: Carolyn Berger

From: Elvis Dieguez

CC: Larry Overbay, George Robitaille, Anne Andrews, Jeff Fairbanks, Mike Tuley

Date: 1/15/2004

Re: Untangling ambiguous matches

Theoretical Algorithm for Untangling Ambiguous Matches

1. Sort signals from strongest to weakest, and sort the ground-truth so ordnance items appear before clutter items. Begin assigning matches with the strongest signal.

2. Find all signals that do not match any ground-truth item – these signals can immediately be "tossed" into a "background container." The remaining signals match to one or more ground-truth items.

3. You cannot have multiple signals assigned to one ground-truth item or one signal assigned to multiple ground-truth items. There must be a one-to-one match between signals and ground-truth items.

4. If the signal lies within the halo of N clutter items and M ordnance items, assign the signal to the nearest of the M ordnance items.

5. If the signal lies within the halo of N clutter items and 0 ordnance items, assign the signal to the nearest of the clutter items.

6. If the signal lies within the halo of 0 clutter items and M ordnance items, assign the signal to the nearest of the ordnance items.

7. After repeating steps 4 – 6 for all signals (beginning from the strongest and working down to the weakest), then any signals not assigned must have been an additional detection of an item already assigned a stronger signal. These signals are not counted when generating the ROC and they are excluded from all statistics calculated. In essence, they are removed from the detection list submitted by the vendor whenever a statistic is calculated.

Practical Application of Untangling Ambiguous Matches

1. Create a true/false table with 'signal' being across the columns and 'ground-truth' down the rows (the signals are sorted from strongest to weakest and the ground-truth from ordnance to clutter). Initialize the table so every cell is 'false.'

2. Beginning with the first column, move down the row checking if the signal lies within the halo of the item given at that row. If it does, change the cell value to 'true.' Do this for the entire table.

3. Beginning with the first column, check if the entire column is labeled 'false.' If it is, then that signal does not match any ground-truth item. Move the signal to a 'background container' and delete the column from the table. Do this for the entire table.

4. After removing signals that do not match any ground-truth item, you should have a table with at least one cell labeled 'true' in any given column.

5. Beginning with the first column, work down through the cells valued 'true.' If any 'true' cell refers to an ordnance item, stop at the final ordnance item (no need to check the clutter items). If multiple ordnance items are labeled 'true' – choose the item nearest to the signal. If none of the ordnance items are labeled 'true', choose the clutter item nearest to the signal. After you find the item nearest to the signal, change all other cells in that column to 'false.'

6. When you move to the column N (N > 1), before assigning signal N to an item in row M, check if any of the previous 1…M-1 rows have already assigned a stronger signal to that item. If 'yes', change the cell to 'false' and continue on to the next nearest signal. If all items have already been assigned a stronger signal, the entire column of the Nth signal will eventually be labeled 'false.'

7. After checking all the signals, any columns in the table that are completely labeled 'false' constitute additional detections of an item already assigned a signal. They are not to be used when calculating performance statistics.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED *(From–To)*
January 2007	Final	March 2006 – May 2006

4. TITLE AND SUBTITLE

Interpreting Results from the Standardized UXO Test Sites

5a. CONTRACT NUMBER
DASW01 04 C 0003/W74V8H 05 C 0042

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Michael May
Michael Tuley

5d. PROJECT NUMBER

5e. TASK NUMBER
AM-2-1528

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, VA 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER

IDA Document D-3280

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

DUSD(SERDP/ESTCP)
901 N. Stuart Street
Suite 303
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited. (8 June 2007)

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This study interprets results from the Army's Standardized UXO Test Sites by treating likely-to-be-found and difficult targets separately. We determine probabilities of detection and background alarm rates considering the effect of clustered targets, those with nearby obstacles, and those that are deeper than the US Army Corps of Engineer's 11-times diameter rule-of-thumb depth. We find two common reasons for missing targets at the Standardized Sites: overlapping signatures which are difficult to resolve into separate targets and missing the target by slightly more than the fixed distance defined to be the "hit" radius. Ambiguities inherent in the scoring system and the small number of identical encounters make precise estimates of probabilities of detection impossible. Generally, for shallow munitions that are larger than about 60mm in diameter, the well-implemented sensors that were demonstrated at the sites found in excess of 90% of the munitions.

15. SUBJECT TERMS

UXO, munitions management, Standardized UXO Test Site, range clearance, electromagnetic induction sensor, cesium vapor magnetometer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Anne Andrews
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	102	19b. TELEPHONE NUMBER *(include area code)*
Uncl.	Uncl.	Uncl.			703-696-3826

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