Copy _____

IDA PAPER P-3114

DEMONSTRATOR PERFORMANCE AT THE UNEXPLODED ORDNANCE ADVANCED TECHNOLOGY DEMONSTRATION AT JEFFERSON PROVING GROUND (PHASE I) AND IMPLICATIONS FOR UXO CLEARANCE

Thomas W. Altshuler
Anne M. Andrews
Regina E. Dugan
Vivian George
Michael P. Mulqueen
David A. Sparrow

October 1995

Prepared for
Deputy Under Secretary of Defense for Environmental Security

INSTITUTE FOR DEFENSE ANALYSES
1801 N. Beauregard Street, Alexandria, Virginia 22311-1772

IDA Log No. H 95-47047

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

DEFINITIONS

IDA publishes the following documents to report the results of its work.

Reports

Reports are the most authoritative and most carefully considered products IDA publishes. They normally embody results of major projects which (a) have a direct bearing on decisions affecting major programs, (b) address issues of significant concern to the Executive Branch, the Congress and/or the public, or (c) address issues that have significant economic implications. IDA Reports are reviewed by outside panels of experts to ensure their high quality and relevance to the problems studied, and they are released by the President of IDA.

Group Reports

Group Reports record the findings and results of IDA established working groups and panels composed of senior individuals addressing major issues which otherwise would be the subject of an IDA Report. IDA Group Reports are reviewed by the senior individuals responsible for the project and others as selected by IDA to ensure their high quality and relevance to the problems studied, and are released by the President of IDA.

Papers

Papers, also authoritative and carefully considered products of IDA, address studies that are narrower in scope than those covered in Reports. IDA Papers are reviewed to ensure that they meet the high standards expected of refereed papers in professional journals or formal Agency reports.

Documents

IDA Documents are used for the convenience of the sponsors or the analysts (a) to record substantive work done in quick reaction studies, (b) to record the proceedings of conferences and meetings, (c) to make available preliminary and tentative results of analyses, (d) to record data developed in the course of an investigation, or (e) to forward information that is essentially unanalyzed and unevaluated. The review of IDA Documents is suited to their content and intended use.

The work reported in this document was conducted under contract DASW01 94 C 0054 for the Department of Defense. 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 that Agency.

IDA PAPER P-3114

DEMONSTRATOR PERFORMANCE AT THE UNEXPLODED ORDNANCE ADVANCED TECHNOLOGY DEMONSTRATION AT JEFFERSON PROVING GROUND (PHASE I) AND IMPLICATIONS FOR UXO CLEARANCE

Thomas W. Altshuler
Anne M. Andrews
Regina E. Dugan
Vivian George
Michael P. Mulqueen
David A. Sparrow

October 1995

INSTITUTE FOR DEFENSE ANALYSES
Contract DASW01 94 C 0054
Task T-AM2-1282

PREFACE

This paper was prepared by the Institute for Defense Analyses under a task for the office of the Deputy Under Secretary of Defense (Environmental Security) under a task entitled, "Technical Analyses and Evaluations of the Jefferson Proving Ground Demonstration of Detection, Identification and Remediation of Unexploded Ordnance."

The authors wish to thank the following people for their help in the preparation of this document. Robert C. Oliver and Christine Jordan provided insightful and thorough reviews of the document. After their reviews, we significantly revised the document; it is substantially improved as a result of their comments and suggestions. Kurt Thomsen, from PRC, also reviewed the document. We would like to thank him for correcting minor errors in the description of the test and demonstrators' experiences. Tom Milani skillfully edited the document under tremendous pressure. Anthea DeVaughan and Susan Taylor did much of the typing, typing, and retyping. Anthea also assisted with the overall transformation of various paragraphs, figures, and tables into a document with clarity and unity.

CONTENTS

FIGURES

TABLES

LIST OF DEMONSTRATORS, TECHNOLOGIES, AND DEMONSTRATOR IDENTIFICATION NUMBERS

40-acre, ground-based system demonstrators:

Demonstrator ID number	*Demonstrator*		*Technology*
31	ADI	M	Hand-held and surface towed GT-TM4 magnetometer, GT odometer, rope and tape for navigation.
19	Arete	MG IC	Geonics IC, Schonstedt gradiometer, man-portable GeoDAPS control system. Trimble DGPS.
16	Battelle/OSU	GPR	Surface-towed GPR and rope/tape/odometer for navigation.
7	Chemrad/EG&G	GPR/IC	Gulf-applied GPR and Pulse Tech IC, acoustic USRADS surveying tool.
10	Chemrad/G-822L	M	822L magnetometer, USRADS acoustic positioning system for navigation.
6	Chemrad/GSM-19	MG	GSM-19 magnetometer/gradiometer, USRADS acoustic positioning system.
23	Coleman	GPR/IC	Towed multisensor array system (TOMAS), GPR and IC. GPS for navigation.
36	Dynamic Systems	M	Man-portable system, Billigsley magnetometer, Foerster magnetometer, TopCon 302 Survey instrument.
29	Ensco	GPR	Several GPRs on sled pulled by modified golf cart. Survey wheel and laser Track.
25	EODT	M/IC	Schonstedt magnetometer and EM-31 conductivity sensor, GEODAPS DGPS.
44	Foerster	MG IC	Ferex (Mark 26) Standard Sensor, Ferex Deep Search Sensor, Mines 2FD Standard Sensor. Ferex—gradient magnetometers, Minex—IC. Surface-towed. DGPS.
2	GDE	GPR	Prototype surface-towed imaging GPR rope/tape/odometer navigation system.
1	Geocenters	MG	STOLS: 2 Gmtrcs/Scntx magnetometers and one Foerster hybrid magnetometer/ gradiometer.
43	Geometrics	M	Prototype Geometrics MagDIS man-portable system: 5 Cesium-vapor magnetometers. DGPS used for navigation.
42	Geo-radar	GPR	Preproduction model of GeoRadar 1000A man-portable GPR.
22	Jaycor	GPR	Two GPRs mounted on a golf cart. Navigation using existing markers.
33	Metratek	GPR IC	Prototype model 200 stepped-frequency GPR mounted on sled pulled by four-wheel-drive vehicle and man-portable Geonics EM61 IC. DGPS used for navigation.
24	SRI	GPR	Trailer mounted GPR system. Navigation by placing stakes every 100 ft in both directions to use as guides. Precise positions determined using DGPS.
13	UXB	M	Hand-carried Schonstedt GA-52B and Foerster Ferex magnetometer. GPS for navigation.
37	Vallon GmbH	M	Hand-held, towed, and gradiometer magnetometers were used. SEPOS rope/tape/ odometer navigation system.

KEY: M = magnetometer, G = gradiometer, IC = induction coil, GPR = ground-penetrating radar

80-acre, airborne system demonstrators:

Demonstrator ID number	*Demonstrator*		*Technology*
9	Airborne Environmental Surveys	GPR IR	Wideband radars centered at 500 MHz and 3 GHz, FLIR 2000F infrared imager on helicopter, DGPS navigation
17	Battelle/OSU	GPR	50–750 MHz GPR mounted on aerial platform (cherry picker), Trimble GPS navigation
35	Geonex Aerodat	MG IC	IC, cesium vapor magnetometer/gradiometer on boom towed below helicopter, DGPS navigation
30	Oilton	IR	FLIR 2000AB infrared imager, AIRDS navigation
20	SRI (Fixed Wing)	GPR	Bistatic GPR with SAR processing mounted on Beechcraft, DGPS navigation
21	SRI (Rotary Wing)	GPR	Ultrawide Band GPR operating in UHF and VHF bands mounted on helicopter, GPS navigation

KEY: M = magnetometer, G = gradiometer, IC = induction coil, IR = infrared, GPR = ground-penetrating radar

EXECUTIVE SUMMARY

In this report we present our best understanding of the performance demonstrated in the 1994 Unexploded Ordnance (UXO) Detection, Identification and Remediation Advanced Technology Demonstration at the Jefferson Proving Ground, Indiana. To arrive at this understanding, we have critically examined the parameters used to evaluate performance, as well as the limitations imposed by the test itself. We have also illustrated the importance of good performance to the effectiveness or efficiency of real world UXO cleanup. Finally, with test limitations, observed performance, and sensor phenomenology in mind, we have determined what guidance the results can provide for future experiments and system development.

A. BACKGROUND

It is inevitable that some ordnance does not explode as intended. This UXO remains as the legacy of past testing, training, and wartime activities. UXO contamination and the resulting humanitarian and economic impacts have intensified the need for systems to detect, identify, and remediate unexploded ordnance worldwide.

Current methods for clearing unexploded ordnance from contaminated land are labor-intensive, hazardous, and costly. Furthermore, a great deal of controversy exists concerning the capabilities of systems in current use or proposed for use. To address this issue, and to encourage the timely development and demonstration of technology to detect and remediate UXO contamination, the United States Congress mandated funds for a technology demonstration *to identify and evaluate innovative and cost effective systems for the detection, identification, and remediation of sites contaminated with subsurface unexploded ordnance.*

B. PROGRAM GOAL

The intent of the 1994 demonstration at Jefferson Proving Ground (JPG) was to establish a controlled site where the performance of technologies for the detection, identification and remediation of unexploded ordnance could be measured. Two controlled sites were established: a 40-acre site used for ground-based detection systems and an 80-acre

site for airborne systems. A variety of inert ordnance items were emplaced in surveyed locations on these sites. In all, 27 contractors, 1 DoD laboratory, and 1 DOE laboratory demonstrated their systems at JPG. Twenty detection systems on ground-based (surface-towed or hand-carried) platforms, six detection systems on airborne platforms, and three remediation systems were demonstrated. These sites continue to be in use for a 1995 series of demonstrations.

C. DEMONSTRATION SITES

Two sites, a 40-acre and an 80-acre, were chosen for the UXO Technology Demonstration Program. For the demonstration program, inert ordnance was emplaced on each site and the orientation, horizontal locations, and burial depths were documented. Ordnance representative of the contamination at JPG was emplaced and ranged from items as small as 60-mm rounds to items as large as 2,000-lb bombs. Other items, including metal fragments and nonordnance items such as beakers, building materials, glass bottles, metal cans, and drums, were also emplaced in recorded locations. Additionally, some holes were dug and refilled without the emplacement of any item.

D. CAVEATS FOR EXTRAPOLATING RESULTS

The goal of the site preparation was to emplace ordnance, ordnance-related, and nonordnance items in as realistic a manner as possible. However, there remain differences between the demonstration sites and actual ordnance-contaminated sites as well as differences between the demonstration procedures and actual remediation procedures that limit our ability to extrapolate the performance at JPG to other sites.

- Demonstrators were permitted only a single sweep and had limited information regarding the type of ordnance contamination.
- The geologic conditions were challenging.
- Other potential sources of signals may not have been identified and thus not included in the emplaced item list.
- No explosive simulants were used; thus, no systems that detect explosives (as opposed to metal) could be tested.
- The baseline ordnance distribution created ambiguity in the interpretation of the data.
- The limited number of ordnance items introduce statistical uncertainties in the measurement of demonstrator performance.

- The relative difficulty of areas searched made it difficult to compare demonstrators that searched different areas of the site.
- The local conditions are not representative of many possibly contaminated sites.

Nevertheless, the results of the JPG demonstration provide unique insight into the state of the technology for the detection of UXO.

E. SCORING DEMONSTRATOR PERFORMANCE

In a test such as the JPG demonstration, analysis begins with a series of demonstrator declarations, i.e., information that describes where the demonstrator thinks buried objects are, and ground truth, i.e., knowledge of where the emplaced objects actually are. The first step in assessing the performance of demonstrators at JPG is to determine which demonstrator declarations match the emplaced baseline items. This is the most fundamental part of any analysis, since matched declarations count toward detection capability and unmatched declarations count toward the false alarm rate. Detection capability and false alarm rate are the two principle measures of demonstrator performance at JPG.

The process of establishing matches is nontrivial. We calculated the impact of various matching schemes and other parameters on the measures of performance for JPG. The method of assigning matches and the robustness of various methods affected the results significantly enough to change our understanding of the performance.

The measure of detection capability that most accurately reflects the demonstrators' performance is the ability of demonstrators to detect groups of emplaced ordnance items on the area searched by each demonstrator, with the number of detections corrected for "lucky" matches of false alarms to undetected baseline items. We have removed plastic antipersonnel mines from consideration. The false alarm rate is calculated as the number of false alarms per unit area, where the specific unit of area is unspecified. The probability of false alarm is calculated as the fraction of the site covered by circles of radius 2 and 5 m, for the 40- and 80-acre sites respectively, around each demonstrator false alarm.

A comparison among demonstrators based on detection capability alone is inherently flawed. Further, it is difficult to compare the performance of a demonstrator with high detection capability and a high false alarm rate to a demonstrator with a low detection capability and a low false alarm rate. Differences in the performance may be directly attributable to different threshold settings or might be the result of real differences

in system capabilities. Standard approaches have been developed to quantify the relationship between the probability of detection and the probability of false alarm. These approaches allow us to determine the relative ability of demonstrators using a single parameter that includes consideration of both the probability of detection and the false alarm rate.

F. RESULTS

1. Demonstrators of Ground-Based Detection Systems

Table S.1 summarizes our best understanding of the performance of demonstrators of ground-based detection systems at the 1994 JPG demonstration.

2. Demonstrators of Airborne Detection Systems

Table S.2 summarizes performance of demonstrators of airborne detection systems.

G. MAIN FACTORS AFFECTING PERFORMANCE AT JPG

Instrument phenomenology, operating conditions and target emplacement affected the successes and limitations of sensor performance at JPG. Below we summarize the most important parameters for each technology.

1. Magnetometers

Magnetometers detected buried ordnance with some success at JPG. This is probably in part due to the relatively favorable fall-off of signal with separation distance between the sensor and the ordnance item. However, because magnetometers rely on the magnetic signature of ordnance that can be different for each individual item depending on size, history, and orientation relative to the earth's magnetic field, detection can be difficult, even for large magnetic objects. Background magnetic conditions can obscure the magnetic signature. This problem would be amplified if the site requiring remediation is located in a region with a high density of magnetic minerals. Finally, the detector system design can strongly influence the ease of interpretation of magnetic signatures.

2. Induction Coils

Detection by induction coils at JPG was moderately successful. Detection of deep objects was suppressed, but not eliminated, by the rapid fall-off in signal with increasing

Table S.1. Demonstrators of Ground-Based Detection Systems: Summary of Performance

Demonstrator		Technology	Best estimate of detection capability,* %	False alarm rate*	Typing	Acres	Notes
ADI	***M***	***Hand-held and surface towed GT-TM4 magnetometer, GT odometer, rope and tape for navigation.***	***67***	***0.43***		40	GPR system proposed, elected not to use—soil conductivity too high. Report all anomalies >100 g. (Pd and FAR adjusted to remove fence line.)
Arete	***MG IC***	***Geonics IC, Schonstedt gradiometer, man-portable GeoDAPS control system Trimble DGPS.***	*31*	***0.32***		25	
Battelle/ OSU	GPR	Surface-towed GPR and rope/tape/odometer for navigation.	0	*2.2*	all O	2.3	Experimental, laboratory version of the radar system. High in clay content.
Chemrad/ EG&G	GPR IC	Gulf-applied GPR and Pulse Tech IC, acoustic USRADS surveying tool.	11	2.3	all O	16	Navigation accuracy 6 in.
Chemrad/ G-822L	*M*	*822L magnetometer, USRADS acoustic positioning system for navigation.*	*40*	*1.9*	all NO	40	Magnetometer range of 20–25 ft for detection.
Chemrad/ GSM-19	MG	GSM-19 magnetometer/gradiometer, USRADS acoustic positioning system.	4	*1.5*	all O	40	
Coleman	*GPR IC*	*Towed multisensor array system (TOMAS), GPR and IC. GPS for navigation.*	*52*	*0.57*	all O	36	No IC is mentioned in proposal. All detections declared as ordnance. High confidence declarations only.
Dynamic Systems	***M***	***Man-portable system, Billigsley magnetometer, Foerster magnetometer, TopCon 302 Survey instrument.***	*47*	***0.5***	all O	5.5	Sensors have range 15–20 ft.
Ensco	GPR	Several GPRs on sled pulled by modified golf cart. Survey wheel and laser Track.	0	4.8	all O	10	
EODT	M IC	Schonstedt magnetometer and EM-31 conductivity sensor, GEODAPS DGPS.	14	***0.42***	all O	11	Schonstedt range of 2–5 ft, conductivity sensor range 10–15 ft.
Foerster	*MG IC*	*Ferex (Mark 26) Standard Sensor, Ferex Deep Search Sensor, Mines 2FD Standard Sensor. Ferex—gradient magnetometers, Minex—IC. Surface-towed. DGPS.*	*57*	3.23		24	Terrain required sensor to be located ~ 12 in. above ground.
GDE	GPR	Prototype surface-towed imaging GPR rope/tape/odometer navigation system.	0	29.7	all O	6.9	Demonstrator noted large diameter clods of earth, standing puddles of water, and mud holes. (Corrected for random hits.)
Geo-Centers	***MG***	***STOLS: 2 Gmtrcs/Scntx magnetometers and one Foerster hybrid magnetometer/gradiometer.***	*66*	*1.33*		40	Gmtrcs/Scntx range > 25 ft, Foerster range = 5–10 ft. Areas inaccessible with STOLS used hand-carried.
Geo-metrics	*M*	*Prototype Geometrics MagDIS man-portable system: 5 Cesium-vapor magnetometers. DGPS used for navigation.*	*30*	***0.43***		35	Demonstrator notes: rough terrain—large clumps of earth and cut-off/turned up roots. (After removing declared pipes, trenches,...)
Georadar	GPR	Preproduction model of GeoRadar 1000A man-portable GPR.	11	2.6	all O	2	Range of detection 5–10 ft. Demonstrator notes GPRs troubled by wet, clay soils.
Jaycor	GPR	Two GPRs mounted on a golf cart. Navigation using existing markers.	0	***0.81***	all O	20	
Metratek	*GPR IC*	*Prototype model 200 stepped-frequency GPR mounted on sled pulled by four-wheel-drive vehicle and man-portable Geonics EM61 IC. DGPS used for navigation.*	*45*	*1.95*		5	System has 12-ft swath and frequency band of 0.2–2.0 GHz. Demonstrator note: muddy in low lying areas, dried over week. Conductivity of soil was high—deep targets not reachable.
SRI	GPR	Trailer mounted GPR system. Navigation by placing stakes every 100 ft in both directions to use as guides. Precise positions determined using DGPS.	0	*1.95*	all NO	13	Demonstrator notes that resistivity would result in attenuation losses through soil such that maximum penetration of radar would be less than 2 m.
UXB	***M***	***Hand-carried Schonstedt GA-52B and Foerster Ferex magnetometer. GPS for navigation.***	*64*	*1.51*	all NO	30	Proposal indicates that GA-52B sensor work to depth of 3 m and Ferex sensors to depth of 5.8 m. Conditions at JPG "more than meet" the ideal conditions.
Vallon GmbH	M	Hand-held, towed, and gradiometer magnetometers were used. SEPOS rope/tape/ odometer navigation system.	*65*	11.7	all O	12	Proposal states objects up to 7 m deep can be detected and location accuracy is 5 cm. (With declarations labeled fence post removed.)

KEY: M = magnetometer, G = gradiometer, IC = induction coil, GPR = ground-penetrating radar, all O or all NO: all declarations were typed as either ordnance or nonordnance.

* Best estimate of performance is described in Chapter 4. Bold italic numbers indicate the high category of detection capability or false alarm performance. Italics only indicate the second category. Demonstrators with names and technologies listed in bold italics are in the top category of the receiver operator curve characteristic, *d*, which accounts for both detection capability and false alarm rate. Italics only are in the second category. See Chapter 5.

Table S.2. Demonstrators of Airborne Detection Systems: Summary of Performance

Demonstrator		*Technology*	*Best estimate of detection capability,* %*	*False alarm rate**	*Typing*	*Acres*	*Notes*
Airborne Environ. Surveys	GPR IR	Wideband radars centered at 500 MHz and 3 GHz, FLIR 2000F infrared imager on helicopter, DGPS navigation	0.01	0.11		80	
Battelle/ OSU	GPR	50–750 MHz GPR mounted on aerial platform (cherry picker), Trimble GPS navigation	0	0.28	all O	29	Vehicle was limited to roadway and radar was likely ineffective at distances > 500 ft
Geonex Aerodat	MG/ IC	IC, cesium vapor magnetometer/ gradiometer on boom towed below helicopter, DGPS navigation	0.03	0.39	all O	80	GPS failed, navigation limited to survey lanes marked on ground, winds caused boom to sway, 31 hr downtime from weather and equipment failures
Oilton	IR	FLIR 2000AB infrared imager, AIRDS navigation	0	1.7	all NO	80	High vegetation hampered detection, optics/flight path not optimized for imaging ordnance
SRI (Fixed Wing)	GPR	Bistatic GPR with SAR processing mounted on Beechcraft, DGPS navigation	0	0.36	all O	80	Detection hampered by wet soils
SRI (Rotary Wing)	GPR	Ultrawide Band GPR operating in UHF and VHF bands mounted on helicopter, GPS navigation	0.02	0.22	all O	80	Detection hampered by wet soils (maximum penetration 10 m in dry, sandy soils)

KEY: M = magnetometer, G = gradiometer, IC = induction coil, GPR = ground-penetrating radar, all O or all NO: all declarations were typed as either ordnance or nonordnance.

* Best estimate of performance is described in Chapter 4. Bold italic numbers indicate the high category of detection capability or false alarm performance. Italics only indicate the second category. Demonstrators with names and technologies listed in bold italics are in the top category of the receiver operator curve characteristic, *d*, which accounts for both detection capability and false alarm rate. Italics only are in the second category. See Chapter 5.

depth. The size and orientation of the ordnance item influence detectability by induction coil systems. Soil attenuation was not a limiting factor in the performance of these systems at JPG, nor is it likely to be elsewhere.

3. Ground-Penetrating Radars (GPRs)

Detection using GPRs was unsuccessful at JPG. Radar detection of the emplaced mines at JPG was prevented by the inability to discriminate objects of interest from background clutter of comparable size. High-frequency radars are needed in order to have appreciable response to the relatively small mines. These radars also scatter significantly from stones or other small discontinuities in the soil. At high frequencies, the radar waves are rapidly attenuated by any moisture in the soil. Discrimination between mines at a given depth and stones of comparable size at the same or shallower depth will be impossible.

For larger objects, the radar response in the 50–500 MHz region is sufficiently large for detection. At JPG, however, there is significant attenuation even at these low frequencies due to the high ground conductivity. It is not surprising, given the high conductivity, that essentially nothing was found by the radars. Better performance at detecting large objects, if they are not too deep, might be expected in other regions if the conductivity is lower.

4. Infrared Detectors

Infrared detection of buried objects is challenging under even the most favorable conditions. For the heavily vegetated terrain at JPG, it might well have been impossible with any available infrared system. The surprising result was that the surface ordnance was undetected. This was due to use of optics and flight altitudes that required sub-pixel detection. We do not believe that the contrasts were likely to be high enough to allow sub-pixel detection, particularly in the presence of the vegetation. However, even if sub-pixel detection had been possible, the location errors associated with the airborne platforms would have rendered statistical determination of detection impossible. In the absence of recognizable imagery, therefore, it is impossible to establish any evidence for detection of ordnance by the demonstrated infrared systems. It is not likely that infrared techniques will allow detection of buried ordnance in vegetated areas, unless the burial was recent enough to allow detection of the scar.

H. CONCLUSIONS

1. Demonstrators of Ground-Based Detection Systems

The performance data allows the following conclusions to be drawn regarding the technologies demonstrated under the conditions at JPG:

- Magnetometer systems, on average, exhibited the best performance in this demonstration, with probabilities of detection of nearly 70 percent using the best estimate of performance. The number of emplaced items found by the demonstrators using magnetometers indicate that all the magnetometers were capable of detecting some ordnance.
- Stand-alone GPRs were unsuccessful at detecting buried unexploded ordnance.
- Induction coil metal detectors paired with GPRs did not perform as well as the best magnetometers, but nonetheless, as a group, systems with induction coils displayed moderate detection capabilities. Comparison to the detection capabilities of stand-alone GPRs suggests that the induction coil may be responsible for the detection capabilities displayed by these adjunct systems.
- Most demonstrators exhibited little or no ability to distinguish ordnance from nonordnance, or to identify one ordnance item from another.
- Most demonstrators reported multiple false alarms per ordnance item detected.
- False alarms and inaccuracies in locating ordnance items would cause even the best demonstrators to disturb 5 times the minimum amount of surface area

required to recover a buried item; most demonstrators disturb between 10 and 100 times the minimum surface area required. The effect of false alarms is in many cases an order of magnitude more detrimental than location inaccuracies.

- Demonstrators with extensive field experience performed better than those with little.

Different detection systems might be expected to perform with different levels of success depending on the size, type or depth of ordnance. Therefore, detection capabilities were calculated on subsets of the emplaced items sorted by these attributes. This subdivision produces a much smaller number of items in each category, thereby increasing statistical uncertainties. With this limit on the validity of the conclusions in mind, the data shows the following trends.

- Bombs are found with the highest reliability, with three demonstrators using magnetometers finding greater than 80 percent of the bombs. Further, there is a significant difference in performance between the magnetometers and the induction coils, with the magnetometers detecting bombs at about twice the rate of induction coils.
- For mortars, which are generally nearer to the surface, detection capabilities for magnetometer and induction coil technologies were indistinguishable.
- Not surprisingly, neither magnetometer nor induction coil technologies were able to detect the plastic antipersonnel mines.
- Magnetometers were more proficient at detecting ordnance items buried below 6 ft, where, in general, the larger targets were located. The ability of the induction coil systems, on the other hand, to detect ordnance falls off considerably in the "below 6 feet" range.
- When detection capabilities are sorted by size, induction coils perform best at locating the medium-size targets. The magnetometers, on the other hand, performed best at detecting the medium and large targets but not the small targets.

2. Demonstrators of Airborne Detection Systems

The results from the 80-acre site demonstration at JPG lead to the following conclusions:

- There is no evidence that airborne systems detected any ordnance. Thus, it is difficult to support any contention that rapid, accurate characterization of large tracts of land contaminated with subsurface UXO is feasible with current technology.

- Accurate navigation and mapping to allow matching of demonstrator detections with emplaced items was likely a limiting factor for the airborne systems demonstrated at JPG. This illustrates the important distinction between sensors that are able to provide images of surface UXO and systems that are able to support clearance activities.
- Integrating sensors on airborne platforms for the purpose of detecting subsurface UXO is likely to be unproductive with our current level of understanding. Any sensors proposed for integration should first be tested on targets with calibrated signatures and then on sites with a realistic placement of ordnance under field conditions such as at JPG. These tests of a stand-alone sensor should be a prerequisite for consideration of integration on an airborne platform.

3. Implications for Research and Development Efforts

The JPG demonstration has produced summary level performance data with implications for future R&D efforts into UXO detection. The demonstration was carried out at a single site, by a relatively small number of demonstrators using only a subset of the technologies that could be applied to UXO detection. These demonstrators varied in nature from firms that find UXO commercially to nonprofit, academic technology developers. The conclusions one infers from the data are limited by the above considerations. Nevertheless, this data is a unique resource for R&D planning in UXO detection. Our conclusions about the implications of JPG data for R&D follow.

- Most demonstrators reported more false alarms than detections of ordnance items. This suggests that improving sensor sensitivity alone is not likely to improve performance. Better discrimination is required. Improvements in discrimination will require better understanding of the background features interfering with the detection of targets of interest and generating false alarms. This is probably the most important area for research.
- There is a set of targets in the JPG demonstration that were found by none of the demonstrators. All of these targets fall into one of three categories: (1) small ordnance items in close proximity to larger items, where the signatures of the larger items may mask the signatures of the smaller; (2) medium-sized mortars buried more than 3 feet deep; and (3) ordnance items near trees or survey stakes that may obstruct access to the ordnance items. Future efforts should determine whether these targets represent a general feature of buried UXO.
- The results from the JPG demonstration suggest that false alarms might be reducible using sensor fusion. It may even prove possible to reduce false

alarms with multiple sensors of the same technology. Unfortunately, this reduction of false alarms requires modest reductions in the detection capability of single sensors. Reducing false alarms through fusion, while retaining detection of the rarely detected emplaced items to improve detection capability over the single sensor values, will require much improved discrimination, if it is possible at all.

- The superior performance of commercial firms suggests that experience is an important factor in determining performance. Studying the knowledge base of these firms may provide valuable information to complement the technical data of sensor developers.

1. INTRODUCTION

The objective of this paper is to describe our best understanding of the performance demonstrated in the 1994 Unexploded Ordnance (UXO) Detection, Identification, and Remediation Advanced Technology Demonstration at the Jefferson Proving Ground, Indiana. To arrive at this understanding, we have critically examined the parameters used to evaluate performance, as well as the limitations imposed by the test itself. We have also illustrated the importance of good performance on the effectiveness or efficiency of UXO cleanup. Finally, with test limitations, observed performance, and sensor phenomenology in mind, we have determined what guidance the results can provide for future experiments and system development.

The report is organized as follows:

- Chapter 2 provides background information on the nature and extent of the UXO problem in the United States and worldwide, demonstrating the motivation of Congress in establishing the Jefferson Proving Ground (JPG) demonstration. It goes on to describe JPG, the set-up of the demonstration, and the limitations that the particulars of the demonstration place on extrapolating the results to other sites.
- Chapter 3 describes the measures that were used to evaluate the performance of the demonstrators and discusses the impact that good or poor performance in each of these metrics might have on clean-up activities. This information is likely to be of particular interest to those involved in the evaluation of technologies for UXO detection.
- Chapter 4 critically examines the measures of performance as they relate to evaluating and comparing the demonstrators. Many aspects of the scoring of demonstrators require careful examination. Specific parameters used in calculating the measures of performance affect the outcome. These parameters are explored in detail and the impact on demonstrator performance is quantified. Chapter 4 provides insight applicable to system developers and test designers.
- Chapter 5 presents our best understanding of demonstrator performance and comparisons among demonstrators and technologies, considering the factors discussed in Chapter 4. It contains a summary of the results from JPG as well as a ranking of demonstrators.

- Chapter 6 provides a brief description of each of the technologies and describes the parameters that most likely led to the success or failure of the technologies at JPG. Further, the performance at JPG is used to frame recommendations for longer term research and development efforts.
- Chapter 7 contains the major results and conclusions that can be reached from the JPG data.

2. UXO CONTAMINATION AND THE JEFFERSON PROVING GROUND DEMONSTRATION

It is inevitable that some ordnance does not explode as intended. This unexploded ordnance remains as the legacy of past testing, training, and wartime activities. Unexploded ordnance contamination and the resulting humanitarian and economic impacts have intensified the need for systems to detect, identify, and remediate UXO worldwide.

The scope of the UXO-contamination problem is staggering. The use of ordnance in conflicts has increased dramatically over time, largely driven by the advent of the cluster bomb and by the plethora of ordnance items available at relatively low cost. Ordnance contamination may be on the surface and hence visible, or it may be buried. Within the United States, greater than 11 million acres of government lands are potentially contaminated (Ref. 1). UXO contamination within the United States is predominantly the result of testing and training activities to prepare for war; however, in some rare cases, ordnance greater than 100 years old remains from the Civil War and earlier military activities. UXO contamination in the United States is found on active ranges, on bases slated for realignment and closure, on formerly used defense sites (FUDS), or on other government lands such as Department of Interior or Fish and Wildlife lands. The United States Department of Defense has the responsibility for clearance of domestic land contaminated with UXO.

UXO contamination is also a worldwide humanitarian concern. In many countries, there is a significant fatality rate caused by UXO that remains after military conflicts. In 1991 in France, 36 farmers were killed while tilling their land (Ref. 2). There is one casualty per day in Khe Sanh, Vietnam (Ref. 3), and more than 14,000 Poles died in the 36 years immediately following World War II (Ref. 2). In the United States, on the other hand, UXO contamination is typically limited to well-defined areas on test ranges and training bases, and as a result, rarely poses a danger to the public. Nevertheless, accidents do occur. In Tierrasante, CA, two young boys were killed and several others were injured when an old 37-mm round exploded. This accident and other incidents, such as the discovery of UXO from a World War I munitions facility in a residential neighborhood of Washington, DC, have raised consciousness within the United States regarding the potential danger of UXO on land in the public sector.

The impact of UXO contamination is also economic. In many countries, unexploded ordnance limits access to farm land and major economic installations such as power plants and water treatment plants. Many years after a war or conflict, UXO may inhibit development. As an example, during the construction of a rail bed in France, deminers were on constant duty to remove ordnance; 5 tons of bombs were removed on a typical day. In addition, four front end loaders and several earth movers were destroyed over the course of the project (Ref. 2). On government lands in the United States, UXO contamination inhibits the transfer of land from the custody of the government to the private sector during consolidation efforts and base closure and realignment. Cost estimates for the surface and subsurface remediation of UXO-contaminated land vary, but in the United States alone, most are in the tens or hundreds of billions of dollars.

Current methods for clearing unexploded ordnance from contaminated land are labor-intensive, hazardous, and costly. Furthermore, a great deal of uncertainty exists concerning the capabilities of systems in current use or proposed for use. To address this issue, and to encourage the timely development and demonstration of technology to detect and remediate UXO contamination, the United States Congress mandated funds for a UXO technology demonstration. Subsequently, the Army Environmental Center (AEC), as program manager, and the Naval Explosive Ordnance Disposal Technology Division (NAVEODTECHDIV), as technical lead, conducted a large-scale UXO technology demonstration at the Jefferson Proving Ground, Indiana.

A. THE JEFFERSON PROVING GROUND TECHNOLOGY DEMONSTRATION PROGRAM

In 1994, a series of demonstrations of the detection, identification, and remediation of unexploded ordnance was held at JPG. The objective of the JPG unexploded ordnance technology demonstration program was *to identify and evaluate innovative and cost effective systems for the detection, identification, and remediation of sites contaminated with subsurface unexploded ordnance.* Proposals were solicited from technology developers and vendors both within and outside of the United States, and a cross section of technologies was selected for the demonstration program. The Army, Navy, and Air Force, one of the DOE national laboratories, and more than 25 private contractors participated, in various capacities, in the demonstration. Summary statistics describing performance have recently been published, along with some analysis of the results (Refs. 4, 5). Tables of demonstrators, identification numbers, and technologies used throughout the report are located immediately following the Table of Contents at the front of the report.

B. THE TEST SITE

JPG is located 5 miles north of Madison, Indiana, on an area approximately 5 miles wide and 19 miles long, covering about 55,624 acres. The primary mission of JPG was to conduct production acceptance tests, reconditioning tests, surveillance tests, and other studies of ammunition and weapons systems. Test firing of munitions began at JPG on May 10, 1941, and a variety of munitions ranging from 20-mm rounds to 2,000-lb bombs were tested during the approximately 50 years of operations at JPG. The main firing line at JPG runs east and west near the south end of the base with firing lanes to the north. The land consists of gentle rolling hills of mostly poorly drained silty loam, and the area surrounding JPG is principally agricultural.

Two sites, a 40-acre and an 80-acre, were chosen for the UXO technology demonstration program. Both sites are located north of the main firing line, outside the safety fan surrounding the adjacent impact ranges. Thus, these sites are thought to be clear of ordnance and other potentially hazardous substances or wastes. Both sites had surface vegetation consisting mainly of grass, bushes, shrubs, and small trees with only a few large trees scattered at each site. The 40-acre site was used for the demonstration of hand-carried or surface-towed sensor systems, and the 80-acre site was used for airborne sensor systems. Remediation demonstrations were conducted on both sites.

For the demonstration program, inert ordnance, representative of the contamination at JPG and other sites, was emplaced on each site and the orientation, horizontal locations, and burial depths were documented. This ordnance ranged from items as small as 60-mm rounds to items as large as 2,000-lb bombs. Other items, including metal fragments and nonordnance items such as beakers, building materials, glass bottles, metal cans, and drums were also emplaced in recorded locations. In addition, some holes were dug and refilled without the emplacement of any item. Prior to the emplacement of items, an explosive ordnance disposal (EOD) team swept the sites to identify potential magnetic signal sources. The EOD team used handheld magnetometers as well as the government prototype Surface Towed Ordnance Locator System (STOLS). As a result of the sweep, several anomalies were added to the emplaced item list as nonordnance.[1] This ground-truth serves as the baseline data necessary to evaluate the performance of technologies demonstrated in the program. The baseline data was not released to the demonstrators and continues to be held by the government.

---

[1] Details can be found in Refs. 4 and 5.

The method of ordnance emplacement and the ordnance layout plan was determined in consultation with former- and active-duty explosive ordnance disposal (EOD) personnel and used data including:

- ordnance testing and training records,
- ordnance delivery and implementation characteristics,
- numerous technical and UXO site reports, and
- area geology information.

Data from clean-up operations at FUDS were obtained from the Army Corps of Engineers and used to select various items commonly responsible for false alarms. After all items were emplaced, both sites were mowed to remove most of the small vegetation, and the areas were "disked" to present a uniform appearance.

C. CAVEATS REGARDING THE INTERPRETATION OF DATA FROM THE JPG DEMONSTRATION

The goal of the site preparation was to emplace ordnance, ordnance-related, and nonordnance items in as realistic a manner as possible. However, there remain differences between the demonstration sites and actual ordnance-contaminated sites as well as differences between the demonstration procedures and actual remediation procedures. As an example, each system demonstration typically involved a single pass on the prepared site; remediation of an actual site would likely involve multiple passes and the removal of items between passes. In this section we discuss some of the most important limitations of the JPG demonstration that both affect the ability to interpret the data accurately and limit the extrapolation of the performance at JPG to sites with widely differing conditions. These limitations should be understood to apply to all of the observations and conclusions presented throughout this report.

Limitations on sweep methodology and lack of information regarding type of ordnance contamination. Many demonstrators indicated that more knowledge of the types of ordnance items and their likely distribution would have helped in properly "tuning" their systems for the specific application. Namely, magnetometer system settings and sweep procedures would be quite different for small, shallow objects than for large, deep objects. Demonstrators indicated that even limited knowledge regarding the ordnance types (this corresponds to knowledge of past usage of the site) may have resulted in an increase in the probability of detection for various categories of ordnance. On the other hand, the demon-

strator work plan supplied before the demonstration contained a list of the items that would be emplaced on the site.

In an actual cleanup operation, it would be possible to perform multiple passes on a particular site and to remove items detected on each pass prior to the next pass. Thus, the system could be adjusted to respond to the different ordnance types expected at the site, i.e., large ordnance at large depths, or small ordnance at shallow depths. Even at sites where little historical information is available, multiple passes on the site at different system parameter settings may result in the detection and removal of more ordnance overall, especially if detected items are removed after each pass.

Geologic conditions challenging. The geologic conditions at JPG are challenging in several respects. On several days there was standing water on some portions of the site, and vegetation on the 80-acre site was reported to be as high as 48 inches. The standing water limited the access of several systems to various areas of the site and surface vegetation limited the effectiveness of infrared systems. Further, the depth of penetration of ground-penetrating radar (GPR) systems is severely limited by high soil conductivity and soil moisture content. Some demonstrators with multi-sensor platforms opted not to use their GPR sensors upon measurement of these parameters at the site. Chapter 6 provides more details.

Presence of other sources of signals. Although the site was cleared before items were emplaced in the ground, the sweeps may not have identified all possible sources of signal. After the testing of systems at JPG, a fence line was discovered and additional analyses performed to quantify the resulting effects on demonstrator performance. Even so, it is likely that not all sources of signals were identified. As a result, the division of "false alarms" measured at JPG into those arising from a deliberately emplaced man-made nonordnance item and those arising from an unknown source may be artificial.

No explosive simulants were used. Nuclear, chemical, and biological (i.e., dogs) sensors that detect high concentrations of explosive material could not be demonstrated at JPG since only inert ordnance was emplaced and no explosive simulants were used. These systems will have different detection capabilities and may provide information for discrimination purposes.

Baseline ordnance distribution. The most significant aspect of the test design that affected the ability to interpret the results is the distribution of baseline items. A nearest neighbor ordnance distribution for the 40-acre and 80-acre sites shows that a large fraction

of the ordnance items are separated from their nearest neighbors by only a few feet or less. This pattern may or may not be typical of the patterns of contamination in real-world sites. If this separation is at or near the limit of the sensor's ability to resolve nearby items, then the measurement of detection capabilities will be affected by the particular pattern of ordnance distribution and the sensor resolution. Chapter 4 details the effect of closely spaced targets on measured performance.

Relative difficulty of areas searched. Many demonstrators did not search the entire site. Because the emplaced ordnance and the local environment vary across the site, it is difficult to compare demonstrators that searched different portions of the site.

Local Conditions. System performance will be affected by climate, weather, soil, and other such conditions that are peculiar to the JPG site. Thus, caution should be used when extrapolating the results to predict performance at different sites.

Statistical Uncertainties. Because a finite number of items were emplaced at JPG, there are statistical uncertainties in the measures of performance. In some cases, these uncertainties can be large enough to reverse the order in which one would rank demonstrator performance.

It should be noted that despite these limitations, the JPG demonstration, albeit at one location with specific geological conditions, procedures, ordnance types, and ordnance distributions, provided an opportunity to compare many systems in a blind test. The results of the JPG demonstration provide unique insight into the state of UXO detection technology.

3. MEASURES OF PERFORMANCE AND IMPLICATIONS FOR UXO CLEANUP

There is great debate regarding the efficacy of currently fielded UXO detection, identification, and remediation systems. There is a similarly fierce debate regarding the most appropriate methods for evaluating the performance of these systems. In this chapter we discuss the measures used to evaluate the performance of systems demonstrated at JPG. Since few remediation systems were proposed for testing at JPG, we focus predominantly on systems for detection and identification.

A. MEASURES OF PERFORMANCE

The efficacy of UXO cleanup is driven largely by detection capability. Systems with a high probability of detection will locate more ordnance than systems with a low probability of detection. In this regard, a high probability of detection is desirable. Decades of experience with sensors used for the detection and identification of a myriad of items have shown that it is possible to increase the probability of detection simply by lowering the threshold of the system. This, however, also increases the false alarm rate. Therefore, it is important to consider both detection capability and the corresponding false alarm rate. If the number of false alarms is sufficiently high, the usefulness of the sensor is obviated since the cost and time of remediation will be driven by digging holes that do not yield ordnance items. The remaining performance parameters—target position and depth accuracy, target classification capability, survey rate, and survey costs—were included to allow more refined comparisons between systems of similar performance.

Below we provide a summary and brief description of measures for evaluating the performance of systems for detection and identification:

1. Detection capability

   Detection capability is the ability of a demonstrator to detect ordnance items. At JPG, this is measured as the fraction of emplaced ordnance items detected. Detection capability was determined in several subcategories, for example, the ability of a demonstrator to detect bombs.

2. False negatives

   A false negative is a demonstrator declaration of ordnance that cannot be associated with any object emplaced at the site.

3. False positives

   A false positive is a demonstrator declaration of ordnance that corresponds to an emplaced nonordnance item.

4. Target position and depth accuracy

   The average radial distance between the demonstrator declaration and a matched emplaced item is called target position accuracy or location accuracy. An analogous definition applies to target depth accuracy.

5. Classification capability

   Seven classification ratios were created to measure the demonstrator's ability to correctly classify detected items. In this report, this measure is calculated by dividing the number of items of a class detected and correctly classified, by the total number of items of that class detected.

6. Survey rate

   The survey rate is measured as the portion of the site visited in the allotted 40 hours.

7. Survey costs

   Costs are based on firm, fixed prices from the demonstrator proposals.

Using these measures, we next discuss the implications of various levels of performance for UXO cleanup.

B. IMPLICATIONS FOR UXO CLEANUP

We consider primarily the ground-based detection and identification systems demonstrated on the 40-acre site at JPG. Although the performance parameters discussed are also broadly applicable to airborne systems, the significance of the results for the airborne systems is not discussed in detail. The performance of airborne systems demonstrated on the 80-acre site was poor. Thus, a detailed discussion of the implications of this performance for the cleanup of subsurface UXO is not likely to yield useful insight. The level of performance indicates that the use of airborne platforms for rapid, wide area surveys of land contaminated with subsurface ordnance is not currently viable. Chapter 6, Technology and System Considerations, discusses several possible reasons for the observed poor performance of airborne systems. Performance of airborne systems for the detection of surface ordnance remains to be quantified in a blind test.

Remediation systems are not addressed in this chapter since the role of these systems in cleanup, and hence the performance metrics, are substantially different than those for identification and detection systems. Further, only three remediation systems were demonstrated at JPG and two of the three were developmental in nature. Hence, conclusions regarding remediation may not be representative of the currently available technology.

1. Detection Capability

Detection capability is expressed as the probability of detection, defined as the number of items detected divided by the total number of emplaced items.[1] As such, the ability to mitigate risk is directly proportional to this first measure of performance. Consider an impact area where 1 million rounds have been fired during testing. It is common to use 10 percent as an estimate of the "dud rate," or the fraction of ordnance items that fail to explode. Hence, 100,000 rounds remain as potential explosive hazards. A sensor system with a probability of detection equal to 90 percent, would allow 90,000 of the 100,000 potentially hazardous rounds to be located, leaving 10,000 such rounds undetected. A system with a probability of detection equal to 70 percent would leave three times as many ordnance items undetected. At JPG, the detection capability of all available systems demonstrated was less than 70 percent, and the average detection capability was 30 percent.[2] Thus, systems with the highest probability of detection demonstrated at JPG would leave approximately 30,000 of the 100,000 potentially hazardous ordnance items on the hypothetical impact area described above; a probability of detection equal to 30 percent would leave 70,000. Despite caveats regarding the ability to extrapolate this performance to other sites, this performance is well below the advertised capability of these systems.

The expense of ordnance remediation has driven efforts to include future land use considerations in the decision to remediate UXO-contaminated land. In cases where the land must be remediated, restrictions on future land use might be considered in the decision regarding the depth of clearance. For example, if the land were to be used as a national

---

1 Demonstrators were scored using several different measures of detection capability; these are described in detail in Chapter 4. For our purposes in this section, these distinctions are not critical.

2 Many demonstrators did not search the entire 40-acre site; therefore, the probability of detection quoted here is defined as the number of ordnance items detected divided by the number of emplaced ordnance items on the area searched only. Plastic antipersonnel mines have been removed from the emplaced item list. Details are provided in Chapter 4. The average detection capability is the average of the detection capabilities of the individual demonstrators on their area searched as opposed to the total number of detections divided by the total number of opportunities on the cumulative area searched.

park, clearance to a depth of 5 m (15 ft) would be unnecessary. Determining the relative ease with which ordnance can be detected at various depths would be useful for framing such policy decisions. In Chapter 5 we present the probability of detection demonstrated at JPG as a function of the depth of the emplaced item.

2. False Alarm Rate

As mentioned above, it is possible to increase the number of detections obtained by any sensor by lowering the threshold level at which a detection is declared. In the limit of near-zero threshold, this endeavor would create an area virtually littered with declarations. This approach would also require the removal and sifting of all soil at a site to the maximum depth of remediation desired. Under such circumstances, the sensor ceases to be useful. Of course, this is an extreme example; nevertheless, it illustrates the point that for a system to be considered viable, not only must the detection capability be high, but the false alarm rate must be manageable. Figure 5.1, in Chapter 5, shows how the demonstrators on the 40-acre site at JPG fared with regard to both probability of detection and false alarm rate.

In the case of UXO detection, a manageable false alarm rate is one that can be tolerated given operational constraints and resource limitations. Namely, if a detection-driven remediation approach is taken (as opposed to a strip mining approach), then the cost of remediating a segment of UXO-contaminated land will be dominated by, and hence proportional to, the number of detections that must be investigated. Digging holes that yield ordnance items is a necessary part of the remediation cost, but resources are wasted by digging holes that do not yield ordnance items.

In the vernacular of the JPG demonstration, there are two types of false alarms, false negatives and false positives. A false negative is a declaration that cannot be associated with any object emplaced at the site; i.e., a false negative may be the result of a geologic feature, system noise, or any man-made object not deliberately emplaced as part of the demonstration. False negatives that result from geologic features or system noise may have serious implications for the remediation timeline in that extended search efforts may be required for remediation personnel to be convinced that there is not an ordnance item or other object present. False positives are declarations thought by the demonstrator to be ordnance but that are other emplaced man-made debris. It should be noted a false positive may be more readily verified (i.e., an object is found) than a false negative. In this

respect, a false positive may not affect remediation timelines or costs as significantly as a false negative.[3]

Some in the sensor-development community object to describing the detection of man-made objects and geologic features as false alarms since they are clearly associated with objects that would be expected to generate a signal. For example, an induction coil would be expected to detect an aluminum soda can and a magnetometer would be expected to find a highly magnetic rock. Others argue that any detection identified with a non-ordnance item would lead to remediation efforts that do not reduce the UXO risk and hence should be viewed as a false alarm. Independent of the definitions, the goal is to remove the hazard associated with UXO; both false negatives and false positives impede the ability to accomplish this objective effectively and efficiently.

Greater than half of the demonstrators had more than 10 times as many false negatives as false positives. It should be noted that a false positive can only be assigned in cases where a declaration corresponds to a *known* emplaced nonordnance item. There may be unverified man-made objects in the locations designated as false negatives. However, if the demonstrator false negatives corresponded to subsurface items, then we would expect a correlation of false negatives between demonstrators. In some cases where several demonstrator false alarms corresponded to each other, an investigation was conducted and if an item was located, this item was added to the baseline. This could not be accomplished in all cases; we assume that the preponderance of the remaining false alarms were not caused by unverified manmade objects.

As an example, let us take a demonstrator to the hypothetical impact area described above in Section 1 Detection Capability. We will use a demonstrator with performance similar to that of an above-average demonstrator on the 40-acre site at JPG. Recall that this impact area contains approximately 100,000 potentially hazardous ordnance items. As a conservative estimate, we assume that each of the 900,000 additional ordnance items that exploded as intended results in one object in the impact area that would cause a detection.[4]

---

3 Of course, it must be verified that the object recovered is, in fact, the object responsible for the signal.

4 If we assume that ordnance items of widely different sizes have been used on the impact area, for example, 20-mm rounds and larger projectiles or bombs, then this assumption seems reasonable. The fragments of large exploded ordnance items would be comparable in size to small ordnance items. Some of these small fragments would remain in the ground even after a surface clearance. If, on the other hand, the impact area was used for only large ordnance items, then sensors may be able to adequately discriminate small, shallow debris from large, deep ordnance items. In the latter case, assuming one debris item per exploded ordnance item fired would artificially inflate the estimated false positive rate.

For convenience, we will use an ordnance probability of detection equal to 50 percent and we will assume that the probability of detection for nonordnance is also equal to 50 percent. The number of false negatives is calculated by estimating the size of the impact area from other impact area ordnance densities and using the false alarm rate demonstrated at JPG on this area. We assume no ability to distinguish between ordnance and nonordnance items, i.e., no classification capability. (This last point is discussed below.) This combination of the probability of detection and false alarm rates would yield the following remediation on the hypothetical impact area:

Number of Holes	*Found in the Hole*	*Percentage of Total Number of Holes*
50,000	Hazardous ordnance item	10 %
450,000	Ordnance-related item, "false positive"	89 %
3,000	Nothing, "false negative"	1 %
Total = 503,000		

The number of holes that yield ordnance-related items far exceeds the number of holes that yield intact, hazardous ordnance items. Further, despite digging approximately 500,000 holes, this hypothetical demonstrator would leave 50,000 potentially hazardous ordnance items remaining on the site. For a demonstrator with a significantly higher false negative rate, the results would be yet more disturbing.

Of course, many UXO-contaminated sites are not impact ranges. These areas may be lightly contaminated with ordnance but will not be heavily contaminated with ordnance debris. At these sites, ordnance debris will not be the driving factor in the generation of false alarms. Instead, most of the effort will be in exploring miscellaneous sources of signals such as those generating false negatives in the JPG demonstration. As an example, during a pipeline construction in Chocolate Mountain, CA, several hundred signals had to be investigated to find four 500-pound bombs (Ref. 6).

3. Combining Detection Capability and False Alarm Rate

It is difficult to compare the relative performance of demonstrators with high probabilities of detection and high false alarm rates to those with low probabilities of detection and low false alarm rates. In Chapter 4, we discuss in detail the use of a receiver operating characteristic (ROC) curve based on gaussian noise to compare systems with widely different performance. A ROC curve illustrates diminishing returns: at some point

along the curve, a modest increase in detection capability is associated with a disproportionately large increase in false alarms. In this chapter, we compare systems based on the effect of demonstrator performance on UXO remediation. To accomplish this, we need a trade-off curve, analogous to the ROC curve, but that shows how increased probability of ordnance recovery is related to the percentage of the site dug up.

Figure 3.1 shows, for a single demonstrator, the probability of detection[5] against the fraction of site surface area that must be disturbed to recover ordnance items. The fraction of the site dug up is calculated as the fraction of the site surface area covered by circles of radius R_{crit} drawn around each demonstrator declaration. The points on the curve are determined by progressively increasing R_{crit} until the entire site is covered and calculating the number of ordnance items recovered at each fraction of the site dug up. This provides a curve with "probability of detection" varying between 0 and 1. It should be noted that at large values of the radius, additional items are not truly detected but rather are uncovered at random. The effect of overlapping circles is estimated as

$$\text{fraction of site area} \approx 1 - e^{-x}$$

$$x = \frac{N\pi r^2}{A_{site}}$$

where

N = number of holes

r = radius of hole

A_{site} = area of entire site.

(This estimate breaks down for large values of r; therefore, we also calculated exactly the fraction of the site covered. We found no appreciable difference in the results.)

In practice, the probability of detection would likely be increased by lowering the threshold of the sensor and tolerating more false alarms rather than by digging progressively larger holes around existing declarations. This would also increase the fraction of the site that must be disturbed, although the functional behavior would be different. Nevertheless, this approach is useful for comparing the performance of different demonstrators at their chosen operating conditions at JPG.

---

5 The pattern of ordnance emplacement on the JPG site prompted us to define several measures to bound the probability of detection. For this calculation, the probability of detection is estimated as the fraction of ordnance items that would be uncovered in holes of a specified radius R_{crit}. For detailed definitions refer to Chapter 4.

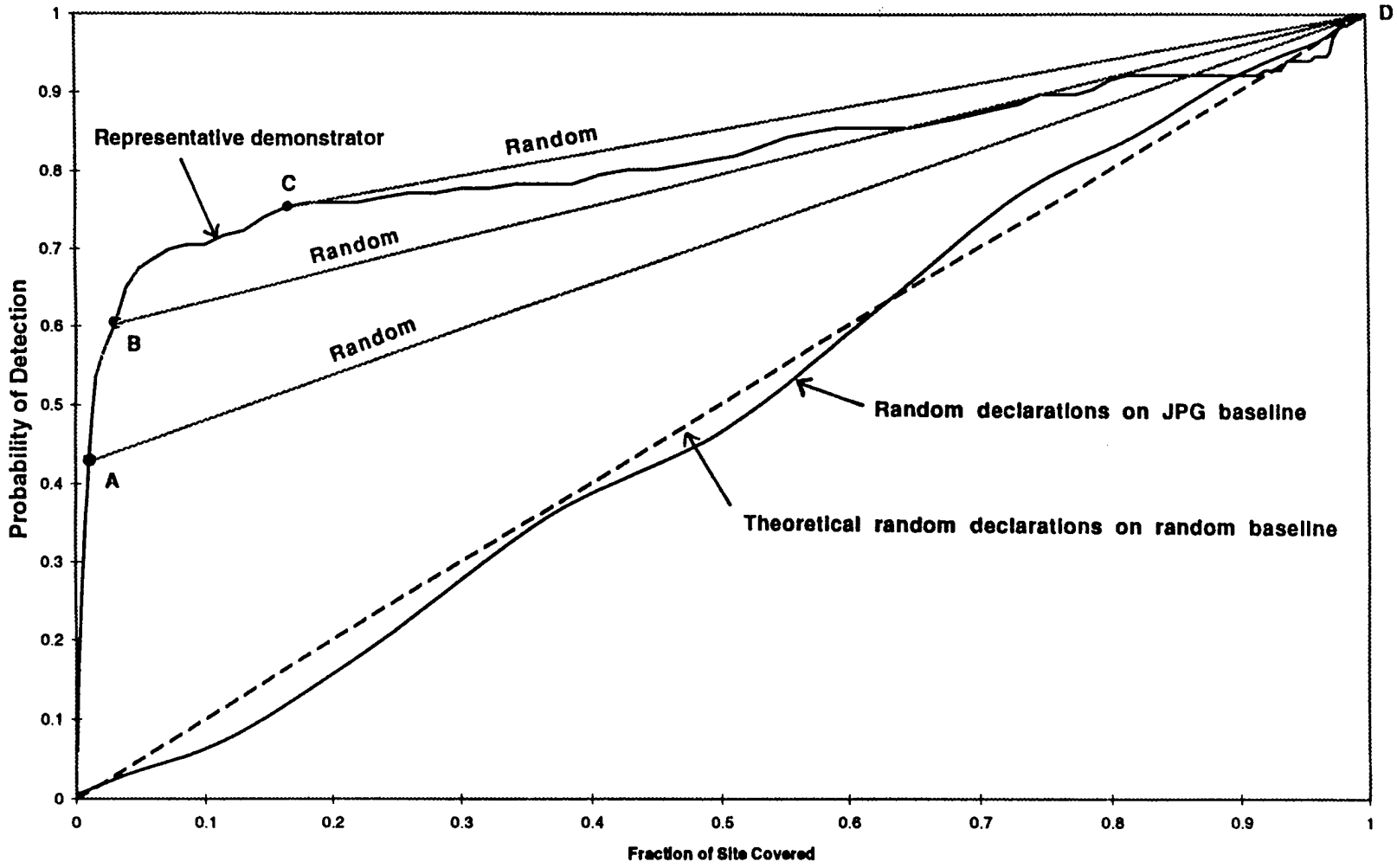


Figure 3.1. Probability of Detection [P_{near}(ord)] vs. Fraction of Site Covered Calculated by Varying R_{crit}

The performance of any demonstrator can be measured against the performance that would be achieved by randomly placing declarations. The dashed line in Fig. 3.1 shows the theoretical performance of a demonstrator with random declarations on a hypothetical site where the items have been emplaced randomly. This shows that *overall* the use of the sensor by the representative demonstrator allows more ordnance to be located at each fraction of the site covered. However, it is more useful to compare the added benefit of the sensor against random declarations *at each point along the curve.* At any point along the representative demonstrator curve in Fig. 3.1, a straight line to a probability of detection and fraction of site covered equal to one (point D on Fig. 3.1) shows the functional behavior of further "detections" due to randomly uncovering further items in progressively larger holes. This allows the sensor performance to be compared against "random hole digging." For example, at points A and B the sensor curve slope is greater than that of lines A-D and B-D, indicating that the sensor is still performing better than random. At these points, for a given increase in the probability of detection, the sensor allows a smaller increase in the fraction of the site covered than would be expected by random declarations. At point C, however, continuing to use the sensor results in performance that is worse than random. An anticorrelation between the emplaced items that were easy to detect versus those that were difficult to detect would create such a result. Such an anticorrelation would require that circles around detections be quite large before the ordnance items remaining at the site were detected. To determine if this anticorrelation is a function of the baseline distribution, we calculated the "probability of detection" for a demonstrator with declarations placed randomly on the JPG baseline. This curve is labeled "Random declarations on JPG baseline" in Fig. 3.1. This curve illustrates that the baseline ordnance distribution does not account for the anticorrelation, since the curve for a demonstrator with random hits does not deviate significantly from the theoretical curve. This anticorrelation is larger for some demonstrators than for others; we have not been able to identify the source of this deviation from random behavior.

As a coarse estimate, let us assume that the cost of site remediation is proportional to the fraction of the site that must be dug up. Figure 3.1 shows that, initially, a modest increase in the radius of the hole dug around each declaration has a large effect on probability of detection. Up to a probability of detection of 70 percent, less than 10 percent of the site needs to be disturbed, and there are clear advantages to using the sensor. As described above, as long as the slope of the curve is greater than the line drawn to point D, the sensor provides an advantage over digging random holes. This initial steep climb of probability of detection is an indication of the demonstrator's location accuracy, and there is

little influence of false alarms, although the size of the holes around the false alarms is growing also. Above a detection capability of 70 percent, the curve flattens. In this flat region, the size of the holes continues to grow, but few additional ordnance items are recovered since all additional ordnance items are uncovered at random. To increase the number of ordnance items recovered to 80 percent would require 40 percent of the site surface area to be disturbed.

Figure 3.2 shows the probability of detection against the fraction of site disturbed for a representative sample of demonstrators. This allows the performance of demonstrators (at the operating conditions chosen for JPG) to be compared over a wide range of probabilities of detection. The demonstrators with initially steep slopes are superior to those with initially shallow slopes, since they allow more ordnance to be recovered at low fractions of site area disturbed. For these demonstrators, even after the performance is random, the initially high probability of detection allows more ordnance items to be recovered with a smaller fraction of site area disturbed. Figure 3.2 shows the value of comparing demonstrator performance over a wide range of probabilities of detection: some of the demonstrators with initially high probabilities of detection are not the best over the entire range; for example, see demonstrators Coleman and UXB.

4. Distance and Depth Accuracy

The ability to accurately locate a subsurface ordnance item both decreases the amount of earth that must be removed to remediate an ordnance or nonordnance item and helps one to choose the proper tools for remediation. The data on distance and depth accuracy achieved by the demonstrators at JPG are provided in Chapter 4.

The volume of earth that must be moved to remediate an emplaced ordnance item increases as the square of the radius of the hole that must be dug.

$$V_{\text{earth removed}} \propto r^2$$

In other words, decreasing the radius of the hole by a factor of two decreases the volume of dirt that must be removed by a factor of four. Therefore, distance inaccuracies can have a significant effect on the amount of dirt that must be removed to locate and remediate subsurface items.

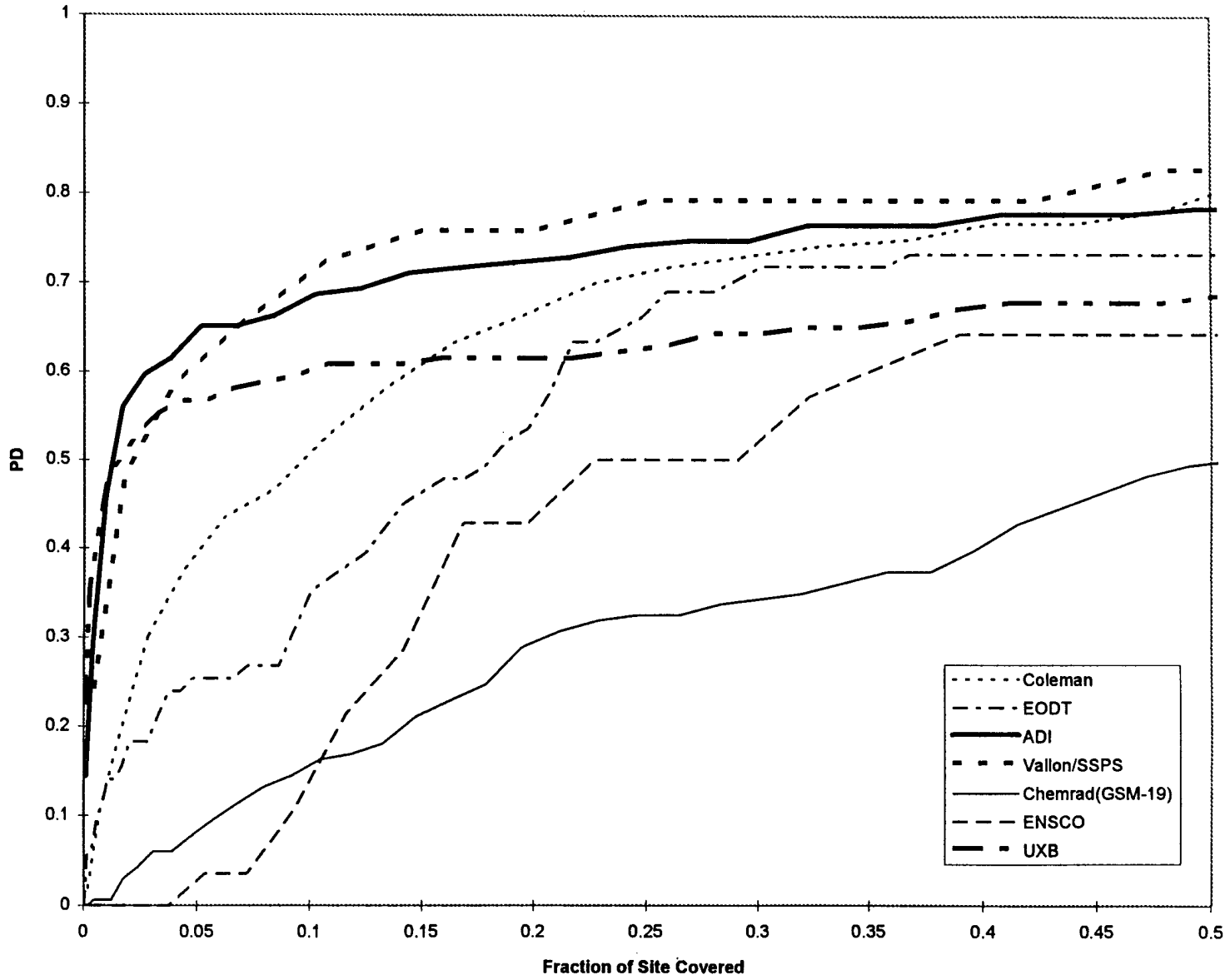


Figure 3.2. P_{near}(ord) vs. Fraction of Site Covered Calculated by Varying R_{crit} for a few Representative Demonstrators

To illustrate the effect of distance accuracy on remediation, we calculated the minimum surface area that must be disturbed to recover all emplaced ordnance items. The minimum surface area required to recover one ordnance item is calculated as the area of a circle with diameter equal to the maximum dimension, l, of an emplaced ordnance item when the dimensions of the ordnance item are projected to the surface. The total minimum area is the sum of the area of all ordnance items detected by a particular demonstrator. The actual area required for remediating these detected items is calculated using the demonstrator's average location accuracy, $\bar{r}$, plus one standard deviation as the radius of a circle.[6] We then also add the portion of the minimum area circle that lies outside the demonstrator's circle. These areas are described by the following equations and Fig. 3.3:

$$A_{min} = \Sigma_{\text{ordnance detected}} \frac{\pi}{4}(l)^2$$

$$A_{loc} = \Sigma_{\text{ordnance detected}} \left[\pi(\bar{r} + 1\sigma)^2 + (\text{portion of } A_{min} \text{ outside } A_{loc}) \right]$$

The ratio of these two numbers reflects the impact of location accuracy on remediation activities. To compare the importance of location accuracy to the impact of false alarms, we also calculated the amount of surface area that must be disturbed when both the location inaccuracy and the false alarms are considered.

$$A_{loc+FA} = \Sigma_{\text{ordnance detected}} \left[\pi(\bar{r} + 1\sigma)^2 + (\text{portion of } A_{min} \text{ outside } A_{loc}) \right]$$
$$+ \Sigma_{\text{FAs}} \left[\pi(\bar{r} + 1\sigma)^2 \right] \quad .$$

Since demonstrators had different detection capabilities, Fig. 3.4 compares the ratio of A_{loc}/A_{min} and $A_{(loc+FA)}/A_{min}$. This figure shows that even demonstrators with above average distance accuracy typically disturb 5 times the minimum amount of surface area required; most demonstrators disturb between 10 and 100 times the minimum surface area required. Nevertheless, Fig. 3.4 shows that at JPG, the effect of false alarms is in many cases an order of magnitude more detrimental than location inaccuracies.

---

6 See Chapter 4 for detailed definition of location accuracy.

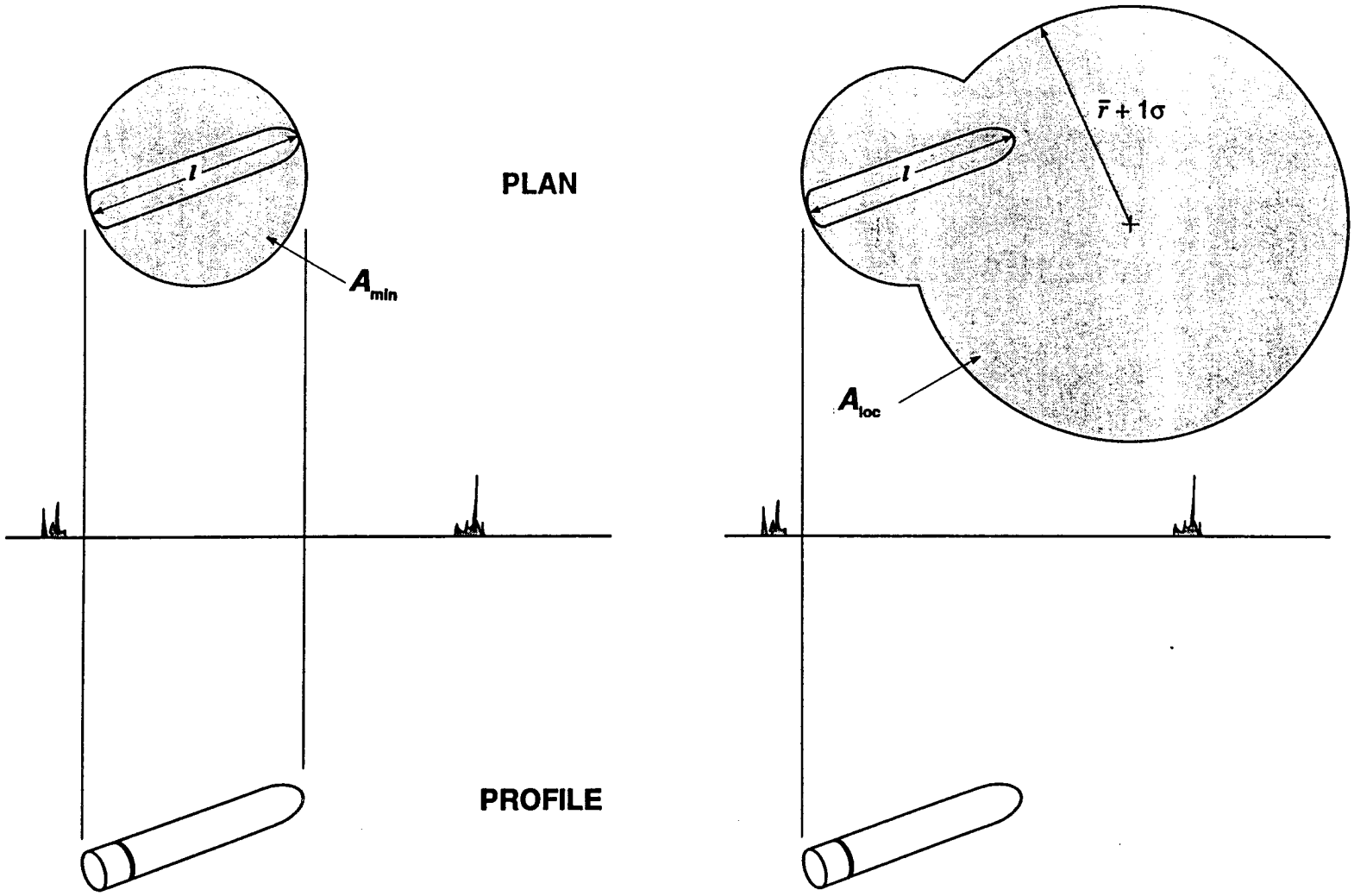


Figure 3.3. Minimum Area Required to Recover an Ordnance Item and Area Required Given Demonstrator Location Inaccuracies

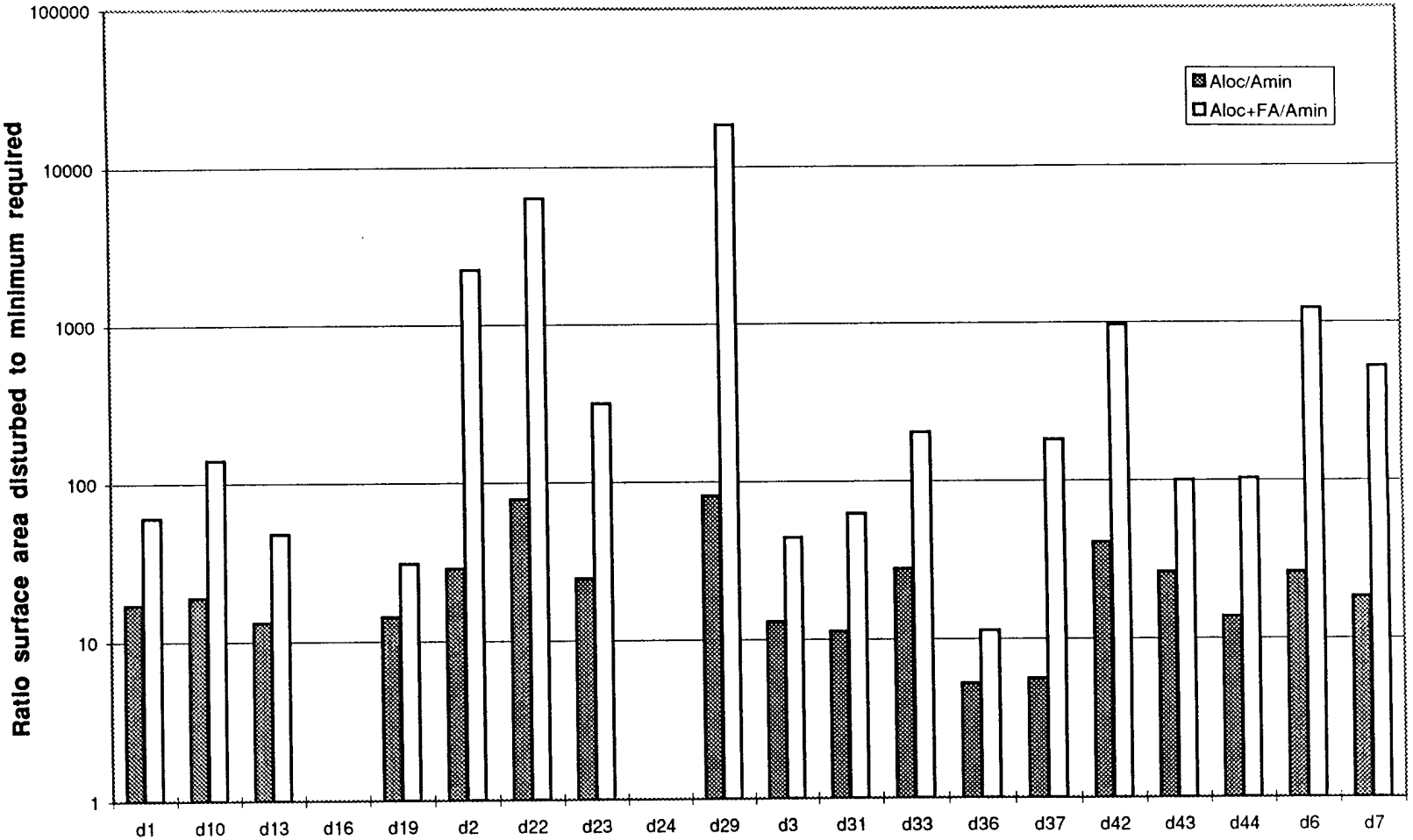


Figure 3.4. Comparison of the Effect of Location Inaccuracies to the Effect of False Alarms on Amount of Area That Must Be Disturbed to Recover Ordnance Items

It would be useful to express these numbers in terms of volume of earth that must be removed rather in terms of surface area. However, it is difficult to anticipate the impact of depth inaccuracy on remediation: even if depth is underestimated, remediation activities will likely continue until the item is found; similarly, if the depth is overestimated, remediation activities would cease once the item is located. Further, the amount of earth that must be removed is not a linear function of the depth since, for safety reasons, it is not possible to dig straight-walled holes.[7] Thus, we make no attempt to quantify the effect of depth inaccuracy on cleanup activities.

5. Classification Capability

Most of the sensors used for the detection of unexploded ordnance capitalize on the metal content of ordnance items. The result of this approach is that all metal objects, such as ordnance fragments and construction materials, which are common on impact areas, are also detected. In the above example, wherein 1 million ordnance items were fired in an impact area, the estimated number of hazardous ordnance items was 100,000. If we estimate the number of non-hazardous ordnance-related items to be the balance, or 900,000, then a sensor with a detection capability of 90 percent would find not only 90,000 of the 100,000 hazardous ordnance items, but also 810,000 of the 900,000 non-hazardous ordnance items. Further increases in detection capability without concomitant increases in discrimination capability would increase the fraction of both hazardous and non-hazardous items detected. Clearly, the cost of remediation would be driven by the examination and removal of non-hazardous items. The ability to distinguish between potentially hazardous ordnance items and other nonhazardous items would reduce both the cost and the time required for remediation efforts.

Both ordnance and nonordnance items were emplaced at JPG. Figure 3.5 illustrates the relative ability of demonstrators to *detect* ordnance and nonordnance items, regardless of their ability to correctly classify the items. This figure shows that the ability of demonstrators to detect ordnance and nonordnance is comparable. Thus, as the above example illustrates, an ability to classify detections as ordnance or nonordnance, would be critical in controlling cleanup costs. Notably, of the 20 demonstrators at the 40-acre site, 12 did not even attempt to distinguish between ordnance and nonordnance, and one found nothing. Of the remaining seven demonstrators, three found only ordnance and two found

---

7 Worker safety considerations influence the size of the hole that must be dug: to avoid cave-ins, holes must have sloping sides.

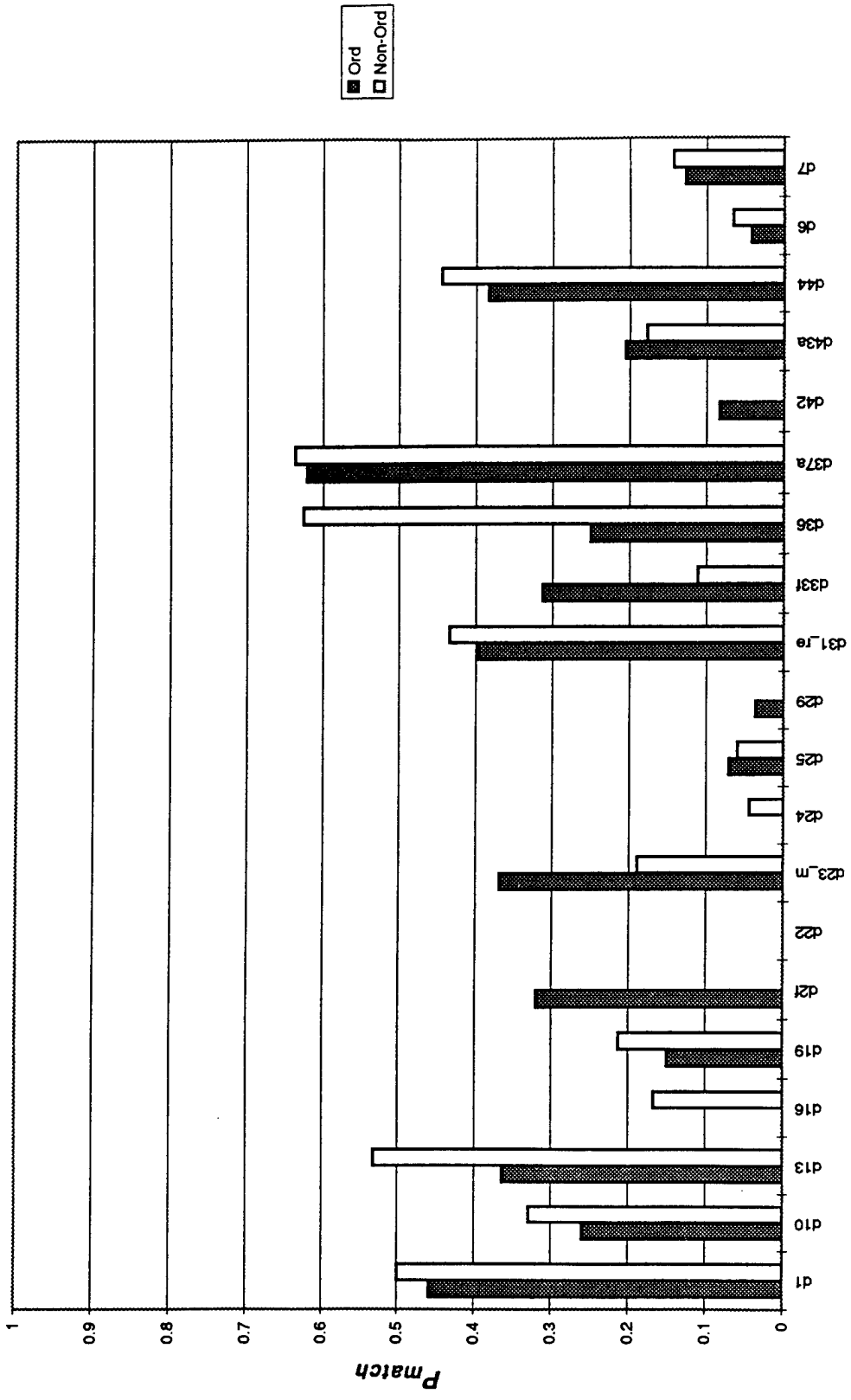


Figure 3.5. P_{match}(ord) and P_{match}(nonord) vs. Demonstrator

only nonordnance. Thus, only two of the demonstrators that attempted to distinguish between ordnance and nonordnance found some of both. For these two demonstrators, the probability of correctly identifying ordnance was greater than 90 percent, but the probability of correctly identifying nonordnance was less than 10 percent. This is a result of the demonstrators declaring nearly all items as ordnance and translates to essentially no classification capability whatsoever.

Ultimately, the ability to more accurately classify buried ordnance is important for the purpose of establishing the level of contamination at a UXO-contaminated site and also for remediation. The classification capability demonstrated at JPG indicates that it is not currently possible to distinguish ordnance from nonordnance and ordnance debris, much less specific ordnance items from one another. Most demonstrators at JPG showed no appreciable ability to provide even coarse classification sufficient to guide remediation equipment choices, i.e, shovel or backhoe.

6. Survey Rate

The prevalence of contaminated land in the United States and abroad requires that UXO detection surveys be conducted as rapidly as possible. One would expect some tradeoff between the speed of survey and the reliability of the survey. The demonstrators with the fastest survey rate were those that searched the entire site in the allotted 40-hour time period. We might expect some of the demonstrators with slower survey rates, i.e., those that did not search the entire site, to have higher detection capabilities than those that searched the entire site. This is not the case. The demonstrators with the highest survey rate are also those with the highest probability of detection. Therefore, to achieve the higher detection capabilities and lower false alarm rate demonstrated at JPG would require 1 hour per acre of land surveyed (40 hours/40 acres); to investigate the 11 million acres of land in the United States potentially contaminated with UXO would require 11 million hours. Assuming a 10-hour work day, a 5-day work week, a 50-week work year, and 100 systems, this corresponds to 44 years. An order of magnitude increase in the number of systems would allow the survey to be completed in 4.4 years. This estimate assumes that all potentially contaminated land can be surveyed at the same rate as JPG, and that surveys can be conducted year round, assumptions that doubtless make the estimated survey time low.

Clearly, it would be desirable to do wide area, rapid surveys using airborne platforms. However, the performance of airborne systems demonstrated at the 80-acre site at

JPG indicates that it is not currently possible to detect and accurately locate subsurface ordnance in this manner. Therefore, some prioritization of sites is required to accommodate more time-consuming ground surveys.

7. Survey Cost

The cost of surveying the 40-acre site is listed by demonstrator in Ref. 4. Figure 3.6 shows the probability of detection (measured as P_{group} (ordnance only); see Chapter 4) on the area surveyed and the cost of each of the demonstrators on the 40-acre site. For the demonstrators with higher detection capabilities, the survey cost ranges between $1,000 and $10,000 per acre. However, for three of the demonstrators with low false alarm rates, the survey cost is approximately $2,000 per acre. If we extrapolate this cost to the 11 million acres potentially contaminated with UXO, the *survey cost alone* would be in excess of $22 billion. Even the least expensive system demonstrated at JPG was approximately $1,000 per acre, which is only a factor of 2 decrease in the above estimated cost. These numbers should be viewed as preliminary since many of the systems demonstrated at JPG were developmental in nature. Alternatively, some of the demonstrators may have partially subsidized their demonstration costs, which would make these estimated costs low. Travel expenses have not been eliminated from the demonstrators' reported total cost at JPG; however, logistical support and surveying stakes were provided at no cost.

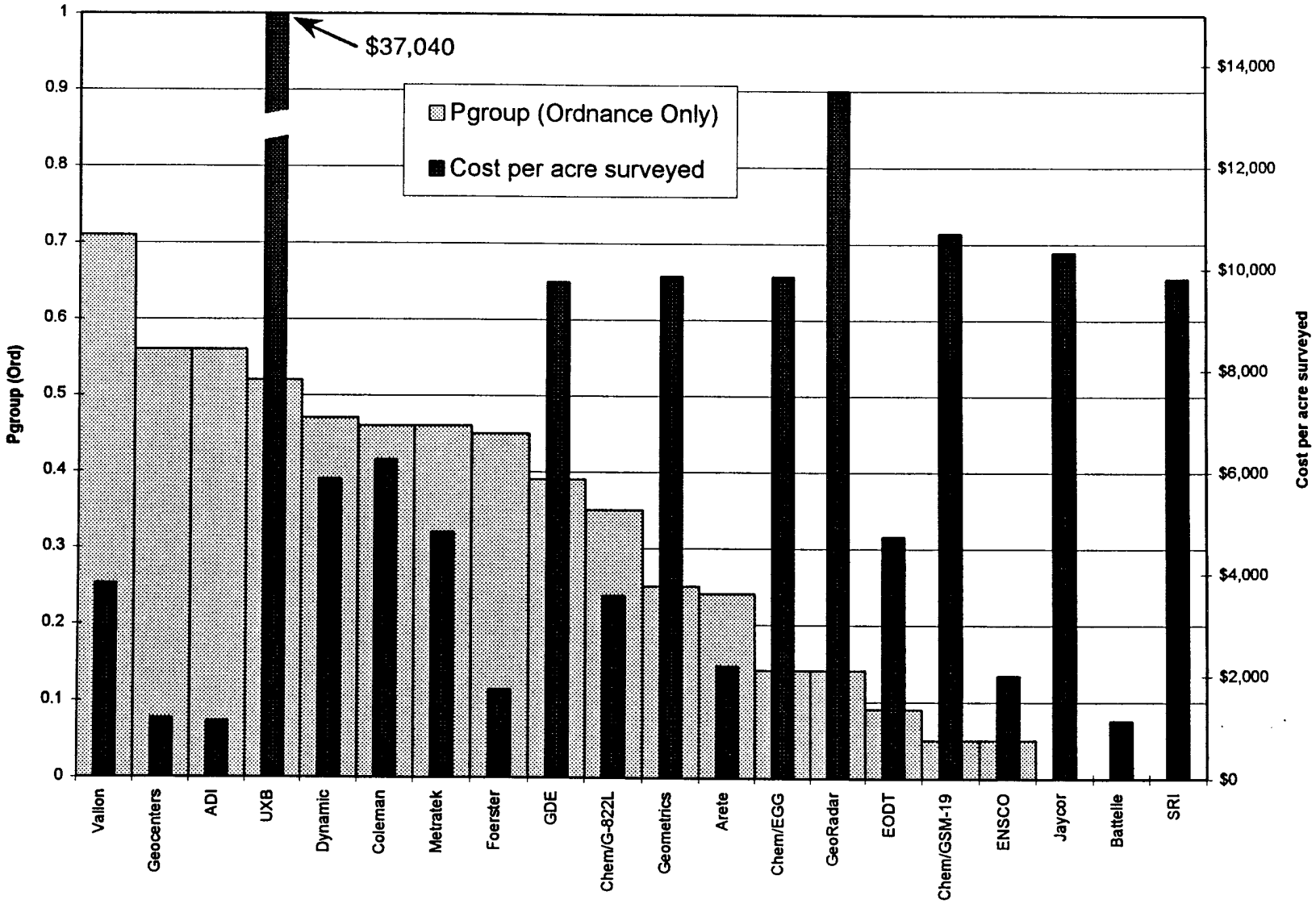


Figure 3.6. P_{group}(ord) and Cost (Both per Acre Surveyed) vs. Demonstrator

4. CALCULATION AND CRITICAL EXAMINATION OF MEASURES OF PERFORMANCE AT JPG

In a test such as the JPG demonstration, analysis begins with a series of demonstrator declarations, i.e., information that describes where the demonstrator thinks buried objects are, and ground truth, i.e., knowledge of where the buried objects actually are. The first step in assessing the performance of demonstrators at JPG is to determine which demonstrator declarations match the emplaced baseline items. This is the most fundamental part of any analysis, since matched declarations count toward detection capability and unmatched declarations count toward the false alarm rate.[1] At the same time, the process of establishing matches is nontrivial. Consider questions such as:

> *How close does a demonstrator declaration have to be to a known emplaced item to be considered a match?*
>
> *If a demonstrator littered an area with declarations, how should the detection capability be calculated given that many of the matches were likely lucky?*
>
> *If there are several closely spaced emplaced items and a demonstrator makes only one declaration in this area, did the demonstrator detect the group of items or only one individual item?*
>
> *If a demonstrator used a sensor that does not respond to the physical characteristics of a certain type of emplaced ordnance items, should these items be counted as missed?*

In this chapter we address these questions and calculate the impact on the measures of performance for JPG. The method of assigning matches and the robustness of various methods is important since it affects the results significantly enough to change our understanding of the performance at JPG.

A. TARGET MATCHING ALGORITHM

To analyze the results of the JPG demonstration, the government contracted with Automated Research Systems, Inc., to write a computer program that would establish

---

1 Results presented in the IDA reports differ slightly from the measures in the Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase 1) report prepared by PRC Inc. (Ref. 5). The reasons for these differences are discussed in Appendix A.

matches between demonstrator declarations and emplaced items. This program is called the Target Matching Algorithm (TMA) and was used to calculate all the results presented in the PRC report (Ref. 2).

The TMA uses a scoring mechanism to arrive at a one-to-one pairing of emplaced baseline targets with demonstrator declarations. The inputs to the TMA include a critical radius, depth, and azimuthal and declination angles. For a demonstrator declaration and an emplaced item to be considered as a potential match, their horizontal separation (projected to the surface) must be less than the critical radius, their depth discrepancy less than the critical depth, and so on. For this demonstration, a horizontal separation less than the critical radius was the only requirement for a match. Other attributes were used only to select a preferred pairing when more than one demonstrator declaration was within the critical radius of an emplaced target or vice versa.

The TMA examines each demonstrator declaration in turn, moving from west to east, finding all emplaced targets that are within the critical radius of each demonstrator declaration, and computing a "score" for each of these potential matches. Points are awarded to a potential pairing for agreement on location, depth, size, classification, type, and angles. The number of points awarded for a match is not affected by the proximity of the match; i.e., all targets within the critical radius are awarded all of the points allotted for a location match. After all demonstrator declarations have been processed, the TMA revisits each declaration to pair it with the emplaced target that has the highest score. At this point, it is possible for a single emplaced target to be matched to more than one demonstrator declaration. To detect such an occurrence, the TMA next visits each emplaced item. If an emplaced item has been matched to more than one demonstrator declaration, the pairing with the highest score is used. The other demonstrator declaration is matched to the emplaced item that has the next highest score. The algorithm repeats this process until there are no more emplaced targets matched to two or more demonstrator declarations. A similar process is used if more than one emplaced item is matched to a demonstrator declaration. Thus, from this algorithm, a set of "optimized" one-to-one matches between baseline targets and demonstrator declarations is constructed.

For two emplaced items within the critical radius, R_{crit}, of a demonstrator declaration, the first tie breaker in the TMA is depth match. For example, consider two baseline items that are within $R_{crit} = 2$ m of a demonstrator declaration; one item is 10 cm from the declaration and the other is 1.5 m away. The depth of the emplaced item 1.5 m away is within the critical depth set in the TMA, but the depth of the emplaced item 10 cm away is

not. In this situation, the farther target will be allotted more points and hence matched to the declaration, even though it is likely that the declaration was prompted by a return from the nearer target. (This, of course, assumes that the demonstrator has good navigation ability.) If both emplaced items match the declaration in critical depth (or if neither match) the tie is then broken on the basis of size. The result of this tie-breaking scheme is that the characterization of demonstrator capability for determining target depth in the TMA is corrupted by preferential depth matching. The depth characterization capabilities reported for the JPG experiment from this algorithm are thus better than they would be if all declarations were matched to the nearest target in horizontal space. In practice, the critical depth selected was very large, and therefore few ties were broken on the basis of depth match. Instead, most ties were broken on the basis of the next criteria, size match. In either case, matching demonstrator declarations to emplaced items that are farther away than the item generating the return will result in a measure of demonstrators' location ability that is worse than it would be if declarations were matched to the nearest emplaced item. The tie-breaking scheme of the TMA will also affect measurement of the classification ratios, typing, and sizing abilities.

As mentioned above, the TMA has distinct limitations. In later sections, we consider alternative algorithms for matching baseline items with demonstrator declarations as a means of quantifying the limitations of the TMA. To distinguish among these alternatives, the probability of detection determined using the TMA is referred to as P_{match}. Although we do not consider P_{match} to be the best representation of probability of detection as it is usually meant, analysis and interpretation using the numbers derived from the TMA provides consistency with the PRC report.

Figures 4.1 (a) and (b) show the probability of detecting ordnance items calculated using the TMA, P_{match}(ord), for each demonstrator on the 40-acre site. Figure 4.1 (a) shows the number of matches over the whole site regardless of the fraction of the site searched by the demonstrator. As most of the demonstrators searched only a portion of the site, Figure 4.1 (b) shows the probability of detecting ordnance only on the area searched. It should be noted that the reported searched area may not exactly correspond to the grid square boundaries that are used to determine the baseline data sets for demonstrators that did not search the entire site, and the relative difficulty level of a particular demonstrator's searched area is not accounted for.

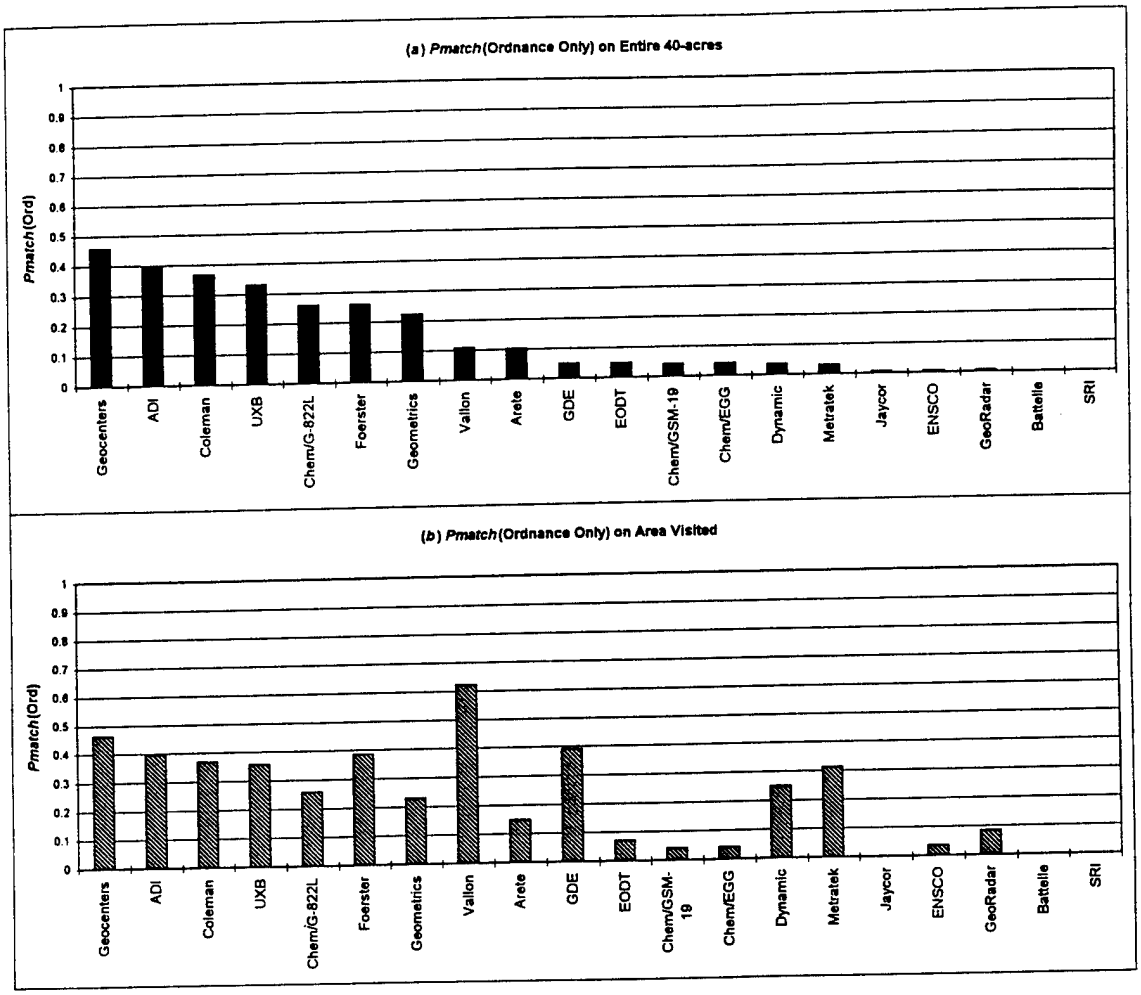


Figure 4.1. Performance of Ground-Based Detection Systems

B. CHOOSING A PROXIMITY REQUIREMENT FOR MATCHES

A demonstrator declaration is considered to match an emplaced target if the horizontal separation of the two is less than some critical radius, R_{crit}. Because the probability of a match will increase with R_{crit}, the selection of R_{crit} will influence the results. Further, as the critical radius becomes large, more demonstrator declarations will fall inside the region of consideration for a match. Especially for demonstrators with a high number of false alarms, this artificially inflates the calculated detection capability. We refer to these types of matches as "lucky" matches, and at large values of the critical radius, the detection capability of some demonstrators needs to be adjusted for these occurances.

As a practical matter, the effect of the selection of R_{crit} on detection capability would not be an issue if false alarms had been sparse, so long as the R_{crit} chosen was large

enough to account for location inaccuracies.[2] It is an issue for the JPG data because there are multiple false alarms per ordnance item detected, and false alarm rates vary widely among demonstrators. If R_{crit} is set too high, a demonstrator with poor detection capability and a high false alarm rate might be credited with a larger number of matches than a demonstrator with a good detection ability but a low false alarm rate. If R_{crit} is too small, location inaccuracy dominates the ability to establish matches, and a demonstrator that successfully finds emplaced items, but determines their exact locations inaccurately, will be ranked below a demonstrator that is not as proficient at finding emplaced items but has more precise location measurements.

To avoid such biases in the measurement of demonstrator performance, the detection capability must be determined by a means that is relatively unaffected by a high false alarm rate, and stable against changes in R_{crit}. This is accomplished with a judicious choice of the critical radius. First, R_{crit} should be chosen at a point where the change in detection capability with R_{crit} is slow. Second, in some cases a correction to the detection capability that accounts for lucky matches is necessary.

The number of matches (and hence the detection capability) is a monotonically increasing function of the critical radius. However, the number of matches increases differently for each demonstrator, depending in part on the total number of declarations. As a notional example, Figure 4.2 shows the number of matches versus R_{crit} for two imaginary demonstrators. Demonstrator B, with good (horizontal) location accuracy, has a greater initial slope than demonstrator A, with poor location accuracy. In the region of large R_{crit}, for demonstrator A, with a high number of false alarms, the number of matches will increase faster with R_{crit} than for demonstrator B, with fewer false alarms. Thus, the relative ranking of demonstrators on detection ability can be directly affected by the selection of the critical radius. Further, Figure 4.2 illustrates that the ranking on detection capability alone, without considering location accuracy and false alarm rate, is not meaningful. For example, if we consider demonstrators A and B, who declared 600 and 250 targets, respectively, at points where they would uncover equal numbers of emplaced items, we observe the following:

- At the point where each would uncover 80 targets, demonstrator A would have 520 false alarms and the holes dug would be of radius 2 m. On the other hand,

---

2 By location inaccuracies we mean the inability to pinpoint the location of an item. For most demonstrators, we believe that location accuracy is limited primarily by navigation.

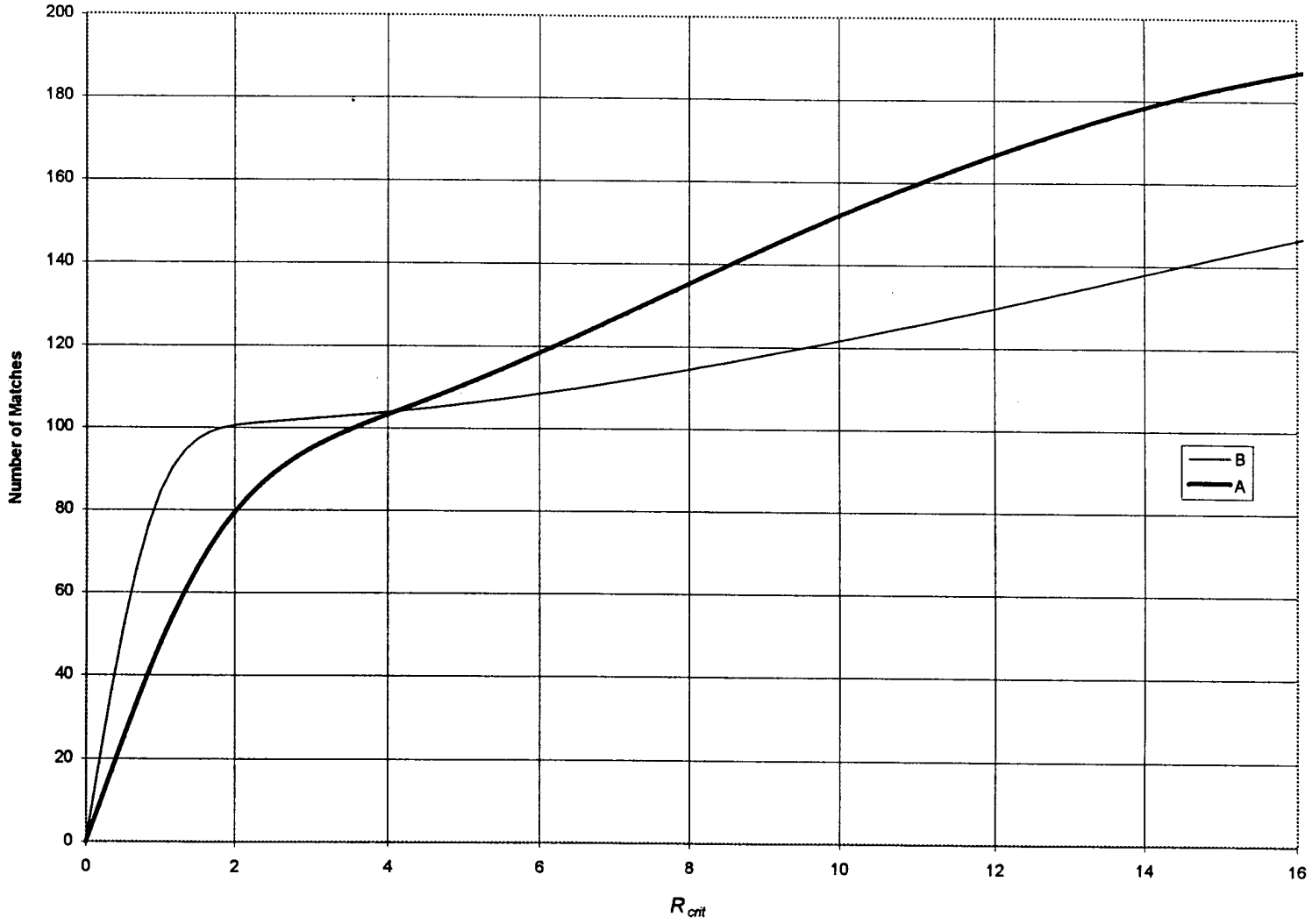


Figure 4.2. Number of Matches vs. R_{crit} for Two Notional Demonstrators. Demonstrator B has fewer false alarms and better horizontal location accuracy than Demonstrator A.

demonstrator B would have 170 false alarms of radius 1 m. Despite the equal number of matches, demonstrator B would be a superior choice for clean-up activities.

- At the point where the curves cross, both demonstrators would uncover 103 emplaced items. Demonstrator A would have 497 false alarms and demonstrator B would have 147 false alarms. The holes dug for each demonstrator would be 4 m in radius. Again, despite equal matches, demonstrator B's performance is superior.
- At the point where both demonstrators are uncovering 120 emplaced items, demonstrator A would have 480 false alarms of radius 6 m, and demonstrator B would have 130 false alarms of radius 9 m. Here the choice is less obvious. Although demonstrator A has more false alarms, the holes dug are of much smaller radius. In this regime, however, all additional "matches" are the result of uncovering undetected items at random in greater volumes of earth excavated. The hole size for both demonstrators is unreasonably large for use in clearance activities. Clearly, one does not want to be in this regime of detection capability versus R_{crit} in rating performance or in making decisions about digging holes.

Beyond the initial region where detection capability depends on location accuracy, we can estimate a detection capability that is nearly independent of the critical radius. To accomplish this, we model the increase in P_{match} with R_{crit} and apply a correction to P_{match} that accounts for lucky matches. To account for the effect of lucky matches on the detection capability, we must reduce the P_{match} calculated using the TMA by the probability of finding an emplaced item at random. In other words, we estimate the number of matches that arise from false alarms that happen to lie within R_{crit} of an undetected item. To determine the probability of finding an emplaced item at random, we multiply the fraction of the site area covered by circles of radius R_{crit} around each demonstrator declaration by the total number of emplaced items. This serves as an estimate of the number of matches that would be found by randomly placing an equal number of declarations on the site. Appendix B describes in more detail the method used to estimate the effect of lucky matches. This effect is important for only two demonstrators of ground-based systems, GDE and Vallon. For the remaining demonstrators, correction for random matches does not change significantly detection capability.

With respect to the effect of both location inaccuracy and lucky matches, the corrected P_{match} is relatively independent of R_{crit} for most demonstrators in the region $1\text{ m} < R_{crit} < 3\text{ m}$. Thus, although we have not eliminated the need for an R_{crit} selection, we have desensitized the analysis to this selection, and arrived at a value of R_{crit}

large compared to any expected sensor errors and common to all demonstrators despite the wide variations in false alarm rates. Unless otherwise noted, a standard R_{crit} value of 2 m was used throughout the analyses for all of the 40-acre site demonstrators.[3] We have confidence that this value of the critical radius captures all the non-accidental detections. For demonstrators with very high false alarm rates, the corrected P_{match} gives a more accurate indication of performance.

C. THE EFFECT OF MATCHING ON OTHER MEASURES OF PERFORMANCE

If remediation activities are to be based on detection of ordnance items with these remote sensors, it is important as a practical matter to have an accurate assessment of the sensor's ability to pinpoint the target both in x-y location and in depth. As discussed in Section A above, the location and depth accuracies calculated using the TMA are contaminated by the tie-breaking scheme. Figure 4.3 shows the location and depth accuracies based on an alternative matching scheme that matches the *closest* declaration to the emplaced item, i.e., the nearest x-y location match. The figure shows that five of the six[4] demonstrators with significant detection capability (marked with **) exhibited location accuracy of better than 3 ft (1 m); three of these five were more in the range of 2 ft (0.7 m). Of the six demonstrators, five showed depth accuracy of better than 2 ft (0.7 m), and one showed depth accuracy of about 3 ft (1 m).

D. THE EFFECT OF CLOSELY SPACED ITEMS IN THE BASELINE

In addition to the above concerns, the distribution of baseline items at JPG affects the ability to use the TMA with confidence. Figures 4.4 and 4.5 show the nearest neighbor distribution for the 40- and 80-acre sites. For both, a large fraction of the ordnance items are separated from their nearest neighbors by only a few feet or less. This has implications for the matching of demonstrator declarations with emplaced items.

The TMA relies on one-to-one matches of demonstrator declarations with baseline items. Such a matching scheme assumes sensor resolution that is sufficient to distinguish

---

3 R_{crit} = 2 m also represents a desired upper limit for excavation purposes.

4 Although P_{match} prior to correction for random matches indicates that GDE showed substantial detection capability, this demonstrator had a false alarm rate that was sufficiently high that nearly all of the detections can be attributed to random matches of false alarms and undetected baseline items. Therefore, this demonstrator is not included for this analysis.

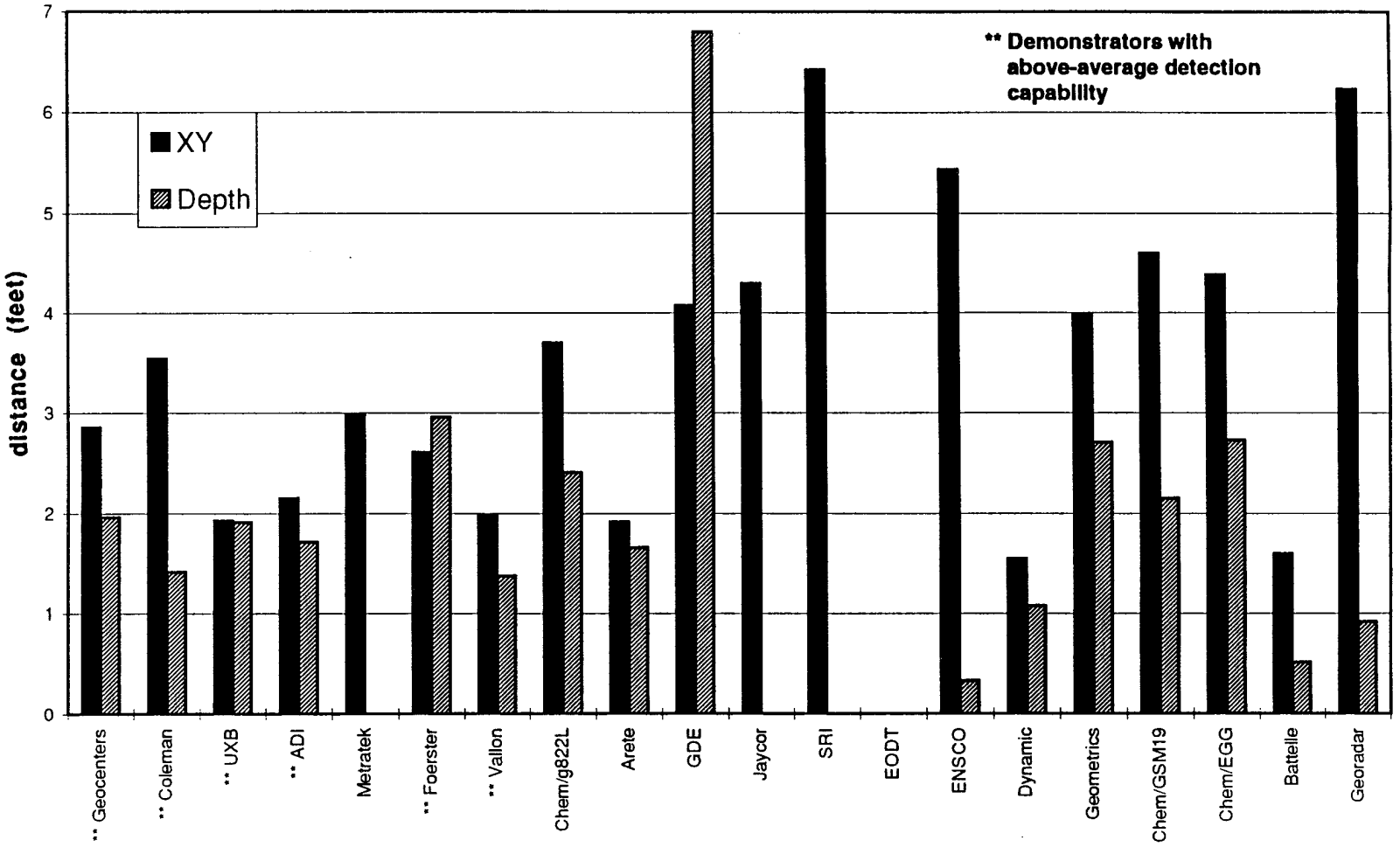


Figure 4.3. Location and Depth Accuracies From Pairing Emplaced Items With Nearest Demonstrator Declaration

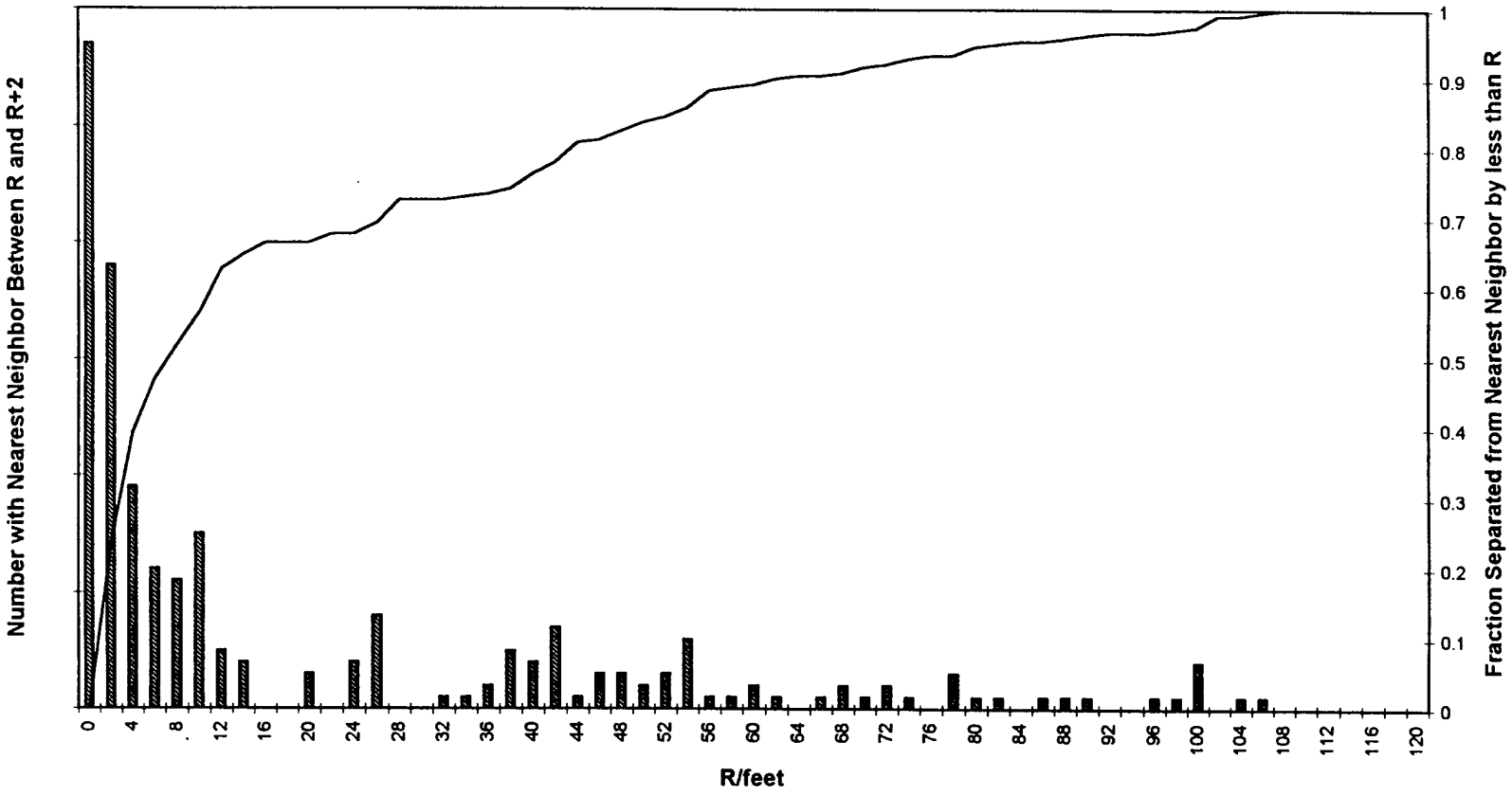


Figure 4.4. Baseline Nearest Neighbor Distribution for the 40-Acre Site. Bars show the number of baseline items with nearest neighbors bertween R and R + 2 feet. The solid line shows the fraction of baseline items separated from their nearest neighbor by less than R feet.

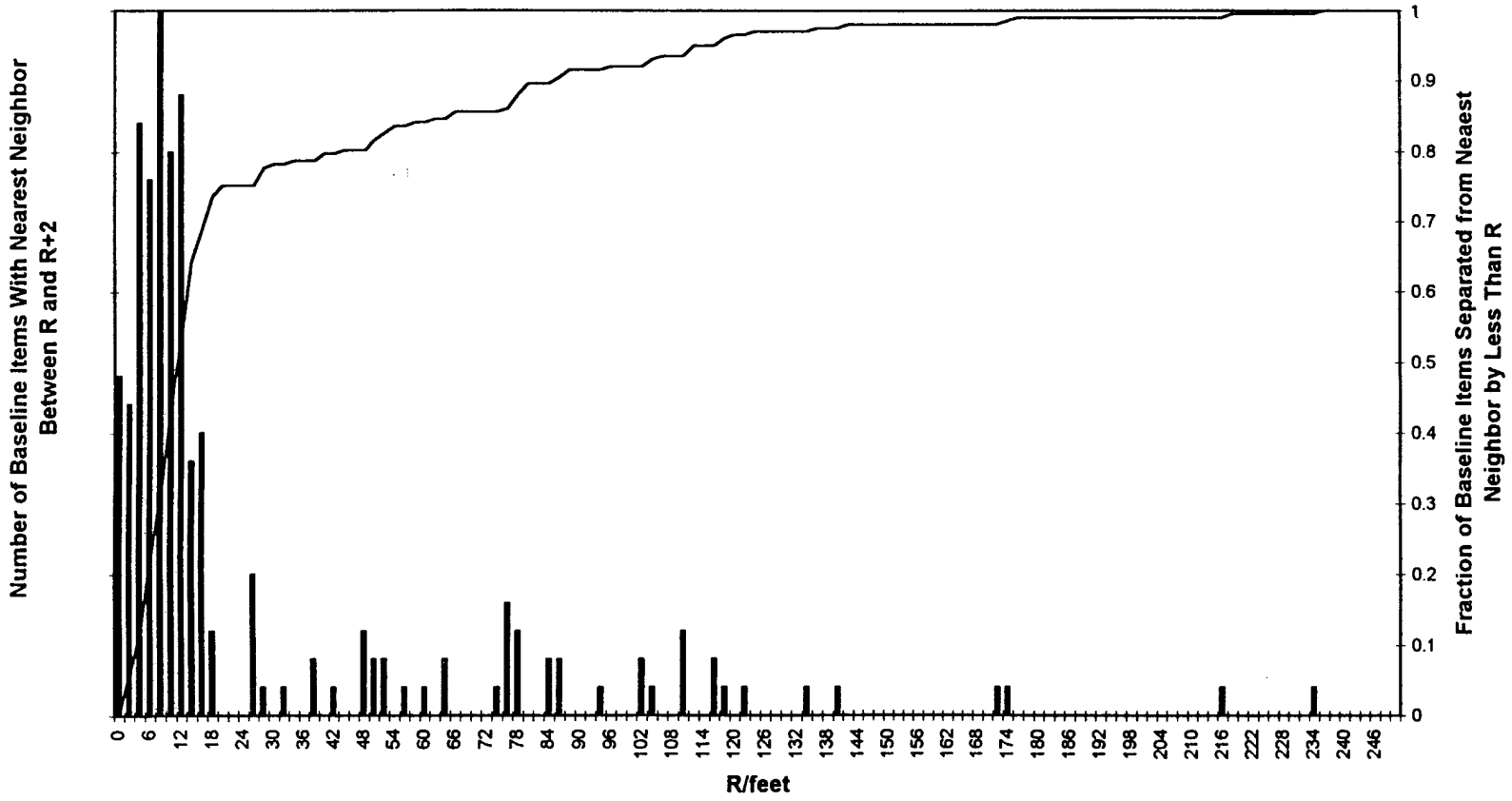


Figure 4.5. Baseline Nearest Neighbor Distribution for the 80-Acre Site. Bars show the number of baseline items with nearest neighbors bertween R and R + 2 feet. The solid line shows the fraction of baseline items separated from their nearest neighbor by less than R feet.

closely spaced targets present in the baseline target set. If sensors are not able to resolve closely spaced targets, the implications for assessing demonstrator performance may be significant, since a large portion of the baseline targets are separated from their nearest neighbors by distances that are small compared to an R_{crit} equal to 2 m. The relatively large number of ordnance items that were emplaced close to other ordnance items creates ambiguity in establishing matches. For example, if a demonstrator searched a region containing several closely spaced ordnance items, various outcomes might result. If the demonstrator has insufficient resolution, only one declaration would be reported for a group of unresolved targets; in reality, the demonstrator did not miss several of the individual items in the group but detected only the aggregate group. If the demonstrator has sufficiently high resolution to establish the presence of multiple ordnance items, several distinct declarations, one for each item detected, might be reported. Alternatively, the same demonstrator may simply mark the area with a single declaration indicating a region of concern. Insufficient information is available about the methods used by the demonstrators to assign declarations to determine *a priori* which is a more accurate reflection of demonstrator capabilities. In fact, different demonstrators will likely have different resolution abilities, and any one demonstrator may be able to resolve some groups of baseline targets but unable to resolve other groups. Therefore, in this section we describe various methods of assessing probability of detection. In this way, we provide worst and best case interpretations of the data regarding demonstrator probability of detection.

Figure 4.6 illustrates three ways of scoring one demonstrator declaration within R_{crit} of three closely spaced baseline targets. The one-to-one matching method of the TMA, P_{match}, would conclude that the demonstrator located the one target that was the "best" match to the declaration, and that the other two targets were undetected. The second possibility, P_{near}, counts one detection for each target within R_{crit} of the declaration

$$P_{near} = \frac{|N_n|}{|B|} = \frac{\text{baseline items within } R_{crit} \text{ of a declaration}}{\text{total baseline items emplaced}}$$

and would conclude that the demonstrator located three of three possible emplaced items. The third possibility, P_{group}, would credit the demonstrator with detecting one *group* of emplaced items, where a group is defined as all items within R_{crit} of a demonstrator declaration, and not charge it for any missed detections. Likewise, the undetected emplaced items are divided into groups and the probability of detection is calculated as the number of groups matched to demonstrator declarations divided by the total number

of groups.[5] This approach assumes that the demonstrator has insufficient resolution to identify the individual items in the group; therefore, if a demonstrator has multiple declarations near the same group, only one of the declarations is counted toward the detections. Figure 4.7 shows the impact of these various matching algorithms on the calculation of probability of detection.

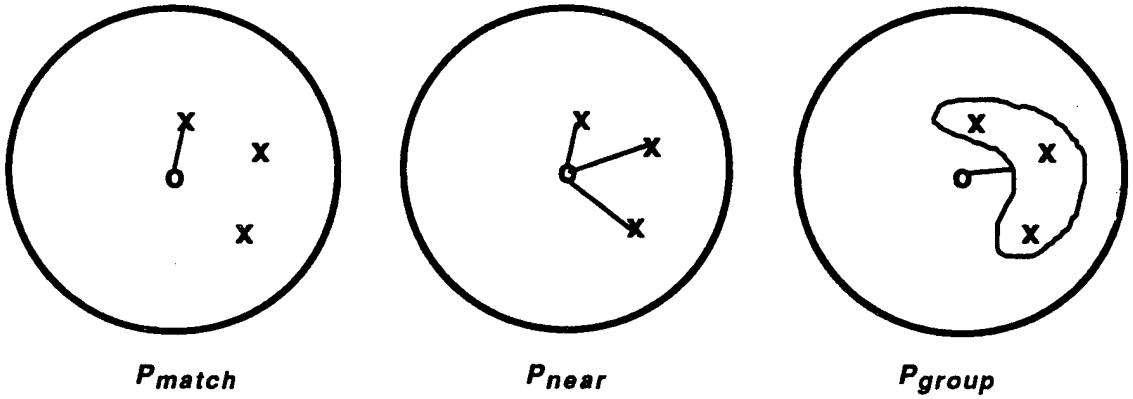


Figure 4.6. Three Methods of Scoring Three Emplaced Items Within R_{crit} of One Demonstrator Declaration

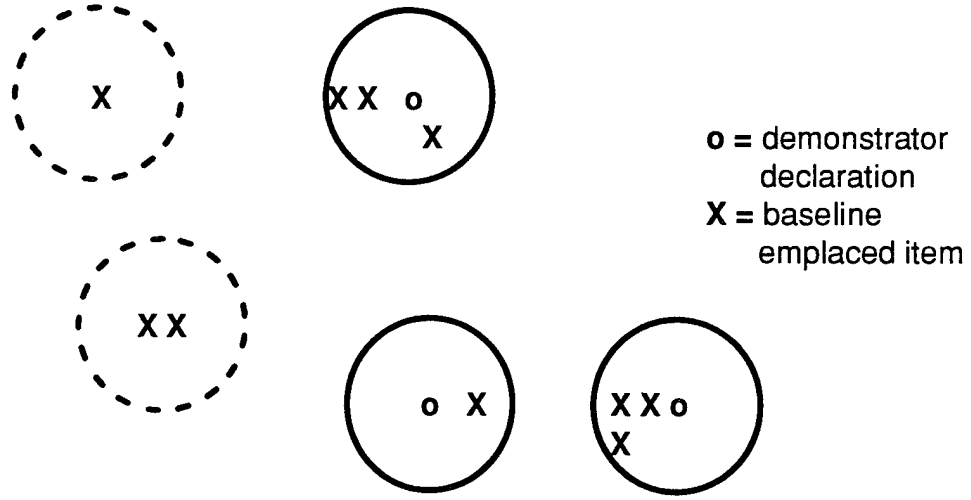


Figure 4.7. Calculation of P_{match}, P_{near}, and P_{group} in a Hypothetical Situation

---

5 This method of calculating P_{group} differs from the method used for the P_{group} calculations reported in Reference 1 in that the undetected baseline items are also grouped. Therefore, some of the numbers are slightly different.

Interpreting the algorithms. Each of the calculations of the probability of detection described above is imperfect; none of the measures of detection capability directly corresponds to the true ability of a demonstrator to find ordnance on an actual site. Nevertheless, these various measures can be used to put upper and lower bounds on performance.

The P_{match} scheme will likely underestimate the detection capability of the sensor, because it will determine that the sensor is unable to detect targets in situations where it was simply unable to resolve them from nearby targets. The P_{near} method overstates detection capabilities because it measures the fraction of baseline items that would be recovered if a hole of radius R_{crit} were dug at each demonstrator declaration. As such, P_{near} allows multiple detections to be credited from a single demonstrator declaration. This measure of detection capability is strongly a function of the distribution of emplaced test articles. If the test articles are emplaced in groups with spacing that is close relative to R_{crit}, then many emplaced items would be recovered serendipitously in holes of radius R_{crit}. Thus, P_{near} cannot be used reliably to predict performance at other sites where the ordnance distribution is not known. P_{group} gives the most accurate representation of detection capability. However, it is difficult to rigorously compare different demonstrators, who will have different resolution capabilities, without knowing the details of the instrument resolution and the processing algorithms. Such information is required to make judgments about whether the demonstrator declared one target where there are multiple items because the return in fact resulted from only one of the targets, or because of inadequate sensor resolution. Further, it is likely that any single demonstrator will have declarations that fall into both categories. It should be noted that this approximation may also overstate demonstrator detection capabilities, since it relies on the assumption that a single demonstrator declaration in the vicinity of a group of baseline items results from lack of resolution rather than lack of detection.

Figure 4.8 shows each of the measures of detection capability for each demonstrator.

E. THE IMPACT OF MINES ON MEASURED PERFORMANCE

Among the ordnance items emplaced at JPG were a number of plastic antipersonnel mines. Magnetometers would certainly not be among the detectors considered for locating plastic mines. In this sense, it is unfair to grade magnetometers on their ability to locate

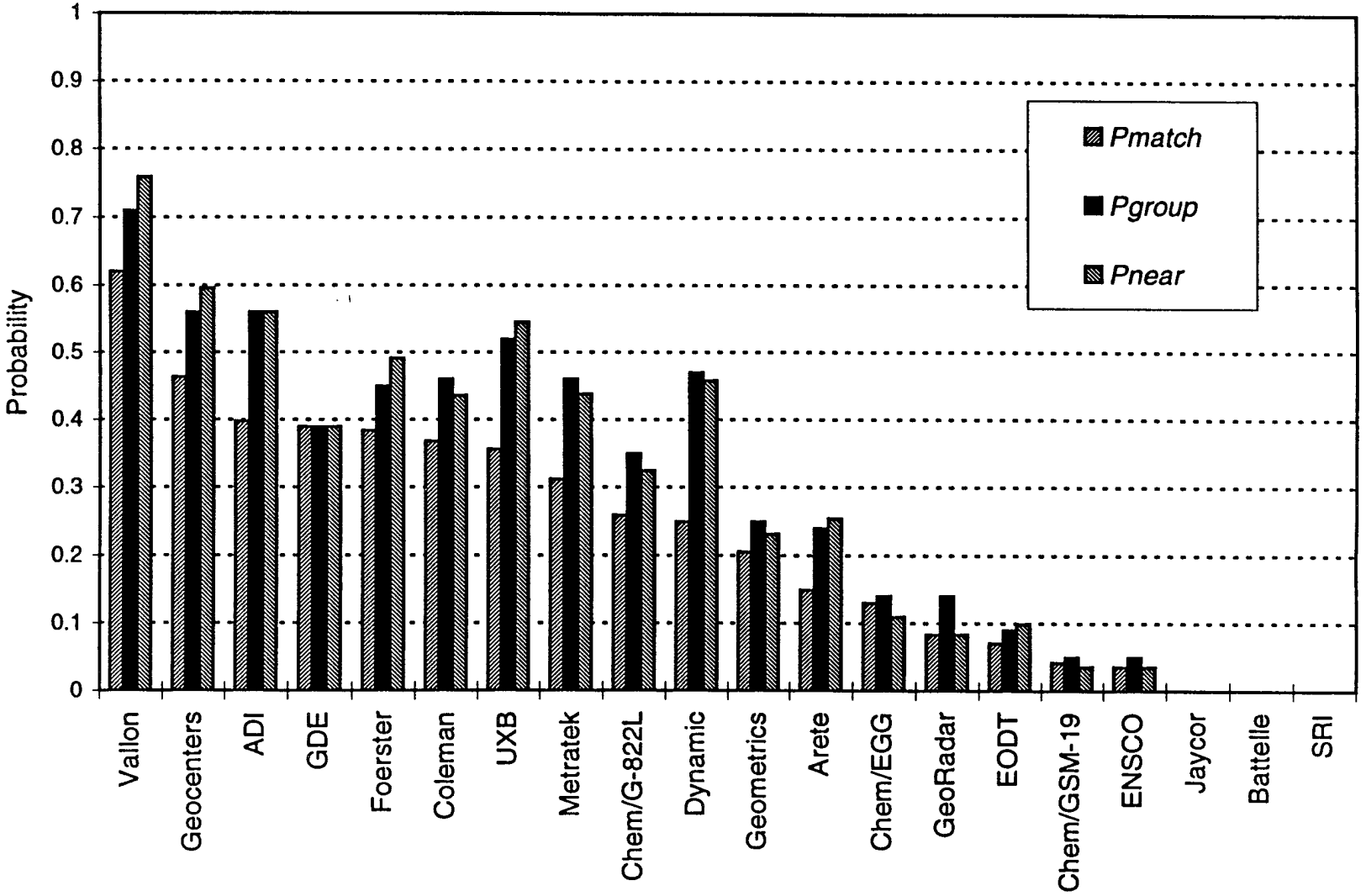


Figure 4.8. Comparison of Demonstrator Detection Capability as Measured by P_{match}, P_{group}, and P_{near} for Ordnance Items Only on the Area of the 40-Acre Site That Was Searched

items that it is physically impossible for them to detect and for which they make no claim of detection capability. Induction coils theoretically could respond to non-ferrous metal in the firing mechanism of the mines. However, this is usually a very small amount of metal, and any sensor response is likely to be masked by clutter returns caused by numerous natural and man-made objects that would produce a similarly low level of sensor response. Of the sensors demonstrated at JPG, ground penetrating radar is the most likely to respond to the plastic mines. However, at JPG either the wavelengths were too long compared to the physical size of the mines or too short to propagate even a few inches without attenuation. (This is discussed in detail in Chapter 6.)

Considering the physical limitations of the sensor responses to the emplaced mines, it is not surprising that no system demonstrated at JPG showed any ability to detect these targets. The calculation of detection capability for ordnance items not including the plastic mines may give a more accurate impression of a sensor's ability to detect UXO characteristic of what would be found on a test or training range. However, if what is desired is to clear sites that were once minefields or ranges on which ordnance that disperses plastic mines has been tested, it must be realized that the sensors used at JPG have demonstrated no ability to detect such targets. Figure 4.9 compares detection capability, as measured by P_{group} on the area searched, corrected for lucky matches, with and without mines included.

F. COMBINED MEASURE OF DETECTION CAPABILITY AND FALSE ALARMS

The two principal measures of demonstrator performance at JPG are detection capability and false alarm rate. However, it is difficult to compare demonstrators with widely different detection capabilities and false alarms rates. For example, it is difficult to compare the performance of a demonstrator with high detection capability and a high false alarm rate to a demonstrator with a low detection capability and a low false alarm rate. Differences in the performance may be directly attributable to different threshold settings or might be a result of real differences in system capabilities. Thus, any discussion of detection capability without false alarms is inherently flawed.

It is well recognized that measures of detection capability and false alarms are inextricably linked. Standard approaches have been developed to quantify the relationship between the probability of detection, P_d, and the probability of false alarm, P_{fa}. (Ref. 7). A P_d versus P_{fa} curve can be drawn by keeping the signal-to-noise ratio constant and

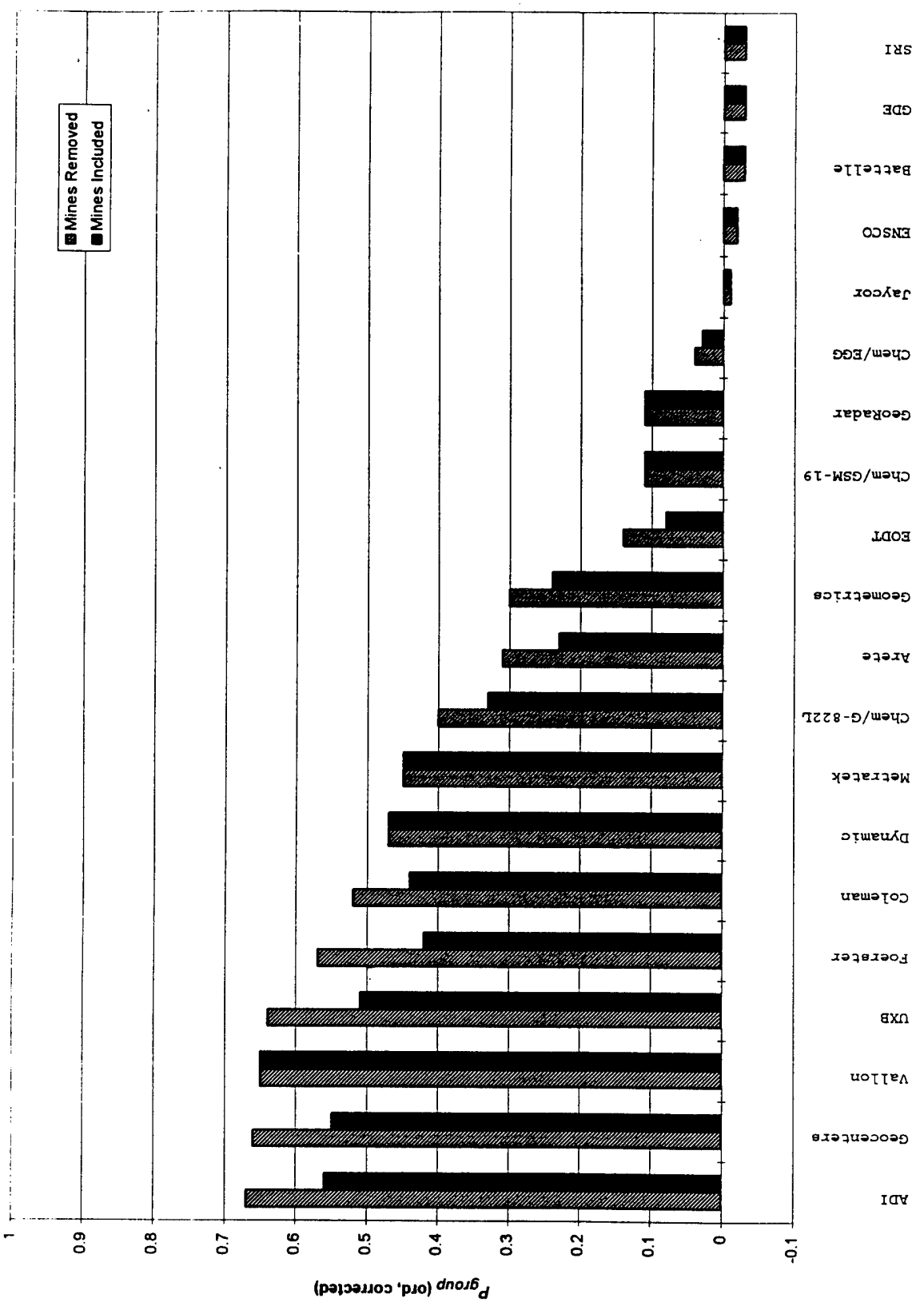


Figure 4.9. P_{group}(ord) vs. False Alarm Rate With and Without Plastic Mines in the Baseline

varying the threshold for a detection. These curves are known as receiver operating characteristic (ROC) curves. Two sensors that are described by the same ROC curve are equally capable of distinguishing signal from noise, and any differences in performance are attributable to differences in threshold setting.

The ROC curve is governed by a variable *d*, illustrated in Figure 4.10, which measures the distance between the centroids of the target population and the background population in units of the standard deviation for a gaussian noise distribution.[6] This *d* is a statistical parameter that indicates how well a particular measurement distinguishes signal from noise. The dashed lines in Fig. 4.10 indicate various possible threshold settings. Higher *d* values indicate better ability to separate signal from noise. The curves in Figure 4.11 show *d* values of 0, 1, 2, and infinity, where 0 represents no ability to separate signal from noise and infinity represents perfect separation. It is possible to compute a *d* value from the demonstrator's P_d and P_{fa} values at a single point. This measure allows for comparison of demonstrators with widely different detection capabilities and false alarm rates.

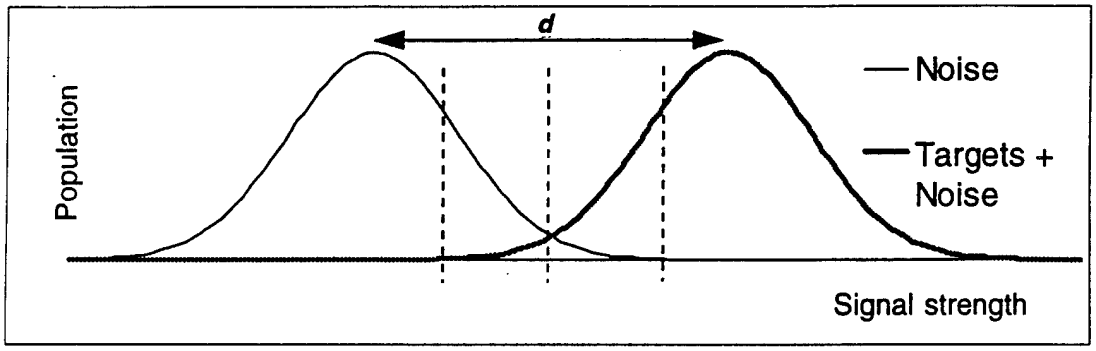


Figure 4.10. Illustration of the Target Population and Background Population That Define the ROC Curve

There are some shortcomings to this approach. To accurately characterize the sensor, the entire ROC curve is desired. The mathematical model used here is imperfect even in the case of noise-limited radar detection, for which a gaussian distribution is a reasonable approximation of system noise. The clutter statistics, which currently dominate the false alarms for UXO detection, are not random and certainly not gaussian, and further are unlikely to be the same for the different sensors. Nevertheless, the approach provides a

---

6 For simplicity we assume that the baseline item and false alarm populations have the same standard deviation. If the standard deviations are different, an ROC curve of a somewhat different shape results.

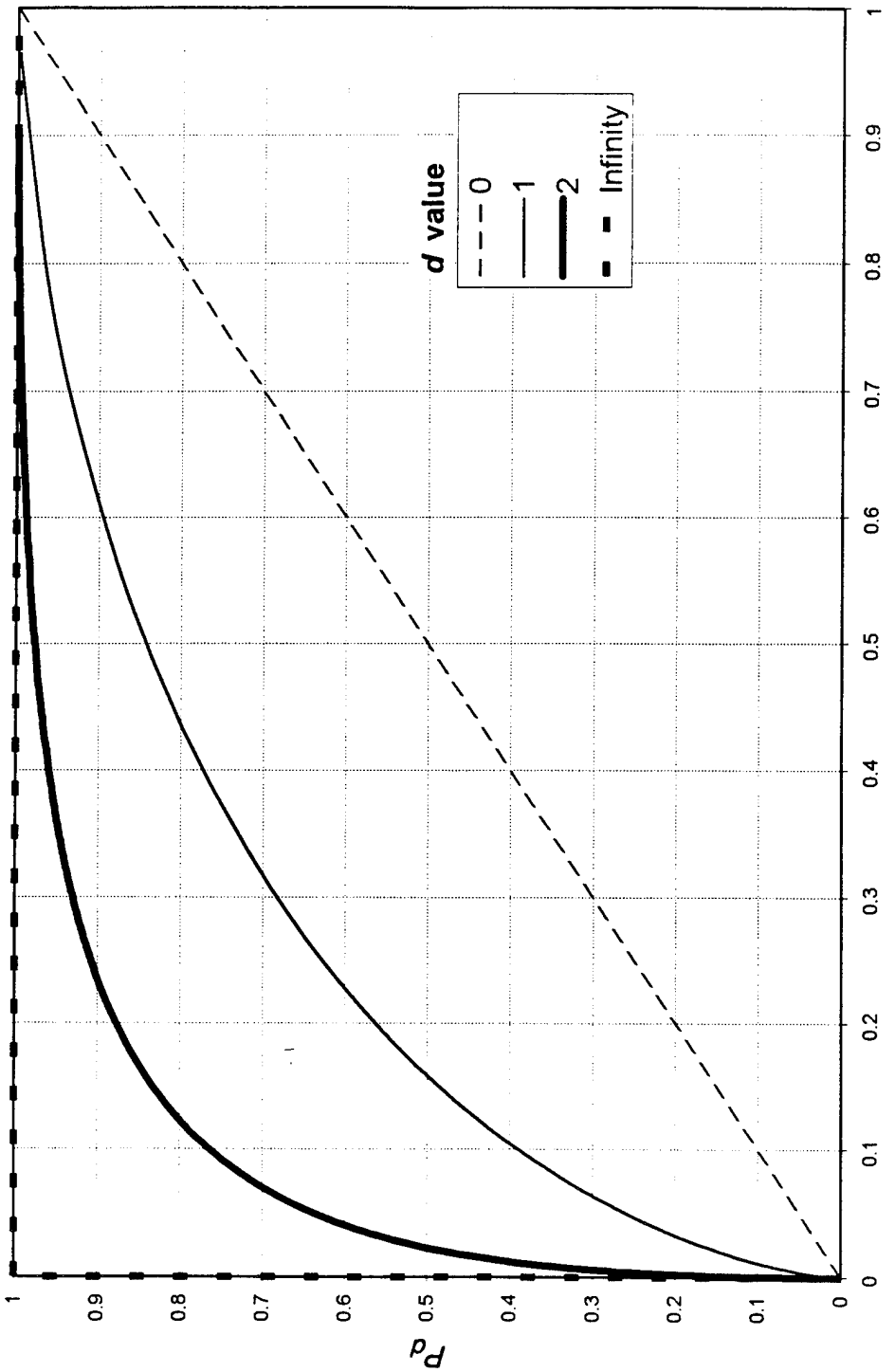


Figure 4.11. Sample Receiver Operating Characteristic Curves for d = 0,1,2, and Infinity

well-defined measure, in use throughout the sensor community, that allows comparison among demonstrators. Furthermore, for the JPG demonstration, the d values are small, so deviation from gaussian statistics will be less important than for large values of d. Large values of d will be highly sensitive to the tails of the distributions, where differences from gaussian statistics will be greatest. Small values of d, indicating that the populations are not well separated, are determined from the more highly populated areas of the distributions and are less sensitive to the details of the distribution model.

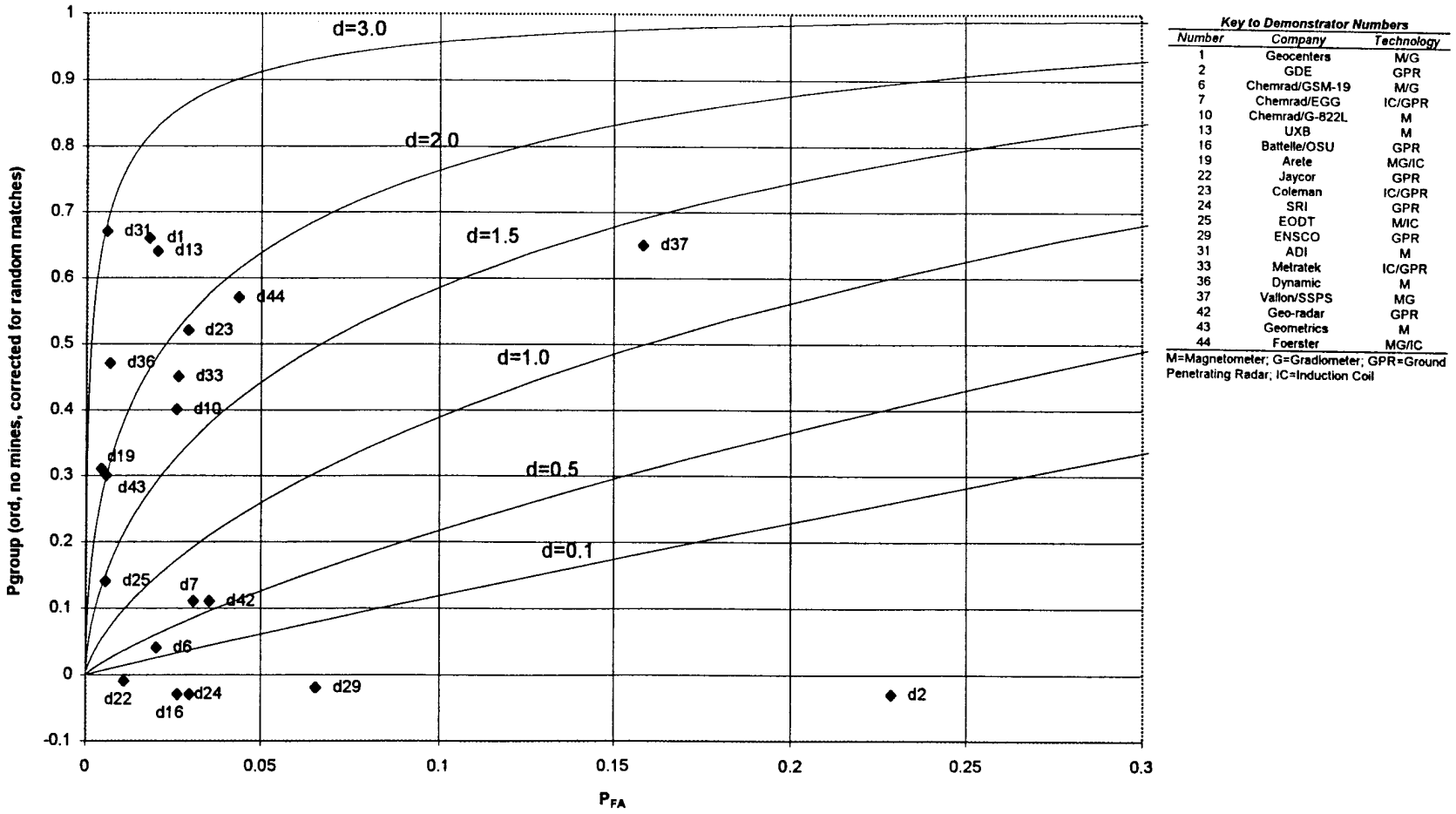
For the JPG demonstration, each demonstrator is assigned a P_d and P_{fa} as follows. The probability of detection is estimated as P_{group} on area searched only. We consider only ordnance and have removed mines from the baseline. The probability of false alarm, P_{fa}, is estimated as the fraction of the site surveyed covered by false alarms when a circle of radius R_{crit} is centered on each false alarm declaration. Figure 4.12 shows P_d versus P_{fa} for each demonstrator at the 40-acre site. Various ROC curves are overlaid on this plot. This figure shows the value of the ROC curve parameter for comparing the performance of demonstrators. Demonstrators Coleman and Geometrics have significantly different P_d and P_{fa} values, but are seen to represent equivalent capability to separate signal from noise based on a gaussian noise model.

G. BEST ESTIMATE OF PERFORMANCE

In this chapter we have shown the impact of the matching algorithm, a high false alarm rate, and baseline emplaced item type and distribution on the measures of performance at JPG. We conclude that the best estimate of ordnance detection capability for demonstrations on the 40-acres site is represented by P_{group} on the area searched by each individual demonstrator with mines eliminated from the baseline emplaced ordnance set. We have also corrected these detection capabilities to account for the effect of lucky matches.

False alarms include all demonstrator declarations that are not within R_{crit} of an emplaced ordnance or nonordnance group. The false alarm rate is calculated as the number of false alarms per unit area, where the specific unit of area is not specified. The probability of false alarm is calculated as the fraction of the site covered by circles of radius R_{crit} around each demonstrator false alarm.

To compare demonstrators with widely different detection capabilities and false alarm rates, we use the ROC curve parameter, d, described above in Section F.



Key to Demonstrator Numbers

Number	*Company*	*Technology*
1	Geocenters	M/G
2	GDE	GPR
6	Chemrad/GSM-19	M/G
7	Chemrad/EGG	IC/GPR
10	Chemrad/G-822L	M
13	UXB	M
16	Battelle/OSU	GPR
19	Arete	MG/IC
22	Jaycor	GPR
23	Coleman	IC/GPR
24	SRI	GPR
25	EODT	M/IC
29	ENSCO	GPR
31	ADI	M
33	Metratek	IC/GPR
36	Dynamic	M
37	Vallon/SSPS	MG
42	Geo-radar	GPR
43	Geometrics	M
44	Foerster	MG/IC

M=Magnetometer; G=Gradiometer; GPR=Ground Penetrating Radar; IC=Induction Coil

Figure 4.12. P_{group}(ord) on Area Searched vs. $P_{false\ alarm}$

5. RESULTS AND COMPARISONS AMONG DEMONSTRATORS

Prior to the first demonstrations, it was believed in some quarters that detection capabilities would be high and false alarms would be sparse. Under such conditions, demonstrators with similar detection capabilities would have been distinguished by their ability to identify ordnance from nonordnance, ability to determine ordnance type, location accuracy, survey rate, and cost. As a practical matter, the demonstrators did not exhibit capabilities that required this level of fidelity for separation, nor were the discrimination capabilities sufficient to make these tests meaningful.

As described in the preceding chapter, the specifics of the matching algorithm, a high false alarm rate, and baseline emplaced item type and distribution all influence the calculation of the measures of performance at JPG. Among the possible alternatives, we chose to estimate detection capability as the ability of demonstrators to detect groups of emplaced ordnance items on the area searched by each demonstrator with the number of detections corrected for lucky matches of false alarms to undetected baseline items.[1] We have removed plastic antipersonnel mines from consideration. Further, a comparison among demonstrators with regard to detection capability alone is inherently flawed; a meaningful comparison requires including the false alarm rate as well. Thus, when comparing demonstrators, both the detection capability and the false alarm rate are used. To compare demonstrators with widely different detection capabilities and false alarm rates, we use the receiver operator characteristic curve parameter, d, which is described in Chapter 4. A high value of d is preferable.

This chapter provides a concise description of our best estimate of the performance of demonstrators at both the 40- and 80-acre sites.

---

1 In correcting for random matches, we subtract the number of matches expected by randomly placing demonstrator false alarms from the total number of matches of baseline items to demonstrator declarations.

A. DEMONSTRATORS OF GROUND-BASED SYSTEMS (40-ACRE SITE)

1. Probability of Detection and False Alarm Rate

The bar chart shown in Figure 5.1 compares detection capabilities and false alarm rates for the various demonstrators on the 40-acre site. The top panel shows the probability of detection for groups of ordnance (mines excluded), corrected for random matches, and uses as a baseline items emplaced on the entire 40-acre site. Since most demonstrators indicated the ability and the intention to search the entire site, it is reasonable to compare how much of the total emplaced ordnance each demonstrator detected in the allotted 40 hours. The center panel shows the probability of detection on only the area searched. This measure gives information on the detection capabilities of the demonstrator, without regard to time constraints. The bottom panel shows false alarm rates calculated by dividing the number of false declarations by an unspecified unit of area.[2] Since credit is given for an ordnance detection regardless of the demonstrator's ability to identify the item as ordnance, the false alarms include declarations (classified as ordnance or nonordnance) that are not matched to baseline items. The order of the demonstrators, sorted by the fraction of the ordnance on the whole site detected, is the same for each of the panels. The figure shows that the demonstrators with the highest detection capability were in the 50-70 percent detection range. The comparison among the panels emphasizes the point that one cannot consider detection capability without also considering false alarm rate.

The problem of weighing an increase in detection capability against an increase in the false alarm rate remains. In other words, how may demonstrators with significantly different detection capabilities and false alarms be compared? Figure 5.2 shows the probability of detection and the probability of false alarm values for the 40-acre demonstrators with a family of receiver operator (ROC) curves overlaid. Demonstrators that fall on the same ROC curve demonstrate equivalent capabilities to distinguish the presence of an ordnance item from a false alarm. (See Chapter 4 for details.)

---

2 The area is in unspecified units to prevent demonstrators from using the information to calculate the numbers and types of targets emplaced on the site.

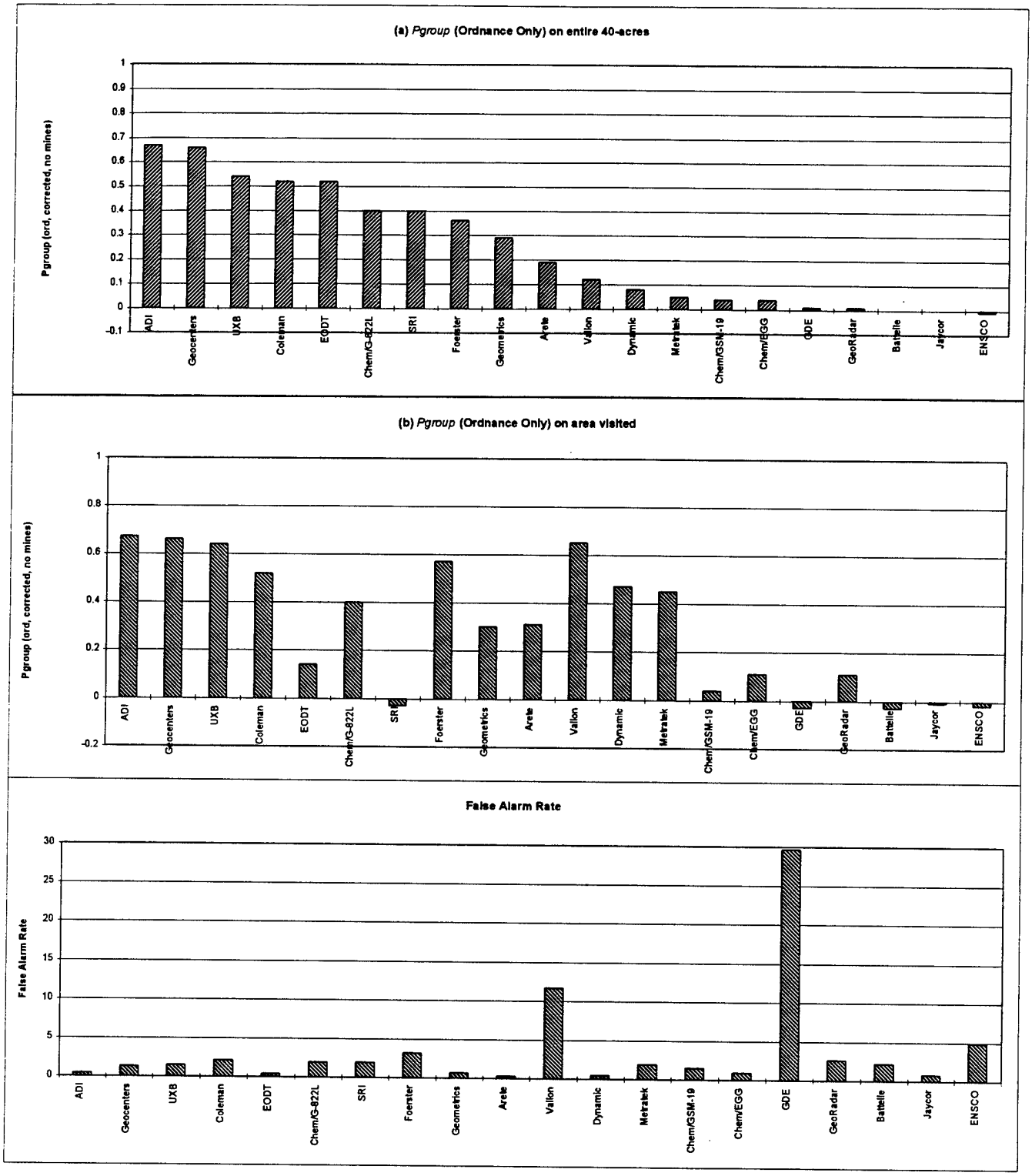
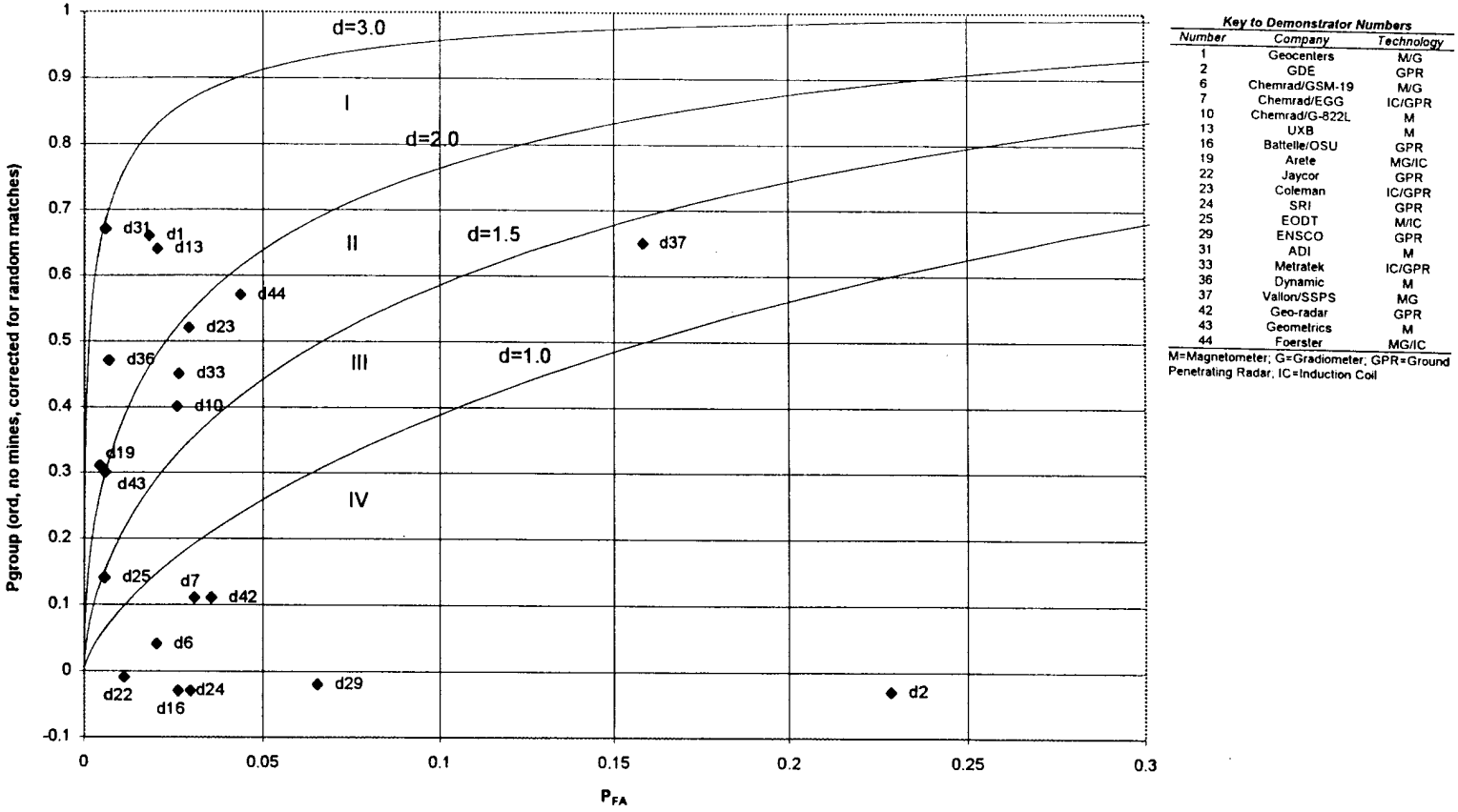


Figure 5.1. Summary of Performance of Ground-Based Detection Systems on the 40-acre Site



Key to Demonstrator Numbers

Number	*Company*	*Technology*
1	Geocenters	M/G
2	GDE	GPR
6	Chemrad/GSM-19	M/G
7	Chemrad/EGG	IC/GPR
10	Chemrad/G-822L	M
13	UXB	M
16	Battelle/OSU	GPR
19	Arete	MG/IC
22	Jaycor	GPR
23	Coleman	IC/GPR
24	SRI	GPR
25	EODT	M/IC
29	ENSCO	GPR
31	ADI	M
33	Metratek	IC/GPR
36	Dynamic	M
37	Vallon/SSPS	MG
42	Geo-radar	GPR
43	Geometrics	M
44	Foerster	MG/IC

M=Magnetometer; G=Gradiometer; GPR=Ground Penetrating Radar, IC=Induction Coil

Figure 5.2. Scatter Plot and ROC Curves for Systems on the 40-Acre Site

A demonstrator with a high probability of detection that falls on a lower-value ROC curve is inferior to a demonstrator with a low probability of detection that falls on a higher-value ROC curve. The rationale for this conclusion is that sensors may be operated at many threshold settings for different applications. Therefore, two demonstrators with seemingly widely different capabilities could be using the exact same sensor with different threshold settings. These two demonstrators, however, would fall on the same ROC curve. Thus, ROC curves allow the detection capability and the false alarm rate to be weighed against each other in a consistent manner. Of note, the best *d* value for a system demonstrated on the 40-acre site at JPG is approximately 3. By way of comparison, a high performance surveillance radar would be designed to have a *d* of 20 for the smallest targets of interest at the longest ranges of interest.

This data allows the following conclusions to be drawn regarding the performance of the technologies demonstrated on the 40-acre site under the conditions at JPG:

- Magnetometer systems, on average, exhibited the best peformance, with probabilities of detection near 70 percent using the best estimate of performance. All demonstrated the magnetometers were capable of detecting some ordnance.
- Stand-alone ground-penetrating radars (GPRs) were unsuccessful at detecting buried unexploded ordnance.
- Induction coil metal detectors paired with GPRs did not perform as well as the best magnetometers; nonetheless, as a group, systems with induction coils displayed moderate detection capabilities. Comparison to the detection capabilities of stand-alone GPRs suggests that the induction coil may be responsible for the detection capabilities displayed by these adjunct systems.
- Most demonstrators reported multiple false alarms per ordnance item detected.
- Demonstrators with extensive field experience performed better than those with little.

2. Detection Capability by Size, Depth, and Ordnance Type

As discussed in Chapter 3, an understanding of the ability to detect various sizes and types of ordnance items at specific depths would help in framing policy decisions regarding UXO cleanup activities. For example, future land use considerations may influence decisions regarding the depth of remediation. Therefore, in this section, we provide information on the performance of systems demonstrated at JPG as a function of these parameters.

Detection capabilities were sorted by the size, type, and depth of emplaced ordnance, as well as by the demonstrator's ability to correctly determine these attributes. The ability of demonstrators on the 40-acre site *to detect items* in the various classifications *without regard to their ability to correctly identify items* is shown in the Figs 5.3–5.5. Although the best estimate of performance uses groups of emplaced items to evaluate detection capability, for these figures we calculate detection capability based on individual items. It is necessary to use a one-to-one matching between a demonstrator declaration and a single emplaced ordnance item with defined attributes of size, depth, and type. It should be noted that size, type, and depth are correlated in this test, with the larger items buried deeper.

The type of technology is indicated in Figs 5.3–5.5. Before any comparisons are made on the basis of technology, it should be stressed that the data sets from which the numbers are calculated are small, so statistical uncertainties in the numbers are significant. With this limit on the validity of the conclusions in mind, the data shows the following trends.

- Bombs are found with the highest reliability with three demonstrators using magnetometers finding greater than 80 percent of the bombs. Further, there is a significant difference in performance between the magnetometers and the induction coils, with the magnetometers detecting bombs at about twice the rate of induction coils.
- For mortars, which are generally nearer to the surface, the detection capabilities between magnetometer and induction coil technologies were indistinguishable.
- Not surprisingly, neither technology was able to detect the plastic anti-personnel mines.
- Magnetometers were more proficient at detecting ordnance items buried below 6 ft, where in general the larger targets were located. On the other hand, the ability of the induction coil systems to detect ordnance falls off considerably in the "below 6 feet" range.
- When detection capabilities are sorted by size, induction coils perform best at locating the medium size targets. The magnetometers, on the other hand, performed best at detecting the medium and large targets but not the small targets.

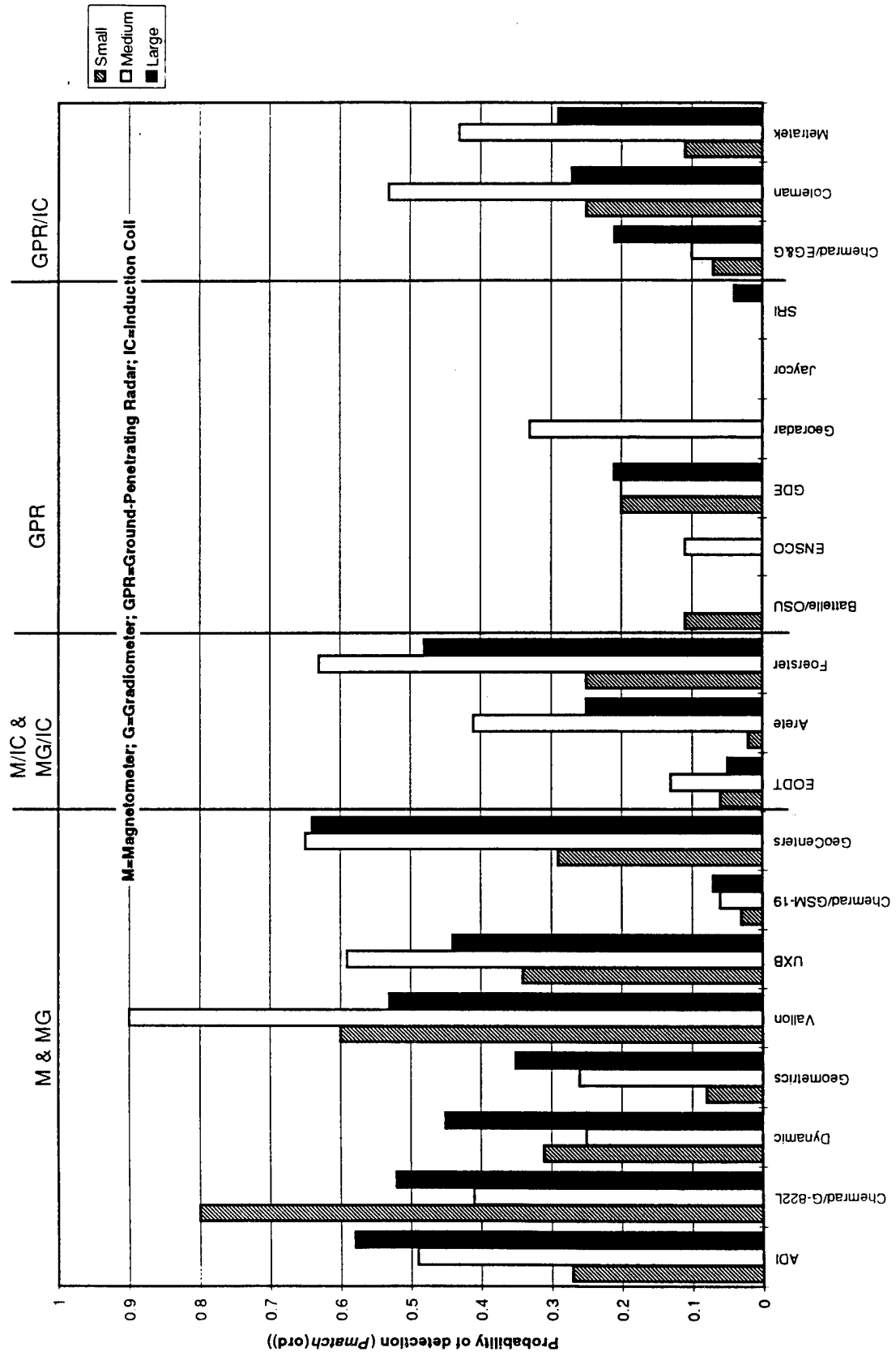


Figure 5.3 Probability of Detection as a Function of Size

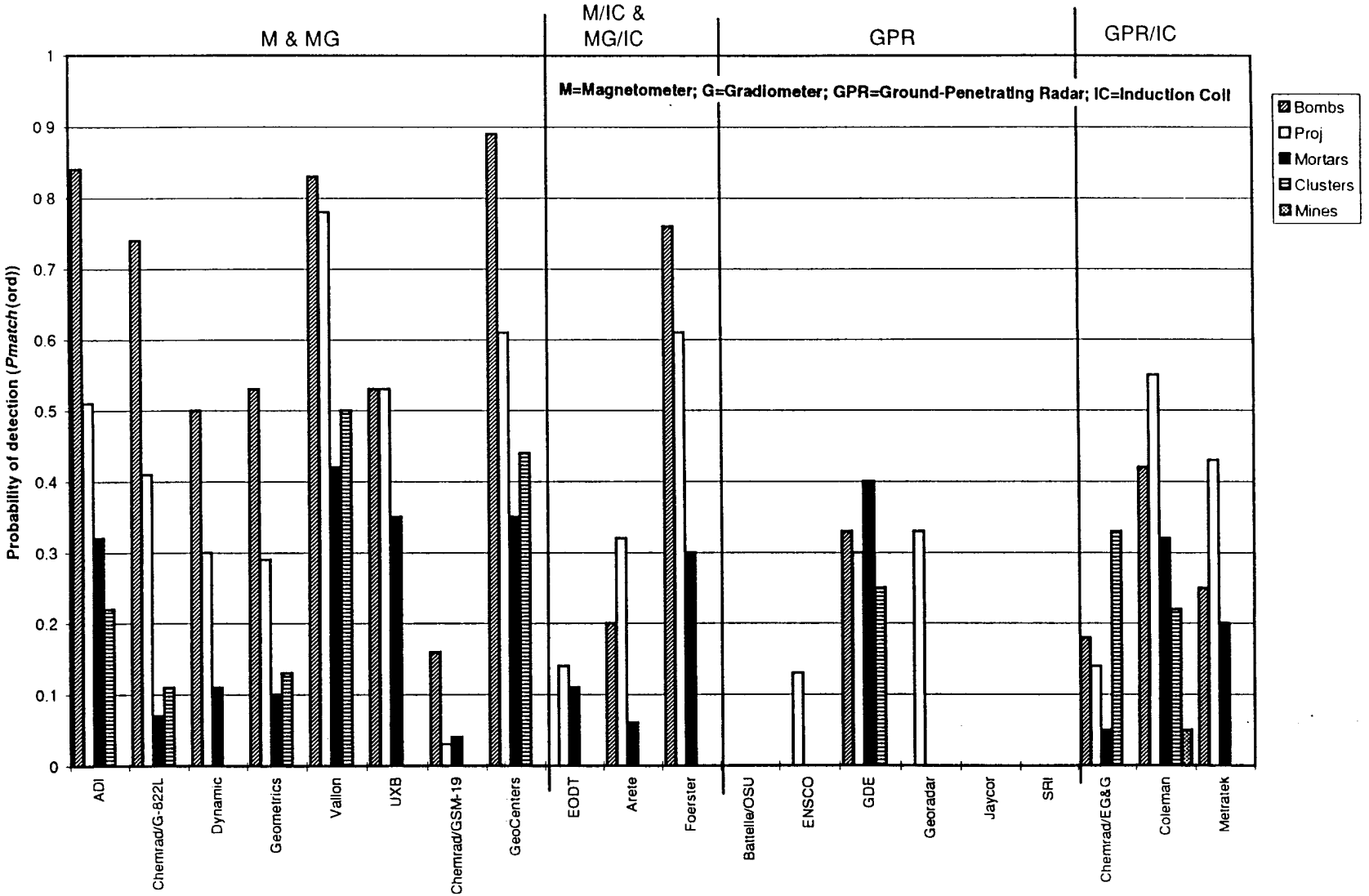


Figure 5.4 Probability of Detection as a Function of Ordnance Type

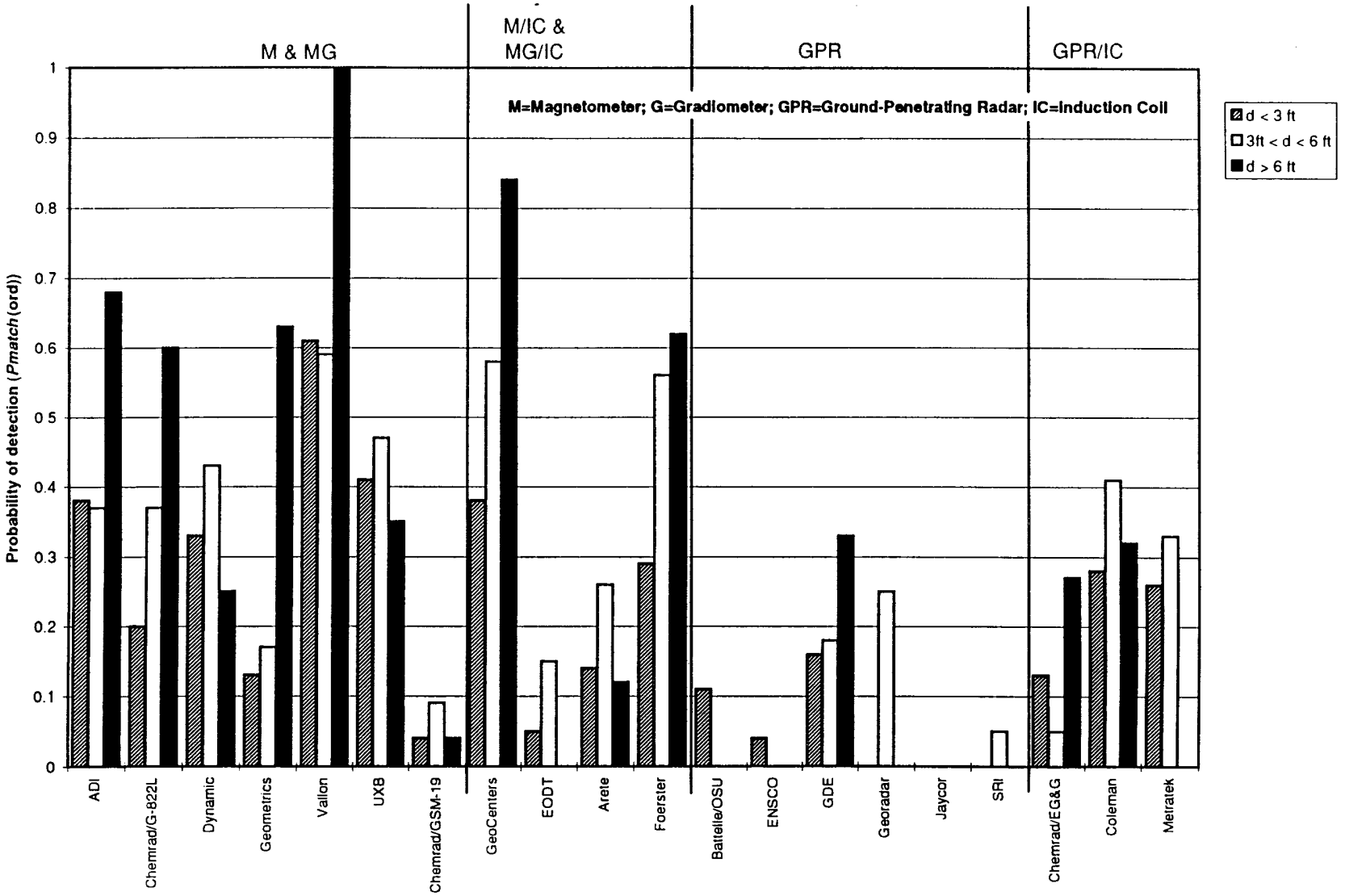


Figure 5.5 Probability of Detection as a Function of Depth

Finally, a comment on the meaning of the "clusters" category for ordnance type is in order. It is not obvious how to interpret the relatively poor performance in this category that is implied by the data. The emplaced item database contained groups called clusters. There were also locations where multiple targets were buried in close proximity to one another that were not deemed clusters; rather, they were entered in the database separately, i.e., one mortar and one bomb separated by a few inches to a few feet. Therefore, one should not conclude that clusters of items in some way inhibit a sensor's detection capability. Although it is possible that this is true, the ambiguity in the data classifications does not permit this conclusion to be drawn from this data.

B. DEMONSTRATORS OF AIRBORNE SYSTEMS

None of the target declarations by any of the demonstrators of airborne systems can be attributed to a return from a single emplaced item. The number of matches, using a 5 m proximity requirement,[3] for the airborne systems is consistent with the number that would be expected from placing the same number of declarations at random. Figure 5.6 shows probability of detection for ordnance items before and after corrections for lucky matches. Note that the demonstrated detection capabilities scatter around zero. We also calculated the location accuracy that would be expected based on random matches. The results of these calculations are shown in Appendix D and further support the contention that no demonstrator declarations were actually caused by individual emplaced ordnance items.

The overall results of the airborne demonstrations on the 80-acre site are summarized in Figure 5.7.

The results from the 80-acre site demonstration at JPG lead to the following conclusions:

- There is no evidence that airborne systems detected any ordnance.
- Integrating sensors on airborne platforms for the purpose of detecting subsurface UXO is likely to be unproductive with our current level of understanding.
- Any sensors proposed for integration should first be tested on targets with calibrated signatures and then on sites with a realistic placement of ordnance under field conditions. These tests of a stand-alone sensor should be a pre-requisite for consideration of integration on an airborne platform.

---

3 See Chapter 4 for a description of the proximity requirement.

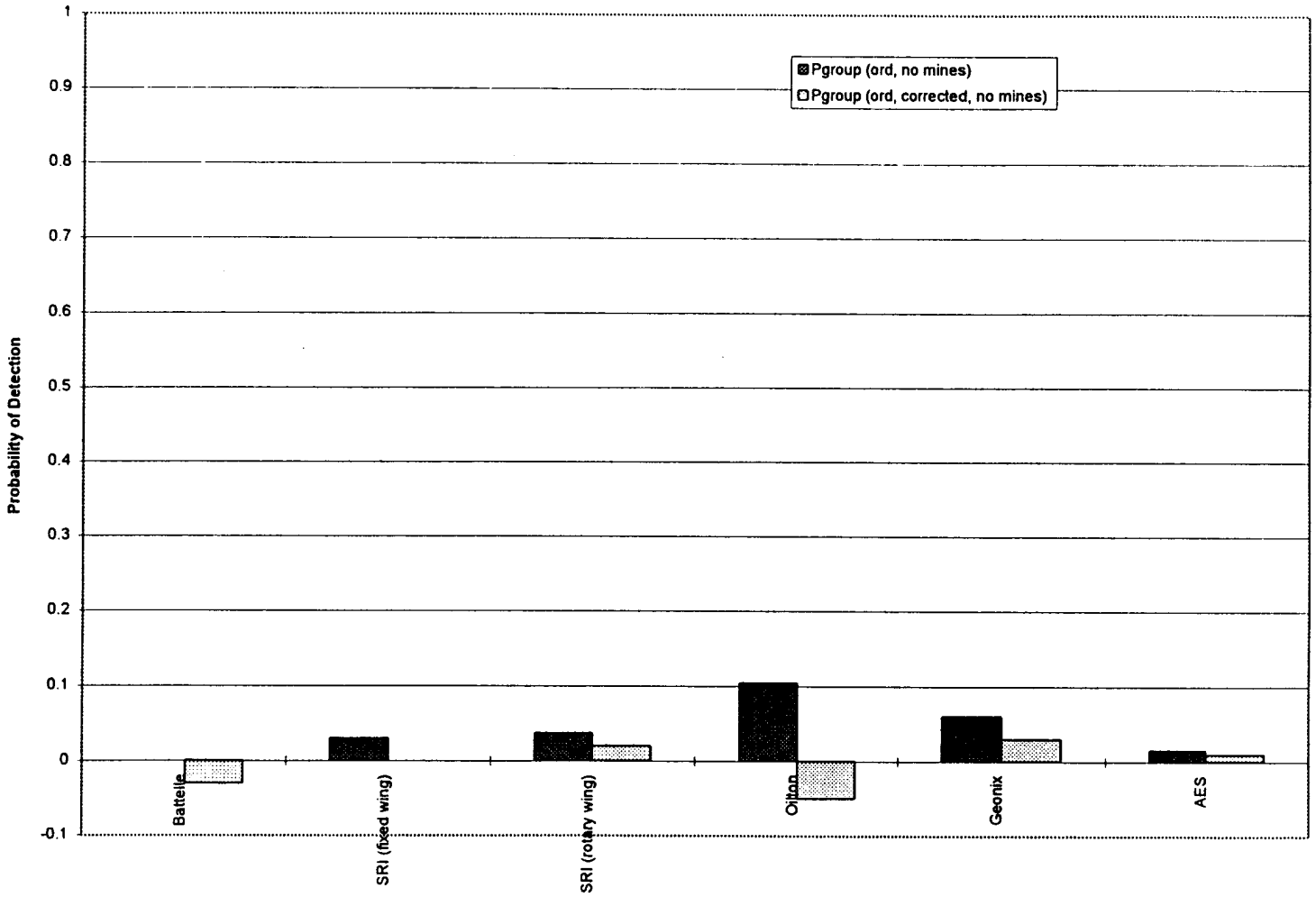


Figure 5.6. P_{match}, Uncorrected and Corrected for Lucky Matches, at R_{crit} = 5 m for Airborne Detection Systems. Negative corrected P_{match}(ord) indicates that more items are expected to be found in an equal number of declarations placed at random.

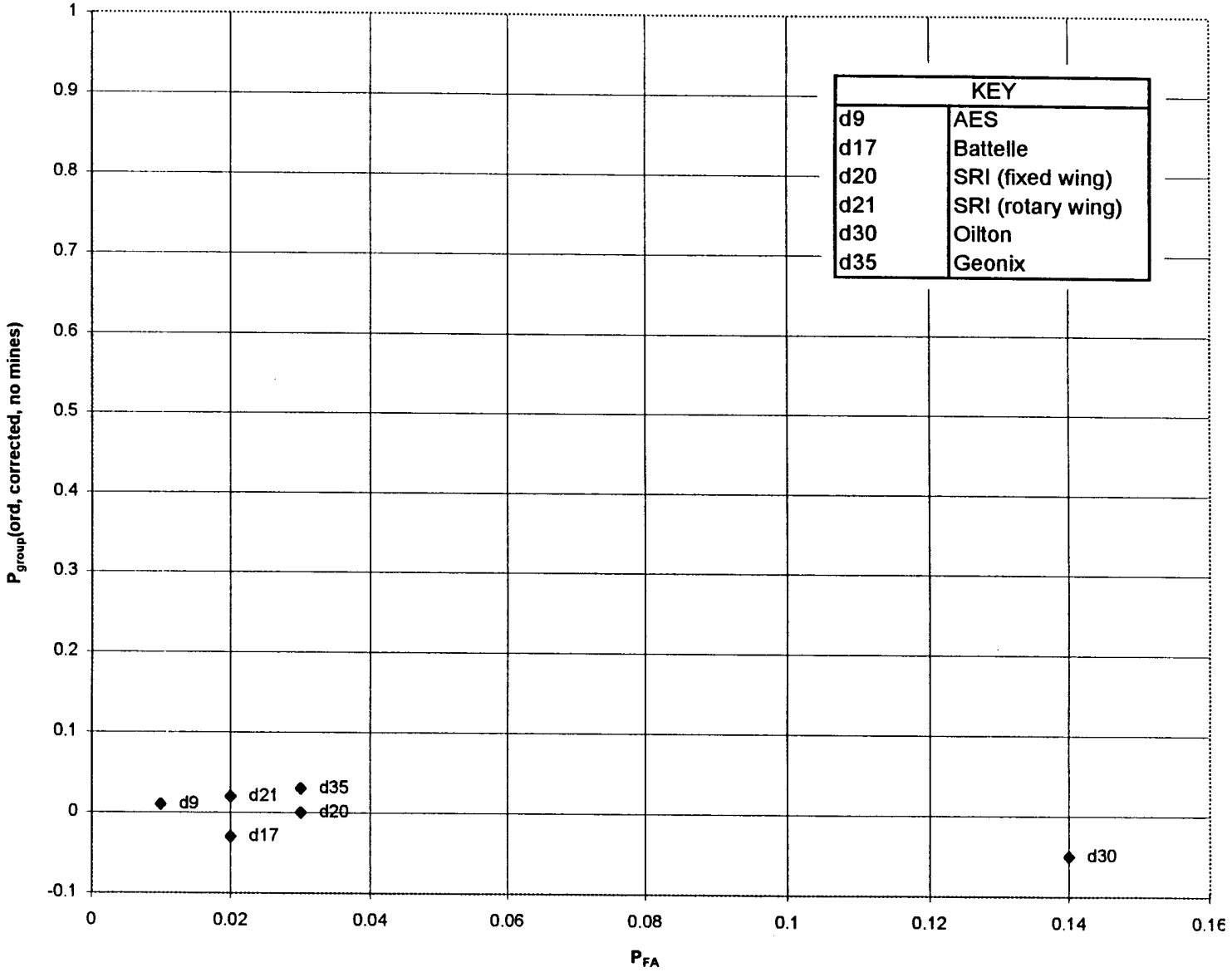


Figure 5.7. Demonstrators of Airborne Detection Systems: Summary of Performance

C. RELATIVE RANKING OF DEMONSTRATORS — BINNING

Because there are a finite number of baseline targets, the measured probability of detection will have a statistical uncertainty associated with it. It was determined that, for several demonstrators, the relative ranking could not be determined to the 95 percent confidence level. That is, if the demonstrators are ranked 1 through 20, the statistical error in the measurement is large enough to invert the ranking order for some demonstrators. In addition, as previously discussed, several factors could change the relative performance ranking among demonstrators. (See Chapter 4.) We do not attempt to provide any ranking for the airborne demonstrators. Additional tests, not discussed in Chapter 4, on the accuracy of the assessment and the effects on demonstrator rank are discussed in Appendix C.

For these reasons an exact ranking of demonstrators cannot be made with a high level of confidence. Therefore, demonstrators were placed in "bins" on the basis of our best estimate of detection capability[4] and false alarm rate. The bins for detection capability are shown in the table below. The average detection capability was 30 percent.

Binning of demonstrators on the 40-acre site on the basis of detection capability

Best estimate of detection capability	*Bin*	*Number of Demonstrators*
> 0.50	I	6
0.20 – 0.50	II	5
< 0.20	III	9

These categories were chosen with the idea that demonstrators with a detection capability that exceeds 50 percent have show some promise for ordnance detection. Those with a detection capability less than 20 percent, on the other hand, show little promise. Although this broad binning system for ranking demonstrators is not without flaws, it is not as misleading as ranking the demonstrators 1 through 20, when the statistical uncertainties could move a demonstrator by several places in ranking order.

---

4 See Chapter 4 for a description of the best estimate of detection capability.

The bins for false alarms are shown in the table below. The probability of false alarm is defined as the fraction of the site area that would have to be dug up to investigate the false alarms reported by each demonstrator.

Binning of Demonstrators on the 40-acre Site on the Basis of False Alarms

Fraction of site area covered by false alarms	*Bin*	*Number of Demonstrators*
0 – 0.01	I	6
0.01 – 0.03	II	8
0.03 – 0.08	III	4
> 0.08	IV	2

Figure 5.8 shows the best estimate of detection capability versus the probability of false alarm with the performance bins indicated.

The statistical uncertainties discussed above also preclude ranking the demonstrators on the basis of the receiver operating characteristic separation measure,[5] *d*, so again a binning system, shown in the following table, is employed.

Binning of Demonstrators on the 40-acre Site on the Basis of *d* (Separation of Signal and Noise)

d	*Bin*	*Number of Demonstrators*
> 2.0	I	5
1.5 - 2.0	II	5
1.0 - 1.5	III	2
< 1.0	IV	8

Figure 5.2 shows the best estimate of detection against the probability of false alarm with bins of constant *d* indicated. A high *d* value is desirable.

5 See Chapter 4 for more details.

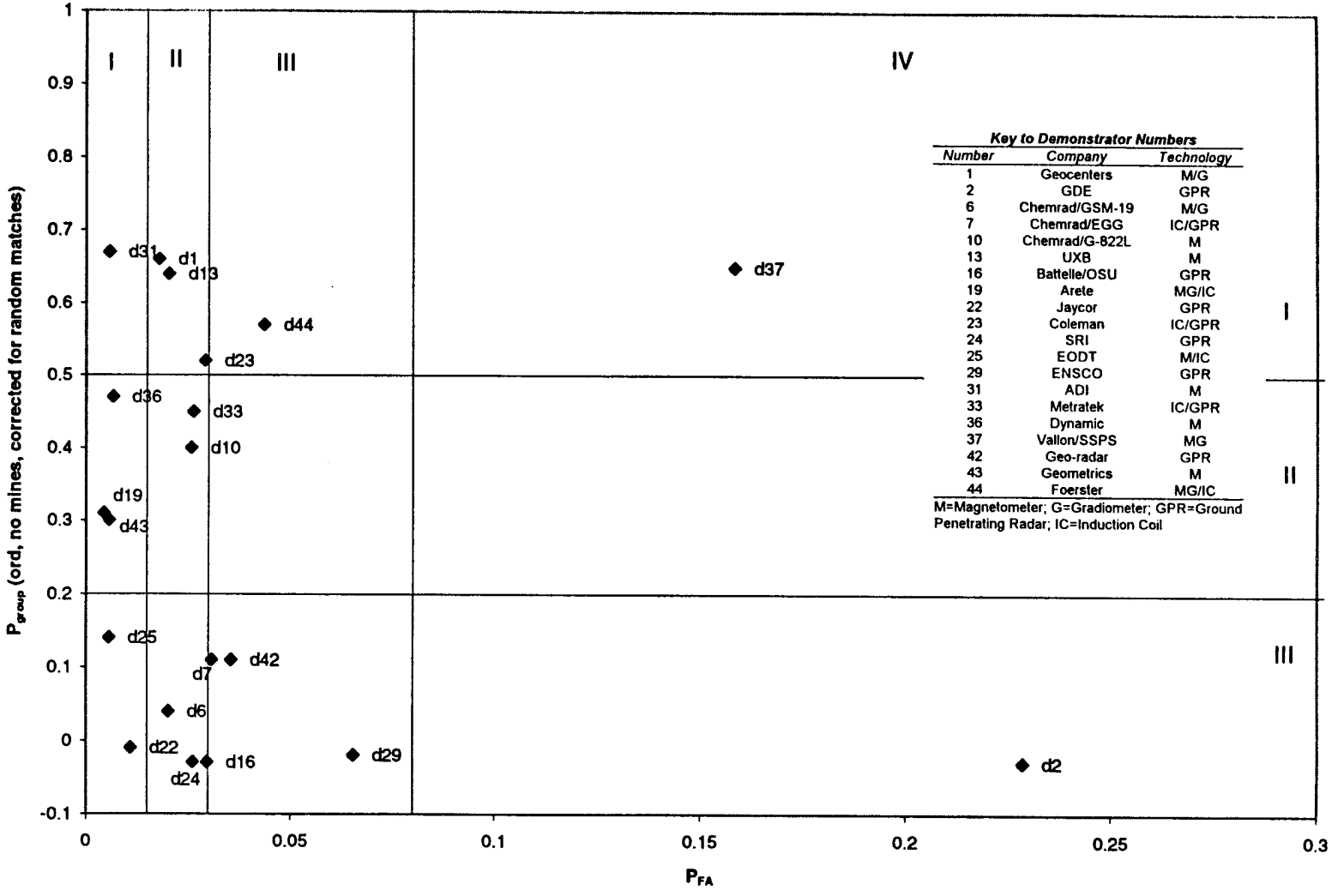


Figure 5.8. P_{match}(ord) on Area Searched vs. $P_{false\ alarm}$ With Performance Bins Indicated

6. TECHNOLOGY AND SYSTEM CONSIDERATIONS

A. INTRODUCTION

The Jefferson Proving Ground Demonstration was conducted to assess the performance of systems available for UXO location and remediation. The design of the demonstration was independent of the developers or the current owners of the equipment tested. As such, the demonstration provides information about available capabilities that is relevant to government users and decision makers responsible for clearance activities. It was not intended to provide system developers with sophisticated technical performance data. Thus, the measures of effectiveness were operational rather than developmental. For example, the probability of detection of mortars was measured, rather than the signal-to-noise level measured by a sensor 2 m from a mortar. Thus, this demonstration was a field test focusing on operational utility rather than a technical test focusing on hardware performance.

Although recovery is a major part of the UXO cleanup effort, there were few proposals to the program to demonstrate remediation capabilities. The three remediation demonstrators selected included only one commercial firm, Benthos. DOE and DoD were represented by Sandia National Laboratory and the Air Force Wright Laboratories, respectively. The small set of demonstrators limits the technology and system considerations and the implications for future R&D work on remediation systems.

We stress that the emphasis throughout the R&D community has been on detection, rather than remediation. This is in part because detection is important for both decision making, i.e., "Should we remediate the land?" and also for the effectiveness of remediation, i.e., "How much of the ordnance will actually be removed?" Nevertheless, even with 100 percent detection capability, in the end an enormous number of items must be excavated. Therefore, it is likely that equally or more important opportunities exist in improving underlying technologies or integrated systems for the remediation part of the UXO clearance.

The goal of this chapter is to discuss technology and system effects on performance at JPG. We also identify some implications for R&D toward improving the detection,

discrimination, and false alarm performance. To accomplish this goal, we must first understand the critical features determining the performance of different technologies during the JPG demonstration. This allows the applicability of the results to be estimated for other settings. It is only after some estimate of the universality of the JPG results is made that general conclusions can be drawn.

For the detection demonstrations, demonstrators reported only the locations thought potentially to contain ordnance, and some features of that ordnance, including size, type, and orientation. This information was used to determine overall system performance, but the contributing effects of sensor sensitivity, navigation accuracy, or signal processing could not be separated. As a result, the usefulness of the data for indicating the most promising directions for R&D is limited. Nevertheless, the results support some general guidance.

In addition to the data reported to the government, many of the demonstrators also collected and retained detailed data about the ground properties and the signals observed. In the companion volume to this report, the locations and types of a number of remediated ordnance items from the 40-acre and 80-acre sites are listed (Ref. 4). This information has been shared with the demonstrators for use in refining threshold settings, processing schemes, or other aspects of their systems.

B. SENSOR TECHNOLOGIES USED IN THE DETECTION OF UXO

One striking feature of the current approach to unexploded ordnance detection is that it focuses almost entirely on the detection of metal, usually magnetic, structural casing materials, rather than the detection of energetic compounds. Because of the preponderance of metallic debris on impact areas, systems that can detect high concentrations of explosives may provide opportunities for significantly reducing false alarms, so that limited recovery resources can be applied to activities that will reduce hazards. Of course, a technology that exploits other physical properties will be subject to other types of false alarms as yet uncharacterized. No techniques for directly detecting explosives are currently available for rapid field surveys.

The sensor technologies demonstrated at JPG included magnetometers, induction coils, and GPRs demonstrated on both airborne and ground platforms, and infrared imagers demonstrated on two of the airborne platforms. Other approaches used in UXO detection include visible imaging and LIDAR. A variety of acoustic, chemical, nuclear, and biological approaches are being studied for future application. These alternative and

developing approaches, which were not represented at JPG, are not discussed here. The following sections focus on sensor technologies demonstrated at JPG. Table 6.1 summarizes the most important parameters for each technology.

1. Magnetometers

a. Detection of Ordnance by Magnetometers

Detection of buried ordnance using a magnetometer system relies on the existence of a magnetic signature that is large compared to system noise and distinguishable from background magnetization. The magnetic signature of the ordnance derives from two primary contributions that can be of comparable magnitude. First, since ordnance items are commonly fabricated with steel, which contains iron, the placement of buried ordnance in the earth's magnetic field contributes to the overall magnetic signature. In the vicinity of ferromagnetic ordnance, the local magnetic field will exhibit a slight deviation from the background geomagnetic field. The net field deviation near any ferromagnetic item from the background geomagnetic field is strongly dependent on the size, shape, orientation, magnetization, and location of the item.

Second, ordnance items exhibit a weak remanent[1] magnetization that is dependent on the manufacturing processes. The magnetic field is "frozen into" the ordnance. During the manufacturing process, the earth's magnetic field causes a preferential orientation of the magnetic dipoles within the steel, leaving the object with a remanent magnetization. The magnitude and orientation of the remanent magnetization are influenced by metallurgical properties of the steel: grain size, chemical composition, the physical orientation during formation, and the types of heavy machining required to produce the finished product. In addition, firing and ground impact of the ordnance can alter its magnetization and thus its intrinsic magnetic signature. Although the remanent magnetic signature is generally smaller than the field perturbation effects, it cannot, in general, be ignored.

The leading term in both of these contributions is the magnetic dipole field. Thus, in our estimates, we have assumed that the total magnetic signature of ordnance is represented by the magnetic dipole field given by the sum of the two terms and thus has a

---

1 Remanent magnetization is defined as the magnetization that remains in a magnetic material after it has been magnetized to a level below saturation. See Ref. 8.

Table 6.1. Summary of Parameters Affecting Performance of Detection Technologies

Technology Demonstrated at JPG	Some Advantages/ Disadvantages	Optimal Site Critical Parameters	Optimal Ordnance Critical Parameters	Comments on JPG
GPR (Airborne/ Ground-Based)	A: High rate of site coverage possible (airborne) A: Imaging possible with post-processing of data A: May detect non-metallic items, i.e., mines D: Difficult optimization of resolution vs. depth of penetration	Low DC Soil Conductivity Low Soil Water Content High Homogeneity Minimal Surface Clutter	Large Relative Dielectric Constant compared to soil Large Radar Cross Section relative to radar wavelength (size/ orientation)	Soil conductivity high Soil water content high Most large objects not positioned optimally—(depth/orientation) Very Poor GPR conditions
IR (Airborne)	A: High rate of site coverage possible (airborne) A: Imaging possible with post-processing of data A: May be able to detect recent surface penetration D: Low image resolution D: Cannot detect deeply buried items	Large Thermal Loading Minimal Vegetation Smooth Terrain Low Soil Conductivity Low Soil Water Content	Large Thermal Mass Shallow Depth Large Physical Cross Section (size/ orientation)	Size insufficient for multiple pixel imaging Thermal contrast insufficient for single pixel detection High soil water content Modest thermal loading
Magnetometers (Airborne/ Ground-Based)	A: High signal to instrumentation noise A: Gradual signal drop off D: Low rate of site coverage D: Finds ferromagnetic objects	Low Geophysical Magnetic Noise Minimal Magnetic Minerals (either shallow small or deep large features)	Ferromagnetic Large Magnetic Metal Mass/Size High Inherent Magnetization	High probability of detection of large objects Moderate background magnetic noise and magnetic clutter produced high false alarm rate
Induction Coils (Airborne/ Ground-Based)	A: High signal to noise A: Finds all conductive metal objects A: Sensitive to small shallow objects D: Rapid signal drop off D: Low rate of site coverage	Low Soil Conductivity (only causes change in time constant required for inductive pulse)	Conductive Objects Shallow Depth Favorable Size/ Orientation (for eddy currents)	Minimal detection below 1.5 meter depth because of $1/r^6$ signal fall off

A = Advantage; D = Disadvantage

$1/r^3$ field strength dependence, where r is the distance from the dipole to the observation point. This dependence imposes some constraints on airborne detection. Magnetic fields are not attenuated by soil,[2] so the possibility of detecting deep objects exists as long as the $1/r^3$ field dependence does not cause the magnetic signature to become undetectable by the magnetometers. Current ground-based systems are limited primarily by clutter and discrimination rather than sensitivity. The difficulty in detecting ordnance using its magnetic signature is that the magnetic field of the earth is large compared to the small localized variations in the total field caused by the ordnance item. (See Table 6.2 for some characteristic numbers associated with the earth's field, a geological anomaly 1 km away, and a small ordnance item.) Similar local variations can be caused by naturally occurring phenomena or by manmade magnetic items that are not of interest. Gradient measurements are often used effectively to eliminate the geomagnetic field.

Table 6.2. Characteristic Vertical Magnetic Field and Field Gradient Estimates

	Geomagnetic Field	*Magnetic Anomaly*	*76-mm Projectile at 2 m*
Magnetic Field (γ)*	50,000	1,000	5
Vertical Gradient (γ/m)	0.01	3	3.5

* $1\ \gamma = 10^{-5}$ Oe, where the oersted is the cgs unit of magnetic induction.

Two different types of magnetometers were used in the demonstration at JPG: fluxgate and cesium vapor magnetometers. The basic fluxgate magnetometer consists of a ferromagnetic material (core) encircled by a pair of coils. The primary coil is driven with an alternating current. The secondary coil measures the time-varying magnetization of the core. If a static or low frequency external magnetic field has a component parallel to the axis of the core and coils, the induced voltage across the secondary coil is asymmetric, and the magnitude of the asymmetry is directly related to the external magnetic field. In most cases, the fluxgate magnetometer consists of two cores of high permeability material with two primary coils wound in opposite directions. No net voltage develops across the secondary coil or coil set if the external field is zero. For non-zero external fields, the secondary coil has a non-zero voltage directly related to the external field. Fluxgate magnetometers measure the component of the field parallel to the axis of the coils, and can be sensitive to fields as small as 0.1 γ.

---

2 This statement holds true for static magnetic fields. Time varying magnetic fields can be attenuated in soil.

The cesium vapor magnetometer is an optically pumped magnetometer that utilizes the quantum mechanical energy level splitting caused by weak magnetic fields. A filtered light source is used to excite electrons in the cesium from the ground state to excited states. The process of re-emission and absorption causes a specific energy level of the cesium vapor to become populated. As this process proceeds, the cesium vapor becomes transparent to the light. Next, a tunable RF signal is used to de-excite the electrons. This causes the cesium vapor to begin absorbing the filtered light again. The RF frequency for which absorption of the filtered light is a maximum is directly proportional to the magnitude of the external magnetic field.

The cesium vapor magnetometer measures the magnitude of the external field. This is different from the fluxgate magnetometer, which measures the component of the magnetic field for each measurement axis. It is important to note that the magnetic field is a vector quantity for which the individual field contributions can be superimposed, or added together. If the earth's magnetic field of 50,000 γ and the field from a buried ordnance item on the order of 5 γ are parallel, then the total magnetic field magnitude is equal to 50,005 γ. However, if the two fields are perpendicular, then the magnitude of the total magnetic field increases by only 2.5×10^{-4} γ. Thus, for a single measurement, an ordnance item that has its field contribution perpendicular to the earth's magnetic field at the measurement point would not be distinguisable from the earth's field using a cesium vapor magnetometer. Careful mapping of the full magnetic field vector in the vicinity of the ordnance would be required to resolve the item.

A fluxgate magnetometer is sensitive to the direction of the field. Thus, knowledge of the orientation of the magnetometer is required. If the magnetometer axis is one degree out of parrallel with the earth's magnetic field, a change in field of on the order of 5 γ is expected. If the magnetometer axis is nearly perpendicular to the earth's magnetic field, a rotation of 5×10^{-3} degrees results in a 5 γ change in the measured field. This strong directional sensitivity can lead to difficulty in interpretation of data if the magnetometer orientation is not well controlled. A fluxgate magnetometer system can be constructed using multiple sensors in an orthogonal array to give multi-axis measurement of the magnetic field. The multi-axis measurement provides additional data to assist with a reconstruction of the magnetic field. Still, as was stated before, the orientation of a fluxgate magnetometer system must be known in order to interprete the magnetic signal measured by the magnetometer.

Gradiometers can be used to mitigate some of the difficulties associated with measuring changes of a few γ caused by ordnance items in a background of 50,000 γ from the earth's magnetic field. Multiple cesium vapor or fluxgate magnetometers can be arrayed to obtain field gradient measurements. The geometry of the sensor array can be tuned to maximize sensitivity for general magnetic signature and to assist in the coverage of large areas. Since both the fluxgate and cesium vapor magnetometers measure the magnetic field at the observation point, the gradient measurement requires subtracting two independent magnetic field measurements made using two magnetometers separated by some known distance. Thus, the gradient measurement is a result that is determined in addition to the magnetic field strength. Although the gradient measurement permits the rejection of the constant component of background geomagnetic field, it will still be susceptible to the spatial changes in the geomagnetic field caused by local magnetic anomalies as shown in Table 6.2. Errors are introduced by any imbalance between the response functions of the two sensors and individual sensor noise.

The instrumentation errors introduced into the gradiometers described above can be substantially reduced by using a superconducting quantum interference device (SQUID) magnetometer, in a gradiometer configuration. This type of system employs a superconducting transformer to couple the magnetic field to the sensors. The superconducting transformer can be balanced to permit static magnetic field rejection to 1 part in 10^6 to essentially eliminate errors in the field gradient measurement due to uncanceled static field contributions. Commercial SQUID system sensitivity is better than 10^{-4} γ, more than two orders of magnitude better than the cesium vapor magnetometer. We do not feel that this level of sensitivity is needed for detection of UXO. However, its availability may offer improved system design. SQUID systems are commercially available presently, but none were demonstrated at JPG.

b. Critical Features Affecting the Detection of Ordnance by Magnetometers

The ability to detect magnetic ordnance items depends on many parameters. This technique is passive. The magnetic signature falls off as $1/r^3$, and is highly dependent on the orientation of the ordnance item in the background earth's magnetic field as well as the remanent magnetization. In the absence of background noise, the $1/r^3$ field dependence permits detection of magnetic ordnance to depths much greater than other techniques demonstrated at JPG. On the other hand, naturally occurring magnetic minerals, as well as magnetic man-made nonordnance items can cause magnetic signal clutter. Background

magnetic field anomalies at JPG have been measured in excess of 1 γ/m, which is the same order of magnitude expected for ordnance sizes and depths found at JPG. This level of field variation can cause difficulties discriminating ordnance from other objects.

Additional difficulty in detection of ordnance can occur because it is possible for the two magnetic components (remanent magnetism caused in manufacturing and variation in the earth's magnetic field caused by the permeability of the ordnance item) to lie in an orientation that is unfavorable for detection, or even partially cancel, depending on the orientation of the ordnance's remanent magnetization relative to the earth's magnetic field. This is one example of how the orientation of the ordnance can affect the probability of detection.

The design of the magnetometer system can have a marked effect on the overall performance. Since the geomagnetic field is much larger than the magnetic signatures of buried ordnance, gradiometers with different geometries can be employed to maximize the probability of detection in different situations. We will categorize gradiometers as vertical if the two magnetometers are placed one on top of another and as horizontal if the magnetometers are placed side by side. These gradiometer geometries were used at JPG. To estimate the effects of gradiometer geometry for a fluxgate magnetometer on the detectability of magnetic ordnance, a simple dipole model for the magnetic signature is used. To model a 155-mm projectile, three 500 emu-cm magnetic dipoles are oriented colinearly with orthogonal magnetization (see Fig. 6.1). The dipole array is constructed such that all dipoles lie in a plane parallel to the ground with a separation between dipoles of 30 cm, to approximate the length of a 155-mm projectile.[3] The vertical component of the magnetic field is calculated to determine the signature measured by both vertical and horizontal gradiometers.

Figure 6.2 shows the gradient field (γ/m) for two different separations (0.5 and 1.0 m) of the magnetic coils as the gradiometer passes to the side of the dipole array. Here the gradiometer is at a height of two meters above the center of the dipole array, the distance *d* from the center of the dipole array to the path traversed by the gradiometer is 1.5 m, and the horizontal distance ranges from 5 to 0 m. Also shown in Fig. 6.2 is the response for the same geometry with *d* equal to 0.5 m. The vertical gradiometer is very

---

[3] The dipole array is oriented to give a broad magnetic signature.

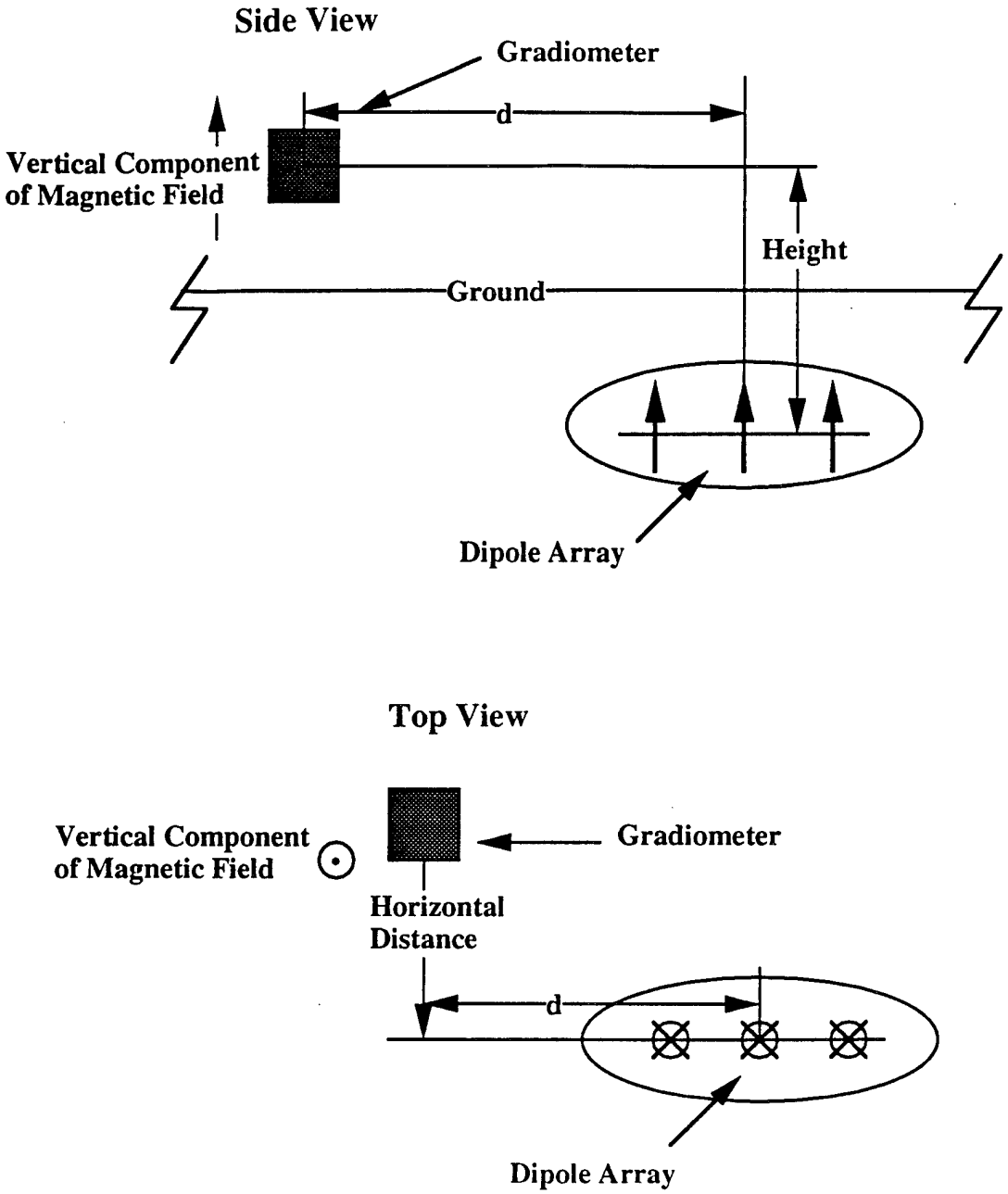


Figure 6.1. Dipole Placement for Gradiometer Calculations

sensitive in close proximity to the target, but at distances greater than a couple of meters away from the point of emplacement on the surface, the vertical magnetic field gradient from the buried ordnance is masked by the background field.

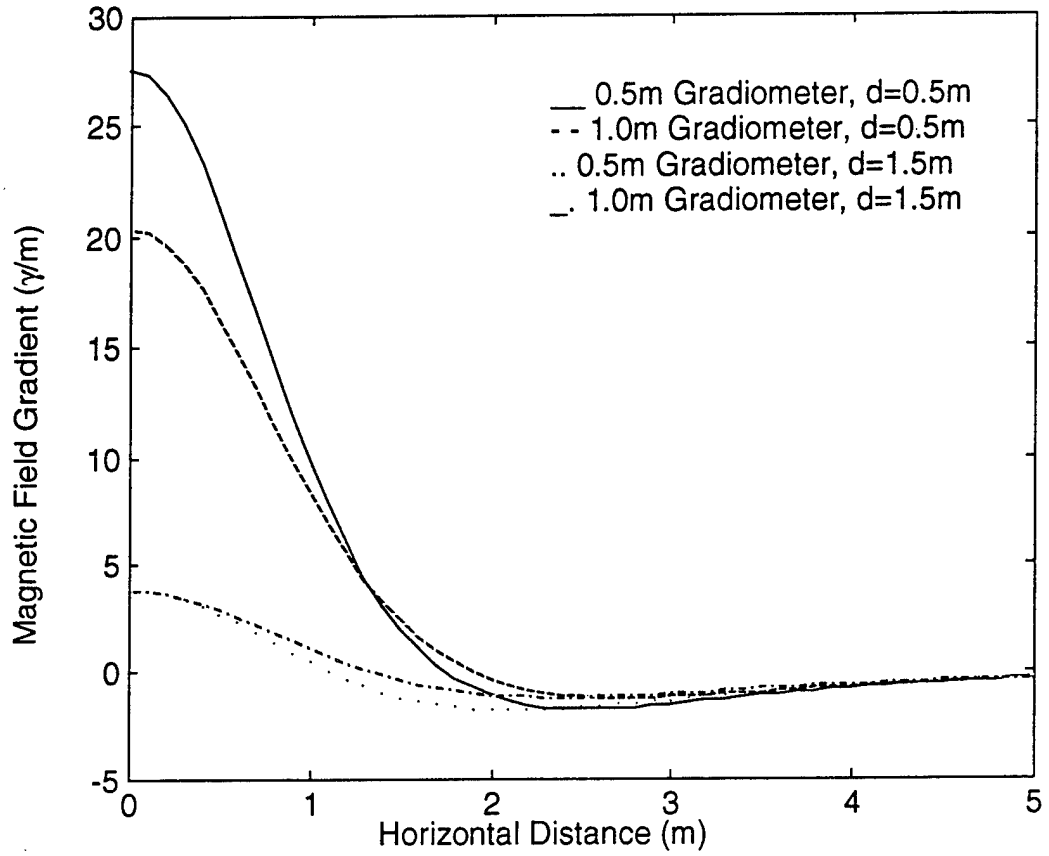


Figure 6.2. Vertical Gradiometer for Triple Dipole (155-mm Ordnance Model)

Calculations for a horizontal gradiometer are also presented. Figure 6.3 shows the calculated field gradient for the same dipole array used above. *d* is the distance from the dipole array center to the closer of the two magnetometers in the gradiometer array. In this case, the gradiometer gives improved sensitivity compared to the vertical gradiometer when the horizontal gradiometer is far from the target. Horizontal spatial gradients at the JPG 40-acre site exceeded 1.0 γ/m in certain areas. This gradient would exceed the signal of a buried ordnance item at distances greater than approximately 3 to 4 m horizontally from the emplacement point on the ground. The vertical gradiometer has better performance near the target.

These two examples illustrate that sensor geometry must be designed to address the specific conditions present in ordnance detection. Small separation vertical gradiometers

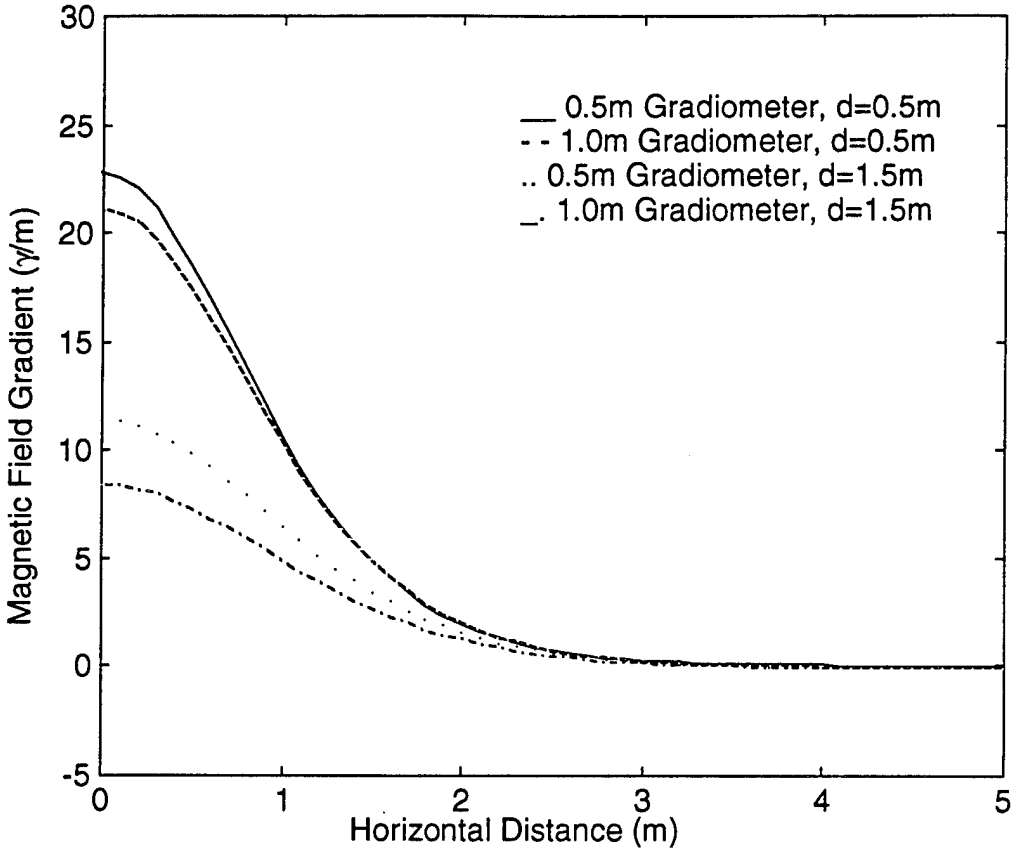


Figure 6.3. Horizontal Gradiometer for Triple Dipole (155-mm Ordnance Model)

are useful in location of ordnance close to the surface. Horizontal gradiometers are more sensitive in detecting objects at greater depths and also increase the surface area scanned for each measurement. This point is critical when attempting to search for magnetic signals over large areas.

c. Observed Performance of Magnetometers at JPG

Magnetometers were the most successful technology on the 40-acre site at the JPG demonstration. This success may be partly attributed to the fact that magnetometers are unaffected by high ground conductivity and have a relatively favorable $1/r^3$ signal dependence. In addition, the magnetometers were more often demonstrated by firms that find ordnance commercially, whereas other technologies were more often demonstrated by system developers. Comparison among demonstrators of the same technology indicates that field experience is an important factor in performance. For example, of the demonstrators that used cesium vapor magnetometers, those that detect UXO commercially performed better than developers of these systems.

The magnetometer represents the only technology for which there are multiple demonstrators with a significant number of detections. We have looked at correlations among the magnetometer detections by three of the better demonstrators that searched all or most of the 40-acre site. The examination of correlations produces a somewhat striking result. The magnetometer with the highest P_{match} by definition detects the most targets. One might have expected the magnetometer with the second highest P_{match} to detect most of those targets but none that the best magnetometer missed, and so on down the line. In this model, the targets can be ranked by level of difficulty, and the better sensors detect targets further down the list, but all the sensors detect from the top down with no gaps. At the opposite extreme, one could postulate that all the detections, even using the same technology, are independent. In this case, the probability of each demonstrator detecting any one target is the same as the probability of detection on the whole site. This probability is unaffected by whether or not other demonstrators found the target.

Assuming no correlations among detections, it is possible to predict, from the P_{match} values alone, the number of targets one would expect to be found by all, two, one, or none of these three demonstrators. When the calculated numbers are compared with the data, one finds that more targets are found by all three or by none of the demonstrators than would be expected assuming the detections are uncorrelated. It would appear that about 15 percent of the targets are easy to find (everyone finds them),[4] and about 5 to 10 percent of the targets are hard to find (no one finds them). For the remaining targets, those of medium difficulty, the pattern of detection is consistent with the detections being independent among the demonstrators.

The "hard-to-detect" targets were individually examined. It was observed that all of these targets fall into one of three categories: (1) small ordnance items in close proximity to larger items, where the signatures of the larger items may mask the signatures of the smaller; (2) medium-sized mortars buried more than 3 feet deep; and (3) ordnance items near trees or survey stakes that may obstruct access to the ordnance items. It is possible that these circumstances of ordnance emplacement explain the inability of demonstrators to find the undetected targets. However, available data about sensor performance is insufficient to draw this conclusion with high confidence. Further, these circumstances may, and

---

4 The percentage of targets is estimated by using a reduced data set, based on an area searched in its entirety by four of the better demonstrators. For the "easy-to-find" items, 15 percent is the excess of these targets beyond what would be found by all four if detections were uncorrelated. Similarly, for the "hard-to-find" targets, 5 to 10 percent is the excess of items found by no demonstrators. Inclusion of poorly performing demonstrators would make the analysis impossible.

in fact seem likely to, arise in actual ordnance-contaminated sites, so sensor performance in these situations is likely to be important. Additional investigation is required to determine if this is a general feature of buried UXO not specific to the JPG demonstration. If so, from the point of view of increasing cleanup effectiveness, some R&D effort should be focused on items that fall in the above three mentioned categories of hard-to-detect targets.

2. Induction Coils

a. Detection of Ordnance by Induction Coils

Induction coils can be used to detect the metal associated with most ordnance items. In an induction coil, a time-varying electric current in a transmitting coil produces a time-varying magnetic field. This fluctuating field will induce a voltage resulting in eddy currents in conductive material. When the transmitting field is turned off, the decay of the eddy current will, in turn, induce a voltage in a receiver coil. Because the induction system relies on induced currents, this system can be used to detect any conductive materials and is not limited to ferrous metals. However, an induction coil will also be susceptible to false alarms arising from all nearby nonordnance items made of conducting materials.

Induction coils are generally pulsed with a repetition rate in the kHz range. The decay time of the induced current depends on the geometry, magnetic permeability, and conductivity of the medium in which the current is induced. Since the conductivity of most natural objects and even the most conductive soils is orders of magnitude below the conductivity of metal, time gating the receiver coil can exploit the difference in decay time between highly conductive items and surrounding naturally occurring conductors in the soil.

b. Critical Features Affecting the Detection of Ordnance by Induction Coils

The transmitter coil of an induction system produces a magnetic field that falls off with the cube of the distance between the coil and the object of interest. As an example, Fig. 6.4 shows the field at a depth r from a transmitting loop of radius 0.5 m with a current of 10 amps. The field produced by the induced current in the conductor, which is the source of the voltage induced in the receiver coil, will fall off by the same $1/r^3$ relation to distance. The signal in the receiver coil, therefore, will fall off as $1/r^6$, where r is the distance from the induction coil system to the buried item.

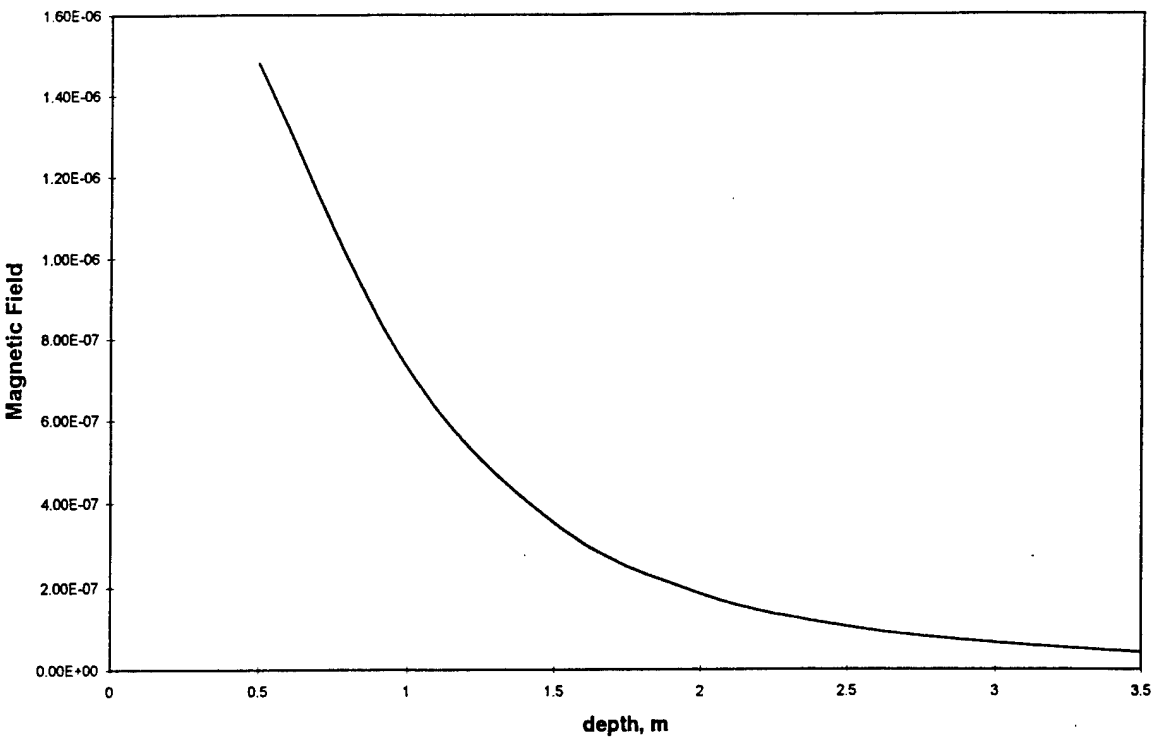


Figure 6.4. Magnetic Field Versus Distance From Coil

The soil conductivity at JPG was reported by one of the demonstrators as 0.1 mho/m, corresponding to a resistivity of 10 ohm-m. This can be compared to the resistivity of steel, 2×10^{-7} ohm-m, and of aluminum, 2×10^{-8} ohm-m. The current in a conductive target falls off exponentially in time. This fall-off is faster for items with low conductivity (high resistivity). Thus, the induced current in a metal ordnance item remains appreciable at times on the order of 1 msec, while the current in naturally occurring conductive materials in the surrounding soil will fall off several orders of magnitude more quickly. This short time constant is caused by the large resistivity of the soil compared to the metal and is obviated only slightly by the geometries of the affected region of soil relative to the metallic ordnance.

Figure 6.5 shows the expected voltage in the receiver for a stainless steel conducting loop 20 cm in diameter with a cross-sectional diameter of 1 cm. This could represent a small ordnance item. The transmit loop is assumed to have a peak current of 10 amps with a rapid characteristic turnoff time.

Figure 6.5 indicates that a detector operating in the 10 kHz range with sensitivity in the nV region will be able to detect the sample object at depths up to about 1 to 1.5 m. Below that depth, the expected voltage at the receiver coil for an object of this size is below

the sensitivity of most detectors. This discussion models ordnance items in any orientation as a horizontal loop. A more accurate model of the geometry would change the estimates of response by less than an order of magnitude.

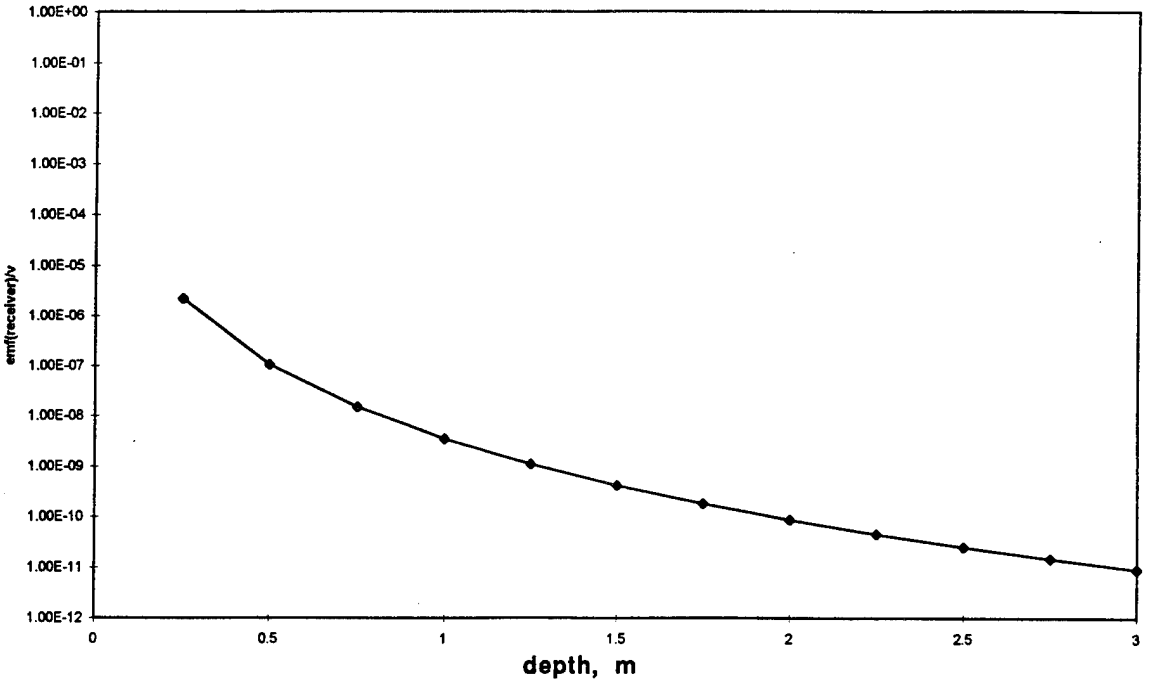


Figure 6.5. Voltage at the Receiver Coil From the Ordnance Model Described in the Text for a Notional Transmitter Coil

So far we have neglected any signal loss due to attenuation of the electromagnetic field through the soil. Although this source of losses is an important consideration in the performance of GPRs (discussed below), at the frequencies typically used for inductive measurements the effect of soil attenuation is expected to be small in comparison to the fall-off of the field with distance in free space. For the resisitivity measured at JPG and a 10-kHz frequency, the one-way attenuation is expected to be less the 20 percent at 3-m depth. The attenuation will be greater for higher frequencies and lower for lower frequencies. At sites with lower conductivity, the attenuation will be less at all frequencies. Thus, attenuation caused by soil conductivity is not likely to explain the poor performance of induction coil devices for the detection of deeply buried ordnance items. Rather, this is likely a result of the $1/r^6$ fall-off of the voltage induced in the receiver coil. For example, as shown in Fig. 6.6, attenuation through 1 m of soil reduces the field by only a few percent for the frequencies of interest, where the $1/r^6$ fall-off from 1 m to 2 m reduces the signal by a factor of 64.

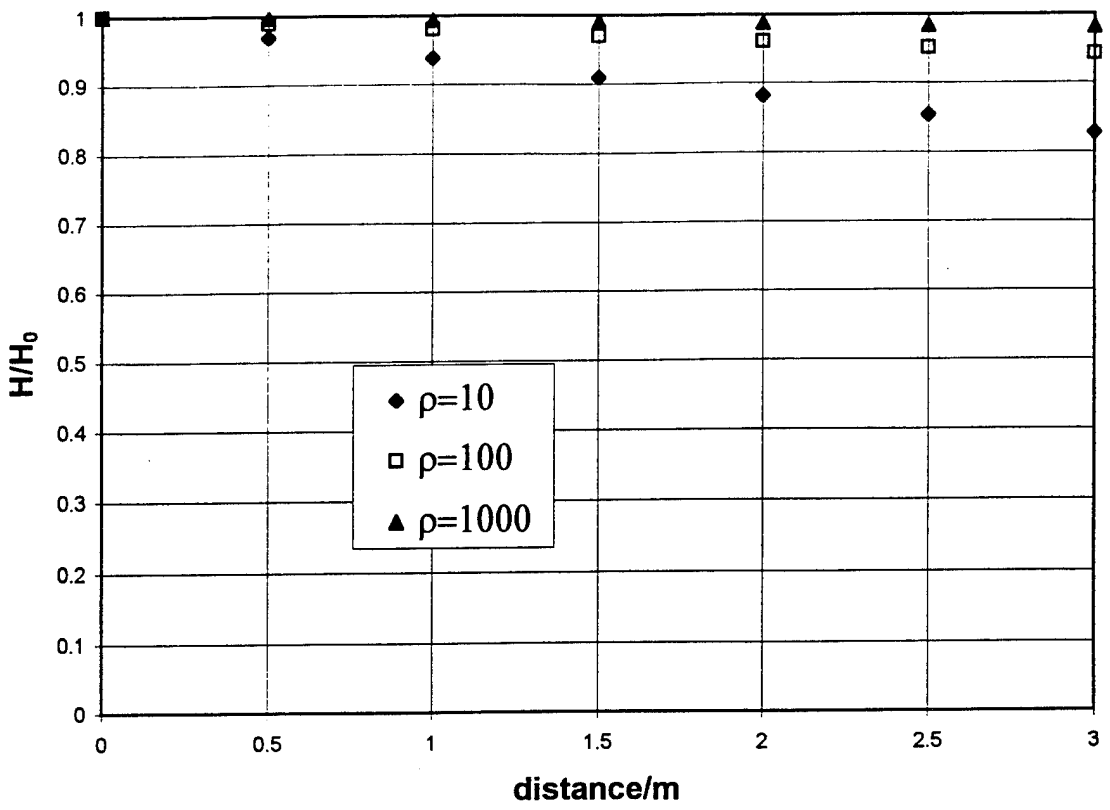


Figure 6.6. Attenuation of a 10-KHz Electromagnetic Field Through Soil Conditions Measured at JPG (ρ = 10) and Representative High Conductivity and Moderate Conductivity Soils. (The resistivity, ρ, is measured in units of ohm-m.)

c. Observed Performance of Induction Coils at JPG

None of the contractors at JPG demonstrated a pure induction coil system; instead, all of the induction coils were part of multiple sensor systems. Therefore, it is difficult to isolate the performance measured by induction coils and draw conclusions about the implications of the JPG results for the development of such systems for UXO detection. Foerster, Areté, and EODT demonstrated induction coils paired with magnetometers or gradiometers. These demonstrators did not separate their target declarations into those arising from the induction coils and those from the magnetometers. Since the magnetometers performed relatively well, separating the declarations after the fact is impossible. Two demonstrators, Coleman and Metratek, used systems that combined induction coils and GPRs. Since the stand-alone radars failed to locate any ordnance at all, we can attempt to isolate induction coil performance by assuming that the induction coils were responsible for all detections of these combined systems. In fact, Metratek separated declarations by detector type, and 80 percent of the ordnance items located were detected

with the induction coil, indicating that this is probably a reasonable assumption. The data does not allow the false alarms to be separated by technology type.

As might be expected from the above discussion, the results in Chapter 5 show that the induction coils did not perform as well as the magnetometers at the greater depths. Since this is probably a result of the fall-off of magnetic fields with distance, the depth coverage might be improved by increasing the sensitivity of the receiver coils. However, increased sensitivity will also exacerbate the false alarms, so it is not obvious that increasing sensitivity without improving discrimination ability would be of great benefit. In both cases where induction coil results could be isolated, the demonstrators detected nonordnance items at appreciable rates; neither attempted to classify declarations as ordnance or nonordnance. The detection rates will be affected by the relative number of ordnance and nonordnance items, as well as the relative number of shallow and deep ordnance items, emplaced at JPG, so we are reluctant to draw quantitative conclusions from this observation. However, it illustrates the need for better discrimination ability with induction coil techniques.

Finally, it is interesting to note that the approximately 15 percent of the targets that the best magnetometers all failed to find were also not detected by the induction coils. The phenomenology of the two techniques is sufficiently different that this result is not necessarily expected. Hard-to-detect targets are either close to larger items, deeply buried medium size mortars, or have obstructed access (see above discussion regarding magnetometers); these features would similarly affect detection by induction coils. Investigation of the nature of these difficult targets and determining whether such targets are typical on UXO-contaminated sites may also provide insight for research and development activities.

3. Ground-Penetrating Radar

a. Detection of Ordnance by Ground-Penetrating Radar (GPR)

The sequence of events in radar detection of underground objects is as follows: First, a radar pulse, usually of very short duration, is emitted. If the antenna is not in electrical contact with the ground, the radar wave propagates through the air to the ground. At the air-ground interface some energy is reflected; the amount depends upon the angle of incidence, the wavelength, the dielectric constant of the medium, and the surface roughness. Some of the energy penetrates the ground. If the antenna is in electrical contact with the ground, there are minimal losses at the air-ground interface. The radar pulse propagates through the ground and is scattered as a result of changes in the dielectric constant. These

can be continuous changes in the dielectric constant arising from changes in soil content, or boundary changes due to the presence of buried ordnance, benign man-made objects, rocks, or other inhomogeneities in the soil. The amount of backscattering depends upon the magnitude of the change in dielectric constant, the size and shape of the object creating this change, and the radar wavelength.

While the pulse propagates, it is attenuated by the medium through which it is traveling. Significant energy is lost through this attenuation, which is a strong function of the frequency. In addition, differences in dielectric constant as a function of frequency spread the pulse in time. Specifically, the speed of light in a dielectric medium depends on the dielectric constant, which varies with frequency. Hence, radar waves of different frequencies have different velocities in the medium, and the wave spreads out.

The energy that is backscattered is collected by the antenna. For the simplest case, the antenna is in contact with the ground, transmitting an approximate plane wave into the ground. The time required for the return of the short pulse is used to determine depth of the object. The resolution of this depth determination is approximately equal to half the pulse duration multiplied by the speed of light. The 40-acre site can be thought of as divided horizontally into parcels the size of the antenna's beam spot on the ground and divided vertically into sections each with the depth of the range resolution. Detection of ordnance items with high probability and low false alarm rate requires that the returns from any of these "boxes" containing ordnance be large compared to the returns from "boxes" with no ordnance and compared to system noise.

In general, there is a tradeoff between cross section and attenuation for the detection of any underground objects. The analysis below indicates that, for conditions at JPG, soil attenuation was sufficiently great even at frequencies of a few hundred MHz that detection of ordnance items would be difficult, if not impossible, with GPR. This occurs because with low frequency GPRs that have relatively good soil penetration (even at JPG), the scattering cross section of ordnance objects will be low. For higher frequency GPRs where the scattering cross section of ordnance objects is high, the soil penetration will be poor because of a combination of high conductivity and soil moisture. Calculations and analyses are also presented below for soil conditions different from JPG but more suitable for GPR.

The airborne systems suffer additional losses due to reflection at the air-ground interface and greater stand-off range. Furthermore, for the side-looking airborne radars, as opposed to the down-looking surface radars, there are discrete clutter objects on the surface

that appear in the same range gates as the buried target, and discrimination for all but the largest objects is effectively impossible.

b. Critical Features Affecting the Detection of Ordnance by GPRs

There are many parameters that affect the performance of radars in general and GPRs in particular. Understanding the results obtained by radars hinges on understanding the combined impacts of the suppression of radar cross section at long wavelength and the attenuation of the waves at short wavelength. The critical parameters, therefore, are the radar cross section and the attenuation through soil, both viewed as a function of frequency (or wavelength). Simple models for these parameters are presented below, and are used for quantitative estimates of detection ability in the next section.

(i) A Simple Model of Radar Cross Sections

Our calculations require a model for the radar cross section of a variety of objects over a wide range of frequencies. We will begin by considering the solution for scattering from a conducting sphere (see Fig. 6.7, taken from Ref. 9) as a basis. The cross section in the long-wavelength Rayleigh region is proportional to the volume squared divided by the wavelength to the fourth power. In the short wavelength optical region the cross section is defined by the presented area. In between, where the wavelength is on the order of 2π times the radius of the sphere, the cross section shows resonance structure.

For the ordnance problem, we will approximate bombs and projectiles as cylinders, and mines as disks. For these shapes, there exist exact formulae for the Rayleigh and optical regions. To keep the analysis simple, we will use the Rayleigh cross section at low frequency, until it intersects the optical limit, and the optical value thereafter. This is illustrated by the dashed line on Fig. 6.7 for a sphere. Neglecting the resonance structure will generally underestimate the cross section. However, for cylinders and disks the presented cross section will depend on orientation as well and we will generally use the maximum cross section orientation for our calculations. Therefore, the errors generated by the two approximations will tend to offset one another. We should note that these are the cross sections for perfect conductors of these shapes, an excellent approximation for metals. The approximation overestimates the cross sections of plastic mines, stones, or other objects with finite dielectric constant, typically by a factor of several. We ignore these effects in the following discussions. Table 6.3 gives the Rayleigh and optical cross

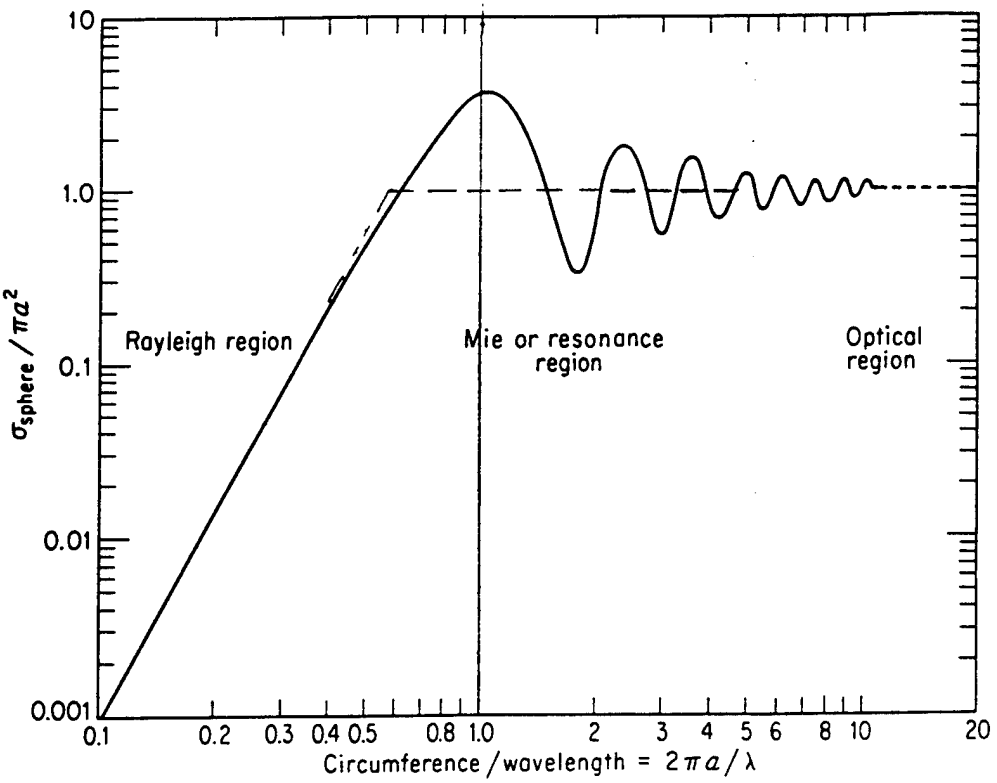


Figure 6.7. Radar Cross Section, σ, of the Sphere. *a* = radius; λ = wavelength. (Source: Ref. 9)

sections for spheres, cylinders, and disks, and the value of the wavelength where they are equal, in terms of the wavelength λ, the volume V, the radius a, and, for the cylinder, the length L. See Reference 10 for a more complete discussion.

Table 6.3. Rayleigh and Optical Limit Cross Sections for Various Shapes

	Rayleigh Limit	*Optical Limit*	*Matching Wavelength*
Sphere	$81\pi^3 V^2(1/\lambda)^4$	πa^2	$\lambda = \sqrt{3}\,(2\pi a)$
Cylinder	$64\pi^3 V^2(1/\lambda)^4 F^2$	$La(2\pi L/\lambda)$ *	$\lambda = 4\pi a(\pi/2F^2)^{1/3}$
	$F = (1 + 4a/3\pi L[e^{-(3L/4a)}])$		
Thin Disk	$64\pi^3 a^2(a/\lambda)^4$ *	$\pi a^2(2\pi a/\lambda)^2$ *	$\lambda = 4a$

* Maximum cross-section orientation.

(ii) A Simple Model of Radar Attenuation

Attenuation of radar waves is driven by two effects. The DC conductivity of the medium is the dominant effect at low frequency. At higher frequencies the water content of the soil leads to resonance attenuation via the same phenomenon that enables a microwave oven to heat anything with water in it. These effects have been both measured and

calculated. We present in Fig. 6.8 calculations for 15 percent water content and a variety of conductivities (Ref. 11). The vertical lines indicate the range of GPR frequencies demonstrated at JPG (100 MHz to 5 GHz).

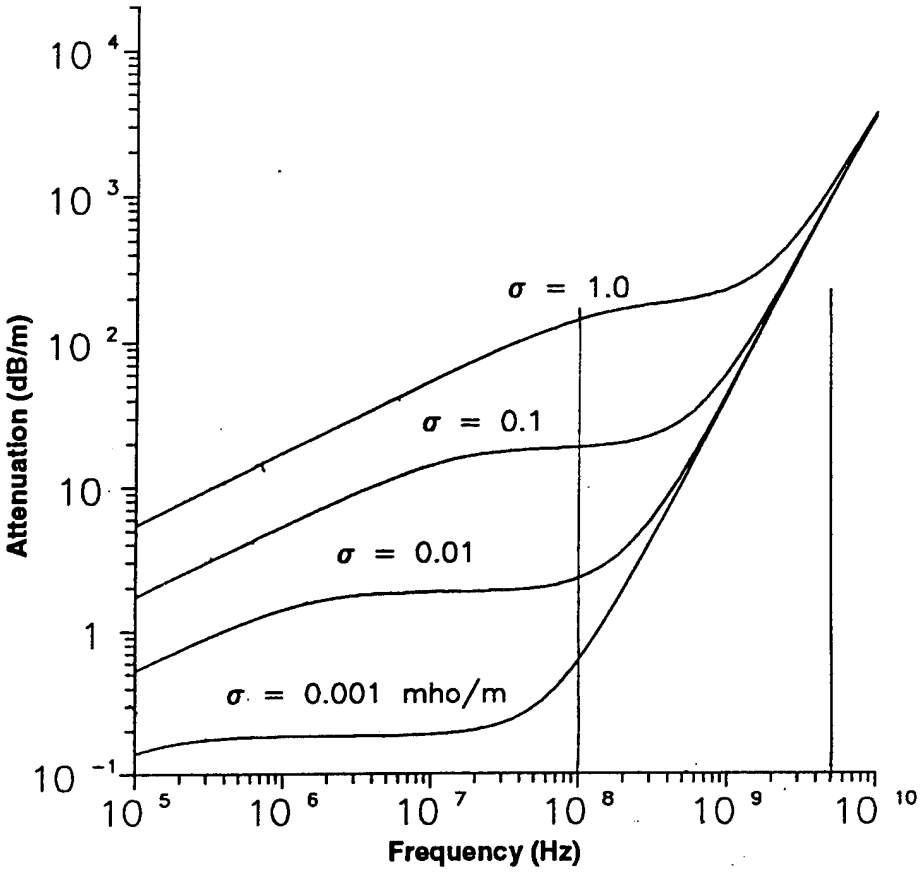


Figure 6.8. Attenuation of Radar Waves as a Function of Frequency in Wet Soils (Ref. 10)

(iii) Detectability of Mines at JPG–GPR

For small buried objects such as mines, effective detection will be limited by the ability to distinguish the mines from other small objects, which may be at a shallower depth. Figure 6.9 shows the two-way attenuation through 4.8 in. of soil (the depth of the mines at JPG) for two values of conductivity, 0.01 mho/m and 0.1 mho/m. These values are chosen because they represent the conductivity expected at JPG based on previous measurements (0.01 mho/m) and that observed by some of the demonstrators at the time of the demonstrations (0.1 mho/m). Also shown is the dependence of the free space cross section (in arbitrary units) of a 4.5-in. diameter disk as a function of frequency. The problem with detectability is clear. At higher frequencies (>1,000 MHz), where the free

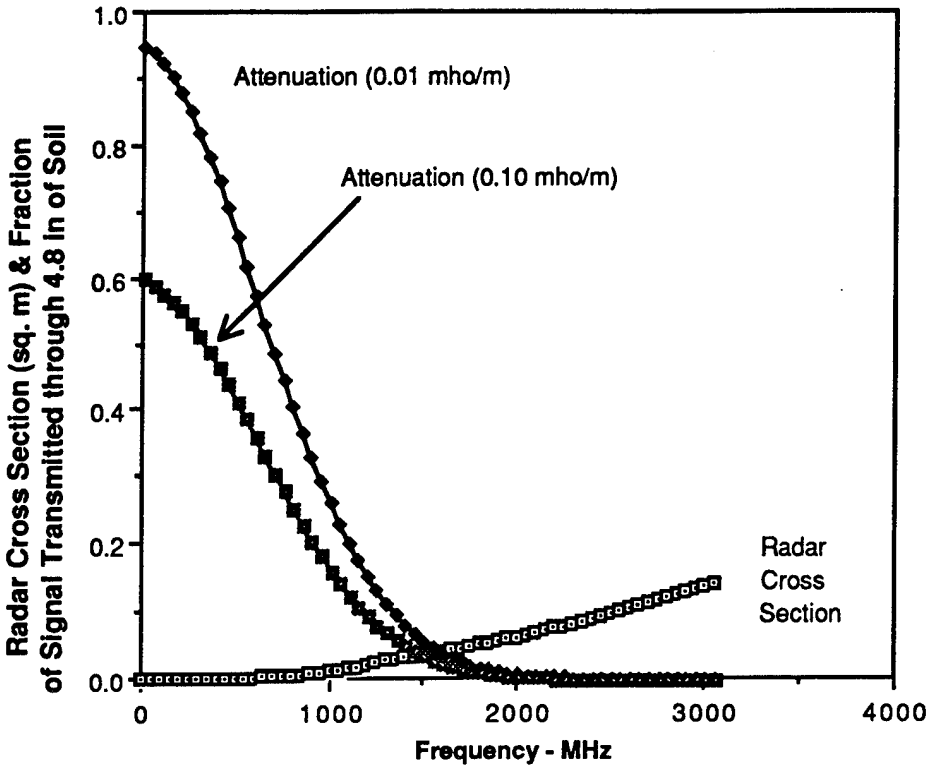


Figure 6.9. Effect of Attenuation Through 4.8 In. of Soil on Signal Transmitted and Radar Cross Section (RCS) of 4.5-In. diameter Mines

space cross section is large, the attenuation will limit detection. At lower frequencies, where the attenuation is low, the free space cross section of a mine-size object is vanishingly small. Another way of viewing this is in Fig. 6.10, where the free space cross section and the attenuated responses (the product of the free space cross section and the attenuation) are plotted as a function of frequency. We note that the response does not depend strongly on the conductivity.

The effective cross sections, although small, are sufficient to significantly exceed system noise at the short ranges involved in a ground-based GPR application. The question is whether returns from mines can be distinguished from the competing signals from background clutter. The bar chart in Fig. 6.11 shows the integrated response, or effective cross section, for a high frequency wide band radar with a uniform power density from 1 to 3 GHz. The response of a 4.5-in. mine buried at 4.8 in., as the mines were at JPG, is comparable to the response from a 3-in. stone at half that depth, a 2-in. stone just under the surface, and is much smaller than the response to a 3-in. stone near the surface, assuming the mines and stones have similar dielectric constants. A threshold setting that allowed

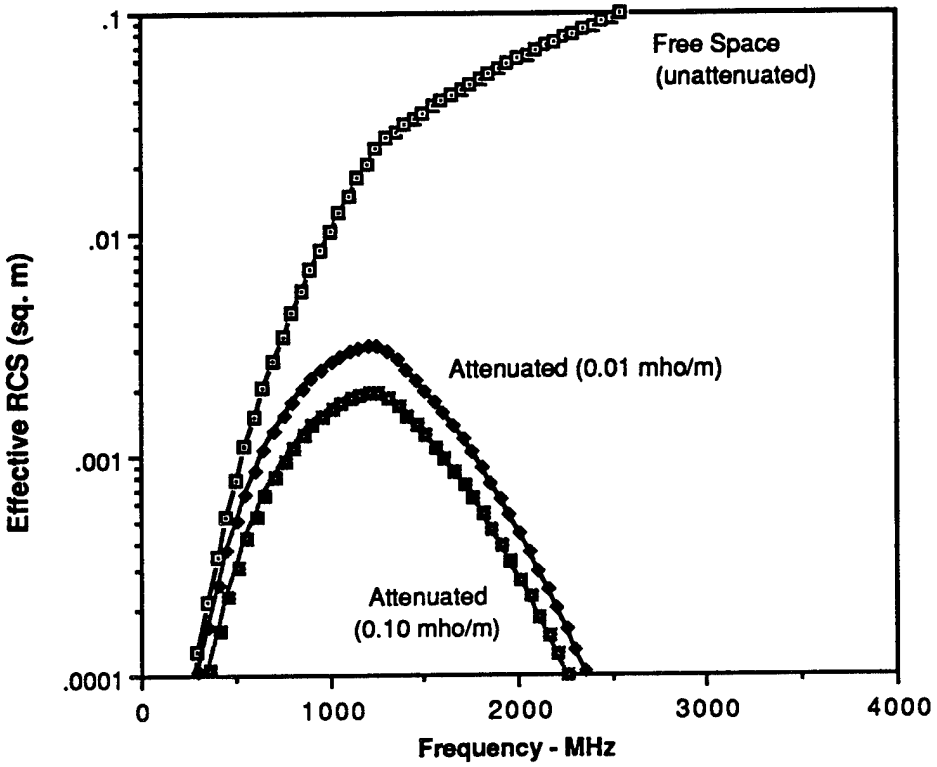


Figure 6.10. Effective RCS for 4.5-In. Mines

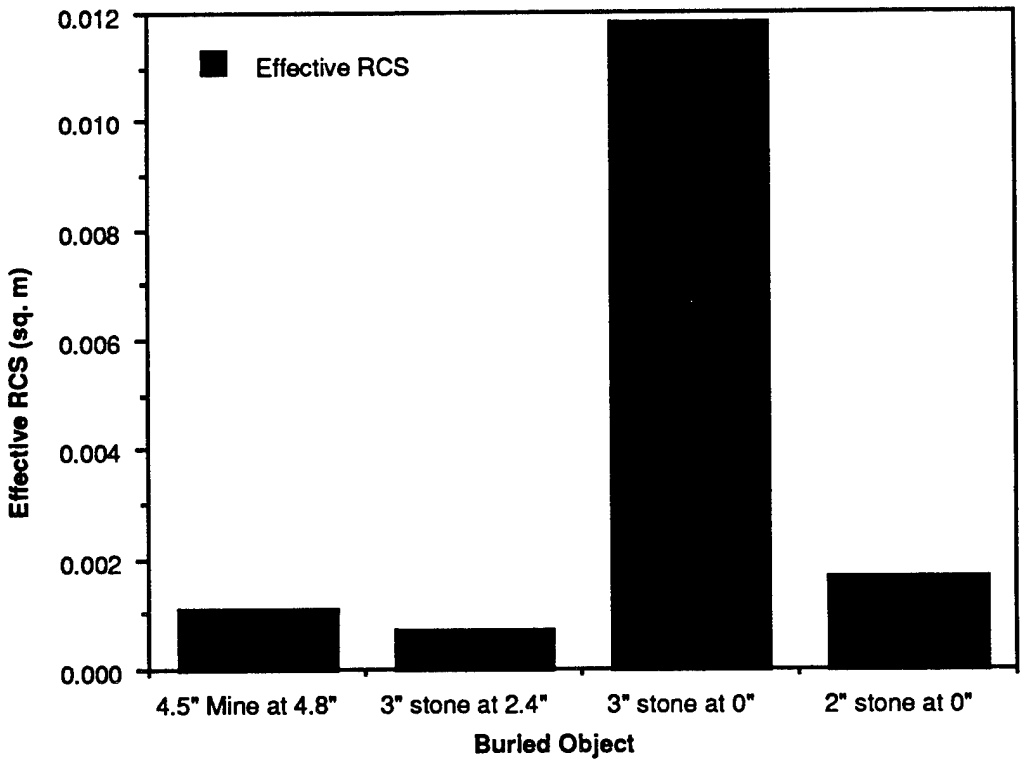


Figure 6.11. Effective Cross Sections for Various Objects

detection of these mines would also report myriad small stones at shallower depths as targets.

Armed with foreknowledge of the size and depth of the mines, it is possible to optimize the radar waveform to discriminate against smaller stones. However, any stones of the same size as the mines but shallower will still give competitive signals. Furthermore, an optimization focused only on discrimination against small stones drives one into the Rayleigh region and significantly reduces the cross section for scattering from the mines. Finally, a technology that requires knowledge of both size and depth of target for reliable performance is unlikely to have much practical value.

We have looked at the frequency ranges of the radars used at JPG. None of those about which we have information is optimized for this particular mine detection problem. It is not surprising that essentially no mines were detected. (The target matching algorithm reports one match with a mine for one demonstrator on the 40-acre site.) The mines were not undetected because of the high ground conductivity, however. It appears that the small size of the mines drives the system to very high frequencies that are rapidly attenuated by any water in the soil. For 4.5-in. buried mines such as those emplaced at JPG, the only prospect for high probability detection with low false alarm would be in a setting that contains very dry soil that is relatively homogeneous. This would include laboratory demonstrations and shallow mines emplaced in some roadbeds or on test ranges in desert regions.

(iv) Detectability of Projectiles at JPG–GPR

Projectiles, unlike mines, will tend to be located at significant depth. For deep projectiles, the attenuation will eliminate any practical possibility of detection with GPR. For projectiles relatively near the surface, we must again consider the discrimination problem. Consider the response of a 155-mm artillery shell, oriented horizontally and buried at a depth of 0.5 m. This is a relatively shallow depth to find a 155-mm shell but is also a characteristic range resolution cell size. High P_d, low false alarm detection of 155-mm projectiles will depend on how their response compares to that of natural objects closer to the surface.

In Fig. 6.12 we present the attenuation for two values of the conductivity and the cross section, in arbitrary units, for a cylinder of length 620 mm and diameter 155 mm oriented perpendicular to the radar beam. Another way of viewing this is in Fig. 6.13,

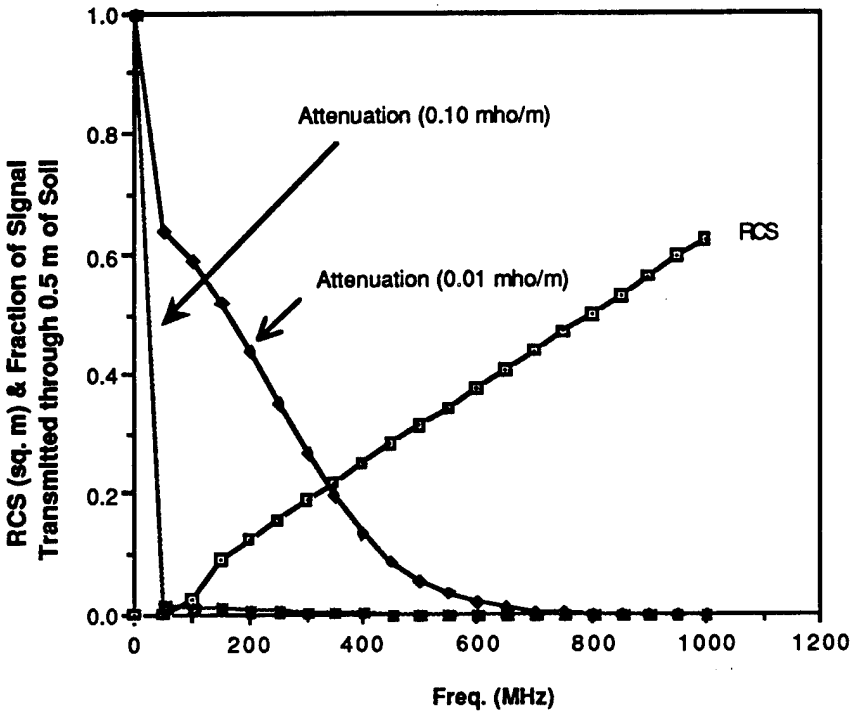


Figure 6.12. Effect of Attenuation Through 0.5 m of Soil on Signal Transmitted and RCS of 155-mm Projectile

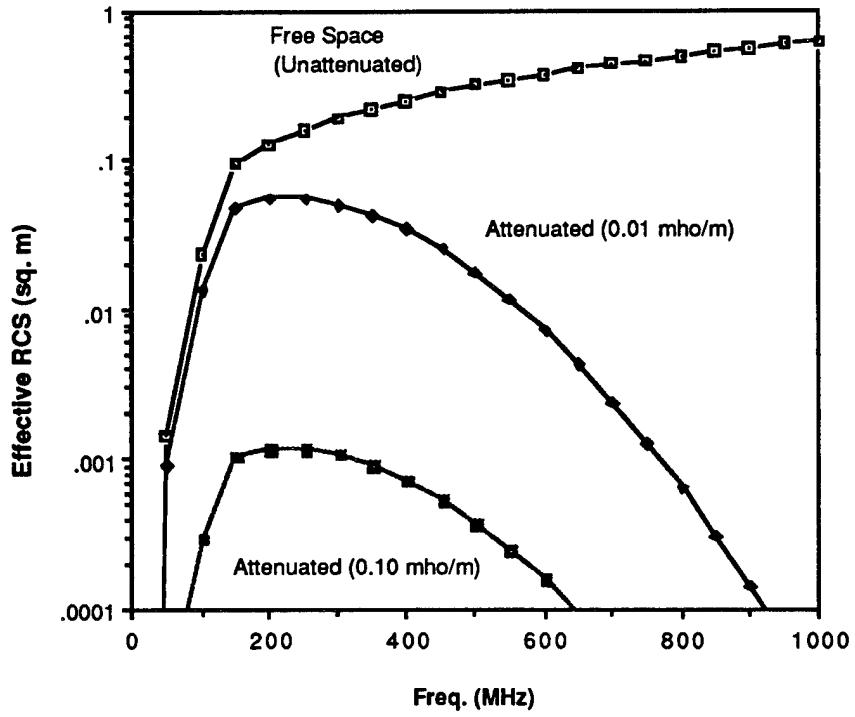


Figure 6.13. Effective RCS for 155-mm Cylinder

where the free space and attenuated responses are plotted as a function of frequency. From these figures we see that the expected 0.01 mho/m conductivity would not have had significant impact upon the detectability of these shells, but the 0.1 mho/m conductivity measured at JPG at the time of the demonstrations reduces the signal by two orders of magnitude. For the radars used at JPG about which we have data, the responses to a 155-mm shell modeled this way are less than or equal to the response to a 20-cm flat rock near the surface. In most terrain, this is a common enough occurrence to cause significant background problems. However, as the figure indicates, for lower conductivities in the 0.01 mho/m range, radars operating below around 500 MHz show only modest reduction (less than an order of magnitude) of the full free space response to shells of this size. Hence, the false alarms will be driven by rocks the same size as the shells. Rocks this size are perhaps a rare occurrence, but probably one that is inescapable.

For mines, recall the long wavelength response is too small to be useful and the short wavelength response is attenuated by water in the soil. Artillery shells, on the other hand, are large enough to scatter longer wavelength radar waves. The attenuation of these waves is less than that of short waves and depends upon DC conductivity, rather than soil moisture. At JPG, where the conductivity is very high, the shells were too deep to be reliably detected and discriminated from background objects. At another site with lower conductivity, the same shells at the same depths might well be distinguishable from underground clutter by a radar with a well chosen band.

(v) Detectability of Bombs at JPG–GPR

In the discussion so far, the difficulties with detection have been primarily due to clutter. For deeply buried objects such as bombs, the attenuation may be so great that system noise is the limiting feature. The depths at which this will happen depend upon details of the radar systems to which we do not have access.

For bombs that are close to the surface, the arguments from the section on projectiles largely apply. However, many bombs are significantly larger than even a 155-mm projectile, hence, there would be significantly fewer competing clutter signals for a large bomb lying flat within half a meter of the surface. For bombs that are not nearly horizontal, the presented cross sections are significantly smaller, and detection would be more difficult.

We have examined the probable returns from the emplaced bombs on the 40-acre site at JPG based on their size, depth, and orientation. Most were too deep to be detected

with GPR, given the high soil conductivity. Those near the surface had an unfavorable orientation for detection. Thus, it is not surprising that the stand-alone radars at JPG found no bombs. Location of large bombs within a meter or two of the surface in regions of low conductivity is probably possible with surface-towed GPRs, but other sensors could easily accomplish this job. Remote detection from an airborne platform suffers several additional losses and is unlikely except for very shallow objects or in the very lowest loss soils.

c. Observed Performance of Ground-Penetrating Radars

Since a major problem faced by the GPRs is attenuation while propagating through the ground, one would expect GPRs to preferentially detect objects near the surface. For the stand-alone GPRs the total number of matches on all targets was so small that statistical evaluation to determine if the GPRs are preferentially responding to objects near the surface is not possible.

The high ground conductivity at JPG provides a partial explanation for the poor GPR performance. However, similar difficulties were observed in the 1993 tests of airborne GPRs at YPG. The Yuma tests were technical rather than operational in emphasis, and allowed for detailed analyses of the losses. Detailed studies by Lincoln Laboratories into the source of the difficulty in target detection have been published (Ref. 11). An object with a given radar cross section measured in free space will have its signature reduced by a factor of two or three if placed in contact with, or buried under, an inhomogeneous complex dielectric medium. Losses also occur at the interface between the air and ground, where much of the energy is reflected. There are also two-way propagation losses through the soil. Finally, the return signal, on emerging from the ground into the air, is refracted into a larger solid angle. This last effect will have only modest impact for systems near the ground, but will be significant in the airborne case.

The results from the Yuma experiment are consistent with the JPG findings.

- Small targets near the surface cannot be distinguished from clutter objects because their cross sections are similar.
- Physically larger ordnance items, which tend to be buried deeper, have a large free space cross section, but the interface and propagation losses are so great that these objects also are largely undetectable.

4. Infrared Imaging Detection

a. Detection of Ordnance by Infrared Systems

Infrared systems respond to the thermal radiation emitted or reflected from objects. All objects emit radiation, with both the total amount and spectral distribution depending upon their temperature and emissivity. For the application to ordnance detection, reflected radiation is unlikely to be significant. As a practical matter, only radiation in certain wavelength regions propagates through the atmosphere. Hence, development has focused on detectors that operate in the so-called windows in the 3–5 micron and 8–12 micron wavelength region. Objects at temperatures of a few hundred degrees celsius (C) primarily radiate in the 3–5 micron region, whereas objects at room temperature primarily radiate in the 8–12 micron region. For the detection of ordnance items unlikely to be more than a few degrees different in temperature from their surroundings, the 8–12 micron region is usually used.

Infrared systems can be used for detection in two regimes. For well-resolved targets, images can be formed and examined, provided there is sufficient thermal contrast within the images. Typical modern systems readily form useful images if the contrasts are on the order of tenths of a degree C or larger. Alternatively, very hot objects can be detected by simple thresholding even when they are too small to form recognizable images. The most notable example of the latter is the use of infrared search and track systems to detect and track aircraft at long ranges in clear weather. The distant aircraft occupy a solid angle much smaller than a single element of the image (a picture element or "pixel"), yet the heat from aerodynamic friction and from the engines is sufficiently large that the aircraft are detectable. Often, aircraft can be detected with reasonable false alarm rates even against clouded (cluttered) backgrounds. As a practical matter, in a cluttered background the contrast must be so high that simple thresholding reduces the false alarm problem.

We do not believe that the IR signal from unexploded ordnance heated by the sun is sufficiently high or different from its surroundings to allow sub-pixel detection. Natural variations in temperature, emissivity and heat capacity will overwhelm the thermal signatures of ordnance items with false alarms. However, detection of ordnance items with good false alarm rejection based on multiple pixel images is possible under certain conditions. The critical parameters controlling the performance of multiple pixel detection are discussed below.

b. Critical Features Affecting the Infrared Detection of Ordnance

IR systems may, under optimal conditions, be useful for the detection of multiple-pixel-sized objects. In this case, the contrast requirements are much lower than for subpixel detection. Sufficient thermal contrast is required to exceed system noise and allow clear image formation. Unlike the case for subpixel detection, the contrast need not be great enough to eliminate false alarms by thresholding. Instead, false alarms are reduced or eliminated by recognizing or identifying the target object. There is considerable experience in recognition/identification of combat vehicles with forward looking infrared systems (FLIRs). In fact, recognition and identification have specific meanings, i.e., recognition means distinguishing a tank from a truck and identification means distinguishing one model tank or truck from another, e.g., an M-60 tank from a T-72 tank. Given that the contrast is sufficient so that system noise is not a problem, typically 50 percent probability of correct recognition and identification is achieved with 6 or 8 pixels, respectively, across the smaller presented dimension of the target.

We will use the level of imagery required for recognition, 6 pixels across the small dimension, as a surrogate for detection of ordnance with a reasonable prospect of discriminating against clutter. (For objects with a very high aspect ratio, 6–8 pixels along the large dimension will often suffice.) This surrogate will apply if thermal contrast is adequate. As a practical matter, this requires, among other things, good solar loading and no precipitation. The models in use were developed to compare the performance of different systems averaged over a collection of targets, not to predict the performance of a specific system. This imposes limitations on the use of these models. However, it clearly identifies the approximate number of pixels on target needed for reasonable detection and discrimination performance.

For a buried object, the temperature differences between the object and its surroundings depend upon the relative heat capacities and conductivities of the object and the soil. In general, as the ground warms and cools during the diurnal cycle there will be a lead or a lag in the temperature of the buried object. This underground temperature difference is represented on the surface by a blurred image with a smaller temperature difference. Furthermore, the original temperature difference between the ordnance and the earth is likely to be smaller if the ordnance item is buried, since the item is heated via conduction through the soil, rather than directly via sunlight. All of these effects make the infrared detection of buried objects significantly more difficult than detection of exposed objects.

In Table 6.4 we present the approximate narrow dimension of an ordnance item (in mm) with a reasonable probability of being both detected and correctly discriminated from background clutter for several FLIR systems at several altitudes, including those demonstrated at JPG. The calculations labeled RAH-66 use a 70 μradian detector resolution, based on specifications for the target acquisition FLIR for the RAH-66 Comanche. The corresponding resolutions for the other columns are 160, 940, and 350 μradians, respectively. The Safire and FLIR 2000 F are commercially available from FLIR Systems, Inc. To give a specific example, the RAH-66 FLIR would enable recognition of 105-mm shells at 750 ft and 155-mm shells at a range of 1,000 ft. In contrast, the Safire and the FLIR 2000 F would be unable to recognize a 155-mm shell above 500 ft even in the narrow field of view setting. The footprints even at lower altitude would not allow recognition of the small ordnance items on the surface at JPG if we assume 6 pixels are required for recognition.

Table 6.4. Minimum Dimensions[a] in mm for Detection by Various FLIRs

Slant Range (ft)	*RAH-66*	*Safire*		*FLIR 2000 F NFOV*
		NFOV [b]	*WFOV* [c]	
250	32	73	429	160
500	64	146	859	320
750	96	219	1288	480
1,000	128	293	1717	640

[a] The minimum dimension is taken to be 6 multiplied by the single pixel resolution size on the ground.

[b] Narrow field of view.

[c] Wide field of view.

This situation was exacerbated by the 5-in. to 48-in. height of vegetation reported by the demonstrators over the 80-acre site. This vegetation both obscured the views and interfered with the thermal heating and cooling of the ordnance items. Even if the narrow field-of-view setting had been used at a lower flight altitude, the vegetation would have made detection harder at JPG than at a relatively barren location, such as Yuma.

c. Observed Performance of Infrared Systems

There were two demonstrators at JPG that used IR systems. Oilton used the Safire FLIR exclusively, and Airborne Environmental Surveys used the FLIR 2000F in conjunction with a radar. Both were airborne systems, which, as discussed above, performed poorly. Although it is sometimes possible to detect buried objects with infrared

systems, it is difficult. It is surprising, however, that airborne demonstrators failed to detect and locate surface ordnance.[5] The detection of objects on the surface with infrared systems is not only well established, but is a cornerstone of Army combat procedures. However, at JPG, Oilton flew at 750 ft or 1,000 ft. As the table above indicates, this is too high to allow recognition of small surface ordnance. The 2000F has no recognition capability of small ordnance even at low altitude.

We do not think it likely that the thermal signature from the surface ordnance at JPG was sufficient to permit sub-pixel detection of ordnance. Even given adequate signature, the structure of the JPG test also required accurate location. The errors in location accuracy estimated by Oilton are large compared to the 5-m critical radius used in the TMA for airborne systems. Therefore, we believe there was no real prospect of distinguishing, in a statistical sense, real detections from lucky matches of false alarms to undetected baseline items. Since the subsurface detection problem is yet more difficult, it is not surprising that there was no detection capability for buried ordnance.

At other sites, the requirements for detection of surface ordnance are an adequate thermal driver, usually insolation, and a small enough pixel footprint on the ground, achieved via flight profile and sensor optics. The table above can be used to estimate what sort of flight regimes will be necessary for a given capability FLIR to detect a given size ordnance.

C. DIFFICULTIES WITH THE AIRBORNE DEMONSTRATIONS ON THE 80-ACRE SITE

The sensor technologies demonstrated on airborne platforms were ground penetrating radar, infrared imaging (FLIR), and magnetometers. The key result from the airborne demonstrations was that performance was statistically indistinguishable from detecting nothing, i.e., no evidence emerged to indicate that the demonstrators were locating more ordnance than would be uncovered by digging holes at random.

All the systems experienced difficulties particular to JPG or to the site conditions at the time of the demonstration. The soil conductivity and moisture content are both high at

---

5 The JPG test was aimed primarily at evaluating technologies for the detection of buried ordnance, but some items were placed on the surface of the 80-acre site. For infrared systems, we believe it is important to understand the inability of demonstrators to locate surface ordnance, which should be possible, before any progress can be made in improving the performance on the much more difficult problem of locating buried ordnance.

JPG, increasing the difficulty of GPR detection. Vegetation on the 80-acre site limited useful IR image recording to after sunset and before sunrise, and even then interfered with line of sight and thermal transport. The magnetometer, which was suspended from a helicopter, experienced difficulties from winds that caused the assembly to swing, corrupting the knowledge of the sensor's position, which was needed to locate any ordnance that might be detected. For the electromagnetic sensor, motion of the assembly through the earth's magnetic field probably added a component to the time variation of the magnetic flux, in addition to adding location uncertainty. This would confound attempts to detect the presence of conducting materials using the principle of induction.

There are simple reasons associated with range why ordnance location from an airborne platform is more difficult than from the ground. If the targets are not resolvable, then the signature of interest must compete with larger background signatures as range increases. If the targets of interest are resolvable, for a given angular resolution capability the number of pixels occupied becomes smaller as range increases. Finally, whether or not the target is resolvable, the task of locating the target in some fixed coordinate system becomes more difficult as the target and the sensor are separated.

For the case of GPR, the 1993 demonstration at YPG provides complementary information. At YPG, where the soil characteristics were initially believed to be more favorable, it also proved difficult to locate objects on or under ground using airborne GPRs. Ongoing work in this area is planned at Yuma Proving Ground during the summer of 1995. The results from YPG have been the subject of exhaustive scientific study by Lincoln Laboratory (Ref. 12) into the sources of signal loss. For much of the frequency region of interest there are competing signals from TV, FM radio, and cellular phones, which can be larger than both system noise and the clutter return. Although ground-based systems are close enough to the targets that this does not usually matter, for airborne platforms it is important because of the longer range. Signal processing techniques for canceling these pure tone interference sources exist and are being implemented within the radar community.

For the IR demonstration, we believe the pixel size on the ground was too large. A lower altitude flight or a higher resolution sensor might perform better. For the magnetometer, small surface items were probably lost due to the $1/r^3$ fall-off in magnetic field strength. These conditions, among others, we believe explain the somewhat surprising result that even surface ordnance items were not detected.

The airborne demonstrations indicate that integrating sensors on airborne platforms for the purpose of detecting subsurface UXO is premature at this time. The sensor technology and discrimination approach must be demonstrated with ground-based systems at a calibrated test site and at an independent "realistic" site, such as the one at JPG, before the expense of such an integration would be justified. By "calibrated" we mean a site with buried objects of known signature in well characterized soil for the purpose of determining sensor technical performance. By "realistic" we mean a site with ordnance placed as it is typically found on active ranges, FUDS, and BRAC sites, by someone other than the developer for the purpose of determining system utility. For example, the Yuma 93 effort was calibrated, but had no realistic placement of ordnance, while the JPG effort placed the ordnance more realistically, but placed no targets with calibrated signatures. Both are needed. The results from the ground-based demonstrations indicated that the detection problem is sufficiently challenging at close range that attempting simultaneous sensor development and airborne platform integration for longer range detection is likely to fail.

7. SUMMARY AND CONCLUSIONS

A. RESULTS

1. Demonstrators of Ground-Based Detection Systems

Table 7.1 summarizes our best understanding of the performance of demonstrators of ground-based detection systems at the 1994 JPG demonstration.

2. Demonstrators of Airborne Detection Systems

Table 7.2 summarizes performance of demonstrators of airborne detection systems.

B. MAIN FACTORS AFFECTING TECHNOLOGY PERFORMANCE AT JPG

Instrument phenomenology, operating conditions and target emplacement affected the successes and limitations of sensor performance at JPG.

1. Magnetometers

Magnetometers detected buried ordnance with some success at JPG. Because magnetometers rely on the magnetic signature of ordnance, which can be different for each item based on history, orientation, etc., detection can be difficult even for large magnetic objects. Background magnetic conditions can obscure the magnetic signature. This problem would be amplified if the site requiring remediation is located in a region with a high density of magnetic minerals. Finally, the detector system design can strongly influence the ease of interpretation of magnetic signatures.

2. Induction Coils

Detection by induction coils at JPG was moderately successful. Detection of deep objects was suppressed, but not eliminated, by the $1/r^6$ fall-off in signal. Soil attenuation was not a limiting factor in induction coil performance.

Table 7.1. Demonstrators of Ground-Based Detection Systems: Summary of Performance

Demonstrator	Technology	Best estimate of detection capability,* %	False alarm rate*	Typing	Acres	Notes
ADI ***M***	***Hand-held and surface towed GT-TM4 magnetometer, GT odometer, rope and tape for navigation.***	***67***	***0.43***		40	GPR system proposed, elected not to use—soil conductivity too high. Report all anomalies >100 g. (Pd and FAR adjusted to remove fence line.)
Arete ***MG IC***	***Geonics IC, Schonstedt gradiometer, man-portable GeoDAPS control system Trimble DGPS.***	*31*	***0.32***		25	
Battelle/ OSU GPR	Surface-towed GPR and rope/tape/odometer for navigation.	0	*2.2*	all O	2.3	Experimental, laboratory version of the radar system. High in clay content.
Chemrad/ EG&G GPR IC	Gulf-applied GPR and Pulse Tech IC, acoustic USRADS surveying tool.	11	2.3	all O	16	Navigation accuracy 6 in.
Chemrad/ G-822L *M*	*822L magnetometer, USRADS acoustic positioning system for navigation.*	*40*	*1.9*	all NO	40	Magnetometer range of 20–25 ft for detection.
Chemrad/ GSM-19 MG	GSM-19 magnetometer/gradiometer, USRADS acoustic positioning system.	4	*1.5*	all O	40	
Coleman *GPR IC*	*Towed multisensor array system (TOMAS), GPR and IC. GPS for navigation.*	*52*	*0.57*	all O	36	No IC is mentioned in proposal. All detections declared as ordnance. High confidence declarations only.
Dynamic Systems ***M***	***Man-portable system, Billigsley magnetometer, Foerster magnetometer, TopCon 302 Survey Instrument.***	*47*	***0.5***	all O	5.5	Sensors have range 15–20 ft.
Ensco GPR	Several GPRs on sled pulled by modified golf cart. Survey wheel and laser Track.	0	4.8	all O	10	
EODT M IC	Schonstedt magnetometer and EM-31 conductivity sensor, GEODAPS DGPS.	14	***0.42***	all O	11	Schonstedt range of 2–5 ft, conductivity sensor range 10–15 ft.
Foerster *MG IC*	*Ferex (Mark 26) Standard Sensor, Ferex Deep Search Sensor, Mines 2FD Standard Sensor. Ferex—gradient magnetometers, Minex—IC. Surface-towed. DGPS.*	*57*	3.23		24	Terrain required sensor to be located ~ 12 in. above ground.
GDE GPR	Prototype surface-towed imaging GPR rope/tape/odometer navigation system.	0	29.7	all O	6.9	Demonstrator noted large diameter clods of earth, standing puddles of water, and mud holes. (Corrected for random hits.)
Geo-Centers ***MG***	***STOLS: 2 Gmtrcs/Scntx magnetometers and one Foerster hybrid magnetometer/gradiometer.***	*66*	*1.33*		40	Gmtrcs/Scntx range > 25 ft, Foerster range = 5–10 ft. Areas inaccessible with STOLS used hand-carried.
Geo-metrics *M*	*Prototype Geometrics MagDIS man-portable system: 5 Cesium-vapor magnetometers. DGPS used for navigation.*	*30*	***0.43***		35	Demonstrator notes: rough terrain—large clumps of earth and cut-off/turned up roots. (After removing declared pipes, trenches,...)
Georadar GPR	Preproduction model of GeoRadar 1000A man-portable GPR.	11	2.6	all O	2	Range of detection 5–10 ft. Demonstrator notes GPRs troubled by wet, clay soils.
Jaycor GPR	Two GPRs mounted on a golf cart. Navigation using existing markers.	0	***0.81***	all O	20	
Metratek *GPR IC*	*Prototype model 200 stepped-frequency GPR mounted on sled pulled by four-wheel-drive vehicle and man-portable Geonics EM61 IC. DGPS used for navigation.*	*45*	*1.95*		5	System has 12-ft swath and frequency band of 0.2–2.0 GHz. Demonstrator note: muddy in low lying areas, dried over week. Conductivity of soil was high—deep targets not reachable.
SRI GPR	Trailer mounted GPR system. Navigation by placing stakes every 100 ft in both directions to use as guides. Precise positions determined using DGPS.	0	*1.95*	all NO	13	Demonstrator notes that resistivity would result in attenuation losses through soil such that maximum penetration of radar would be less than 2 m.
UXB ***M***	***Hand-carried Schonstedt GA-52B and Foerster Ferex magnetometer. GPS for navigation.***	*64*	*1.51*	all NO	30	Proposal indicates that GA-52B sensor work to depth of 3 m and Ferex sensors to depth of 5.8 m. Conditions at JPG "more than meet" the ideal conditions.
Vallon GmbH M	Hand-held, towed, and gradiometer magnetometers were used. SEPOS rope/tape/ odometer navigation system.	*65*	11.7	all O	12	Proposal states objects up to 7 m deep can be detected and location accuracy is 5 cm. (With declarations labeled fence post removed.)

KEY: M = magnetometer, G = gradiometer, IC = induction coil, GPR = ground-penetrating radar, all O or all NO: all declarations were typed as either ordnance or nonordnance.

* Best estimate of performance is described in Chapter 4. Bold italic numbers indicate the high category of detection capability or false alarm performance. Italics only indicate the second category. Demonstrators with names and technologies listed in bold italics are in the top category of the receiver operator curve characteristic, *d*, which accounts for both detection capability and false alarm rate. Italics only are in the second category. See Chapter 5.

Table 7.2. Demonstrators of Airborne Detection Systems: Summary of Performance

Demonstrator		*Technology*	*Best estimate of detection capability,* %*	*False alarm rate**	*Typing*	*Acres*	*Notes*
Airborne Environ. Surveys	GPR IR	Wideband radars centered at 500 MHz and 3 GHz, FLIR 2000F infrared imager on helicopter, DGPS navigation	0.01	0.11		80	
Battelle/ OSU	GPR	50–750 MHz GPR mounted on aerial platform (cherry picker), Trimble GPS navigation	0	0.28	all O	29	Vehicle was limited to roadway and radar was likely ineffective at distances > 500 ft
Geonex Aerodat	MG/ IC	IC, cesium vapor magnetometer/ gradiometer on boom towed below helicopter, DGPS navigation	0.03	0.39	all O	80	GPS failed, navigation limited to survey lanes marked on ground, winds caused boom to sway, 31 hr downtime from weather and equipment failures
Oilton	IR	FLIR 2000AB infrared imager, AIRDS navigation	0	1.7	all NO	80	High vegetation hampered detection, optics/flight path not optimized for imaging ordnance
SRI (Fixed Wing)	GPR	Bistatic GPR with SAR processing mounted on Beechcraft, DGPS navigation	0	0.36	all O	80	Detection hampered by wet soils
SRI (Rotary Wing)	GPR	Ultrawide Band GPR operating in UHF and VHF bands mounted on helicopter, GPS navigation	0.02	0.22	all O	80	Detection hampered by wet soils (maximum penetration 10 m in dry, sandy soils)

KEY: M = magnetometer, G = gradiometer, IC = induction coil, GPR = ground-penetrating radar, all O or all NO: all declarations were typed as either ordnance or nonordnance.

* Best estimate of performance is described in Chapter 4. Bold italic numbers indicate the high category of detection capability or false alarm performance. Italics only indicate the second category. Demonstrators with names and technologies listed in bold italics are in the top category of the receiver operator curve characteristic, *d*, which accounts for both detection capability and false alarm rate. Italics only are in the second category. See Chapter 5.

3. Ground-Penetrating Radars

Radar detection of the emplaced mines at JPG was prevented by an inability to discriminate objects of interest from background clutter of comparable size. High-frequency radars are needed in order to have appreciable response to the relatively small mines. These radars also scatter significantly from stones or other small discontinuities in the soil. At high frequencies, the radar waves are rapidly attenuated by any moisture in the soil. Discrimination between mines at a given depth and stones of comparable size at the same or shallower depth will be impossible. This is exacerbated by soil moisture, and is likely to apply almost everywhere.

For larger objects, the radar response in the 50–500 MHz region is sufficiently large for detection. However, at JPG, there is significant attenuation even at these low frequencies due to the high ground conductivity. It is not surprising, given the high conductivity, that essentially nothing was found by the radars. Better performance at detecting large objects, if they are not too deep, might be expected in other regions if the soil conductivity is lower.

4. Infrared Detectors

Infrared detection of buried objects is challenging under all but the most favorable conditions. For the heavily vegetated terrain at JPG, it might well have been impossible with any available system. The surprising result was that the surface ordnance was undetected. We believe this was due to use of optics and flight altitudes that required sub-pixel detection. We do not believe that the contrasts were likely to be high enough to allow sub-pixel detection, particularly in the presence of the vegetation. However, even if sub-pixel detection had been possible, the location errors associated with the airborne platforms would have rendered statistical determination of detection impossible. In the absence of recognizable imagery, therefore, it is impossible to establish any evidence for detection.

On the other hand, currently available commercial systems do have sufficient resolution to permit the recording of recognizable imagery from safe flight altitudes. It remains unclear to what extent vegetation will interfere with the infrared detection of surface ordnance in other locations. To establish the utility of IR sensors for UXO detection, this question warrants investigation. However, we do not regard it as likely that infrared techniques will allow detection of buried ordnance in vegetated areas, unless the burial was recent enough to allow detection of the scar.

C. CONCLUSIONS

1. Demonstrators of Ground-Based Detection Systems

The performance data allows the following conclusions to be drawn regarding the technologies demonstrated under the conditions at JPG:

- Magnetometer systems, on average, exhibited the best performance in this demonstration, with probabilities of detection of nearly 70 percent using the best estimate of performance. The number of emplaced items found by the demonstrators using magnetometers indicate that all the magnetometers were capable of detecting some ordnance.
- Stand-alone GPRs were unsuccessful at detecting buried unexploded ordnance.
- Induction coil metal detectors paired with GPRs did not perform as well as the best magnetometers, but nonetheless, as a group, systems with induction coils displayed moderate detection capabilities. Comparison to the detection capabilities of stand-alone GPRs suggests that the induction coil may be responsible for the detection capabilities displayed by these adjunct systems.

- Most demonstrators exhibited little or no ability to distinguish ordnance from nonordnance, or to identify one ordnance item from another.
- Most demonstrators reported multiple false alarms per ordnance item detected.
- False alarms and inaccuracies in locating ordnance items would cause even the best demonstrators to disturb 5 times the minimum amount of surface area required to recover a buried item; most demonstrators disturb between 10 and 100 times the minimum surface area required. The effect false alarm is in many cases an order of magnitude more detrimental than location inaccuracies.
- Demonstrators with extensive field experience performed better than those with little.

Different detection systems might be expected to perform with different levels of success depending on the size, type or depth of ordnance. Therefore, detection capabilities were calculated on subsets of the emplaced items sorted by these attributes. This subdivision produces a much smaller number of items in each category, thereby increasing statistical uncertainties. With this limit on the validity of the conclusions in mind, the data shows the following trends.

- Bombs are found with the highest reliability, with three demonstrators using magnetometers finding greater than 80 percent of the bombs. Further, there is a significant difference in performance between the magnetometers and the induction coils, with the magnetometers detecting bombs at about twice the rate of induction coils.
- For mortars, which are generally nearer to the surface, detection capabilities for magnetometer and induction coil technologies were indistinguishable.
- Not surprisingly, neither magnetometer nor induction coil technologies were able to detect the plastic antipersonnel mines.
- Magnetometers were more proficient at detecting ordnance items buried below 6 ft, where, in general, the larger targets were located. The ability of the induction coil systems, on the other hand, to detect ordnance falls off considerably in the "below 6 feet" range.
- When detection capabilities are sorted by size, induction coils perform best at locating the medium-size targets. The magnetometers, on the other hand, performed best at detecting the medium and large targets but not the small targets.

2. Demonstrators of Airborne Detection Systems

The results from the 80-acre site demonstration at JPG lead to the following conclusions:

- There is no evidence that airborne systems detected any ordnance. Thus, it is difficult to support any contention that rapid, accurate characterization of large tracts of land contaminated with subsurface UXO is feasible with current technology.
- Accurate navigation and mapping to allow matching of demonstrator detections with emplaced items was likely a limiting factor for the airborne systems demonstrated at JPG. This illustrates the important distinction between sensors that are able to provide images of surface UXO and systems that are able to support clearance activities.
- Integrating sensors on airborne platforms for the purpose of detecting subsurface UXO is likely to be unproductive with our current level of understanding. Any sensors proposed for integration should first be tested on targets with calibrated signatures and then on sites with a realistic placement of ordnance under field conditions such as at JPG. These tests of a stand-alone sensor should be a prerequisite for consideration of integration on an airborne platform.

D. INDICATIONS FOR FUTURE R&D EFFORTS

The JPG demonstration has produced summary level performance data with implications for future R&D efforts into UXO detection. The demonstration was carried out at a single site, by a relatively small number of demonstrators using only a subset of the technologies that could be applied to UXO detection. These demonstrators varied in nature from firms that find UXO commercially to nonprofit, academic technology developers. The conclusions one infers from the data are limited by the above considerations. Nevertheless, this data is a unique resource for R&D planning in UXO detection and should be utilized. Our conclusions about the implications of JPG data for R&D follow.

- Most demonstrators reported more false alarms than detections of ordnance items. This suggests that improving sensor sensitivity alone is not likely to improve performance. Better discrimination is required. Improvements in discrimination will require better understanding of the background features interfering with the detection of targets of interest and generating false alarms. This is probably the most important area for research.
- There is a set of targets in the JPG demonstration that were found by none of the demonstrators. All of these targets fall into one of three categories:

(1) small ordnance items in close proximity to larger items, where the signatures of the larger items may mask the signatures of the smaller; (2) medium-sized mortars buried more than 3 feet deep; and (3) ordnance items near trees or survey stakes that may obstruct access to the ordnance items. Future efforts should address whether these targets represent a general feature of buried unexploded ordnance.

- The results from the JPG demonstration suggest that false alarms might be reducible using sensor fusion. It may even prove possible to reduce false alarms with multiple sensors of the same technology. Unfortunately, this reduction of false alarms requires modest reductions in the detection capability of single sensors. Reducing false alarms through fusion, while retaining detection of the rarely detected emplaced items to improve detection capability over the single sensor values, will require much improved discrimination, if it is possible at all.
- With the possible exception of two demonstrators who also used induction coils, the demonstrators of GPR systems displayed performance that is statistically equivalent to finding nothing. Although this may be due in part to the high soil conductivity at JPG, we suspect that for ordnance detection GPRs will, at most, be useful in highly specialized applications where nonmetallic objects exist near the surface of low-loss uniform soils.
- There is no evidence from the JPG demonstration that the airborne systems detected any ordnance. This being the case, integrating sensors on airborne platforms for the purpose of detecting UXO is likely to be unproductive with our current level of understanding. Any sensors proposed for integration should first be tested on targets with calibrated signatures and then on sites with a realistic placement of ordnance under field conditions, such as at JPG. These tests of a stand-alone sensor should be a prerequisite for consideration of integration on an airborne platform.
- The superior performance of commercial firms suggests that experience is an important factor in determining performance. Studying the knowledge base of these firms may provide valuable information to complement the technical data of sensor developers.

REFERENCES

1. DoD Response to Congressional Request for UXO Contaminated Sites, December 1994.
2. Webster, Donovan, "The Soldiers Moved On. The War Moved On. The Bombs Stayed," *The Smithsonian*, February 1994.
3. Browne, Malcolm, "Battlefields of Khe Sanh: Still One Casualty a Day," *The New York Times*, May 13, 1994.
4. *Evaluation of Individual Demonstrator Performance at the Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase I)*, U.S. Army Environmental Center Report No. SFIM-AEC-ET-CR-95033, prepared by IDA, March 1995.
5. *Unexploded Ordnance Advanced Technology Demonstration Program at Jefferson Proving Ground (Phase I)*, U.S. Army Environmental Center Report No. SFIM-AEC-ET-CR-94120, prepared by PRC, Inc., December 1994.
6. McDonald, J.R., and R. Robertson, "Evaluation of the Trimble DGPS Naviation System: Chocolate Mountain Bombing Range," NRL Report No. NRL/PU/6110-94-260, July 8, 1994.
7. Van Trees, H.L., *Detection, Estimation, and Modulation Theory*, Part I, John Wiley & Sons, 1968.
8. Jiles, D., *Introduction to Magnetism and Magnetic Materials*, Chapman & Hall, 1991.
9. Skolnik, Merrill I., *Introduction to Radar Systems*, Second Edition, McGraw Hill, 1980.
10. *Methods of Radar Cross Section Analysis*, J.W. Crispin, Jr., and K.M. Siegel, eds., Academic Press, 1968.
11. Braunstein, M., IDA internal memorandum. Dr. Braunstein's calculations were used here. See J.I. Davis and A.P. Amman, *Geophysical Prospecting,* **37**, 531, 1989, or Ref. 12 below for more details.
12. Lee, C.F., "Ground Penetrating Radar Phenomenology Investigations," Lincoln Laboratory Report GPR-2,12.

GLOSSARY

Baseline Data:	The data describing location, depth, ordnance type, size and orientation of the deliberately emplaced objects and the objects subsequently identified and added to the data base.
Demonstrator:	One of the selected contractors that participated at JPG.
Demonstrator Declaration:	A declaration by a demonstrator indicating a location thought to contain an ordnance or nonordnance item and possibly describing other features of the object.
Detection Capability:	The ability of a demonstrator to detect emplaced items. Of most interest is the ability to detect ordnance items, but statistics were also compiled on the detection of all baseline items and of nonordnance items in the baseline. Several measures of detection capability, expressed as the fraction of emplaced items detected, were computed. These include P_{match}, $P_{closest}$, P_{near}, and P_{group}. In the chapters describing general applications we have referred to detection capability as probability of detection. However, in the detailed data analysis chapters we have stated which definition we are using.
80-acre Site:	One of two sites established in 1994 at JPG for the UXO tests. In the demonstrations only airborne demonstrators (loosely defined to include one radar system designed for airborne application but deployed from a cherry picker) used the 80-acre site.
Evaluation Criteria:	The seven criteria specified in the request for proposals for use in evaluating demonstrators. In decreasing order of importance, these are Detection Capability, False Negatives, False Positives, Target Position and Depth Accuracy, Classification Capability, Survey Rate, and Survey Cost. As a practical matter, the detection capability and false alarm performance were sufficient for demonstrator evaluation.
False Alarm Rate:	The false alarm rate is the number of false alarms per unit area searched. The size of the unit area is unspecified to prevent deductions about the baseline data.
False Negatives:	A false negative is a declaration of a target, identified by the demonstrator as ordnance, that does not correspond to an emplaced ordnance or nonordnance item. These can be caused by system noise, natural clutter, or unknown man-made objects buried on the site. It should be noted that this is different from the definition of false negative in use in other communities.

False Positives:	A false positive is a declaration of a target, identified by the demonstrator as ordnance, that corresponds to an emplaced nonordnance item (nonordnance items include some anomalies that were subsequently identified by multiple demonstrators and added to the baseline). It should be noted that this is different from the definition of false positive in use in other communities.
40-acre Site:	One of two sites established in 1994 at JPG for the UXO tests. In the demonstrations, only ground-based demonstrators used the 40-acre site.
"Lucky" Matches:	Demonstrator false alarms (sensor responses to system noise or background clutter) that happen to lie within the critical radius of an emplaced baseline item are called "lucky" matches. It is possible to use statistical tests to estimate the number of lucky matches a demonstrator has, but not to identify which of the matches are lucky and which are "true" matches.
Match:	A demonstrator declaration that falls within some specified radius of an emplaced baseline item. A declaration that matches a baseline item is counted toward the demonstrator's detection capability.
Measure of Performance:	A quantitative value describing a demonstrator's capability. The measures of performance include all seven measures from the request for proposals, as well as others developed during our analysis.
$P_{closest}$:	$P_{closest}$ is the probability of detection that uses proximity in the x-y plane to break ties among multiple candidate matches.
$P_{false\ alarm}$:	The percentage of the site covered by randomly placed circles of radius R_{crit} is used as a surrogate for $P_{false\ alarm}$. This is calculated as the number of number of false alarms multiplied by the area of a circle of radius R_{crit}, divided by the area searched. A correction for overlapping of the circles is used.
P_{group}:	P_{group} is the probability of detecting groups of the baseline items that are unlikely to be resolved. Despite uncertainties in the grouping and differences among demonstrators' location accuracy, we believe this is the best measure of detection capability.
P_{match}:	P_{match} is the probability of detection measured as "optimized" one-to-one matches between baseline items and demonstrator declarations. P_{match} is calculated using the ARS target matching algorithm that uses other criteria including depth, size, and type of ordnance to break ties among multiple candidate matches.

P_{near}:	P_{near} is the probability of detection that gives the demonstrator credit for detecting all items within R_{crit} of the declaration. This corresponds to the percentage of items that would be uncovered by digging holes of radius R_{crit} around each declaration and is strongly a function of emplaced item distribution.
R_{crit}:	R_{crit} is a parameter in the analysis used to limit the number of demonstrator declarations that are considered as possible matches to a given baseline item. All candidate declarations must have a location in the x-y plane within a circle of radius R_{crit} around the baseline item. In the absence of multiple candidate declarations or proximity of a declaration to multiple baseline items, this is sufficient to establish a match. If there were multiple candidates for matching, one of the tie-breaking schemes would be invoked.
Target Matching Algorithm (TMA):	The algorithm developed by ARS to identify one-to-one matches between the demonstrator declarations and the emplaced items.
"True" matches:	A true match results from a demonstrator declaration that is paired with the baseline target whose signature prompted the declaration.

ACRONYMS

AEC	Army Environmental Center
BRAC	Base Realignment and Closure
DC	Direct Current
DoD	Department of Defense
DOE	Department of Energy
EOD	Explosive Ordnance Disposal
FLIR	Forward Looking Infrared
FM	Frequency Modulation
FUDS	Formerly Used Defense Sites
GPR	Ground-Penetrating Radar
JPG	Jefferson Proving Ground
IR	Infrared
LIDAR	Light Detection and Ranging
NAVEODTECHDIV	Naval Explosive Ordnance Disposal Technology Division
NFOV	Narrow Field of View
RCS	Radar Cross Section
RF	Radio Frequency
ROC	Receiver Operating Characteristic
R&D	Research and Development
STOLS	Surface Towed Ordnance Locator System
TMA	Target Matching Algorithm
TV	Television
UXO	Unexploded Ordnance
WFOV	Wide Field of View
WWI	World War I
YPG	Yuma Proving Ground

APPENDIX A

DISCUSSION OF DIFFERENCES BETWEEN IDA AND PRC REPORTS

APPENDIX A
DISCUSSION OF DIFFERENCES BETWEEN IDA AND PRC REPORTS

The numbers reported by IDA were calculated using the 16 September 1994 version of the target matching algorithm (TMA) prepared by Automation Research Systems, Limited (ARS), which was tested by IDA, and the emplaced item baseline from 24 January 1995. The PRC report is based on a later release of the TMA and an earlier version of the baseline that were not verified by IDA. Because changes were small, the verification process was not reinitiated. Nevertheless, we felt it important to work with an algorithm and numbers that had been thoroughly checked and exercised. There may be slight differences in the way demonstrators were scored on the portion of the site they visited. In addition, we considered a number of factors related to demonstrator reporting methods—areas visited by the demonstrators, implications of high false alarm rates, and so on—which affect the assessment of demonstrator performance. Some examples include

- A few demonstrators searched the entire site they were assigned to visit, but most demonstrators visited only a portion of their site. Since most demonstrators indicated in their proposals an ability and intention to visit the entire site, it is important to have a measure of the amount of ordnance on the entire site that each demonstrator found in the allotted time. A measure of detection capability on the portion of the site visited provides information about demonstrator performance in the absence of time constraints, but it is difficult to compare the abilities of demonstrators that searched different subsections of the site on this basis. The work plan called for the demonstrators to search the entire site. Demonstrators, upon concluding that they could not properly search the entire site in the allotted time, made individual decisions about whether to search the entire site at lower performance levels than they would have liked or, alternatively, to search only a portion of the site at their best level of performance. Demonstrators that indicated in their proposals the ability to search only a portion of the site are so indicated.
- False alarm rates for some demonstrators were quite high, leading to the concern that a significant number of the baseline items that the demonstrator received credit for finding are a result of fortuitous pairings of false alarms with undetected baseline targets, rather than from target returns. Therefore,

measures of detection capability are corrected for "lucky" matches. The details of this correction are discussed in Appendix B.

- The generally high false alarm rates made the false alarm measures selected prior to the demonstration difficult to interpret. Most demonstrators fell into a small part of the total range for these false alarm measures, obscuring the differences between demonstrators with significantly different abilities. A new measure, the false alarm rate, *FAR*, is defined as the number of declarations at locations containing no known or emplaced items divided by the area searched. This ratio provides a linear measure of false alarms so that large differences in demonstrator performance are evident.

- Many demonstrators failed to distinguish between declarations believed to be ordnance and nonordnance items, making the measures of the demonstrator's ability to discriminate ordnance from nonordnance items meaningless and causing other measures of demonstrator performance to be undefined. These situations have been discussed in an attempt to compare all demonstrators on equal footing.

- Some demonstrators indicated in the comment field of the data input program that detections were believed to arise from fences, pipes, geologic anomalies, and the like, although the declarations were typed as other ordnance or nonordnance. These target declarations were removed from the demonstrator set, and the measures of demonstrator performance were recalculated.

APPENDIX B

CORRECTION FOR RANDOM MATCHES

APPENDIX B
CORRECTION FOR RANDOM MATCHES

To model the effects of the choice of the critical radius, we assumed that the target matching algorithm gave demonstrators credit for two types of matches: "proper matches" and "lucky matches." Proper matches are demonstrator declarations that are the result of returns from an emplaced target and separated from the corresponding emplaced target by a horizontal distance less than R_{crit}. Lucky matches are demonstrator false alarms that happen to fall within the critical radius of an undetected baseline target. The total number of matches reported by the target matching algorithm is the sum of the proper matches plus the lucky matches.

Number of Matches Measured by TMA = "Proper Matches" + "Lucky Matches"

As a notional example, Figure B.1 shows the number of matches from these two terms as a function of R_{crit}. The "proper matches" term rises sharply up to the distance that reflects the demonstrator's location accuracy but does not increase much beyond that point. The "lucky matches" continue to increase with R_{crit}.

True matches are defined as demonstrator declarations that were the result of returns from emplaced targets, regardless of whether they are separated by less than R_{crit}. The number of true matches (N_{TM}) is not a function of R_{crit}. A particular demonstrator declaration is or is not associated with a return from a baseline target, and this fact is not changed by the selection of R_{crit}. However, the subset of true detections that were counted by the TMA, the "proper matches," is a function of R_{crit}. For example, a demonstrator declaration that resulted from a return from an emplaced target can be separated from that emplaced target by more than R_{crit} and therefore not counted by the TMA.

An error function (erf) describes the relationship between the proper matches and the true matches (N_{TM}). This functional behavior assumes that the horizontal distance between the demonstrator declarations and the corresponding baseline targets has a gaussian distribution.

Number of true detections counted by the TMA = $N_{TM} erf(R_{crit}/s)$

The constant s is a scaling factor (with units of length) that is indicative of the proximity of a demonstrator's target declaration set to the emplaced target set. A demonstrator with a small value of s has true detections that are generally closer to the emplaced targets than a demonstrator with a larger value of s. Note that the number of proper matches is always less than or equal to N_{TM}, the total number of true matches.

To calculate the number of "lucky matches," first we determine the fraction of the site that is "covered" by demonstrator false alarms. A fortuitous match will be credited by the TMA if an emplaced item lies within the area of the site covered by disks of radius R_{crit} centered on the false alarms. If the demonstrator declaration density is sparse (i.e., the distance between declared targets is much greater than the critical radius), then there will be little overlapping of the circles and the area "covered" per false alarm is simply πR_{crit}^2. The total area covered by all the false alarms is the number of false alarms multiplied by πR_{crit}^2; i.e., $N_{fa}\,\pi R_{crit}^2$, where N_{fa} is the total number of demonstrator target declarations, N_H, minus the true detections, N_{TM}. Thus, the fraction of the site covered by false alarms becomes $\frac{(N_H - N_{TM})\pi R_{crit}^2}{A}$, where A is the area of the site. If R_{crit} becomes large enough that many of the circles overlap with each other or the edge of the site, then this approximation is not valid and corrections to include overlap and edge effects must be considered.

By assuming that the undetected emplaced targets are randomly distributed, the number of "lucky matches" can be estimated by multiplying the fraction of the area covered by false alarms by the number of undetected baseline targets; i.e.,

$$\text{Number of measured "Lucky Matches"} = \frac{(N_B - N_{TM})(N_H - N_{TM})\pi R_{crit}^2}{A},$$

where $(N_B - N_{TM})$ is the number of unmatched baseline targets. For example, if a demonstrator covers 20 percent of the site with false alarms and there are 50 undetected emplaced targets, then one would expect to get $0.20 \times 50 = 10$ "lucky matches."

The model for "true matches" and "lucky matches" can be combined to obtain the following equation.

Number of Matches from TMA = "Proper Matches" + "Lucky Matches"

$$N_M = N_{TM}\,erf(R_{crit}/s) + (N_B - N_{TM})(N_H - N_{TM})\,\pi R_{crit}^2/A$$

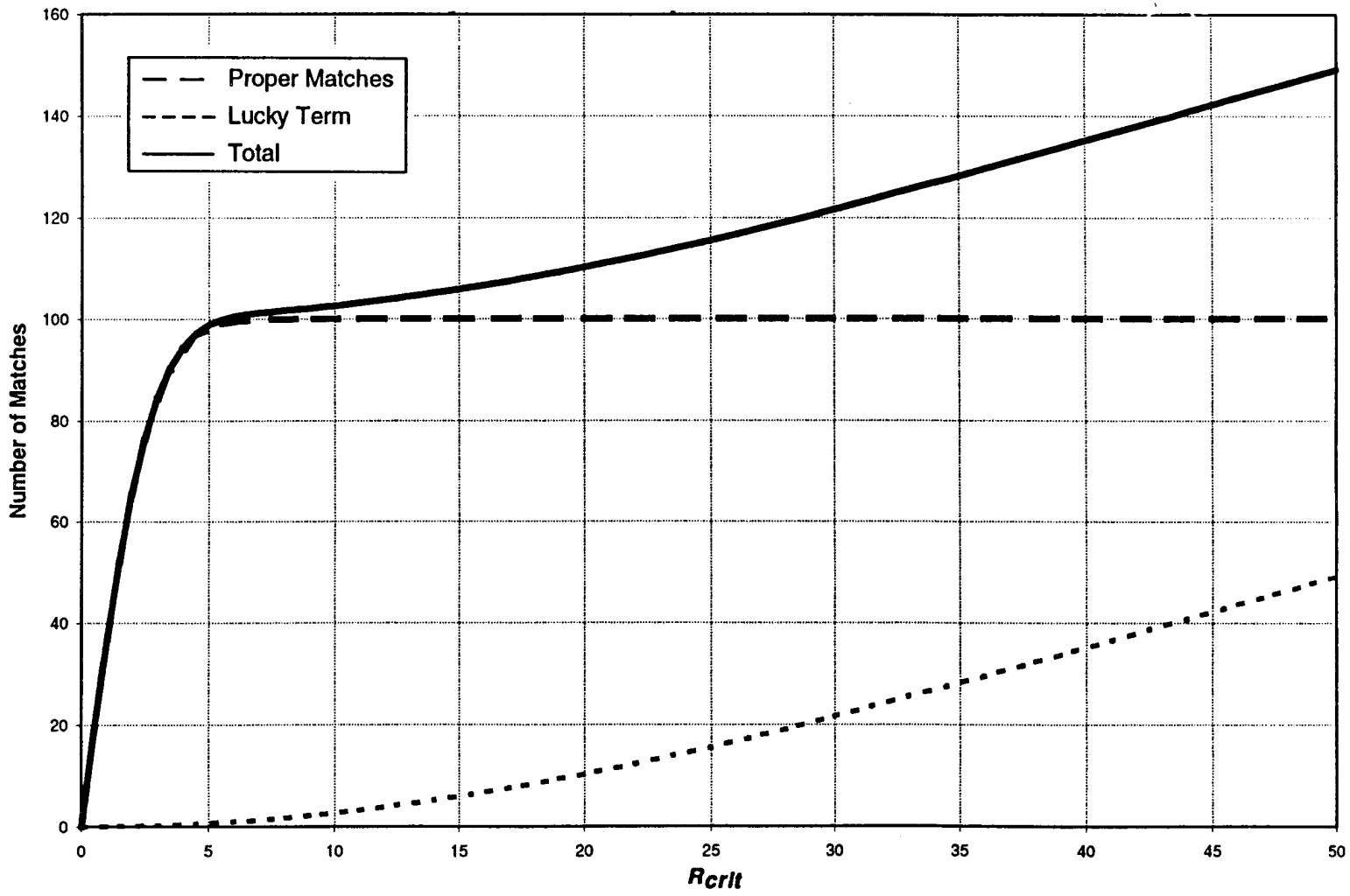


Figure B.1. Number of Matches vs. R_{crit} for a Notional Demonstrator. "Proper" and "lucky" matches have been separated.

where

N_M = Number of matches calculated by the TMA

N_{TM} = Number of true matches

N_B = Number of emplaced items

N_H = Number of demonstrator declarations

A = Area of the test site searched by the demonstrator

s = Free parameter that measures the proximity of the demonstrator's true matches to the emplaced targets.

If R_{crit} is greater than s, yet small enough that the overlap effects do not invalidate the above approximations, erf(R_{crit} /s) can be approximated as unity and the above equation can be solved for N_{TM} directly. A demonstrator's corrected P_{match} can then be calculated as N_{TM}/N_B. Figure B.2 shows the number of true matches (N_{TM}) calculated from the above equation for a sample demonstrator. At small R_{crit}, the error function term dominates. At large R_{crit}, the overlapping of the areas covered by each false alarm leads to an over-correction. Between these two regimes is a section where P_{match} is relatively insensitive to R_{crit}. Since the curve is not perfectly flat, even in the region where the above approximations are good, and since the endpoints of this flat section vary from demonstrator to demonstrator, a specific value of R_{crit} is still needed.

After examination of these curves for several demonstrators, a value of 2 m (6.56 ft) was chosen for all demonstrator evaluations.[1] To assure that the selection of 2 m provided a reasonable basis for the comparison of demonstrators, the value of s was estimated using a least squares fit for four sample demonstrators, and the results are shown in the following table.

Values of s and *erf* for Four Demonstrators at R_{crit} = 2.0 m

Demonstrator	s (ft)	$erf(R_{crit}/s)$
I	2.2	1.0
II	2.5	1.0
III	2.5	1.0
IV	4.6	0.96

[1] P_{match} was calculated for several values of R_{crit} for the eight demonstrators that showed above average detection ability. For all but one of these demonstrators, P_{match} increased slowly with R_{crit} in the region 1–3 m. The demonstrator with a rapidly changing P_{match} had a very high false alarm rate, and the increase in credited detections can be attributed to the lucky matches term.

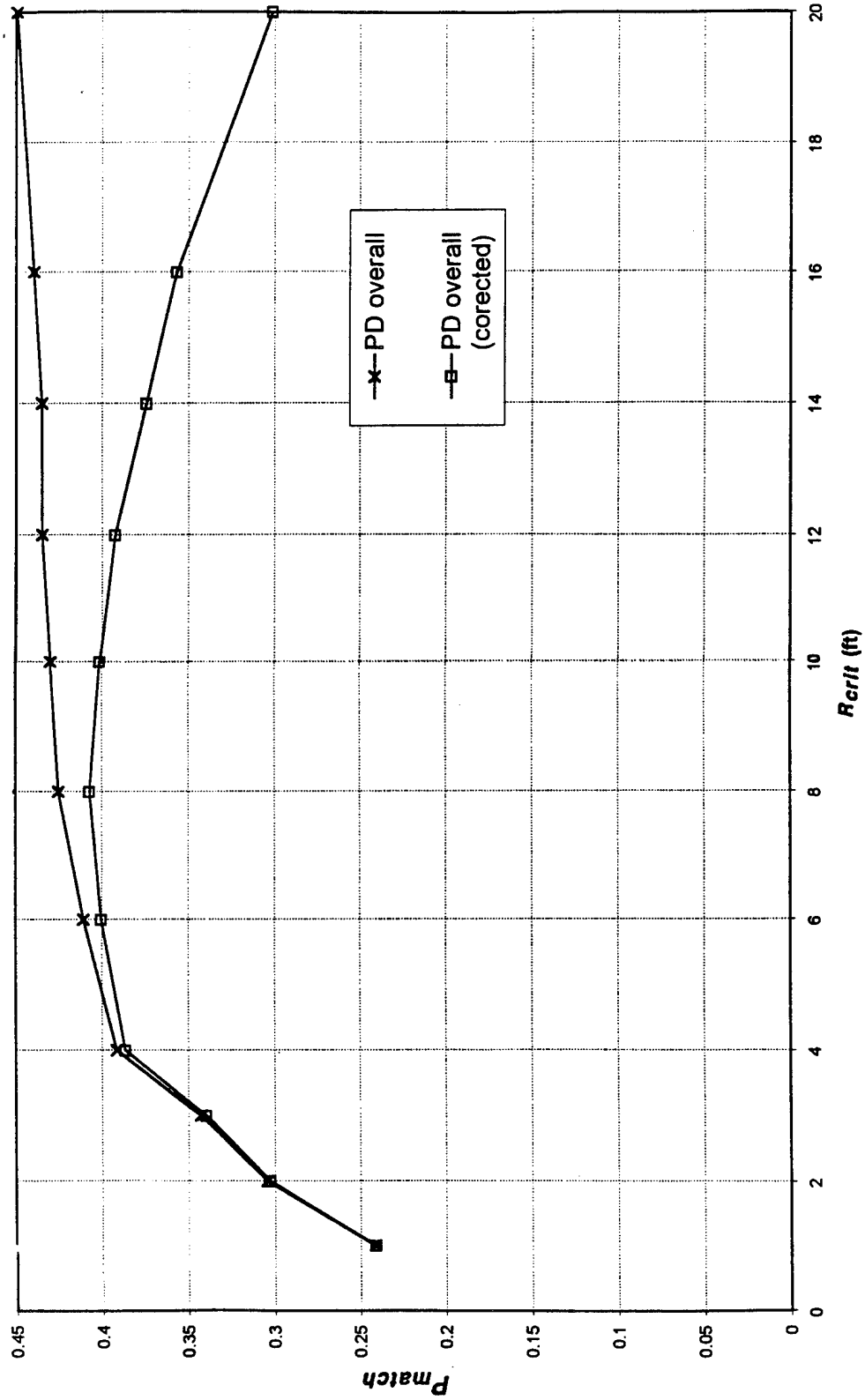


Figure B.2. R_{match} vs. R_{crit} Uncorrected and Corrected for Random Matches, for a Sample Demonstrator

For three of the four cases, s is between 2 and 3 feet, while the fourth demonstrator has a noticeably larger value for s of 4.6 ft. The error function term for R_{crit} = 2 m (6.56 ft) is close to unity for all four cases, which sample the range of location accuracy among demonstrators that detected a significant amount of ordnance.

All demonstrators experienced what we consider a high number of false alarms, compared to both the expectation prior to the demonstration and the number that can be tolerated if a cleanup is to be conducted based on demonstrator declarations. In fact, most demonstrators reported multiple false alarms per ordnance item detected. However, with a critical radius of 2 m, the number of false alarms did not cover a fraction of the site that was large enough to make the corrections for random matches cause a significant decrease in P_{match} for most demonstrators. Only two had enough lucky matches to decrease P_{match} by more than three percentage points. For those two demonstrators, false alarms were extremely dense and covered 23 percent and 16 percent of the area of the site searched for a 2 m R_{crit}.

APPENDIX C

TESTS FOR ACCURACY OF ASSESSMENT

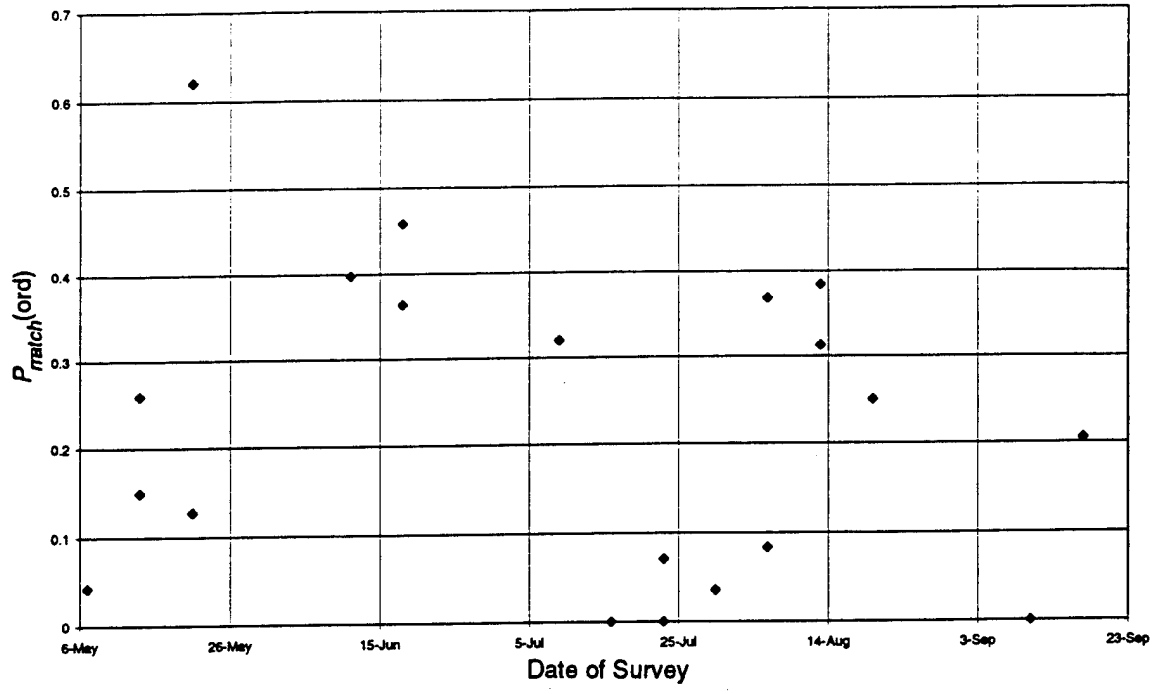
APPENDIX C
TESTS FOR ACCURACY OF ASSESSMENT

The discrepancy between expected and observed performance raised concerns about the accuracy of the assessment of demonstrator performance. The large number of demonstrators that showed detection capabilities that were statistically indistinguishable from placing the same number of declarations at random on the site was particularly worrisome. We performed the following analyses in an attempt to determine if the poor performance could be attributed to some factor associated with the demonstration or the scoring procedure, as opposed to poor sensor performance.

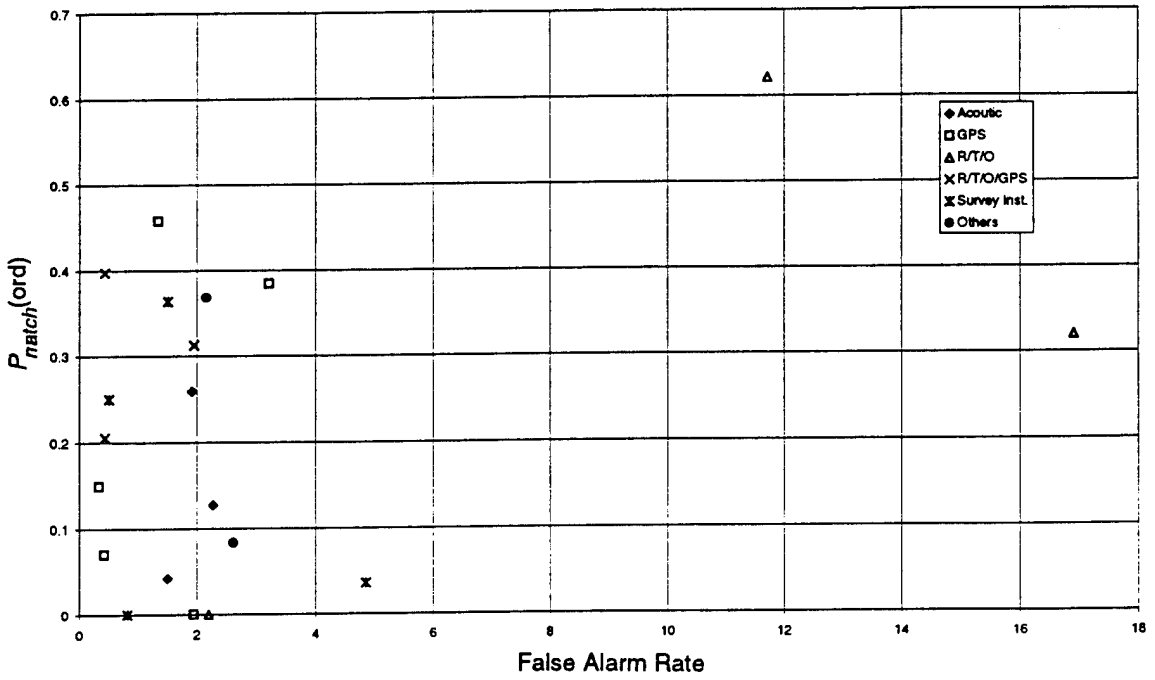
Correlation between performance and demonstration date. Over the course of the JPG demonstrations, the vegetation was permitted to grow, the weather and soil conditions changed, and some of the marker stakes were inadvertently moved. To assess whether these or similar factors had an impact on performance, P_{match} versus date of demonstration was considered. Figure C.1(a) shows that the number of poor performers was fairly consistent throughout. Although there appears to be a downward trend in the performance of the better demonstrators, when the data are fit to a linear model, the slope is found to be statistically insignificant.

Correlation between performance and navigation method. Figure C.1(b) shows no navigation scheme was associated with consistently better or worse performance.

Correlation between declared targets of the various "poor" demonstrators. A common difficulty among demonstrators caused by test artifacts might produce a correlation between the apparent false alarms of the demonstrators that performed poorly. For example, a misplaced survey stake marking one of the grid cell locations would cause all demonstrator target declarations surveyed from that stake to be systematically displaced from their correct locations. The target sets of several of the below-average demonstrators were compared. No such correlation was found.



(a) P_{match}(ord) vs. Date of Survey



(b) P_{match}(ord) vs. False Alarm Rate by Navigation System

Figure C.1. Performance of Ground-Based Detection Systems

Correlation between targets found by above-average demonstrators and targets found by below-average demonstrators. Of the nine demonstrators with $P_{match} < 0.20$, three found no targets and two found only one, probably fortuitously. These demonstrators, all of which used ground-penetrating radars, were disregarded for this analysis. For the other below-average demonstrators, the number of targets identified by one demonstrator that were found by six or more other demonstrators are shown in the table below.

Comparison of Targets Detected by Poor Performing Demonstrators

Demonstrator	*Technology*	*Total Targets*	*Subset of Total Targets Found by Six or More Demonstrators*
A	Magnetometer	12	9 (75%)
B	GPR/IC	11	4 (36%)
C	Magnetometer	27	17 (63%)
D	Magnetometer	10	6 (60%)

About 20 percent of all the emplaced targets were detected by six or more of the demonstrators. It is believed that these items constitute the set of "easy-to-find" targets. The demonstrators that performed poorly can be separated into two groups. For demonstrators A, C and D, the targets detected are concentrated among the frequently found target set. In this group, one has some confidence that the demonstrators' poor performance resulted from insufficient sensitivity (i.e., a threshold that was low enough only to find the easy targets), rather than from some other cause, such as poor navigation. This conclusion can not be drawn for demonstrator B, whose few detections did not fall into this frequently found category. It is difficult to assess this demonstrator, which used a GPR/IC detector. Since most of the demonstrators that showed significant detection capabilities used magnetometers, the "easy" target set does not necessarily contain targets that are easily detected by a GPR/IC system. Therefore, the absence of correlation is not sufficient to conclude that all detections by these demonstrators were random.

APPENDIX D

COMPARISON OF THE AVERAGE RADIAL ACCURACY FOR RANDOM MATCHES AGAINST DEMONSTRATED RADIAL ACCURACY

APPENDIX D
COMPARISON OF THE AVERAGE RADIAL ACCURACY FOR RANDOM MATCHES AGAINST DEMONSTRATED RADIAL ACCURACY

For the airborne demonstrations, corrections for random matches indicate that the detection capability is indistinguishable from zero. To further investigate this point, we performed calculations based on the following hypothesis.

Hypothesis: If the matches are randomly distributed within R_{crit}, then the average location accuracy would be 2/3 R_{crit}.

2/3 R_{crit} is shown in the table below for comparison against the demonstrator location accuracies. The location accuracies of the demonstrators are not significantly different from the radial accuracy that would be expected given random matches.

Location Accuracy[a] for Airborne Demonstrators Compared to 2/3(R_{crit})

	R_{crit}		
	2	5	10
	2/3 R_{crit}		
	1.34	3.32	6.67
AES	1.83	2.34	5.61
Battelle	–	–	8.10
Geonex	–	3.14	5.30
Oilton	1.34	2.68	6.21
SRI-fixed wing	1.40	3.14	6.31
SRI-rotary wing	0.91	3.39	4.70

[a] All dimensions are in meters. Dashes indicate that no emplaced items were detected at this value of R_{crit}.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1995	Final—March 1995–October 1995

4. TITLE AND SUBTITLE
Demonstrator Performance at the Unexploded Ordnance Advanced Technology Demonstration at Jefferson Proving Ground (Phase I) and Implications for UXO Clearance

5. FUNDING NUMBERS
DASW01 94 C 0054
T-AM2-1282

6. AUTHOR(S)
Thomas W. Altshuler, Anne M. Andrews, Regina E. Dugan, Vivian George, Michael P. Mulqueen, David A. Sparrow

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311-1772

8. PERFORMING ORGANIZATION REPORT NUMBER
IDA Paper P-3114

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
COL W. Richard Wright
Chairman, DoD Explosives Safety Board
Hoffman 1, Rm 856 C
Alexandria, VA 22331–0600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 180 words)***

This study describes the performance of detection systems at the 1994 Unexploded Ordnance (UXO) Detection, Identification, and Remediation Advanced Technology Demonstration at Jefferson Proving Ground (JPG), Indiana. Ground-based detection systems demonstrated included magnetometers, stand-alone ground-penetrating radars (GPRs), and induction coil metal detectors paired with GPRs. Airborne detection systems demonstrated included GPRs, infrared imaging systems, and a magnetometer paired with an induction coil. For ground-based systems, magnetometers exhibited, on average, the best performance. Stand-alone GPRs were not successful at detecting buried ordnance. Induction coil metal detectors paired with GPRs displayed moderate detection capabilities. For all airborne systems, the detection capability was statistically indistinguishable from zero; i.e., there is no evidence that any of the systems detected any ordnance. The best demonstrators had detection capabilities of approximately 65 percent with multiple false alarms per detection. No demonstrator was able to distinguish ordnance from nonordnance or debris. Despite caveats regarding the extrapolation of performance demonstrated at JPG to other sites, the demonstration provided an opportunity to compare many systems in a blind test, and the results provide useful guidance for future R&D efforts.

14. SUBJECT TERMS
ground-penetrating radar, GPR, induction coil, infrared, Jefferson Proving Ground, JPG, magnetometer, unexploded ordnance, UXO, munitions, base realignment and closure, BRAC

15. NUMBER OF PAGES
145

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR