FINAL REPORT

Feature Extraction and Classification of Magnetic and EMI Data, San Luis Obispo, CA

ESTCP Project MM-0504

JULY 2010

Dr. Stephen Billings
Dr. Len Pasion
Dr. Nicolas Lhomme
Kevin Kingdon
Jon Jacobson
Sky Research, Inc.

Dr. Douglas Oldenburg
Dr. Lin Ping Song
Laurens Beran
UBC-GIF

Approved for public release; distribution unlimited.

Feature Extraction and Classification of Magnetic and EMI Data, San Luis Obispo, CA

ESTCP MM-0504 Practical Discrimination Strategies for Application to Live Sites

Sky Research, Inc.

July 19, 2010
Version 2.0

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
19-07-2010	Draft Demonstration Report V2	August 2009-July 2010

4. TITLE AND SUBTITLE

Draft 2 Demonstration Report
Feature Extraction and Classification of Magnetic and EMI Data
San Luis Obispo, CA
MM-0504: Practical Discrimination Strategies for
Application to Live Sites

5a. CONTRACT NUMBER
W912HQ-05-C-0018

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Stephen Billings, Dr. Len Pasion, Dr. Nicolas Lhomme, Kevin Kingdon, Jon Jacobson (Sky Research, Inc.)
Dr. Douglas Oldenburg, Dr. Lin Ping Song, Laurens Beran (UBC-GIF)

5d. PROJECT NUMBER
ESTCP 0504

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Sky Research, Inc.
445 Dead Indian Memorial Road
Ashland, OR 97520

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
ESTCP Program Office
Attn: Dr. Herb Nelson
901 North Stuart St.
Suite 303
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

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

12. DISTRIBUTION/AVAILABILITY STATEMENT

13. SUPPLEMENTARY NOTES

14. ABSTRACT

An Unexploded Ordnance discrimination study was conducted at the Former Camp San Luis Obispo (SLO), California. The objective was to discriminate potentially hazardous 2.36” rockets and 60 mm, 81 mm and 4.2” caliber mortars from non-hazardous shrapnel, range and cultural debris. In this report, we describe the performance of twelve different discrimination techniques that utilized data from a number of sensors deployed in full coverage and cued interrogation mode. Two important conclusions can be drawn from the results presented here. Firstly, by appropriate use of discrimination metrics applied to production quality EM-61 data, it is possible to significantly reduce the number of clutter items excavated without missing any targets of interest. Secondly, the next generation of EM sensors, when deployed in a cued-interrogation mode, result in significant additional reductions in the number of clutter items excavated. Furthermore, the next generation sensors can usually distinguish different UXO types from one another.

15. SUBJECT TERMS
UXO, Discrimination, Classification, Magnetics, Electromagnetics, Feature Extraction

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Herb Nelson
UU	UU	UU	UU	181	19b. TELEPHONE NUMBER *(Include area code)* 703-696-8726

Reset

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

EXECUTIVE SUMMARY

The demonstration described in this report was conducted at the Former Camp San Luis Obispo (SLO), California, under project Environmental Security Technology Certification Program (ESTCP) MM-0504 "Practical Discrimination Strategies for Application to Live Sites." It was performed under the umbrella of the ESTCP Discrimination Study Pilot Program. The MM-0504 project is attempting to demonstrate the application of feature extraction and statistical classification to the problem of unexploded ordnance (UXO) discrimination. At the SLO site the objective was to discriminate potentially hazardous 2.36" rockets and 60 millimeter (mm), 81 mm and 4.2" caliber mortars from non-hazardous shrapnel, range and cultural debris. In this report, we describe the performance of twelve different discrimination techniques that utilized data from a number of sensors deployed in full coverage (Multi-sensor towed array detection system [MTADS] magnetometer and EM61 arrays, Geonics EM61 cart and Man-Portable Simultaneous EMI and Magnetometer System [MSEMS] cart) and cued interrogation mode (Time-domain electromagnetic towed array detection system [TEMTADS], MetalMapper and Berkeley UXO Discriminator [BUD]). In the blind-test data a total of 209 ordnance items were found including seventy-six 60 mm mortars (many without fins and/or nose-cone), twenty 2.36" rockets, fifty-nine 81 mm mortars, fifty-one 4.2" mortars and one each of 37 mm, 3" and 5" projectiles.

Each of the discrimination techniques utilized features extracted from a phenomenological model that was fit to the observed data around each anomaly. For magnetics the model was a static dipole, while for electromagnetic (EM) a polarization tensor model was used. From the extracted feature vectors the following twelve different prioritized dig-lists were created: (i) Magnetics array ranked by size of dipole moment; (ii) EM61 cart data ranked by time-decay; (iii) MSEMS cart ranked by time-decay; (iv) MTADS EM61 array ranked by time-decay; (v) BUD statistical classification using size and time-decay parameters; (vi) TEMTADS statistical classification using size and time-decay parameters; (vii) MetalMapper statistical classification using size and time-decay parameters; (viii) TEMTADS ranked by fit to library; (ix) MetalMapper ranked by fit to library; (x) TEMTADS ranked by expert opinion; (xi) MetalMapper ranked by expert opinion; and (xii) TEMTADS ranked by match to polarizations in library. All model fits and discrimination analyses were performed using the UXOLab software that was jointly developed by the University of British Columbia – Geophysical Inversion Facility (UBC-GIF) and Sky Research, principally through funding from the United States Army Corps of Engineers Environmental Research and Development Center (USACE-ERDC) and the ESTCP program.

Magnetometer detection and discrimination performance at this site was quite poor. Dig-sheet ranking was based on the size of the dipole moment and required 460 excavations at the operating point (OP) where 196 of 200 targets of interest (TOI) were recovered. To excavate the final four TOI would have required digging most of the remaining non-hazardous items.

The EM-61 production datasets were much more effective than magnetics. Size estimated from the recovered polarizations was not an effective discrimination metric due to the small-size of the 60 mm mortars and the inability to accurately constrain depth. However, the time-decay rate estimated from the recovered polarizabilities provided an effective ranking scheme. The EM61 cart performance was marginally better than the MSEMS cart and MTADS EM61 array. At the

operating point 586 excavations were required and all 208 detected TOI were recovered, along with 378 of 1068 non-hazardous items (35% of the clutter). The TOI included a single 37 mm projectile that was found on the site. The BUD instrument was only deployed to a subset of anomalies detected at the site and performance was only marginally better than the EM61 cart. Digsheet ranking was based on a Probabilistic Neural Network (PNN) classifier applied to a feature space comprising size and time-decay features estimated from the recovered polarizabilities. At the operating point, 139 excavations were required and 58 TOI were recovered along with 81 of 414 non-TOI (19.6% of the clutter). The one false-negative was Master ID 241: a non-hazardous collection of rocket motor pieces that was declared non-TOI by all cued-interrogation methods.

A total of 1282 items were included in the blind-test data for the TEMTADS, with 206 targets of interest (data were not collected over the 37 mm projectile). Four different methods for dig-sheet ranking were used: (i) Statistical classification applied to a 2-D feature space comprising a size and a time-decay feature; (ii) Library method based on comparing the unconstrained polarization tensor fits to polarization tensors constrained by a library of expected ordnance items; (iii) Expert Opinion where initial ranking was based on the statistical classification method, but an "expert" analyst manually removed items in the TOI list that were thought to be non-TOI; and (iv) Polarization Match based on the match between the recovered polarization tensor and pre-stored polarizations representing the expected ordnance types. The Library method was the most effective with 204 of 206 TOI recovered along with 131 of 1076 non-TOI (12.2% of clutter). The two false negatives were the rocket motor pieces (Master ID 241) declared non-TOI by all cued-interrogation methods and a 60 mm mortar with a target response that overlapped with some nearby clutter. The other three methods were also effective although generated between 2 to 4 false-negatives (including Master ID 241). The Expert Opinion method resulted in a significant reduction in the number of non-TOI excavated (from 137 down to 81) but did result in the misclassification of one 60 mm mortar in a multi-object configuration. Ordnance type was predicted by three of the methods. The correct ordnance type was predicted in 179 of 199 cases (90% success rate) for the statistical classification method, for 196 of 200 cases (98% success rate) for the Library method and 185 of 189 cases (98% success rate) for the Expert opinion. All methods had 100% success rate on the 4.2" mortars and only the statistical classifier couldn't achieve 100% success with the 60 mm mortars. The 2.36" rockets and 81 mm mortars were more difficult to distinguish and were occasionally incorrectly assigned to the wrong ordnance type.

A total of 1409 items were included in the blind-test data for the MetalMapper, with 204 targets of interest (including the 37 mm projectile). Three different methods of dig-sheet ranking were used with each very similar to the corresponding method used for TEMTADS: (i) Statistical Classifier; (ii) Library Method; and (iii) Expert Opinion. The Library method was the most effective (after correcting an initial coding mistake with the excavation of 203 or 204 TOI and 175 of 1205 non-TOI (14.5% of clutter). The Expert Opinion again significantly reduced the number of non-TOI excavated (from 166 down to 57) but resulted in a false negative on a 60 mm mortar in a multi-object configuration. Ordnance type was predicted by two of the methods. The correct ordnance type was predicted in 198 of 200 cases (99% success rate) for the statistical classification method, and 184 of 200 cases (92% success rate) for the Library method.

There are two important conclusions from the results presented here. Firstly, by appropriate use of discrimination metrics applied to production quality EM-61 data, it is possible to significantly reduce the number of clutter items excavated without missing any targets of interest. Secondly,

the next generation of EM sensors, when deployed in a cued-interrogation mode, result in significant additional reductions in the number of clutter items excavated. Furthermore, the next generation sensors can usually distinguish different UXO types from one another.

TABLE OF CONTENTS

LIST OF TABLES vi

LIST OF FIGURES vi

ACRONYMS x

1.0 INTRODUCTION 1
1.1 BACKGROUND 1
1.2 OBJECTIVE OF THE DEMONSTRATION 1
1.3 REGULATORY DRIVERS 3
2.0 TECHNOLOGY 4
2.1 TECHNOLOGY DESCRIPTION 4
2.1.1 Creation of a Map of Geophysical Sensor Data 5
2.1.2 Anomaly Selection and Feature Extraction 5
2.1.3 Classification of Anomalies 7
2.1.4 UXOLab Software 9
2.2 PREVIOUS TESTING OF THE TECHNOLOGY 9
2.3 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY 12
3.0 PERFORMANCE OBJECTIVES 13
3.1 OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF MUNITIONS 13
3.1.1 Metric 13
3.1.2 Data Requirements 13
3.1.3 Success Criteria 13
3.2 OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF NON-MUNITIONS 16
3.2.1 Metric 16
3.2.2 Data Requirements 16
3.2.3 Success Criteria 16
3.3 OBJECTIVE: SPECIFICATION OF NO-DIG THRESHOLD 16
3.3.1 Metric 16
3.3.2 Data Requirements 17
3.3.3 Success Criteria 17
3.4 OBJECTIVE: MINIMIZE NUMBER OF ANOMALIES THAT CANNOT BE ANALYZED 17
3.4.1 Metric 17
3.4.2 Data Requirements 17
3.4.3 Success Criteria 17
3.5 OBJECTIVE: CORRECT ESTIMATION OF TARGET PARAMETERS 17
3.5.1 Metric 18
3.5.2 Data Requirements 18
3.5.3 Success Criteria 18
4.0 SITE DESCRIPTION 19
4.1 SITE SELECTION 19
4.2 SITE HISTORY 19
4.3 SITE GEOLOGY 19

4.4 MUNITIONS CONTAMINATION 19
5.0 TEST DESIGN 20
6.0 DATA ANALYSIS AND PRODUCTS 21
6.1 PREPROCESSING 21
6.2 TARGET SELECTION FOR DETECTION 21
6.3 PARAMETER ESTIMATION 21
6.3.1 Data Covariance Matrix V_d 21
6.3.2 Forming the Data Vector d_{obs} 23
6.3.3 Determining the Number of Objects 23
6.3.4 Procedure for Single-Object Inversion 24
6.3.5 Approach for Dealing with Multiple-Objects 26
6.3.6 Quality Control Procedures 29
6.3.7 Magnetics feature extraction procedure 29
6.3.8 Sensor Specific Considerations 30
6.4 TRAINING 31
6.5 CLASSIFICATION 31
6.5.1 Modified TEMTADS strategy 31
7.0 PERFORMANCE ASSESSMENT 35
7.1 ADDITIONAL ANALYSIS OF PRODUCTION DATASETS 41
7.2 ADDITIONAL ANALYSIS OF TEMTADS DATASETS 56
7.2.1 Classification using asymmetry 67
7.2.2 Classification using the moments of the polarizabilities 68
7.2.3 Automatic feature selection 70
7.2.4 Identifying TOI type 73
7.3 ADDITIONAL ANALYSIS OF METALMAPPER DATASETS 74
7.3.1 More aggressive discrimination strategies 79
7.3.2 Modified MetalMapper classifier 82
7.3.3 MetalMapper Library Method after correcting a coding mistake 85
7.4 PERFORMANCE OF THE BUD 86
7.5 SUMMARY OF EACH METHOD 89
7.5.1 MTADS magnetometer array: size-based ranking 89
7.5.2 EM61 cart: time-decay ranking 89
7.5.3 MSEMS cart: time-decay ranking 89
7.5.4 MTADS EM61 array: time-decay ranking 89
7.5.5 TEMTADS cued-interrogation 89
7.5.6 MetalMapper cued interrogation: 90
7.5.7 BUD cued interrogation 91
8.0 COST ASSESSMENT 92
9.0 MANAGEMENT AND STAFFING 94
10.0 REFERENCES 95

LIST OF TABLES

Table 1. Previous Inversion/Classification Testing 10
Table 2. Performance Objectives for this Demonstration (Performance was assessed both including and excluding Master ID 241) 14
Table 3. Summary of classification results at SLO. If Master ID 241 was in-fact a non-TOI then the number of false-negatives (FN) in each of the cued-methods would be reduced by 1. The "TEMTADS Statistical (adjusted)" method has Master ID 16 moved to the "can't decide" class (this was a QC mistake). The "MetalMapper Library (adjusted)" method fixes a coding mistake made in the original submission. 39
Table 4. List of false negatives encountered in each dig-list (not including the MetalMapper Library method where coding mistake resulted in several false-negatives that should have been identified as TOI). 40
Table 5. Position and depth errors of recovered polarizabilities. 41
Table 6. Summary of discrimination performance for TEMTADS classifiers. 58
Table 7. Ability of TEMTADS Expert Interpretation method to correctly label ordnance type. The table includes a column that shows the number of items that were classified as high-confidence UXO by the analyst. 74
Table 8. Ability of TEMTADS Library method to correctly label ordnance type. 74
Table 9. Ability of TEMTADS Statistical Classification method to correctly label ordnance type. 74
Table 10. Ability of MetalMapper Statistical method to correctly label ordnance type. 85
Table 11. Ability of MetalMapper Library method to correctly label ordnance type. 86
Table 12. Cost Breakdown for the San Luis Obispo Discrimination Study. 93

LIST OF FIGURES

Figure 1. A framework for statistical pattern recognition. 8
Figure 2. Nonparametric density estimate using Gaussian kernels. Kernel centers are shown as crosses. A large kernel width produces a smooth distribution (left) compared to a small kernel width (right). 9
Figure 3. SVM formulation for constructing a decision boundary. The decision boundary bisects support planes bounding the classes. 9
Figure 4: Flowchart showing the procedure for parameter extraction. 22
Figure 5. The depth-misfit relationship, an indirect indicator of the of the depth-size ambiguity for a buried object. 25
Figure 6. Format for the prioritized anomaly list that will be submitted by each classification demonstrator. 31
Figure 7. Size and decay features for training and test data. Concentric ellipses show 1,2, and 3 standard deviation ellipses for each TOI class, estimated from the training data alone. Test feature vectors connected by a line indicate multi-target inversions. 32
Figure 8. Size and decay features for training and test data. Concentric ellipses show 1,2, and 3 standard deviation ellipses for each TOI class, estimated from the training and test data sets using expectation maximization (EM) algorithm. 33

Figure 9. Decision statistic for TEMTADS classification. Contour shows the decision boundary and features vectors in red indicate test feature vectors flagged for digging..... 34
Figure 10. ROC curves provided by IDA for the production datasets. Each of these ROC curves includes both the test and the training data..... 35
Figure 11. ROC curves provided by IDA for the TEMTADS datasets. Each of these ROC curves includes both the test and the training data..... 36
Figure 12. ROC curves provided by IDA for the MetalMapper and BUD datasets. Each of these ROC curves includes both the test and the training data. 37
Figure 13. Location and depth estimation for EM61 cart, MSEMS and MTADS EM61 data sets. For depth estimates, dashed lines represent errors of +/- 15 cm..... 42
Figure 13. (cont'd) Location and depth estimation for EM61 cart, MSEMS and MTADS EM61 data sets. For depth estimates, dashed lines represent errors of +/- 15 cm..... 43
Figure 14. Training and test sets for EM61 cart, MSEMS, and MTADS EM61 data, see legend in Figure 13..... 44
Figure 14. (cont'd) Training and test sets for EM61 cart, MSEMS, and MTADS EM61 data, see legend in Figure 13. 45
Figure 15. ROC curves for TOI classes, EM61 cart, MSEMS, MTADS EM61 and magnetics. Circles indicate operating points..... 46
Figure 15. (cont'd) ROC curves for TOI classes, EM61 cart, MSEMS, MTADS EM61 and magnetics. Circles indicate operating points. 47
Figure 16. Master ID 775 which comprises a 60 mm mortar body in the bottom of the hole and two tail fins. 47
Figure 17. Misfit versus depth curves for inversions of Master ID 775..... 48
Figure 18. Comparison of fits over Master ID 775 (60 mm mortar). Line profile is black line in gridded images, showing observed and predicted data at channels 1 and 4..... 49
Figure 19. Comparison of fits for Master ID 775. Left: original fit, Right: retrospective fit with tight mask. Plotted line profile is adjacent to the line profile shown in figure 6(b). 50
Figure 20. Change in estimated features for Master ID 775 after retrospective inversion with a tight mask..... 51
Figure 21. Master ID 1444, a deep 60mm body with unfavorable orientation..... 53
Figure 22. Comparison of fits over Master ID 1444 (60mm mortar). Line profile is black line in gridded images, showing observed and predicted data at channels 1 and 4..... 54
Figure 22. (cont'd) Comparison of fits over Master ID 1444 (60mm mortar). Line profile is black line in gridded images, showing observed and predicted data at channels 1 and 4..... 54
Figure 23. Comparison of model and data based decay features for MTADS test data..... 56
Figure 24. Analysis of TEMTADS data. Feature vectors connected by a line are multi-target inversions..... 57
Figure 25. Receiver operating curves for TOI classes, TEMTADS classifiers. Circles are specified operating points, diamonds show the location of Master ID 241 (2.36" rocket) in the diglists..... 59
Figure 26. Master ID 241..... 60
Figure 27. Legend for TEMTADS results plots. 60
Figure 28. Fit to TEMTADS data for Master ID 103 (60 mm). Legend in Figure 27 explains layout of the inversion results plot..... 61
Figure 29. Median SNR versus estimated decay rate of primary polarization for TEMTADS test data. Low SNR targets are circled. 62

Figure 30. Fits to estimated polarizations (dots) for two targets. Dashed line is unweighted fit, solid line is weighted fit. Vertical line indicates the channel used to compute the polarization decay parameter. 63
Figure 31. Comparison of primary polarization decay rate estimates without noise floor (left) and with noise floor (right). 64
Figure 32. TEMTADS test data with features from retrospective multi-target inversion of Master ID 1285. 65
Figure 33. Left: retrospective PNN decision surface with TEMTADS test data. Training feature vectors are outlined in black. Right: ROC curves for library method, statistical classification, and retrospective PNN. Markers indicate the point on each ROC curve at which the last TOI is identified (Master ID 241 is not considered a TOI in this ROC analysis). 66
Figure 34. Left: retrospective PNN decision surface with TEMTADS test data. Training feature vectors are outlined in black. Right: ROCs for library method, statistical classification, and retrospective PNNs. Markers indicate the point on each ROC at which the last TOI is identified. 67
Figure 35. Moments of TEMTADS primary polarization. Left: second versus zeroth moments. Right: moment ratio (second divided by zeroth), versus zeroth moment. 69
Figure 36. Retrospective PNN classifiers applied to moment-derived features. “primary moments” denotes a PNN trained on zeroth through second moments of the primary polarization, “all moments” is a PNN trained on zeroth through second moments of all polarizations (a nine-dimensional feature space), “moment 0 moment ratio” is a PNN trained on primary zeroth moment and moment ratio. 70
Figure 37. Smoothed polarizabilities for TEMTADS training data. Shaded rectangles indicate 20 time channels selected with recursive feature elimination. 72
Figure 38. Retrospective analysis of TEMTADS test data using nonlinear SVM in a feature space identified with RFE. 72
Figure 39. Receiver operating curves for TOI classes, MetalMapper classifiers. Open circles are specified operating points. Master ID 241 was not labeled as a TOI when generating these figures. 75
Figure 40. Training (a) and test-data (b) for the MetalMapper statistical classification method, along with a plot of recovered versus actual depths (c). 77
Figure 41. Polarization recovered over Master ID 775 using MetalMapper data inverted as a single object and as a multi-object scenario. 78
Figure 42. Data fit and polarizabilities extracted from Master ID 1315. Left data plots show results for z-axis excitation, middle for x-axis and right for y-axis. Blue is z-axis receiver, green is x-axis and red is y-axis. Solid line is fitted model, dots are observed data. 79
Figure 43. Asymmetry versus size feature space for the MetalMapper. The top-rows uses I23/I12 as the measure of asymmetry while the bottom row uses I23/I1. See the legend in Figure 40 for an explanation of symbols. 81
Figure 44. Plot of misfit versus size for MetalMapper feature vectors. 82
Figure 45. Expectation Minimization classifier applied to the MetalMapper data: (a) training data; and (b) test-data. The white-line is a contour of the decision surface. 83
Figure 46. MetalMapper ROC curves for original PNN classifier and a revised classifier. 84
Figure 47. Part of the MetalMapper ROC curve for the submitted library and expert opinion methods, along with a corrected library based method. The blue dot shows the operating point for the original library method. 85

Figure 48. Feature vectors and classification boundary for BUD training (a) and test (b) data… 87
Figure 49. ROC curve for the BUD compared to the EM61 deployed over the same area..... 88
Figure 50. Project management hierarchy showing Sky Research personnel in blue and UBC-GIF personnel in green. The hierarchy is split between the development and execution components. 94

ACRONYMS

AIC	Akaike Information theoretic Criterion
APG	Aberdeen Proving Ground
AUC	Area Under the Curve
BOR	Body-of-revolution
BUD	Berkeley UXO Discriminator
cm	centimeter
d_{obs}	Data Vector
dB	Decibel
EM	Electromagnetic
EMI	Electromagnetic induction
ESTCP	Environmental Security Technology Certification Program
FAR	False Alarm Rate
FLBGR	Former Lowry Bombing and Gunnery Range
FN	False Negative
FOM	Figures of Merit
FY	Fiscal Year
GPS	Global Positioning System
IDA	Institute for Defense Analyses
IMU	Inertial Measurement Unit
in.	Inches
ITC	Information Theoretic Criteria
m	Meter
mm	Millimeter
M_d	Mahalanobis Distribution
MDL	Minimum Description Length
MPT	magnetic polarizability tensors
MRS	Munitions Response Site
ms	Millisececond
MSEMS	Man-Portable Simultaneous EMI and Magnetometer System
MTADS	Multi-Sensor Towed Array Detection System
mV	MilliVolt
N_{FA}	Number of False Alarms
nT	NanoTeslas
OP	Operating Point
P_{class}	Probability of Classification
pdf	Probability Density Function
PI	Principal Investigator
PNN	Probabilistic Neural Network
POC	Point of Contact
QC	Quality Control
RF	Range Fan
RFE	Recursive Feature Elimination
ROC	Receiver Operating Characteristic

ROI	region of interest
RR	Rocket Range
RTS	Robotic Total Station
SKY	Sky Research, Inc.
SLO	San Luis Obispo
SNR	Signal to noise ratio
STOLS	Surface Towed Ordnance Locator System
SVM	Support vector machine
TEM	Time-domain electromagnetic
TEMTADS	Time Domain Electromagnetic Towed Array Detection System
TOI	Target of Interest
UBC-GIF	University of British Columbia – Geophysical Inversion Facility
USACE ERDC	United States Army Corps of Engineers- Engineering Research and Development Center
UXO	Unexploded ordnance
V_d	Data Covariance Matrix
YPG	Yuma Proving Ground

1.0 INTRODUCTION

1.1 BACKGROUND

The Fiscal Year (FY) 06 Defense Appropriation contains funding for the "Development of Advanced, Sophisticated, Discrimination Technologies for UXO Cleanup" in the Environmental Security Technology Certification Program (ESTCP). In 2003, the Defense Science Board observed: "The … problem is that instruments that can detect the buried unexploded ordnance (UXO) also detect numerous scrap metal objects and other artifacts, which leads to an enormous amount of expensive digging. Typically 100 holes may be dug before a real UXO is unearthed! The Task Force assessment is that much of this wasteful digging can be eliminated by the use of more advanced technology instruments that exploit modern digital processing and advanced multi-mode sensors to achieve an improved level of discrimination of scrap from UXO."

Significant progress has been made in discrimination technology. To date, testing of these approaches has been primarily limited to test sites with only limited application at live sites. Acceptance of discrimination technologies requires demonstration of system capabilities at real UXO sites under real world conditions. Any attempt to declare detected anomalies to be harmless and requiring no further investigation will require demonstration to regulators of not only individual technologies, but an entire decision making process.

The FY06 Defense Appropriation contained funding for the "Development of Advanced, Sophisticated, Discrimination Technologies for UXO Cleanup" for ESTCP. ESTCP responded by conducting a UXO Classification Study at the former Camp Sibert, Alabama. The results of this first demonstration were very encouraging. Although conditions were favorable at this site, including a single target-of-interest (4.2-inch [in] mortar) and benign topography and geology, all of the demonstrated classification approaches were able to correctly identify a sizable fraction of the anomalies as arising from non-hazardous items that could be safely left in the ground. Of particular note, the contractor EM-61-MK2 cart survey with analysis using commercially available methods correctly identified more than half the targets as non-hazardous.

To build upon the success of the first phase of this study, ESTCP expanded the program to include a second study at a site with more challenging topography and a wider mix of targets-of-interest. A range at the former Camp San Luis Obispo (SLO), California, was selected for this demonstration. This demonstration report describes the data processing, feature extraction and classification that were conducted by Sky Research (SKY) and the University of British Columbia (UBC) at SLO.

1.2 OBJECTIVE OF THE DEMONSTRATION

The objectives of this demonstration were to perform data modeling, classification, and discrimination using magnetometer and electromagnetic (EM) data collected by the various data collection demonstrators participating in the study. Specifically, we processed the following datasets collected at SLO:

1) Multi-Sensor Towed Array Detection System (MTADS) magnetometer data;
2) MTADS EM-61 array data;
3) EM-61 cart data;

4) MTADS EM-61 array cooperative inversion with magnetics data;
5) Man-Portable Simultaneous EMI and Magnetometer System (MSEMS) which consists of an EM-61 and magnetometer mounted on a cart;
6) Time Domain Electromagnetic Towed Array Detection System (TEMTADS) cued interrogation array data;
7) Data collected using the Berkeley UXO Discriminator (BUD) deployed in a cued interrogation mode; and
8) MetalMapper cued interrogation data.

Specific processing tasks were as follows:

1) Data modeling:
    a) Dipole fitting of the magnetometer data;
    b) Fitting of 3-dipole beta models to the EM-61 cart, MTADS EM-61 and MSEMS EM61 detection mode data and the TEMTADS, MetalMapper and BUD cued interrogation data; and
    c) We had planned to do cooperative inversion of the EM-61/magnetometer dual-mode and MTADS EM-61 array (both using 3-dipole beta models) using the dipole fits from the magnetometer data to constrain the object's location and depth. However, on inspecting the results on the training data, we decided that the cooperative inversion did not provide an advantage over the unconstrained inversion.
2) Classification and discrimination:
    a) Magnetics size-based: Production of a dig sheet ranked according to size (magnitude of the dipole moment);
    b) MTADS EM-61 statistical: Statistical classification of features derived from the MTADS EM-61 data and the production of a ranked dig sheet;
    c) Cart EM-61 statistical: Same as b) but with features from EM-61 cart-data;
    d) TEMTADS cued interrogation statistical: Same as b) but with the polarizabilities from the TEMTADS array;
    e) BUD statistical: As per b) but with features derived from the BUD; and
    f) MetalMapper statistical: As per b) but with the polarizabilities derived from the MetalMapper data.
    g) TEMTADS library method: We provided an alternative ranking of the TEMTADS based on a library method;
    h) MetalMapper library: As in g) but for the MetalMapper.
    i) TEMTADS "expert" opinion: A third digsheet for TEMTADS was produced based on expert opinion.
    j) MetalMapper "expert" opinion: As in i) but for the MetalMapper.

k) TEMTADS polarization match: A forth dig-sheet was produced for the TEMTADS based on how well the recovered polarization matched the polarizations in a library of ordnance items expected at the site.

Thus we produced a total of eleven ranked dig sheets using a variety of different methods and sensor types.

The first demonstration of the methodology defined in this research project was conducted at the Former Lowry Bombing and Gunnery Range (FLBGR) in Colorado during the 2006 field season. The focus of the FLBGR demonstration was on the verification of the single inversion process used to extract physics-based parameters from magnetic and electromagnetic induction (EMI) anomalies, as well as the statistical classification algorithms used to make discrimination decisions from those parameters.

The second demonstration was conducted as part of the ESTCP discrimination pilot study in 2007 at Camp Sibert, Alabama. The objective was to find potentially hazardous 4.2-in mortars. The demonstration provided another test of the single-target inversion methodology as well as that of the cooperative inversion process. Both cued interrogation and full coverage data collected by different demonstrators were analyzed, allowing the effect of data quality on discrimination decisions to be assessed. For the Camp Sibert discrimination study, the project team created 8 different dig sheets from 6 different sensor combinations: MTADS magnetics; EM-61 cart (classification and size-based); MTADS EM-61 (classification and size-based); MTADS EM-61 and magnetics; EM-63; and EM-63 and magnetics.

Effective discrimination was demonstrated for all sensor combinations, with just one false-negative for the EM-63 when inverted without magnetometer location constraints. The cued interrogation EM-63 data, when cooperatively inverted with the magnetics data, was the most effective discriminator. The MTADS EM-61 was also an effective discriminator, especially when inverted cooperatively with the magnetometer data.

A third demonstration was conducted at the former Fort McClellan, Alabama. During this demonstration, the performance of the Geonics EM-63 was tested when deployed in a cued interrogation mode in a heavily wooded section of the Fort McClellan site with potential items of interest including grenades, 37 millimeter (mm) projectiles, 60 mm mortars, 75mm shrapnel and 3.8-in shrapnel rounds. Because of the heavily wooded characteristic of the demonstration area, traditional positional techniques such as a Global Positioning System (GPS) and Robotic Total Station (RTS) could not be used. Instead, a template constructed from a sturdy pool liner was centered over each anomaly and data were then collected at 55 pre-marked station locations distributed about the center of the template. Except for one 37mm and a number of 60 mm seed items, all munitions encountered at the site were 75mm or 3.8-in shrapnel rounds. The EM-63 surveys were cued off production mode EM-61 data. A feature space comprising the size and the relative-decay rate of the primary polarization was found to be effective for discrimination of the medium caliber projectiles (75mm and 3.8-in shrapnel). All demonstration metrics related to discrimination of these medium caliber projectiles were met.

1.3 REGULATORY DRIVERS

Refer to the Program Office demonstration plan for a discussion of regulatory drivers.

2.0 TECHNOLOGY

2.1 TECHNOLOGY DESCRIPTION

Magnetic and EM methods represent the main sensor types used for detection of UXO. Over the past 10 years, significant research effort has been focused on developing methods to discriminate between hazardous UXO and non-hazardous scrap metal, shrapnel and geology (e.g. Hart et al., 2001; Collins et al., 2001; Pasion & Oldenburg, 2001; Zhang et al., 2003a, 2003b; Billings, 2004). The most promising discrimination methods typically proceed by first recovering a set of parameters that specify a physics-based model of the object being interrogated. For example, in time-domain electromagnetic (TEM) data, the parameters comprise the object location and the polarization tensor (typically two or three collocated orthogonal dipoles along with their orientation and some parameterization of the time-decay curve). For magnetics, the physics based model is generally a static magnetic dipole. Once the parameters are recovered by inversion, a subset of the parameters is used as feature vectors to guide either a statistical or rule-based classifier.

Magnetic and EM phenomenologies have different strengths and weaknesses. Magnetic data are simpler to collect, are mostly immune to sensor orientation and are better able to detect deeper targets. EM data are sensitive to non-ferrous metals, are better at detecting smaller items and are able to be used in areas with magnetic geology. Therefore, there are significant advantages in collecting both types of data including increased detection, stabilization of the EM inversions by cooperative inversion of the magnetics (Pasion et al., 2003) and extra dimensionality in the feature space that may improve classification performance (e.g. Zhang et al., 2003a). However, these advantages need to be weighed against the extra costs of collecting both data types.

There are three key elements that impact the success of the UXO discrimination process described in the previous paragraphs:

1) Creation of a map of the geophysical sensor data: This includes all actions required to form an estimate of the geophysical quantity in question (magnetic field in NanoTeslas [nT], amplitude of EMI response at a given time-channel, etc.) at each of the visited locations. The estimated quantity is dependent on the following:
    a. Hardware, including the sensor type, deployment platform, position and orientation system and the data acquisition system used to record and time-stamp the different sensors;
    b. Survey parameters such as line spacing, sampling rate, calibration procedures etc.;
    c. Data processing such as merging of position/orientation information with sensor data, noise and background filtering applied;
    d. The background environment including geology, vegetation, topography, cultural features, etc.; and
    e. Depth and distribution of ordnance and clutter.
2) Anomaly selection and feature extraction: This includes the detection of anomalous regions and the subsequent extraction of a dipole (magnetics) or polarization tensor (TEM) model for each anomaly. Where magnetic and EMI data have both been collected,

the magnetic data can be used as constraints for the EMI model via a cooperative inversion process.

3) Classification of anomalies: The final objective of the demonstration is the production of a dig sheet with a ranked list of anomalies. This will be achieved via statistical classification which will require training data to determine the attributes of the UXO and non-UXO classes.

The focus of this demonstration is on the further testing and validation of the methodologies for 2) and 3) above that have been developed in UXOLab jointly by Sky Research and the University of British Columbia-Geophysical Inversion Facility (UBC-GIF).

We now describe each of the three key elements of the technology as identified above.

2.1.1 Creation of a Map of Geophysical Sensor Data

Each of the demonstrators will provide filtered, located geophysical data. We do not intend to apply any additional pre-processing to the data.

2.1.2 Anomaly Selection and Feature Extraction

At this point in the process flow, there is a map of each of the geophysical quantities measured during the survey. The next step in the process is detection of anomalous regions followed by the extraction of features for each of the detected items.

Feature Extraction: Time-domain Sensor

In the EMI method, a time varying field illuminates a buried, conductive target. Currents induced in the target then produce a secondary field that is measured at the surface. EM data inversion involves using the secondary field generated by the target for recovery of the position, orientation, and parameters related to the target's material properties and shape. In the UXO community, the inverse problem is simplified by assuming that the secondary field can be accurately approximated as a dipole.

In general, TEM sensors use a step off field to illuminate a buried target. The currents induced in the buried target decay with time, generating a decaying secondary field that is measured at the surface. The time-varying secondary magnetic field $\mathbf{B}(t)$ at a location $\mathbf{r}$ from the dipole $\mathbf{m}(t)$ is:

$$\mathbf{B}(t) = \frac{\mu_o}{4\pi r^3}\mathbf{m}(t)\cdot\left(\hat{\mathbf{r}}\hat{\mathbf{r}} - \mathbf{I}\right) \tag{1}$$

where $\hat{\mathbf{r}} = \mathbf{r}/|\mathbf{r}|$ is the unit-vector pointing from the dipole to the observation point, $\mathbf{I}$ is the 3 x 3 identity matrix, $\mu_o = 4\pi \times 10^{-7}$ H/m is the permittivity of free space and $r = |\mathbf{r}|$ is the distance between the center of the object and the observation point.

The dipole induced by the interaction of the primary field $\mathbf{B}_o$ and the buried target is given by:

$$\mathbf{m}(t) = \frac{1}{\mu_o}\overline{\overline{\mathbf{M}}}(t)\cdot\mathbf{B}_o \tag{2}$$

where $\mathbf{M}$ (t) is the target's polarization tensor. The polarization tensor governs the decay characteristics of the buried target and is a function of the shape, size, and material properties of the target. The polarization tensor is written as:

$$\overline{\overline{\mathbf{M}}}(t) = \begin{bmatrix} L_1(t) & 0 & 0 \\ 0 & L_2(t) & 0 \\ 0 & 0 & L_3(t) \end{bmatrix} \tag{3}$$

where we use the convention that $L_1(t) \geq L_2(t) \geq L_3(t)$, so that polarization tensor parameters are organized from largest to smallest. The polarization tensor components are parameterized such that the target response can be written as a function of a model vector containing components that are a function of target characteristics. Particular parameterizations differ depending on the instrument (number of time channels, time range measured etc) and the group implementing the work. Bell et al. (2001) solves for the components of the polarization tensor at each time channel, and this is the procedure we used for the Geonics EM-61 MKII and the BUD. For the MetalMapper and TEMTADS we used the Pasion-Oldenburg formulation (Pasion and Oldenburg, 2001):

$$L_i(t) = k_i(t + \alpha_i)^{-\beta_i} \exp(-t/\gamma_i) \tag{4}$$

for i={1,2,3},with the convention that $k_1 \geq k_2 \geq k_3$. For a body-of-revolution (BOR), $L_2 = L_3$ for a rod-like object (Pasion and Oldenburg, 2001) and $L_1 = L_2$ for a plate-like object. The reason we will use Pasion-Oldenburg for MetalMapper and TEMTADS is that they cover a sufficiently long time-range so that the two time decay parameters can usually be resolved. For the BUD and EM-61, the time-range is often not long enough to resolve the exponential decay parameter, hence the instantaneous amplitude (beta) formulation is preferred.

Given a set of observations $\mathbf{d}^{\mathbf{obs}}$, we formulate the parameter estimation as an optimization problem through Bayes theorem:

$$p(\mathbf{m} \mid \mathbf{d}^{\mathbf{obs}}) = \frac{p(\mathbf{m})\, p(\mathbf{d}^{\mathbf{obs}} \mid \mathbf{m})}{p(\mathbf{d}^{\mathbf{obs}})} \tag{5}$$

where $\mathbf{m}$ is the vector of model parameters (location, orientation and polarization tensor parameters), $p(\mathbf{m})$ is the probability distribution representing prior information, $p(\mathbf{d}^{\mathbf{obs}})$ is the marginal probability density of the experimental data, and $p(\mathbf{d}^{\mathbf{obs}}|\mathbf{m})$ is the conditional probability density of the experimental data which describes the ability of the model to reproduce the experimental data. The *a-posteriori* conditional probability density $p(\mathbf{m}|\mathbf{d}^{\mathbf{obs}})$ is the probability density we ascribe to $\mathbf{m}$ after collecting the data. The *a-posteriori* conditional probability density encapsulates all the information we have on the model parameters and the model that maximizes it is usually regarded as the solution to the inverse problem. We estimate a value of $\mathbf{m}$ that maximizes the log of the *a-posteriori* conditional probability density:

$$\mathbf{m}^* = \max_{\mathbf{m}} \{\log p(\mathbf{m} \mid \mathbf{d}^{\mathbf{obs}})\} \tag{6}$$

With a single data-set and no prior information on the model parameters (except maybe some bound constraints on the model parameters):

$$\text{minimize } \phi(\mathbf{m}) = \frac{1}{2}\left\| V_d^{-1/2}\left(\mathbf{d}^{\mathbf{obs}} - F(\mathbf{m})\right)\right\|^2, \text{ subject to } m_i^L \le m_i \le m_i^U \ . \tag{7}$$

where $F(\mathbf{m})$ is a vector comprising the forward modeled data at the sampled locations, m_i^L and m_i^U are the lower and upper bounds on parameter i and V_d is the co-variance matrix of the data. Efficient algorithms for the solution of this optimization problem have been implemented for various polarization tensor formulations within UXOLab (including two- and three independent polarization tensors).

<u>Feature Extraction: Magnetics</u>

For magnetics, the physics-based model most commonly used is a dipole:

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_o}{4\pi r^3}\left(3(\mathbf{m}\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \mathbf{m}\right) \tag{8}$$

where the terms were defined earlier. As for the TEM case, a bound-constrained optimization problem is solved to extract feature vectors from each anomaly.

Specific details of the feature extraction methodologies for both magnetic and electromagnetics are described in Section 6.

2.1.3 Classification of Anomalies

At this stage in the process, we have feature vectors for each anomaly and need to decide which items should be excavated as potential UXO. Rule-based classifiers use relationships derived from the underlying physics to partition the feature space. Examples include the ratio of TEM decay parameters (Pasion and Oldenburg, 2001) and magnetic remanence (Billings, 2004). For this demonstration, we focus on statistical classification techniques which have proven to be very effective at discrimination at various test sites (e.g. Zhang et al., 2003b).

Statistical classifiers have been applied to a wide variety of pattern recognition problems, including optical character recognition, bioinformatics and UXO discrimination. Within this field there is an important dichotomy between *supervised* and *unsupervised* classification. Supervised classification makes classification decisions for a *test* set comprised of unlabelled feature vectors. The classifier performance is optimized using a *training* data set for which labels are known. In unsupervised classification there is only a test set; labels are unknown for all feature vectors. Most applications of statistical classification algorithms to UXO discrimination have used supervised classification; the training data set is generated as targets are excavated. More recently, unsupervised methods have been used to generate a training data set that is an informative sample of the test data (Carin et al., 2004). In addition, *semi-supervised* classifiers, which exploit both labeled data and the topology of unlabelled data, have been applied to UXO discrimination in one study (Carin et al., 2004).

Figure 1 summarizes the supervised classification process within the statistical framework. Given test and training data sets, we extract features from the data, select a relevant subset of these features and optimize the classifier using the available training data. Because the predicted performance of the classifier is dependent upon the feature space, the learning stage can involve

further experimentation with feature extraction and selection before adequate performance is achieved.

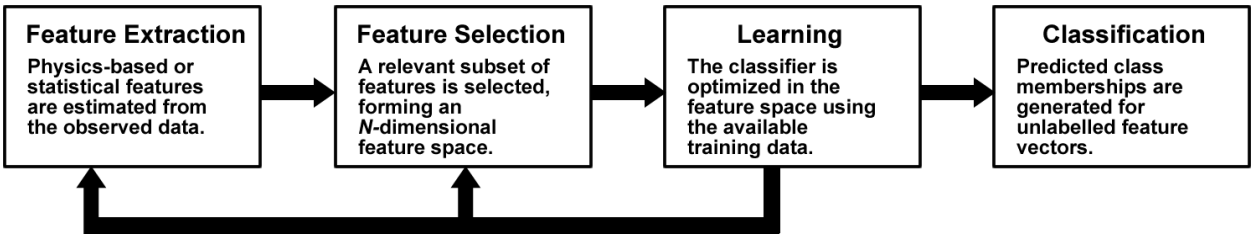


Figure 1. A framework for statistical pattern recognition.

There are two (sometimes equivalent) approaches to partitioning the feature space. The *generative* approach models the underlying probability distributions which are assumed to have produced the observed feature data. The starting point for any generative classifier is Bayes rule:

$$P(\omega_i \,|\mathbf{x}) \;\alpha\; P(\mathbf{x}|\, \omega_i)P(\omega_i). \qquad (9)$$

The likelihood function $P(\mathbf{x}|\, \omega_i)$ computes the probability of observing the feature vector **x** given the class ω_i. The prior probability $P(\omega_i)$ quantifies our expectation of how likely we are to observe class ω_i. Bayes rule provides a mechanism for classifying test feature vectors: assign **x** to the class with the largest *a posteriori* probability. Contours along which the posterior probabilities are equal define decision boundaries in the feature space.

An example of a generative classifier is discriminant analysis, which assumes a Gaussian form for the likelihood function. Training this classifier involves estimating the means and covariances of each class. If equal covariances are assumed for all classes, the decision boundary is linear. While these assumptions may seem overly restrictive, in practice linear discriminant analysis performs quite well in comparison with more exotic methods and is often used as a baseline classifier when assessing performance.

Other generative classifiers assume a nonparametric form for the likelihood function. For example, the probabilistic neural network (PNN) models the likelihood for each class as a superposition of kernel functions. The kernels are centered at the training data for each class. In this case the complexity of the likelihood function (and hence the decision boundary) is governed by the width of the kernels (Figure 2).

The *discriminative* approach is not concerned with underlying distributions but rather seeks to identify decision boundaries which provide an optimal separation of classes. For example, a support vector machine (SVM) constructs a decision boundary by maximizing the *margin* between classes. The margin is defined as the perpendicular distance between *support planes* which bound the classes, as shown in Figure 3. The decision boundary then bisects the support planes. This formulation leads to a constrained optimization problem: maximize the margin between classes subject to the constraint that the training data are classified correctly. An advantage of the SVM method over other discriminative classifiers (e.g. neural networks) is that there is a unique solution to the optimization problem.

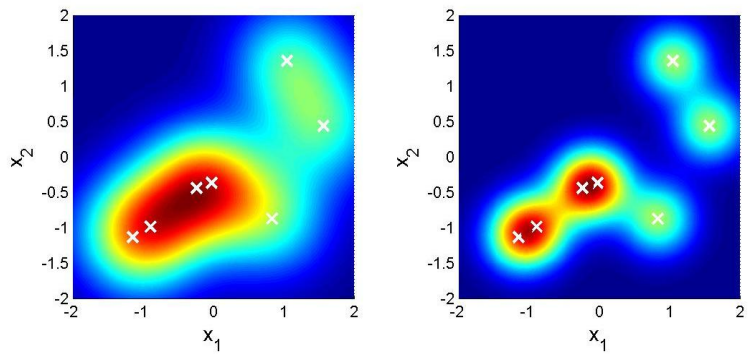


Figure 2. Nonparametric density estimate using Gaussian kernels. Kernel centers are shown as crosses. A large kernel width produces a smooth distribution (left) compared to a small kernel width (right).

With all classification algorithms, a balance must be struck between obtaining good performance on the training data and generalizing to a test data set. An algorithm that classifies all training data correctly may produce an overly complex decision boundary that may not perform well on the test data. In the literature this is referred to as "bias-variance trade-off" and is addressed by constraining the complexity of the decision boundary (regularization). In cases such as linear discriminant analysis, the regularization is implicit in specification of the likelihood function. Alternatively, the complexity of the fit can be explicitly governed by regularization parameters (e.g. the width of kernels in a PNN or Lagrange multipliers in a SVM). These parameters are typically estimated from the training data using *cross-validation*, which sets aside a portion of the training data to assess classifier performance for a given regularization.

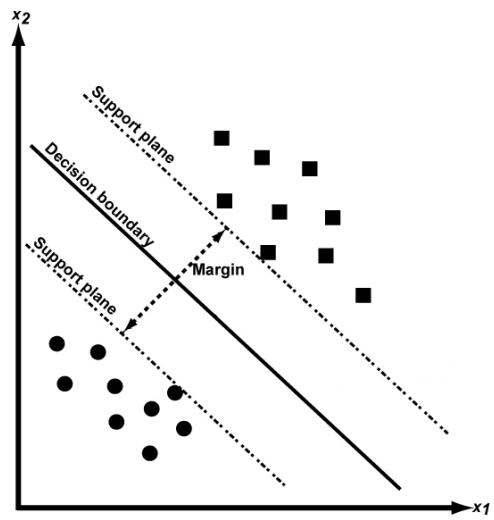


Figure 3. SVM formulation for constructing a decision boundary. The decision boundary bisects support planes bounding the classes.

2.1.4 UXOLab Software

The methodologies for data processing, feature extraction, and statistical classification described above have been implemented within the UXOLab software environment, which was used for this demonstration. UXOLab is a Matlab-based software package developed over a six year period at the UBC-GIF, principally through funding by the United States Army Corps of Engineers-Engineering Research and Development Center (USACE ERDC) (DAAD19-00-1-0120). Over the past five years, Sky Research and UBC-GIF have considerably expanded the capabilities of the software.

2.2 PREVIOUS TESTING OF THE TECHNOLOGY

Table 1 provides a list of some of the previous tests conducted of the underlying data processing and interpretation methodology that will be used in this demonstration.

Table 1. Previous Inversion/Classification Testing

Inversion/Classification Test	Description	Results
Demonstration Site: Aberdeen Proving Ground (APG)/Yuma Proving Ground (YPG)		
Geocenters Surface-Towed Ordnance Locator System (STOLS) EM-61 and magnetometer data	Discrimination ability of the system was marginal due to the following: limitations in positional accuracy (5-10cm), which is inadequate for advanced discrimination); and lack of sensor orientation data; and low Signal-to-noise ratio (SNR). No statistical classification algorithms were applied.	Results contributed to the decision to enhance SKY sensor systems by including the use of RTS for positioning and inertial measurement unit (IMU) for sensor orientation. Demonstrated the feasibility of cooperative inversion of large volumes of data with UXOLab.
Demonstration Site: FLBGR Rocket Range (RR) (8 acres surveyed) and 20mm Range Fan (RF) (2 acres surveyed)		
Geonics EM-61 and EM-63 single inversion, positioned by a Leica TPS 1206 Robotic Total Station (RTS) with orientation provided by a Crossbow AHRS 400 IMU. The RR survey objective was to discriminate a mixed range of projectiles with minimum diameter of 37mm from shrapnel, junk, 20mm projectiles and small-arms. The 20mm RF survey presented a small-item discrimination scenario with survey objective of discriminating 37mm projectiles from ubiquitous 20mm projectiles and 50 caliber bullets.	For the EM-61, 3-dipole instantaneous amplitude models were fit to the available 4 time-channels, while for the EM-63, 3-dipole Pasion-Oldenburg models were recovered from the 26 time-channel data. Parameters of the dipole model were used to guide a statistical classification. Canonical and visual analysis of feature vectors extracted from the test plot data indicated that discrimination could best proceed using a combination of a size and a “goodness of fit” based feature vector. A SVM classifier was then implemented based on those feature vectors and using the available training data.	Two phases of digging and training were conducted at the 20mm RF and three phases at the RR. At the RR, twenty-nine MK-23 practice bombs were recovered, with only one other UXO item encountered (a 2.5 inch rocket warhead). At the 20mm RF, thirty-eight 37mm projectiles (most of them emplaced) were recovered, as were a large number of 20mm projectiles and 50 caliber bullets. For both sites, and for both instruments, the SVM classifier outperformed a ranking based on amplitude alone. In each case, the last detected UXO was ranked quite high by the SVM classifier and digging to that point would have resulted in a 60-90% reduction in the number of false alarms. This operating point is of course unknown prior to digging. We found that using a stop-digging criteria of f=0 (mid-way between UXO and clutter class support planes), was too aggressive and more excavations were typically required for full recovery of detected UXO. Both the amplitude and SVM methods performed quite poorly on two deep (40centimeter [cm]) emplaced 37mm projectiles at the 20mm RF, exposing a potential weakness of the “goodness of fit” metric. Retrospective analysis revealed that thresholding on the size of the polarization tensor alone would have yielded good discrimination performance.

Inversion/Classification Test	Description	Results
Demonstration Site: Camp Sibert		
Geonics EM-61 cart, MTADS EM-61 array, MTADS magnetics array, and EM-63 single and cooperative inversions. EM-63 cued interrogations were positioned by a Leica TPS 1206 RTS with orientation information provided by a Crossbow AHRS 400 IMU. The objective of the surveys was the discrimination of a large target (4.2-in mortars). The site was unusual in that the primary munitions item known to have been used was the 4.2-in mortar, thus providing a site where the discrimination is a case of identifying a single large target amongst smaller pieces of mortar debris and clutter.	For the EM-61, 3-dipole instantaneous amplitude models were fit to the available 3 time-channels, while for the EM-63, 3-dipole Pasion-Oldenburg models were recovered from the 26 time-channel data. MTADS and EM-63 data were also cooperatively inverted. Parameters of the dipole model were used to guide a statistical classification.	The results for all sensor combinations were excellent, with just one false negative for the EM-63 when inverted without cooperative constraints. When inverted cooperatively, the EM-63 cued interrogation was the most effective discriminator. All 33 UXO were recovered with 25 false alarms (16 of these were in the "can't analyze" category). Not counting the "can't analyze" category, the first 33 recommended excavations were all UXO. The MTADS and MTADS cooperatively inverted were also very effective at discrimination, with all UXO recovered very early in the dig list (e.g. for the MTADS cooperative there were just 2 false-positives by the time all 117 "can analyze" UXO were recovered). The MTADS data set suffered from a high number of false alarms due to anomalies with a geological origin (caused by the cart bouncing up and down). In addition, the operating point was very conservative and many non-UXO were excavated after recovery of the last UXO in the dig list. The results from the EM-61 cart were also very good, although 24 false-positives were required to excavate all 105 UXO (that weren't in the "can't analyze" category). The lower data quality of the EM-61 cart resulted in a larger number of "can't analyze" anomalies over metallic sources than the MTADS.
Demonstration Site: Fort McClellan		
Geonics EM-63 deployed in a cued interrogation mode demonstrated. A wide range of potential items of interest of different calibers included grenades, 37mm projectiles, 60 mm mortars, 75mm shrapnel and 3.8-in shrapnel rounds. The EM-63 surveys were cued off production-mode EM-61 data. A template (constructed from a sturdy pool liner) was centered over each anomaly and data were then collected at 55 pre-marked station locations distributed about the center of the template.	Polarization tensor models were fit to each surveyed anomaly. Ground truth information from 60 of the 401 live-site anomalies, along with 18 items in the geophysical proveout and 21 items measured in a test-pit were used to train a statistical classifier. Features related to shape, encapsulated in the relative values of the primary, secondary and tertiary polarizations were unstable and could not be used for reliable discrimination. A feature space comprising the size and the relative-decay rate of the primary polarization was used for discrimination of the medium caliber projectiles (75mm and 3.8-in shrapnel).	All demonstration metrics related to discrimination of these medium caliber projectiles were met. At the operating point, all but 5 of 119 targets of interest were recommended for excavation, with 34 false alarms. If the operating point was relaxed slightly then all medium caliber projectiles would have been recovered with 51 false alarms. Retrospective analysis revealed that excellent discrimination performance could have been obtained by using a feature space comprising an early and late time feature extracted from the object's primary polarization. Furthermore, we found that these feature vectors could be approximated without fitting polarization tensor models to the data, and by using just seven measurement locations around the template center. These approximate early and a late time decay features were extracted from the sounding with the slowest decay (defined as the ratio of the 20th to 1st time-channels).

2.3 ADVANTAGES AND LIMITATIONS OF THE TECHNOLOGY

The main advantage of the technology is a potential reduction in the number of non-hazardous items that need to be excavated, thus reducing the costs of UXO remediation. Advantages of UXOLab and the algorithms within the package include:

- All the functionality required to process raw geophysical data, detect anomalous regions, and perform geophysical inversion and discrimination.
- Algorithms have been developed for inverting magnetic and TEM data sets both separately and cooperatively using a number of different polarization tensor formulations.
- There is an extensive set of algorithms for rule-based and statistical classification.
- Configuration is modular, so that as new sensor technologies become available (e.g. new TEM systems with multi-component receivers etc), the inversion functionality will be immediately available to apply to data collected using those new sensor systems.

The principal disadvantage is that UXOLab is written in Matlab and has not been configured for general use by contractors and non-specialists.

3.0 PERFORMANCE OBJECTIVES

The performance objectives for this demonstration are summarized in Table 2. There are objectives for both the data collection and data analysis demonstrators.

The first three analysis objectives refer to the classification part of the demonstration with the first two referring to the best results from each approach in a retrospective analysis and the third addressing how well each demonstrator is able to specify the correct threshold in advance. The final two objectives also refer to the feature extraction part of the demonstration.

3.1 OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF MUNITIONS

This is one of the two primary measures of the effectiveness of the classification approach. By collecting high quality data and analyzing those data with advanced parameter estimation and classification algorithms we expect to be able to classify the targets with high efficiency. This objective concerns the component of the classification problem that involves correct classification of items-of-interest.

3.1.1 Metric

The metric for this objective is the number of items on the master anomaly list that can be correctly classified as munitions by each classification approach.

3.1.2 Data Requirements

Preparation of a prioritized dig list for the targets on the master anomaly list for each technology demonstrated. Institute for Defense Analyses (IDA) personnel used their scoring algorithms to assess the results.

3.1.3 Success Criteria

The objective was considered to be met if all of the targets of interest (TOI) were correctly labeled as munitions on the prioritized anomaly list.

Table 2. Performance Objectives for this Demonstration (Performance was assessed both including and excluding Master ID 241)

Performance criterion				Production				TEMTADS				MetalMapper			BUD
Performance Objective	**Metric**	**Data Required**	**Success Criteria**	**MTADS Magnetics**	**EM61 cart**	**MSEMS**	**MTADS EM61**	**Statistical**	**Library**	**Polar. Match**	**Expert**	**Statistical**	**Library**	**Export**	**Statistical**
Maximize correct classification of munitions	Number of targets-of-interest retained	• Prioritized anomaly lists • Scoring reports from IDA	Approach correctly classifies all TOI	No 3 FN	Yes	Yes	No 1 FN	No 3* FN	No 2* FN	No 3* FN	No 4* FN	No 2* FN	No 4*FN	No 3* FN	No 1* FN
Maximize correct classification of non-munitions	Number of false alarms eliminated	• Prioritized anomaly lists • Scoring reports from IDA	Reduction of FA > 30% while retaining all TOI	No 13%	Yes 69%	Yes 68%	Yes 51%	No 28% Yes# 51%	No 12% Yes# 81%	No 11% Yes# 58%	No 28% Yes# 51%	Yes 40% Yes# 81%	No 9% Yes# 48%	Yes 40% Yes# 81%	No 0% Yes# 71%
Specification of no-dig threshold	Pd of correct classification and #FA at operating point	• Demonstrator-specified threshold • Scoring reports from IDA	Threshold achieves criteria above	No	Yes	Yes	No	No	No	No	No	No	No	No	No Yes#
Minimize number of anomalies that cannot be analyzed	Number of anomalies that must be classified as "Unable to Analyze"	• Demonstrator target parameters	Reliable target parameters can be estimated for > 90% of anomalies	Yes 90%	Yes 99.8%	Yes 99.9%	Yes 99.7%	Yes 100%	Yes 100%	Yes 100%	Yes 100%	Yes 100%	Yes 100%	Yes 100%	Yes 100%

Performance criterion				Production				TEMTADS				MetalMapper			BUD
Performance Objective	**Metric**	**Data Required**	**Success Criteria**	**MTADS Magnetics**	**EM61 cart**	**MSEMS**	**MTADS EM61**	**Statistical**	**Library**	**Polar. Match**	**Expert**	**Statistical**	**Library**	**Export**	**Statistical**
Correct estimation of target parameters (positions)	Accuracy of estimated target parameters	• Demonstrator target parameters • Results of intrusive investigation	X, Y < 15 cm (1σ) Z < 10 cm (1σ)	NA (x,y) NA (z)	No (x,y) No (z)	No (x,y) No (z)	No (x,y) No (z)	NA (x,y) No (z)=10.7 cm				NA (x,y) Yes (z)=7.1 cm			NA (x,y) NA (z)

FA = False Alarm
FN = False Negative
NA = Not Applicable
Pd = Probability of Detection
* = Including Master ID 241 with questionable ground-truth
# = Excluding Master ID 241 with questionable ground-truth

3.2 OBJECTIVE: MAXIMIZE CORRECT CLASSIFICATION OF NON-MUNITIONS

This is the second of the two primary measures of the effectiveness of the classification approach. By collecting high-quality data and analyzing those data with advanced parameter estimation and classification algorithms we expected to be able to classify the targets with high efficiency. This objective concerns the component of the classification problem that involves false alarm reduction.

3.2.1 Metric

The metric for this objective is the number of items-of-interest on the master dig list correctly classified as non-munitions by each classification approach.

3.2.2 Data Requirements

Preparation of a prioritized dig list for the targets on the master anomaly list for each technology demonstrated. IDA personnel used their scoring algorithms to assess the results.

3.2.3 Success Criteria

The objective was considered to be met if more than 30% of the non-munitions items were correctly labeled as non-munitions while retaining all of the targets-of-interest (TOI) on the dig list.

3.3 OBJECTIVE: SPECIFICATION OF NO-DIG THRESHOLD

In a retrospective analysis performed as part of this demonstration, it was possible to identify the true classification capabilities of a classification procedure based solely on the prioritized dig list submitted by each demonstrator. In a real-world scenario, all targets may not be dug so the success of the approach ultimately depends on the ability of an analyst to accurately specify the dig/no-dig threshold.

3.3.1 Metric

The probability of correct classification, P_{class}, and number of false alarms, N_{fa}, at the demonstrator-specified threshold were the metrics for this objective.

3.3.2 Data Requirements

Preparation of a ranked anomaly list with a dig/no-dig threshold indicated. IDA personnel used their scoring algorithms to assess the results.

3.3.3 Success Criteria

The objective was considered to be met if more than 30% of the non-munitions items were correctly labeled as non-munitions while retaining all of the TOI at the demonstrator-specified threshold.

3.4 OBJECTIVE: MINIMIZE NUMBER OF ANOMALIES THAT CANNOT BE ANALYZED

Anomalies for which reliable parameters cannot be estimated cannot be classified by the classifier. These anomalies must be placed in the dig category and reduce the effectiveness of the classification process.

3.4.1 Metric

The number of anomalies for which reliable parameters cannot be estimated is the metric for this objective.

3.4.2 Data Requirements

Each demonstrator that estimated target parameters provided a list of all parameters as part of results submission along with a list of those anomalies for which parameters could not be reliably estimated.

3.4.3 Success Criteria

The objective was considered to be met if reliable parameters can be estimated for > 90% of the anomalies on each sensor anomaly list.

3.5 OBJECTIVE: CORRECT ESTIMATION OF TARGET PARAMETERS

This objective involves the accuracy of the target parameters that are estimated in the first phase of the analysis. Successful classification is only possible if the input features are internally consistent. The obvious way to satisfy this condition is to estimate the various target parameters accurately.

3.5.1 Metric

Accuracy of estimation of target parameters is the metric for this objective.

3.5.2 Data Requirements

Provide a list of all parameters as part of the results submission. IDA analysts compared these estimated parameters to those measured during the intrusive investigation and determined via subsequent in-air measurements.

3.5.3 Success Criteria

The objective was considered to be met if the estimated βs are within ± 20%, the estimated X, Y locations are within 15 cm (1σ), the estimated depths are within 10 cm (1σ), and the estimated size is within ± 20%.

4.0 SITE DESCRIPTION

The former Camp San Luis Obispo is approximately 2,101 acres situated along Highway 1, approximately five miles northwest of San Luis Obispo, California. The majority of the area consists of mountains and canyons. The site for this demonstration is a mortar target on a hilltop in Munitions Response Site (MRS) 05 (within former Rifle Range #12). See the Program Office demonstration plan (ref) for more details on the site.

4.1 SITE SELECTION

This site was chosen as the next in a progression of increasingly more complex sites for demonstration of the classification process. The first site in the series, Camp Sibert, had only one target-of-interest and item "size" was an effective discriminant. At this site, there are at least four targets-of-interest: 60 mm, 81 mm, and 4.2-in mortars and 2.36-in rockets. This introduces another layer of complexity into the process.

4.2 SITE HISTORY

See the Program Office demonstration plan.

4.3 SITE GEOLOGY

See the Program Office demonstration plan.

4.4 MUNITIONS CONTAMINATION

See the Program Office demonstration plan.

5.0 TEST DESIGN

See the Program Office demonstration plan for a description of the test design for the overall project.

Sky Research/UBC-GIF processed data and delivered the following eleven digsheets:

1) Magnetics size-based: Production of a dig sheet ranked according to size (magnitude of the dipole moment);
2) MTADS EM-61 statistical: Statistical classification of features derived from the MTADS EM-61 data and the production of a ranked dig sheet;
3) Cart EM-61 statistical: Same as b) but with features from EM-61 cart-data;
4) TEMTADS cued interrogation statistical: Same as b) but with the polarizabilities from the TEMTADS array;
5) BUD statistical: As per b) but with features derived from the BUD; and
6) MetalMapper statistical: As per b) but with the polarizabilities derived from the MetalMapper data.
7) TEMTADS library method: We provided an alternative ranking of the TEMTADS based on a library method;
8) MetalMapper library: As in g) but for the MetalMapper.
9) MetalMapper polarization match: We provided an alternative ranking based on matching the recovered polarizations to a template library.
10) TEMTADS “expert” opinion: A fourth digsheet for TEMTADS was produced based on expert opinion.
11) MetalMapper “expert” opinion: As in i) but for the MetalMapper.

6.0 DATA ANALYSIS AND PRODUCTS

In this section we describe our data analysis procedures in general. Details specific to each dataset are provided in a Training Memo that is attached as Appendix A.

6.1 PREPROCESSING

We did not pre-process any of the data delivered by the demonstrators.

6.2 TARGET SELECTION FOR DETECTION

Target selections were made by the data collection demonstrators and/or the Program Office.

6.3 PARAMETER ESTIMATION

A flowchart of the parameter estimation process is shown in MDL = Minimum Description Length Figure 4, with details of each step provided in the following sections. Actions specific to cooperative inversion are encoded in pale orange, while actions specific to the multi-static sensors (TEMTADS, BUD and MetalMapper) are encoded in bright blue. For each dataset, the analyst followed the steps delineated by the flowchart. In addition, a Quality Control (QC) officer was assigned to each dataset and conducted additional visual reviews, re-inverted selected targets, reviewed masks, etc.

6.3.1 Data Covariance Matrix V_d

Our knowledge of the noise levels appropriate to the solution of the inverse problem is encapsulated in the data covariance matrix. We assume independently distributed Gaussian errors and use the following data covariance matrix:

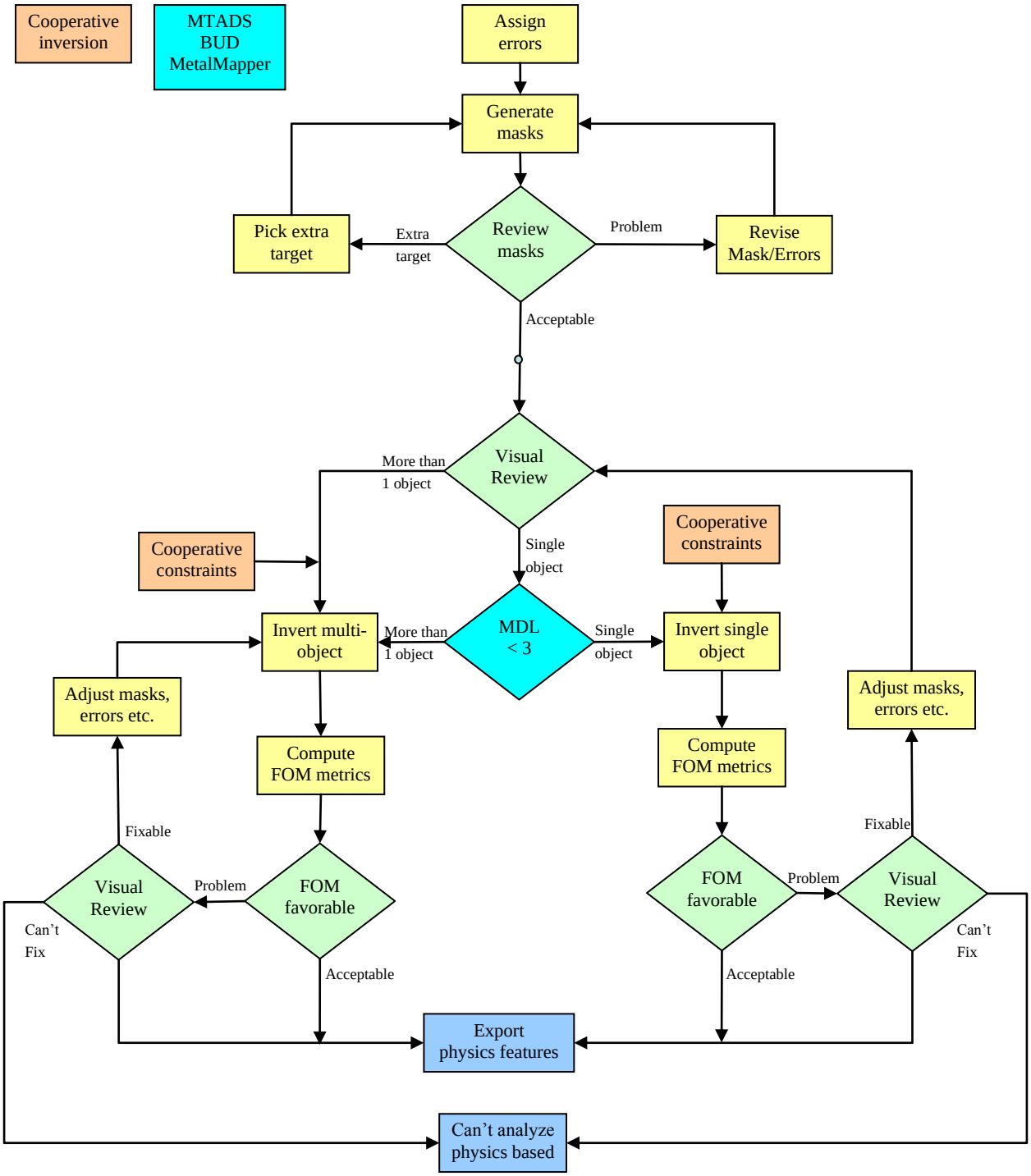
$$\left[V_d^{-1/2}\right]_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ \dfrac{1}{\delta_i + \varepsilon_i} & \text{if } i = j \end{cases} \tag{10}$$

where δ_i is a percentage of the ith datum:

$$\delta_i = \%error \times \left|d_{obs}\right|_i \tag{11}$$

and ε_i is a base level error that is present in the ith datum in the absence of a target. The percentage noise level for each sensor type will be determined by inspection of the inversion results from the test-pit and initial ground-truth data. At Camp Sibert we used % *error* = 10% for both the EM-61 cart and MTADS data and used that value as the starting point at SLO.

Methods to estimate the base-line noise for each sensor type are described in the Training Memo.



MDL = Minimum Description Length

Figure 4: Flowchart showing the procedure for parameter extraction.

For the multi-static data the noise levels were obtained by analyzing the background calibration data that each demonstrator collected several times each day. We experimented with values for the %*error* term using the test-pit and initial ground-truth data.

6.3.2 Forming the Data Vector d_{obs}

Defining the Data to be Inverted 1: Spatial coverage

Once data anomalies are identified, a mask is defined that represents the spatial limits of the data to be inverted. Unlike magnetics data, an unconstrained EMI inversion is very sensitive to adjacent anomalies and to the size of the mask used in areas without nearby anomalies. The masking procedure helps ensure that signal from adjacent anomalies does not affect the inversion results. In addition, from a practical standpoint, inverting the minimum number of observations reduces the computational time.

We employed an advanced masking procedure, which fits an ellipse to contours of the anomalous target. By using an ellipse we recovered a relatively smooth-shaped mask that mimics the shape of the anomaly. The main challenge is to find contours that are both smooth and close to the noise level. Including our background estimates ensures that we choose appropriate starting contour values that are both above the baseline error and that encompass all of the anomalous data.

Defining the Data to be Inverted 2: Time Channels

For the Camp Sibert demonstration, we excluded any channels with a SNR of less than 2 decibels (dB) from the inversion and used the same strategy here.

Visual review of masks and noise levels

The analyst conducted a visual review of all masks, estimated background noise levels and time-channels used in the inversion before the data were submitted to the inversion algorithm. UXOLab has several visualization tools that were utilized in this visual review.

6.3.3 Determining the Number of Objects

Visual Review

In conjunction with the visual review of masks/noise levels, the analyst attempted to visually determine when an anomaly needed to be treated as a "multi-object" scenario. In the first canonical situation, there are two flagged locations that are close together and the analyst needs to decide if:

(1) The two objects are sufficiently far apart that masking and single object inversion will result in an acceptable fit for one or both objects;

(2) The two objects are too close to be inverted using single object inversion, in which case multi-object inversion will need to be attempted (subsequent analysis of the inversion results using figures of merit [FOM] and visual review would be used to decide whether to accept either or both of the inversion results);

(3) There are two flags but just a single object (e.g. two peaks in an EM-61 survey over a horizontal dipole). In that case, the inversion result for one of the flagged items would be copied to the other (ideally one would be deleted from the dig-list as a duplicate but that may be problematic from a scoring perspective).

In the second canonical case there is an additional anomaly not on the master anomaly list that does not allow the original anomaly to be inverted using the single-object assumption. The extra

anomaly might be too close to the original anomaly or might have an amplitude lower than the detection threshold. In that case, the anomaly would be inverted using the multi-object code, and the item with highest UXO likelihood would be used in dig-sheet generation.

<u>Using Minimum Description Length</u>

Multi-object scenarios are not always readily apparent, so we used information theoretic criteria (ITC) to automatically estimate the number of objects in a given mask (see Wax and Kailath, 1985 and reference therein). Given N time-decays D = [d(t1), …, d(tN)] where $\mathbf{d}(t_j)$ is a vector of observed data the Minimum Description Length (MDL) method is employed to minimize a criterion over a family of hypothesized models α_η that have η detectable dipole sources. Namely, η may be from 1 to a maximum number $\eta_{\max}$ and α_η is a vector of model parameters generating the data and is a function of the hypothesized number of sources. The ITC are composed of a data-based log likelihood function for a given model and a penalty function that counterbalances model complexity,

$$\mathrm{IC}(\eta) = -\log f(D|\hat{\alpha}_\eta) + P(\nu(\eta)), \tag{12}$$

where $f(D|\hat{\alpha}_\eta)$ is a density function of data matrix D with the maximum likelihood estimate $\hat{\alpha}_\eta$ under the assumption that η sources are present, and $\nu(\eta)$ the number of the degree of freedom in the hypothesized model. Finally the number of objects is estimated as:

$$\hat{\eta}_{\mathrm{IC}} = \arg\min \mathrm{IC}(\eta). \tag{13}$$

The penalty term $P(\nu(\eta))$ is usually chosen as monotonically increasing function of η and its choice results in a different information criterion, for instance, the Akaike Information theoretic Criterion (AIC) and MDL, given as follows:

$$P(\nu(\eta)) = \begin{cases} \nu(\eta), & \mathrm{AIC} \\ \frac{1}{2}\nu(\eta)\log N, & \mathrm{MDL} \end{cases} \tag{14}$$

Under Gaussian statistics, the AIC and MDL can be implemented rapidly with a closed-form expression for a sequence of assumed sources (Wax and Kailath, 1985)

Figures 51 to 59 in the Training Memo illustrate the application of ITC and multi-object inversion strategies to TEMTADS data.

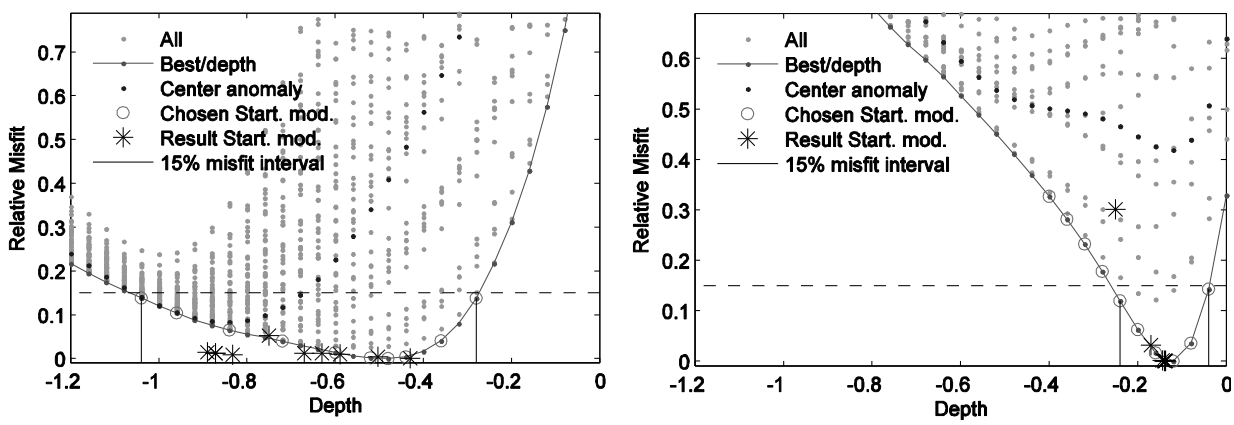
We only used the MDL criterion on the TEMTADS, BUD and MetalMapper data. There are not enough time-channels in the EM-61 (4 channels) for the MDL to work reliably.

6.3.4 Procedure for Single-Object Inversion

The optimization routine we used for inversion is a local Newton-type method that minimizes the least squares objective/misfit. We addressed the problem of local minima and assessed the level of ambiguity in resolving the depth of an item by choosing multiple starting models. We started

each inversion by scanning the subsurface (*x*, *y*, *z*) up to a 1.2 meter (m) depth. At each position we solved for the non-diagonalized polarization tensor[1] for the first time channel (chosen for its superior SNR). For each combination of a position and polarization tensor we computed a data misfit. The depth-misfit curve is defined by the best fit at a given depth (Figure 5, solid line). Starting models for the full inversion of multi-channel data are selected along the depth-misfit curve among the models with relative misfit below a given threshold, here 15% (circles). If the depth-misfit curve contains local minima these are also selected as starting models.

The iterative Newton-type inversion then proceeds with each starting model. A given search stops when the iteration reaches a set threshold (misfit tolerance or number of iterations). A final model is obtained for each of the starting models (black stars in Figure 5, note different misfit because computed on all time channels). In the example of Figure 5a there are final solutions with similar misfit spread over a 0.46 m depth range, which confirms the uncertainty in recovering depth. For comparison we show in Figure 5b the depth-misfit curve for a different target, where the minimum misfit is well defined as a function of depth, and therefore the depth is accurately recovered.



a. Depth-misfit relationship for the inversion of a 4.2-in mortar. Each point corresponds to a different (x,y,z) position. Solutions with similar misfit occur over a wide range of depths.

b. Depth-misfit relationship for the inversion of a 4.2-in mortar, where the initial and final ranges of models with similar misfit is tight, indicating a well defined solution.

Figure 5. The depth-misfit relationship, an indirect indicator of the of the depth-size ambiguity for a buried object.

[1] When the polarization tensor is not explicitly diagonalized, the inverse problem is linear.

Feature Extraction: Cooperative Inversion of TEM and Magnetic Data

In cooperative inversion, multiple data are inverted sequentially with the results of the first inversion used to constrain the second. This prior information can be formally introduced into the Bayesian formulation through the prior $p(\mathbf{m})$. Commonly utilized priors include Gaussian priors and uniform priors (i.e. a constant probability density function [pdf] for a parameter between two limits, and zero probability outside these limits). The solution to the inverse problem that utilizes these priors is:

$$\text{minimize } \phi(\mathbf{m}) = \sum_j \frac{1}{2\sigma_j^2}\left(m_j - \overline{m}_j\right)^2 + \frac{1}{2}\left\| V_d^{-1/2}\left(\mathbf{d}^{\mathbf{obs}} - F(\mathbf{m})\right)\right\|^2, \text{ subject to } m_i^L \le m_i \le m_i^U \quad (15)$$

where j represents the index of parameters whose Gaussian pdf's are assumed to be known. The strategy used here for cooperatively inverting magnetics and EM data was as follows:

1) The magnetics data were inverted for a best fit dipole.

2) The dipole location was used to define $\overline{m}_j$ (for j = 1, 2 and 3 which corresponds to the Easting, Northing and depth of the dipole) and the standard deviation of the parameter uncertainties was used to define σ_j. The estimated model parameter standard deviations can be obtained from the Gauss-Newton approximation to the Hessian at the optimum model location (e.g. Billings *et al*., 2002).

3) The EM data were inverted using the prior obtained from the magnetics data in step 2.

Inevitably, there were anomalies in the TEM data that did not have corresponding magnetic fits and vice versa. Where no constraints from magnetometer data were available, the TEM data were inverted using the same procedure as for single inversion.

6.3.5 Approach for Dealing with Multiple-Objects

Our solution strategy was to decompose the inverse problem into several steps, each of which sought to resolve one major set of model parameters. The procedure is first to solve for non-linear location parameters and subsequently solve for linear polarization parameters. With an optimal estimate of locations and dipolar polarizations, the orientations of each object were extracted from estimated magnetic polarizability tensors (MPT) and then further optimized. For time-domain systems that have sufficient time channels to characterize the decay behavior of the polarizability, we further sought the set of parameters in a parametric model of dipolar polarizations by either fixing the locations and orientations of multiple objects derived in the previous two-step procedure, or incorporating these nonlinear parameters into the inversion for an update. These steps are briefed in the following discussion.

Searching for Optimal Locations

In this inversion step, the primary concern was to find source locations, while the polarization estimate can be followed for a specific set of the location parameters.

To start the inversions, we needed an initial guess for the locations of the η objects. This might be obtained by examining the EMI response distribution assuming that each source contributes to

the field distribution with an associated peak. However this initial guess for a multi-object scenario cannot be guaranteed to be good. In many cases it is challenging to assign the starting point as there is often no knowledgeable way to connect spatial anomalies to each individual unknown object because the field pattern is a complicated function of object locations and orientations as well as transmitter-receiver configuration specific to an instrument and sensor distribution.

To avoid the difficulty arising from selecting one initial point based on field distribution, we proposed a multi-start algorithm as follows. Define a region of interest (ROI) that is sufficiently large to cover the active objects. Within the ROI, a number of points are uniformly or randomly created. Each point consists of the locations for objects and is a potential solution to the object function:

$$\Phi_d(\mathbf{r}) = \sum \left\| W_j \left(\mathbf{d}_{\text{obs}}(t_j) - \tilde{\mathbf{d}}(t_j) \right) \right\|^2, \quad \tilde{\mathbf{d}}(t_j) = \sum_{k=1}^{\eta} A(\tilde{\mathbf{r}}_k)\, \tilde{\mathbf{q}}_k(t_j) \tag{16}$$

where W_j is a data weighting matrix at time t_j and is generally chosen as the inverse of estimated standard error of the data, $\tilde{\mathbf{d}}(t_j)$ is the vector of predicted data, $A(\tilde{\mathbf{r}}_k)$ is a sensitivity matrix connecting k-th source with sensors, and $\tilde{\mathbf{q}}_k(t_j)$ is a 6 x 1 vector composed of the elements of the MPT $P(t)$ and is computed according to equation (23) given below. The vector $\mathbf{r} = \text{vec}[\mathbf{r}_1, \cdots, \mathbf{r}_\eta]$ indicates multi-source locations, where vec [•] represents a vectorization operation, i.e., stacking matrix columns into a single column and the ~ symbol indicates an assumed/estimated quantity. After a forward evaluation using (1), these points are sorted, e.g., in ascending order according to the values of $\Phi_d(\mathbf{r})$. A small population of points with smaller function values $\Phi_d(\mathbf{r})$ is selected as multi-start points. Then we conduct a nonlinear search to determine optimal source locations by updating those selected starting points using a well-developed minimizer like the Levenberg-Marquardt approach or trust region interior point method.

For a given set of locations $\mathbf{r}$, the estimation of dipolar polarizations embedded in the above processes is performed by solving a constrained linear least-squares problem as follows:

$$\tilde{\mathbf{f}}(t_j) = \arg\min_{\mathbf{f}(t_j)} \left\| \mathbf{d}_{\text{obs}}(t_j) - \sum_{k=1}^{\eta} A(\tilde{\mathbf{r}}_k)\, \tilde{\mathbf{q}}_k(t_j) \right\|^2, \text{ s.t. } p_{k,ii}(t) \geq 0,\ |p_{k,ij}(t)| \leq \frac{1}{2}\left[p_{k,ii}(t) + p_{k,jj}(t) \right] \tag{17}$$

where $\mathbf{f}(t) = \text{vec}[\mathbf{q}_1(t), \cdots, \mathbf{q}_\eta(t)]$ and $p_{k,ij}(t)$ is the elements of MPT of the k-th object and ordered into the components of $\mathbf{q}_k(t)$. The constraints in (17) are based on physics: the principal polarizations $L_i(t)$, i.e., the eigenvalues of a MPT, must be positive. This means that the magnetic polarizability tensor $P(t)$ should be physically sought as a (semi)-definite positive one. These are necessary constraints.

In this step, it is important to determine a proper number of sampling points. Its choice is a compromise among computational speed, convergence, and available computing power. Of

course, the greater the number of points sampled in a ROI, the higher the probability that the starting points are close to the global solution we desire. In that context however, a multi-start algorithm has high computing costs and could be very time-consuming for a very large number of sampling points. The selected multi-start nonlinear optimization we proposed has the advantage of reducing the computational cost of the algorithm and while maintaining its interesting capability of global optimization. In our study, it appears reasonable that the number of sampling points is set at 300 and the number of selected starting points for nonlinear updates is 10. Our computational experience shows that the multi-start algorithm outperforms the one with an initial point mentioned above. This multi-start algorithm requires little user involvement and can be parallelized.

Determining Optimal Orientations

Given estimated polarization tensors, we can use the trigonometric expressions of Euler vectors to find the orientation angles $\xi = (\phi, \theta, \psi)$ of an object. With multi-time channel data, we could have a series of estimated ξ_j for each object, which may not be close to each other at all due to noise and model approximation. For the dipole model assumed here (i.e., its principal polarization directions are not varying with time), they might be simply approximated with the orientations obtained at an early or intermediate time channel by consideration of the signal-to-noise ratio or using a joint diagonalization technique to determine an 'average principal direction' shared by these matrices.

However from the first-order perturbation analysis, we know that the perturbation in the estimated magnetic polarizabiltiy tensor can be erroneously propagated into large changes in the determination of the principal directions due to potentially small differences between the principal polarizations. Considering this numerical instability problem, measurement error, and possible tradeoff among the magnetic polarizabiltiy tensors of multiple objects, we used the orientations approximately found using a single time channel or a joint diagonalization technique as starting one and performed a nonlinear update to determine optimal orientations of multi-objects by fixing their locations.

Parametric Model Fitting

In this step, we parameterize the principal polarizations as $L_i(t) = \kappa_i t^{-\beta_i} e^{-t\gamma^{-1}}$ and wish to find the set of $(\kappa_i, \beta_i, \gamma_i)$ parameters corresponding to *i*-th polarization for each object. This is also a nonlinear inverse problem. To do that, the sets of $(\kappa_i, \beta_i, \gamma_i)$ parameters can be first initialized by linearly fitting to the log-transformed discrete polarization $\log L_i(t_j) = \log\kappa_i - \beta_i \log t_j - t_j \gamma_i^{-1}$, $j = 1, \cdots, N$, that are computed in the last step. These initial parameters are further updated nonlinearly by fitting the measured data using fixed locations and orientations derived from the last steps. Finally, we update these parametric models simultaneously with locations and orientations parameters. The parametric solution obtained using the fixed nonlinear spatial parameter (κ, ξ) might be better since in this problem there does not exist a tradeoff between the spatial and polarization parameters sets of (κ, β, γ), e.g., an ambiguity existing between the

depth and κ parameters. As an option, the final parametric solution is picked with a smaller misfit value from above two nonlinear solutions.

6.3.6 Quality Control Procedures

Figures of Merit

At the previous Camp Sibert demonstration so-called Figures of Merit (FOMs) played a significant role in the classification process. FOMs are indicators derived from quantities that affect the quality of data such as signal to noise ratio, anomaly coverage and sources of errors (instrument noise, survey location errors). Key FOM parameters were identified and studied through simulation and field data analysis to assess the reliability of data sets and inversion. Dig lists were established for discrimination with the EM-61 cart and MTADS EM-61 towed-array. Inclusion of FOM in the classification process reduced the number of non-UXO items to excavate.

We had intended to more fully utilize FOMs in this demonstration (e.g. to automate QC, to influence dig/no-dig decisions). However, in the end we found that we mostly used the FOMs to flag anomalies that we should visually inspect and either reinvert, or place in the "can't decide" or "can't analyze" categories. Given the limited use of FOMs in this demonstration, we won't reproduce the details of how to calculate them here, and instead refer the interested reader to Billings, S.D. (2008).

Visual Inspection of Inversion Results

In preparation for the Camp Sibert demonstration we created some new QC views where all relevant information for each anomaly are presented on a single page and exported to a pdf document. The analyst can scroll through each page of the pdf and pass or fail each fit, with the results saved in UXOLab so that only the failed anomalies needed to be reinverted. For this demonstration, additional custom-designed QC views of each anomaly were developed and exported to pdf format.

Role of the Quality Control Officer

Each analyst performed a first pass analysis of the inversion results to determine if there were any obvious problems. In addition, we assigned a QC Officer to each dataset. That person conducted an additional analysis of the data, masks, noise levels, inversion results etc and attempted to identify any problems or inconsistencies. The QC officer either asked the analyst to remedy the problem or addressed the problem themselves.

6.3.7 Magnetics feature extraction procedure

The procedure for magnetics feature extraction is similar to the process illustrated by the flow-chart in Figure 4, with the following exceptions: (1) no data-based feature generation; (2) no MDL; and (3) no calculation of FOM. Salient points about the magnetics feature extraction include: (1) Visual QC is used to assign pass/fail designations to each inversion result; (2) the automatic masking procedure is applied to a total-gradient channel created from the total-field data; (3) the inversion results are largely insensitive to the background noise levels, so we chose

a single global value for all anomalies; (4) only a single start model is required as a good initial guess at the dipole model can be obtained heuristically.

6.3.8 Sensor Specific Considerations

Specific issues/decisions regarding each dataset are as follows:

1) MTADS magnetometer data:
   - Static-dipole fits;
   - Did not attempt multi-object inversions;
2) MTADS EM-61 array and EM-61 cart-data:
   - Three-dipole beta models;
   - Did not attempt multi-object inversions;
   - Did not use MDL analysis to determine number of objects (because the EM-61 doesn't have enough time-gates for the MDL to work reliably);
3) MTADS EM-61 array cooperative inversion with magnetics and MSEMS dual-mode:
   - We tested the cooperative inversion process on the training data and found that it did not improve the feature vectors used to derive the classification strategy. Therefore, cooperative inversion was not used for the test-data.
4) TEMTADS and MetalMapper cued interrogation arrays:
   - Three-dipole beta models that were then fit parametrically with a sum of exponentials model;
   - Multi-object inversions were conducted on items with poor fits or where the MDL indicated that more than one object was in the field of view;
   - Used MDL analysis to determine number of objects;
5) BUD deployed in cued interrogation mode:
   - Three-dipole beta models;
   - Did ***not*** attempt multi-object inversions (as we did not have any test data over multi-object scenarios);
   - Did not use MDL analysis to determine number of objects;

Note that we only attempted to fit overlapping anomalies when there were no more than two overlapping objects in the field of view of the sensor.

6.4 TRAINING

For the statistical classification, we used the data over the test-pit and the initial five training grids released by the Program Office to determine the feature vectors and statistical classifier to use. See the training memo in the Appendix A for more details.

6.5 CLASSIFICATION

We produced a ranked anomaly list for each of the sensor data sets we processed using the format shown in Figure 6. Additional details on the classification method are provided in the training memo reproduced in Appendix. As discussed below, some modifications were made to the TEMTADS strategy after submitting the training memo.

Rank	Anomaly ID	$P_{clutter}$	Comment
1	247	.97	
2	1114	.96	High confidence NOT munitions
3	69	…	
…	…	…	
…	…	…	
…	…	…	Can't make a decision
…	…	…	
…	…	…	
…	…	…	
…	…	…	High confidence munitions
…	…	.03	
…	…	.02	
	…		
	…		
	…		Can't extract reliable features
	…		

Threshold

Figure 6. Format for the prioritized anomaly list that will be submitted by each classification demonstrator.

6.5.1 Modified TEMTADS strategy

To generate a diglist for the TEMTADS data our proposed strategy was to apply quadratic discriminant analysis trained on small (60 mm), medium (81 mm and 2.36"), and large (4.2") TOIs. The latter two classifiers were trained on size, decay and asymmetry, while discrimination for small TOIs was carried out using only size and decay. When applying these classifiers we found that the training data were not particularly representative of the test data. In particular:

- Covariance estimates computed from the training TOI data alone tended to have high eccentricity (the largest principal component was much greater than the others) so that

obvious TOIs in the test data were ranked much lower in the dig-list than desired (Figure 7).

- Obvious clusters of features were present in the test data that were not represented in the training data (Figure 7).

Based on these observations, we chose to improve our estimates of class means and covariances by applying an unsupervised learning algorithm (expectation maximization, or EM) to the test data. This algorithm provides a maximum likelihood estimate of the components of a multivariate mixture of Gaussian distributions. The number of components (i.e. the number of means and covariances estimated) must be specified *a priori*, and the algorithm output can be sensitive to the initial model. While these shortcomings have been addressed using more advanced implementations of EM, here we used a manual initialization the initial model. The resulting means and covariances are shown in Figure 8. We see that clusters 1-3, all corresponding to regions of feature space occupied by TOIs, are a much better representation of the distribution test feature vectors than the covariances estimated with the training data alone (Figure 7).

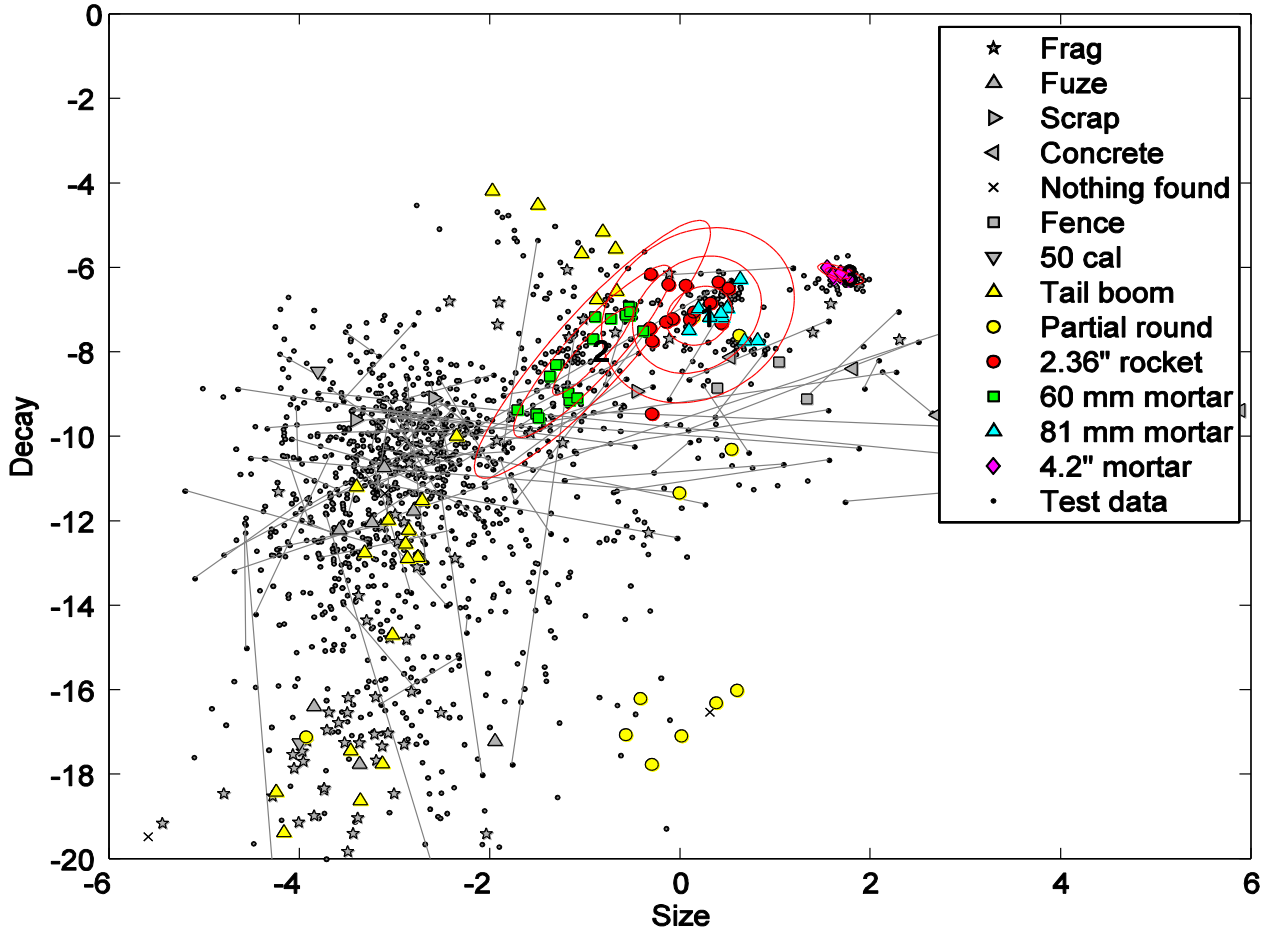


Figure 7. Size and decay features for training and test data. Concentric ellipses show 1,2, and 3 standard deviation ellipses for each TOI class, estimated from the training data alone. Test feature vectors connected by a line indicate multi-target inversions.

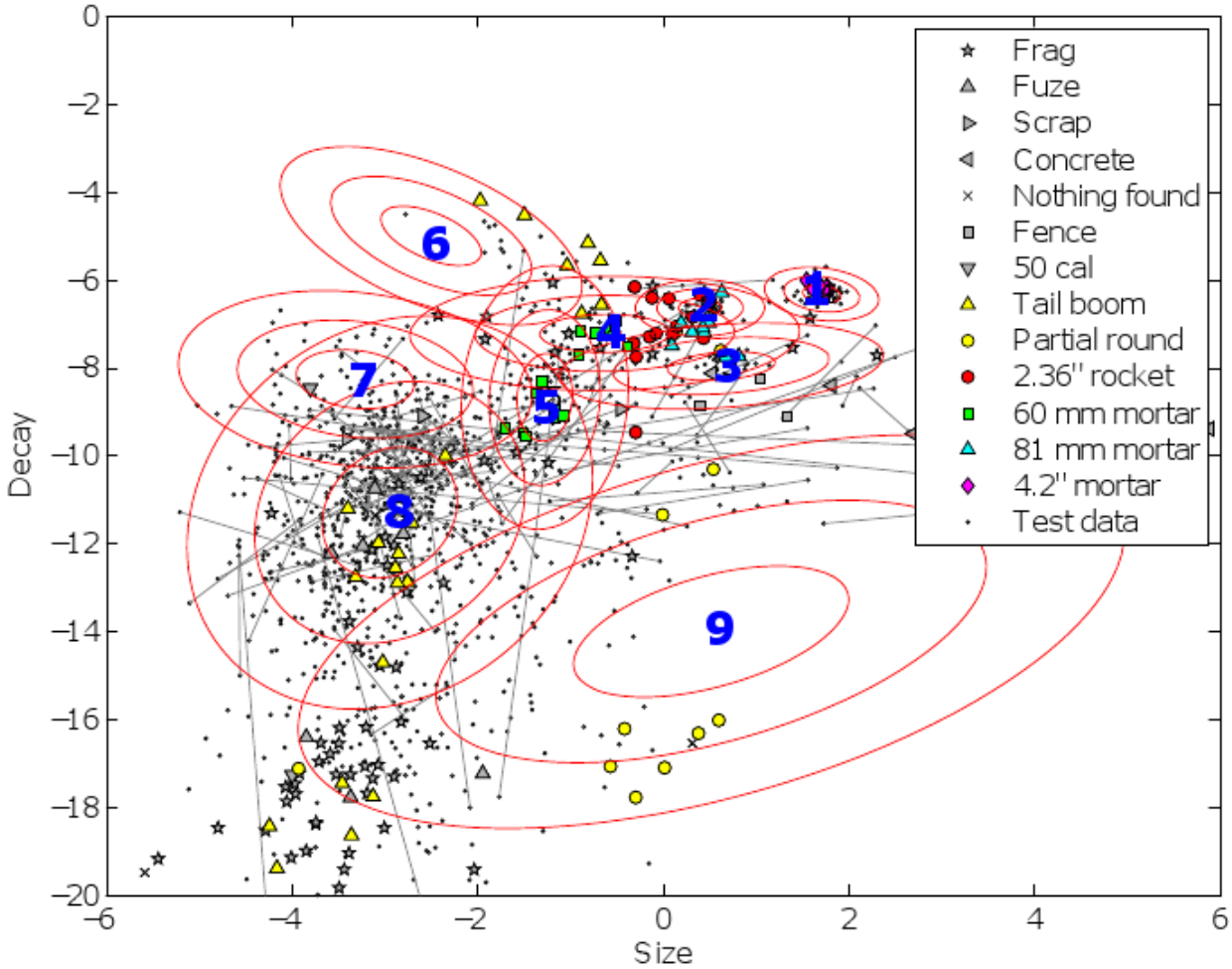


Figure 8. Size and decay features for training and test data. Concentric ellipses show 1,2, and 3 standard deviation ellipses for each TOI class, estimated from the training and test data sets using expectation maximization (EM) algorithm.

To generate a diglist using the clusters estimated with the EM algorithm we used the following procedure:

1. For each test datum compute the Mahalanobis distance (M_d, number of standard deviations) from the test feature vector to each cluster mean.
2. Find the smallest $M_d(TOI)$ for the test vector over all TOI clusters (1-5 in Figure 8) and the smallest $M_d(non\text{-}TOI)$ for the test vector over all non-TOI clusters (6-9 in Figure 8).
3. Compute a decision statistic $d=M_d(TOI) - M_d(non\text{-}TOI)$ for all test vectors, such that $d<0$ indicates that the target is likely a TOI and $d>0$ indicates that the target is likely not clutter

The resulting decision surface is a piecewise quadratic function in the feature space (Figure 9). Some adjustment of the clusters output by the EM algorithm was required to ensure that high

confidence TOI test targets were found, for example cluster 9 (Figure 8) had its mean in the decay parameter shifted slightly downward.

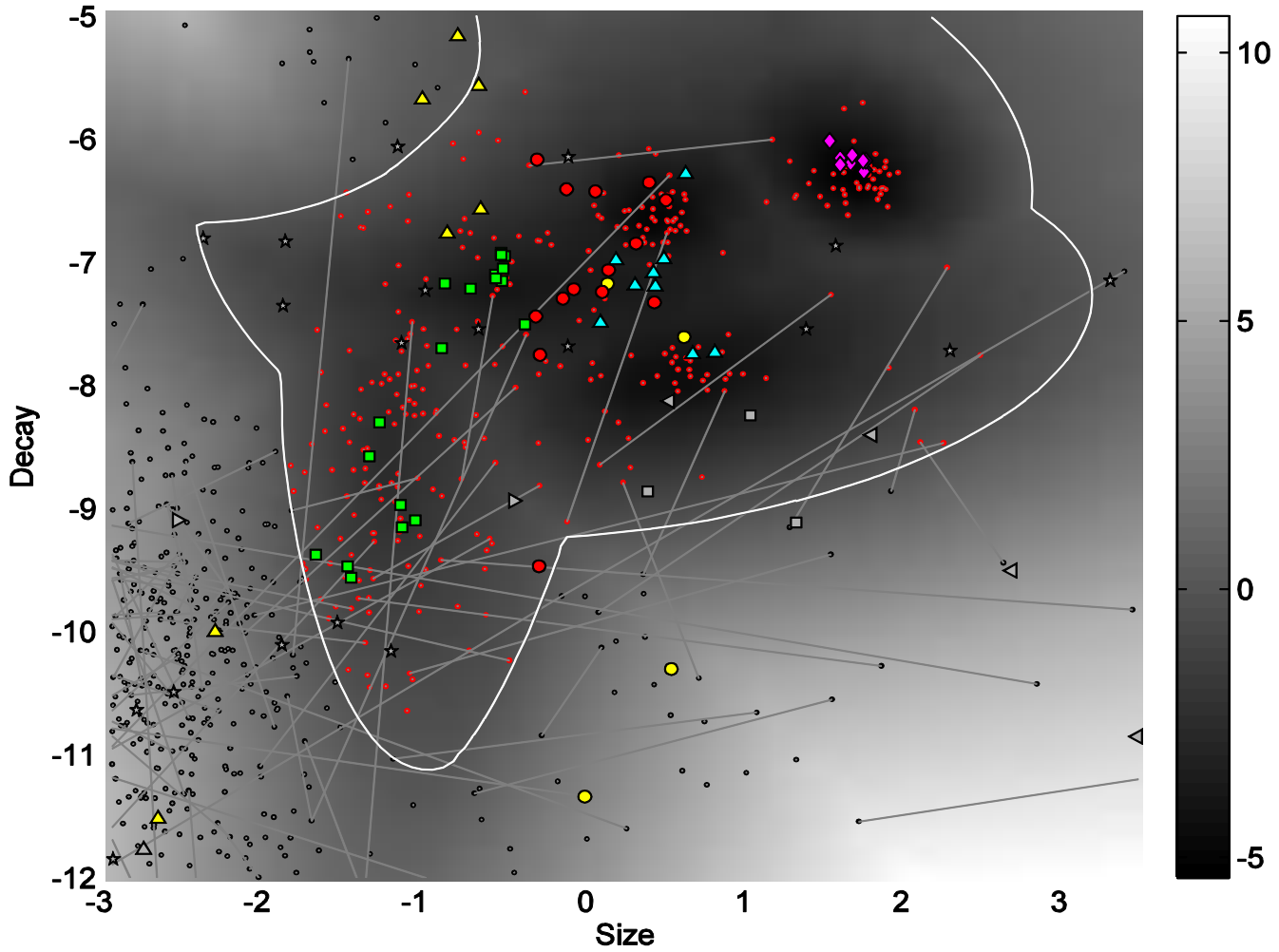


Figure 9. Decision statistic for TEMTADS classification. Contour shows the decision boundary and features vectors in red indicate test feature vectors flagged for digging

7.0 PERFORMANCE ASSESSMENT

IDA scored each of the 11 submissions and created receiver operating characteristic (ROC) curves for each. The IDA-supplied ROC curves are presented in Figure 10 to

Figure 12. Each of the ROC curves includes the anomalies used for training and testing, with the training data occurring first and represented by a black dot (number of TOI and non-TOI in the training data). The grey-shaded region represents the 95% confidence interval on the ROC curve. Green lines represent category 1 (high confidence not TOI), yellow lines category 2 (can't decide) and red lines category 3 (high-confidence TOI), with the gap between the black dot and the start of the red-curve indicating category 4 items (can't analyze).

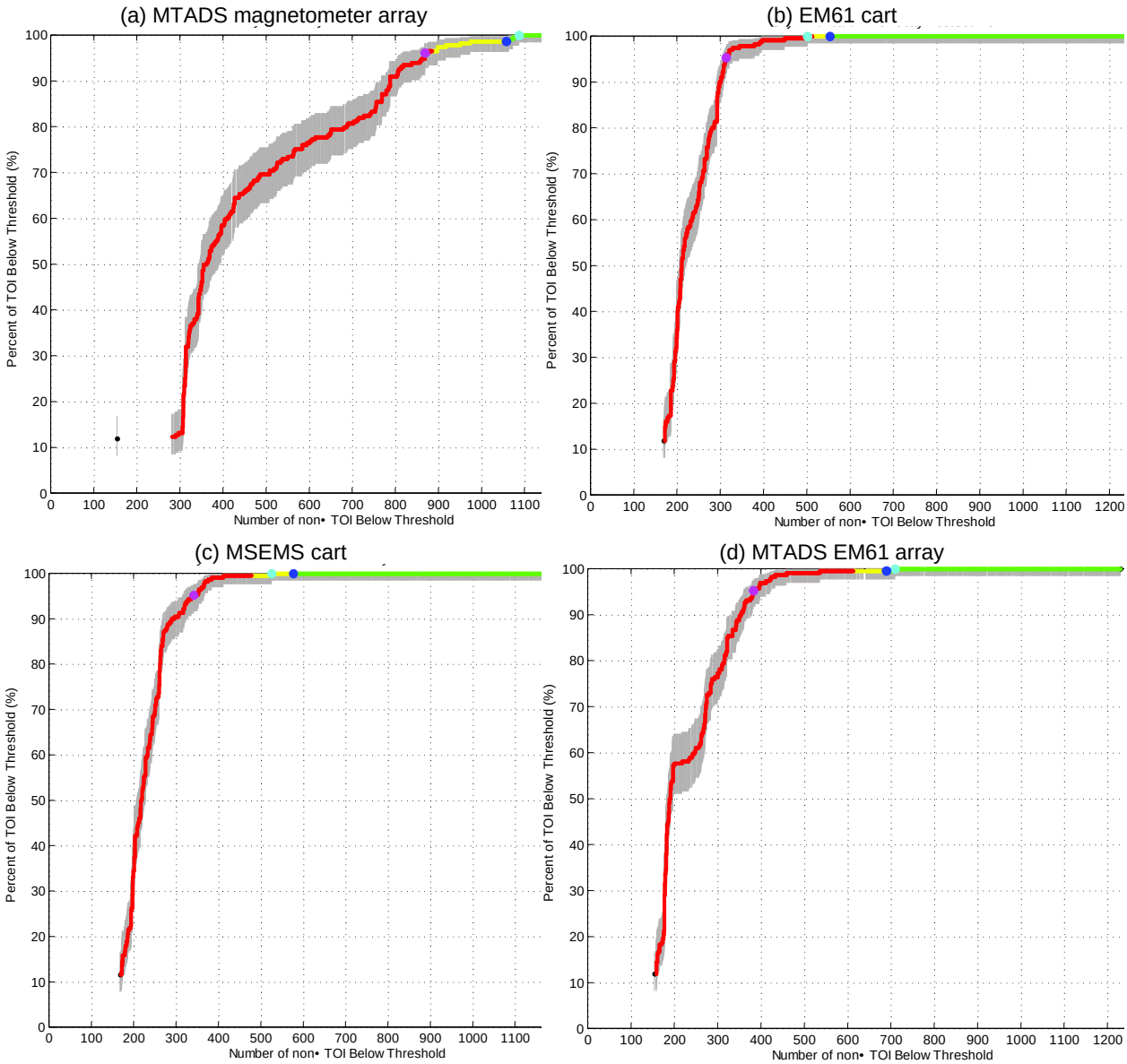


Figure 10. ROC curves provided by IDA for the production datasets. Each of these ROC curves includes both the test and the training data.

Three large dots are plotted on the ROC curve, each specifying one particular dig threshold:

- Dark Blue = the demonstrator's dig threshold;
- Light Blue = the first "best case scenario" dig threshold, that which, in retrospect, would have resulted in the fewest Number of Unnecessary Digs while the Percent of Munitions Dug was 100%; and
- Pink = the second "best case scenario" dig threshold, that which, in retrospect, would have resulted in the fewest Number of Unnecessary Digs while the Percent of Munitions Dug was 95% (or just barely greater than 95%).

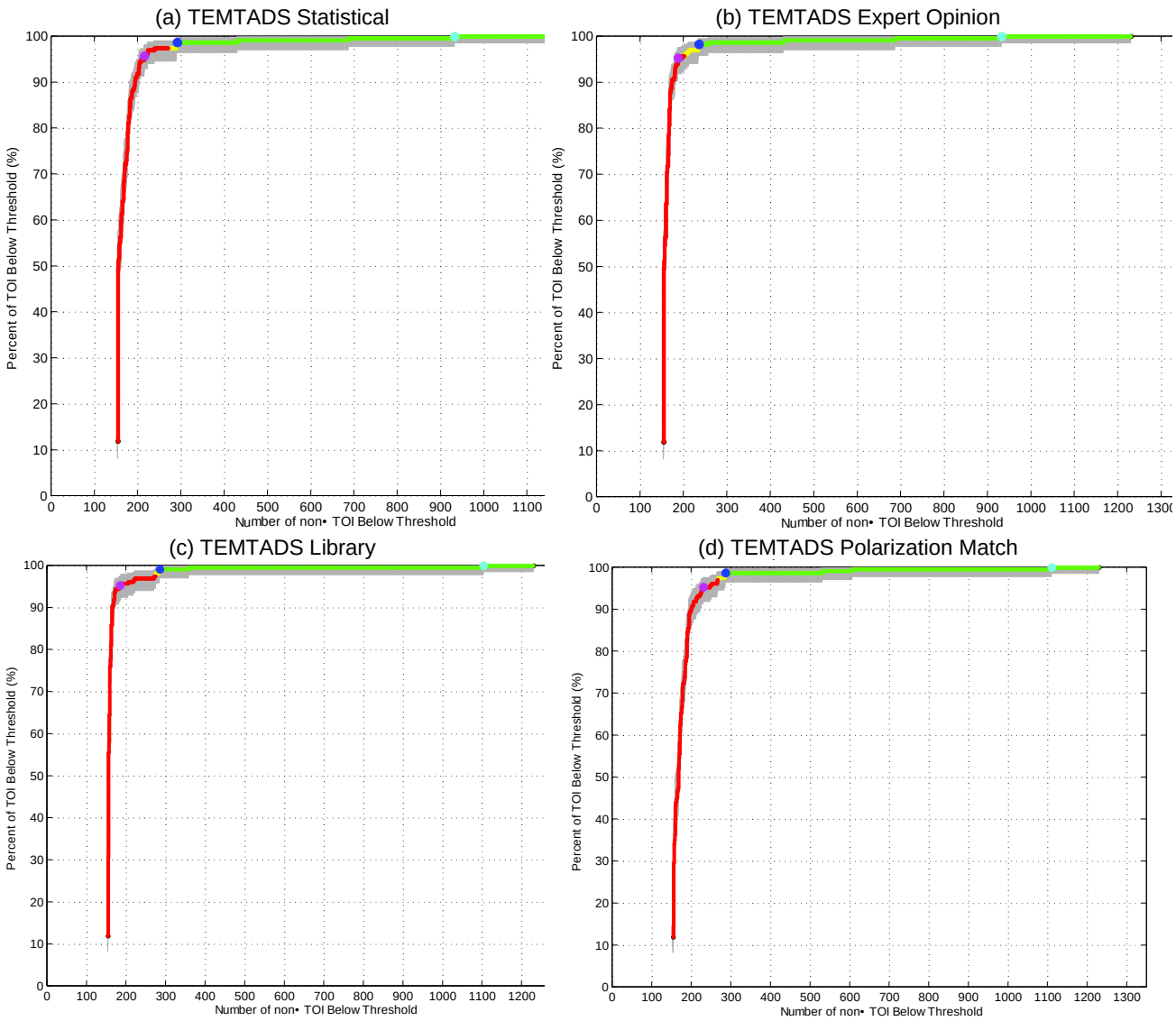


Figure 11. ROC curves provided by IDA for the TEMTADS datasets. Each of these ROC curves includes both the test and the training data.

Summaries of the number of detections, TOI, the operating point (OP), and the numbers of true-positives, false-positives, false-negatives, can't analyze and can't decide anomalies are listed in Table 3. These numbers only include the items in the test dataset (and not the training dataset).

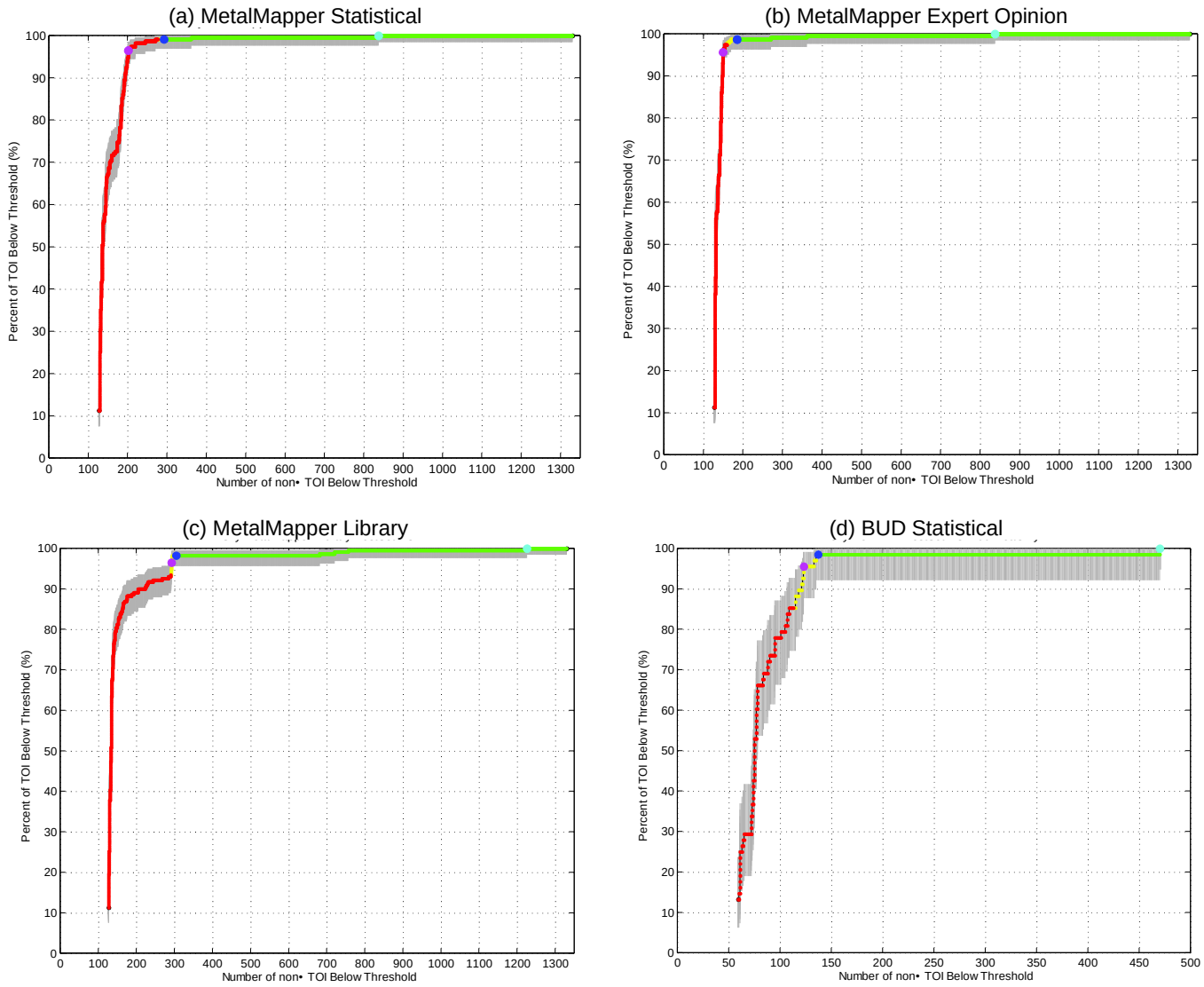


Figure 12. ROC curves provided by IDA for the MetalMapper and BUD datasets. Each of these ROC curves includes both the test and the training data.

The MetalMapper and TEMTADS require excavation of significantly fewer non-TOI (4.7 to 14%) than any of the EM production datasets (35 to 49%). However, both of these advanced sensors created false-negatives which are listed in Table 4. At least one of the false-negatives is questionable (Master ID 241: parts of a 2.36" rocket). For MetalMapper statistical, the only other false-negative was an unexpected 37mm projectile (Master ID 1502), with the expert interpretation creating an additional false-negative (Master ID 775, a multi-object scenario). For the TEMTADS statistical, there were a couple of deep 60mm mortar bodies that ended up as false-negatives (Master IDs 16 and 103). The first of these was actually a QC oversight on our part: the anomaly was identified in the QC spreadsheet as "can't decide" but was not placed in

that category when the final dig-list was prepared. The original submission for the MetalMapper Library contained a coding mistake: when this was corrected only one false negative results (Master ID 241). Table 5 presents the position and depth errors of recovered polarizabilities. Additional analysis of the performance of the different methods is provided in the sections that follow.

Table 3. Summary of classification results at SLO. If Master ID 241 was in-fact a non-TOI then the number of false-negatives (FN) in each of the cued-methods would be reduced by 1. The "TEMTADS Statistical (adjusted)" method has Master ID 16 moved to the "can't decide" class (this was a QC mistake). The "MetalMapper Library (adjusted)" method fixes a coding mistake made in the original submission.

Method	Number of items	Number of TOI	Operating point	True Positives	False Positives	False Negatives	Number of "Can't Analyze"	Number of "Can't Decide"	%TOI dug	%FN dug
Magnetics	1200	200	460	196	264	4	128	176	98.0%	26.4%
EM61 cart	1276	208	586	208	378	0	2	40	100.0%	35.4%
MSEMS cart	1284	205	599	205	394	0	1	99	100.0%	36.5%
EM61 array	1187	206	685	205	480	1	4	76	99.5%	48.9%
TEMTADS Statistical	1282	206	340	203	137	3	0	23	98.5%	12.7%
TEMTADS Statistical (adjusted)	1282	206	341	204	137	2	0	23	98.5%	12.7%
TEMTADS Library	1282	206	335	204	131	2	0	14	99.0%	12.2%
TEMTADS Expert	1282	206	283	202	81	4	0	39	98.1%	7.5%
TEMTADS Polarization Match	1282	206	335	203	132	3	0	23	98.5%	12.3%
MetalMapper Statistical	1409	204	368	202	166	2	0	15	99.0%	13.8%
MetalMapper Library	1409	204	378	200	178	4	0	27	98.0%	14.8%
MetalMapper Library (adjusted)	1409	204	378	203	175	1	0	27	98.0%	14.5%
MetalMapper Expert	1409	204	258	201	57	3	0	27	98.5%	4.7%
BUD Statistical	473	59	139	58	81	1	0	32	98.3%	19.6%

Table 4. List of false negatives encountered in each dig-list (not including the MetalMapper Library method where coding mistake resulted in several false-negatives that should have been identified as TOI).

Method	Master ID	Type	# digs past OP	Depth (cm)	Comment
Magnetics	600	60 mm mortar	33	1	Unfavorable orientation relative to inducing field some remanence
	1108	60 mm mortar	17	29	Small (just the mortar body) deep item with low SNR
	776	60 mm mortar	7	25	Small (just the mortar body) deep item with low SNR
EM61 cart	None				
MSEMS	None				
MTADS EM61	1444	60 mm mortar	21	32	Stop digging too soon. Deep horizontal 60 mm, array doesn't sufficiently excite primary polarizability
TEMTADS Statistical	241	2.36" rocket	644	0	Questionable ground-truth
	16	60 mm mortar	398	42	Poor estimate of decay rate, was caught by QC but category not manually changed to "can't decide"
	103	60 mm mortar	140	35	Poor SNR, should not have relied on polarization fit
TEMTADS Library	241	2.36"	820	0	Questionable ground-truth
	711	60 mm mortar	74	14	Multi-object scenario, made prediction on wrong anomaly
TEMTADS Expert	241	2.36" rocket	644	0	Questionable ground-truth
	16	60 mm mortar	398	42	Poor estimate of decay rate, was caught by QC but category not manually changed to "can't decide"
	103	60 mm mortar	140	35	Poor SNR, should not have relied on polarization fit
	1285	60 mm mortar	22	35	Multi-object scenario but used single object fit
TEMTADS Polarization Match	241	2.36" rocket	827	0	Questionable ground-truth
	103	60 mm mortar	322	35	Poor SNR, should not have relied on polarization fit
	16	60 mm mortar	246	42	Poor estimate of decay rate, was caught by QC but category not manually changed to "can't decide"
MetalMapper Statistical	241	2.36" rocket	547	0	Questionable ground-truth
	1502	37 mm projectile	70	1	Only 37 mm projectile found at the entire site
MetalMapper Expert	241	2.36" rocket	547	0	Questionable ground-truth
	775	60 mm mortar	87	26	Multi-object scenario fit as single object
	1502	37 mm projectile	70	1	Only 37 mm projectile found at the entire site
BUD Statistical	241	2.36" rocket	236	0	Questionable ground-truth

Table 5. Position and depth errors of recovered polarizabilities.

Method	Position error (cm)			Depth error (cm)		
	Bias-X	Bias-Y	Standard deviation	Bias-Z	RMS error	Standard deviation
EM61 cart	7.5	-5.8	19.6	-26	31.2	17.2
MSEMS cart	4.7	-3.7	18.9	-13.8	19.5	13.8
EM61 array	-2.5	2.6	22.7	-32.4	37.8	19.6
TEMTADS statistical				-4.1	11.5	10.7
MetalMapper				-4.5	8.5	7.1

7.1 ADDITIONAL ANALYSIS OF PRODUCTION DATASETS

Figure 13 shows depth and location recovery for each of the EM production data sets (EM61 cart, MSEMS and MTADS EM61 array). In general, estimated depths are deeper than true target depths. Some improvement in recovered depths is seen from cooperative inversion of the MSEMS data. The majority of recovered locations are within 20 cm of the reported ground-truth. Large discrepancies (>50 cm) between estimated and reported locations are likely due to ground-truth being associated with a neighboring anomaly.

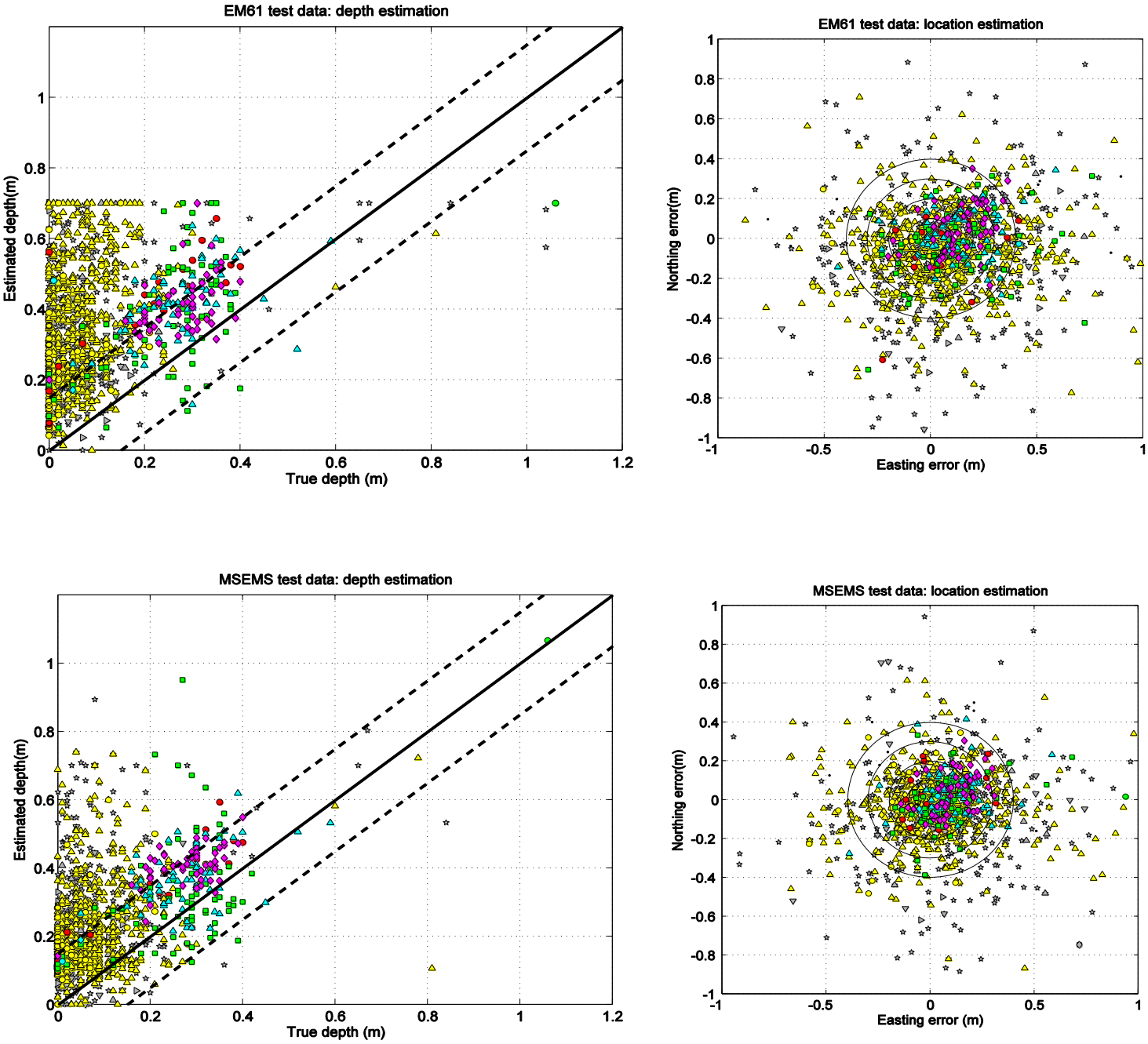


Figure 13. Location and depth estimation for EM61 cart, MSEMS and MTADS EM61 data sets. For depth estimates, dashed lines represent errors of +/- 15 cm.

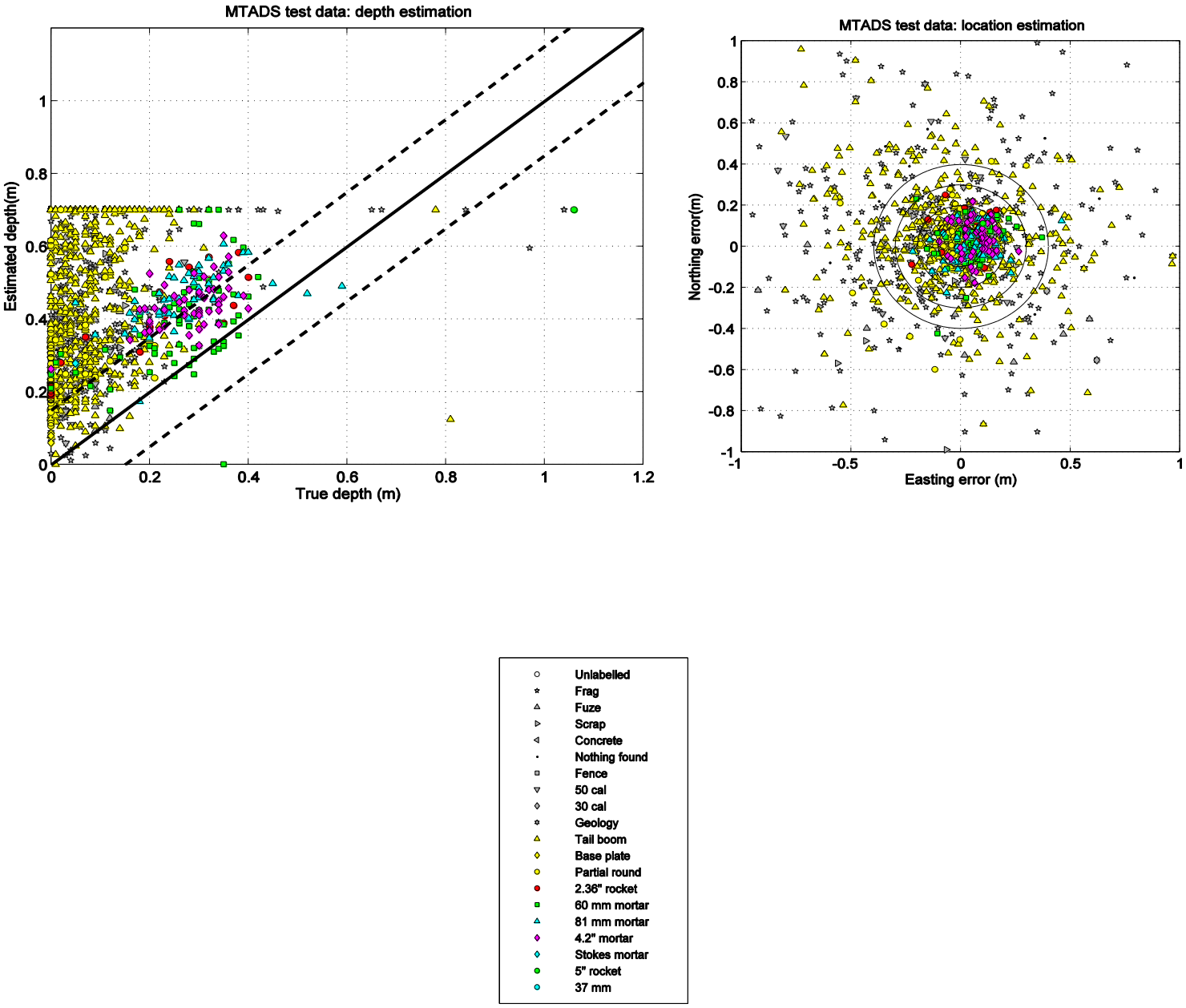


Figure 14. (cont'd) Location and depth estimation for EM61 cart, MSEMS and MTADS EM61 data sets. For depth estimates, dashed lines represent errors of +/- 15 cm.

For EM61 type data sets, we used a threshold on the polarization decay of the induced dipole moment to discriminate between TOI (slow decay) and non-TOI (fast decay). The polarization amplitude and decay were computed as:

$$\text{Polarization amplitude} = (\Sigma\, L_i(t_1)^2)^{1/2} \tag{18}$$

$$\text{Polarization decay} = (\Sigma\, L_i(t_4)^2)^{1/2} / (\Sigma\, L_i(t_1)^2)^{1/2} \tag{19}$$

Figure 15 shows training and test features for EM data sets. Outlying TOI in MSEMS and MTADS data sets have been circled. Because target depth is poorly estimated for these data sets, some TOI have small polarization amplitudes. The test data generally support our hypothesis that TOI are slower decaying than non-TOI.

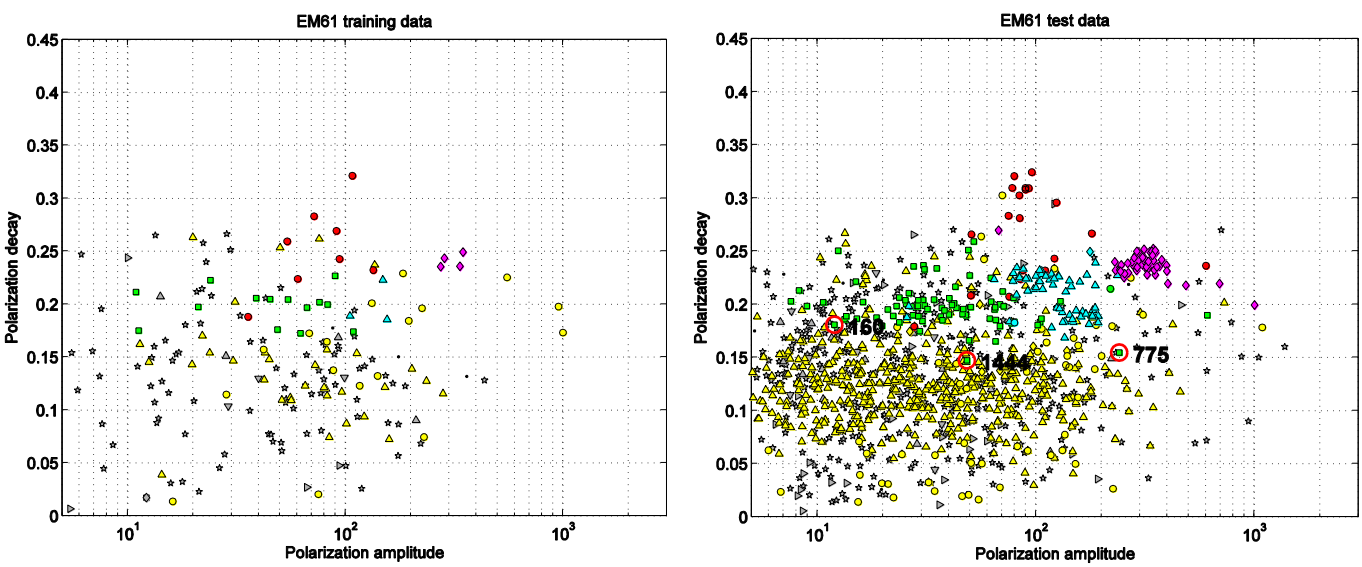


Figure 15. Training and test sets for EM61 cart, MSEMS, and MTADS EM61 data, see legend in Figure 13.

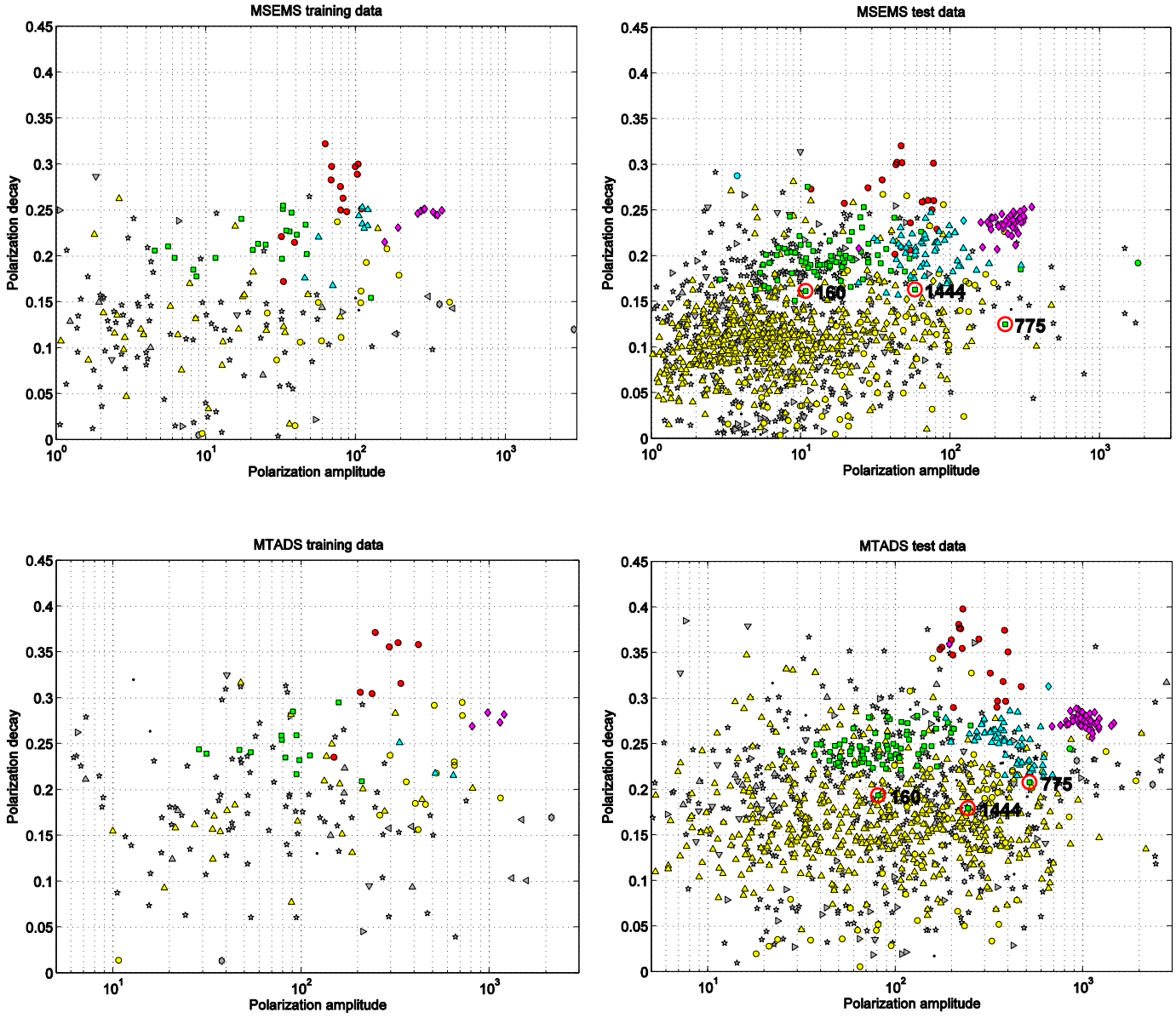


Figure 16. (cont'd) Training and test sets for EM61 cart, MSEMS, and MTADS EM61 data, see legend in Figure 13.

We compare ROC curves for the various TOI classes in Figure 17. The EM61 cart has the best overall performance in terms of area under the curve (AUC) and false alarm rate. All EM production surveys significantly outperform discrimination with magnetics data. 60 mm mortars are consistently the most difficult target class to identify for these sensors.

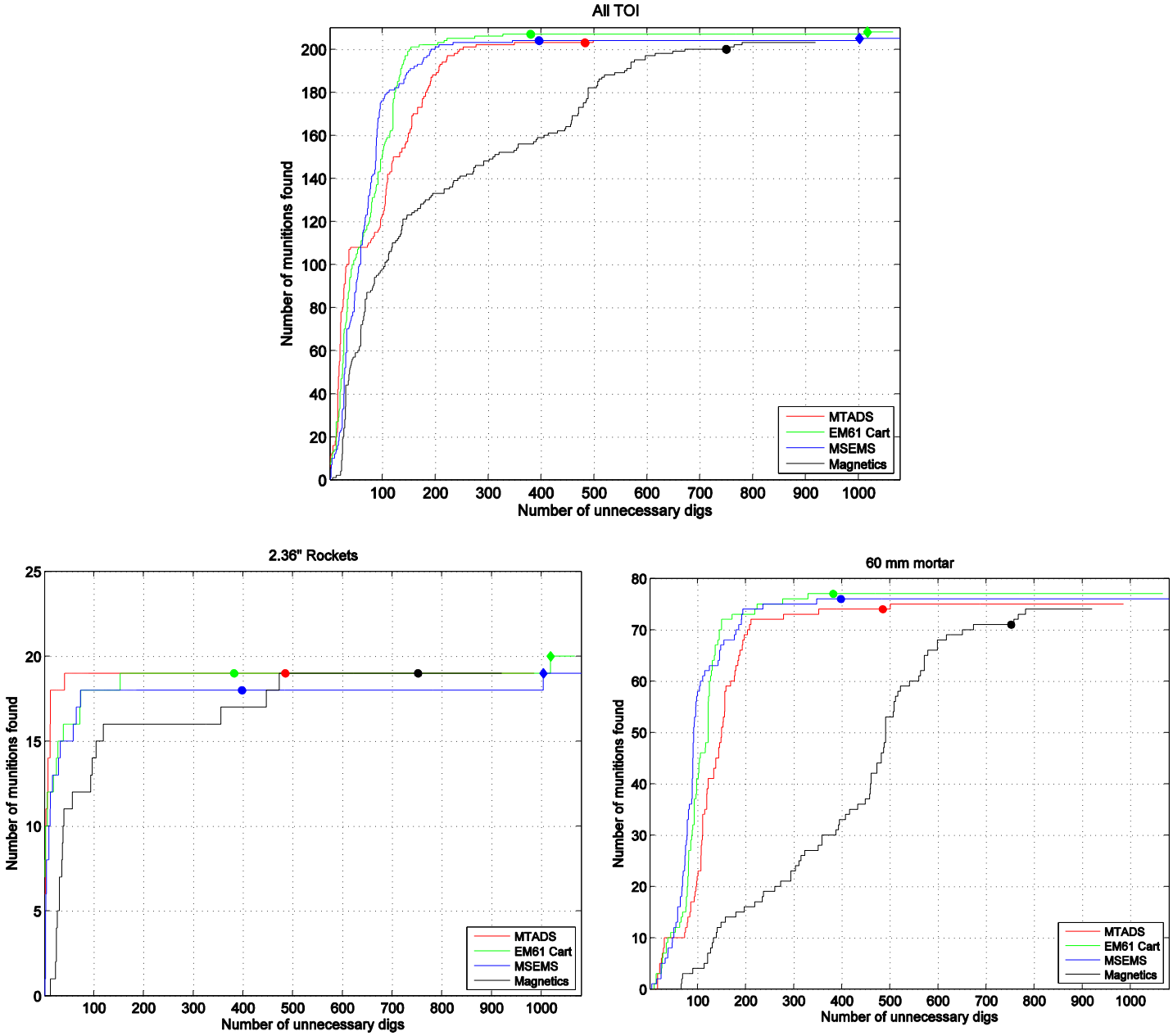


Figure 17. ROC curves for TOI classes, EM61 cart, MSEMS, MTADS EM61 and magnetics. Circles indicate operating points.

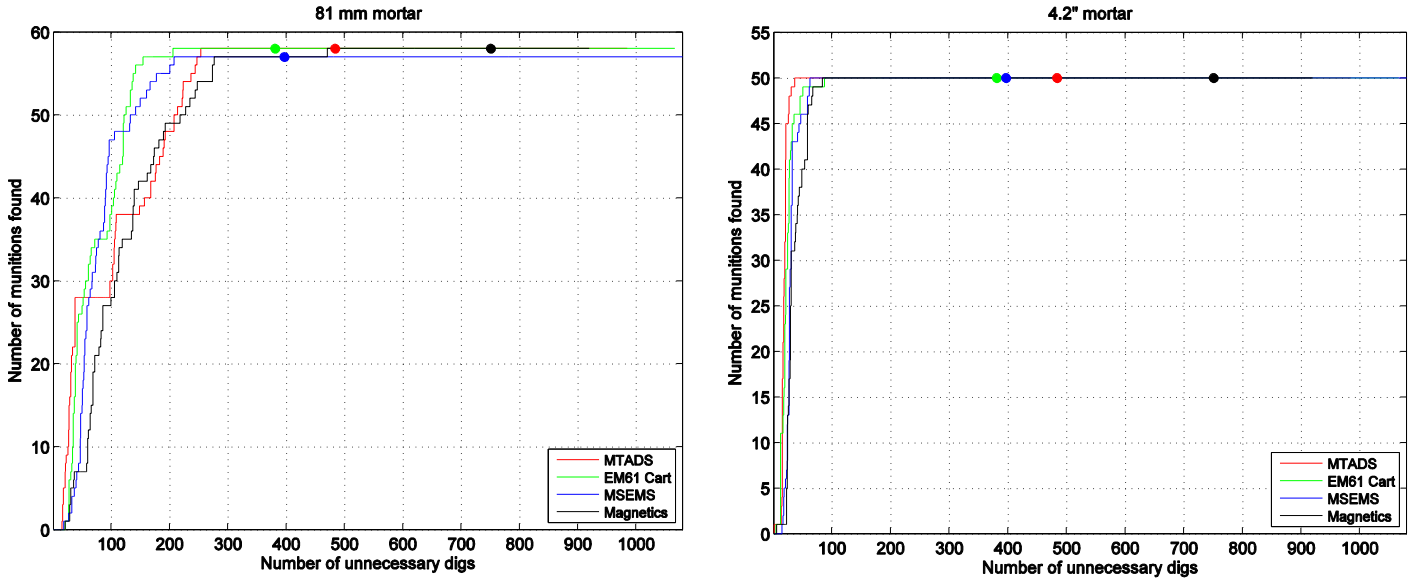


Figure 18. (cont'd) ROC curves for TOI classes, EM61 cart, MSEMS, MTADS EM61 and magnetics. Circles indicate operating points.

To further understand the performance differences for the production data sets, we compare data and fits for the same targets (highlighted in Figure 15) in the three EM production surveys. In Figure 19 we show Master ID 775, which is identified as a 60 mm mortar at 26 cm depth and has an unusually fast decay in the EM61 cart and MSEMS test data (it is the last TOI target found in the MSEMS data).

Figure 19. Master ID 775 which comprises a 60 mm mortar body in the bottom of the hole and two tail fins.

The location of Master ID 775 in the feature spaces in Figure 15 suggests that this target is in fact an 81mm mortar. However, the optimal target depth for all inversions is much deeper than the reported target depth (Figure 20), so that the polarization amplitude - which is positively

correlated with target depth - is also overestimated. Figure 21 compares fits for this target. At first inspection, there is no obvious fault with any of these fits. However, closer examination of the fit for the MSEMS indicates that the peak amplitudes of the data are slightly underfit in the inversion. This suggests that the recovered polarizabilities decay slightly faster than the observed data, so that this target is an outlier in the test data. One solution to this problem is to only fit data within a very tight mask (Figure 22). This emphasizes high SNR data and produces a decay estimate which is more in line with that expected for 60mm targets (Figure 23).

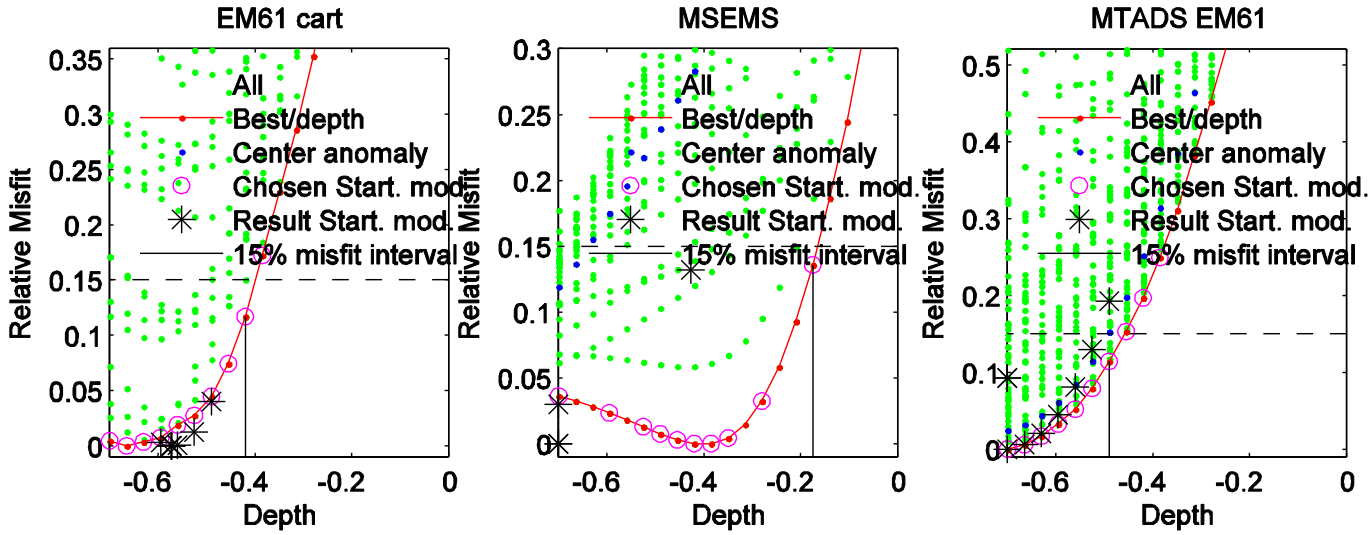
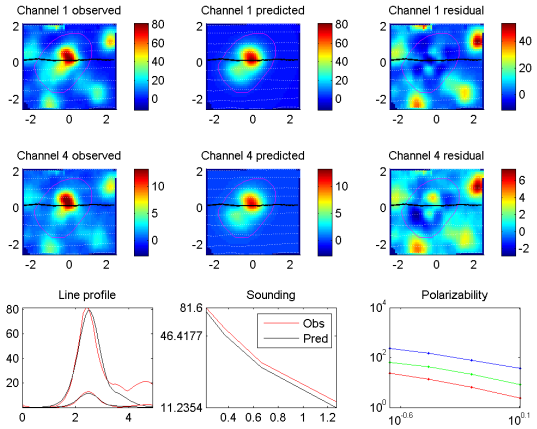
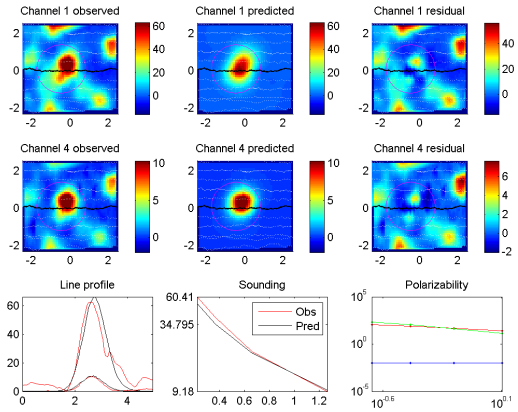


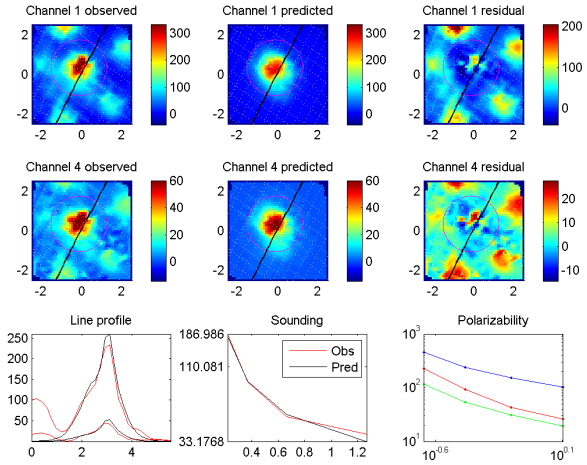
Figure 20. Misfit versus depth curves for inversions of Master ID 775.



(a) EM61 cart



(b) MSEMS



(c) MTADS EM61

Figure 21. Comparison of fits over Master ID 775 (60 mm mortar). Line profile is black line in gridded images, showing observed and predicted data at channels 1 and 4.

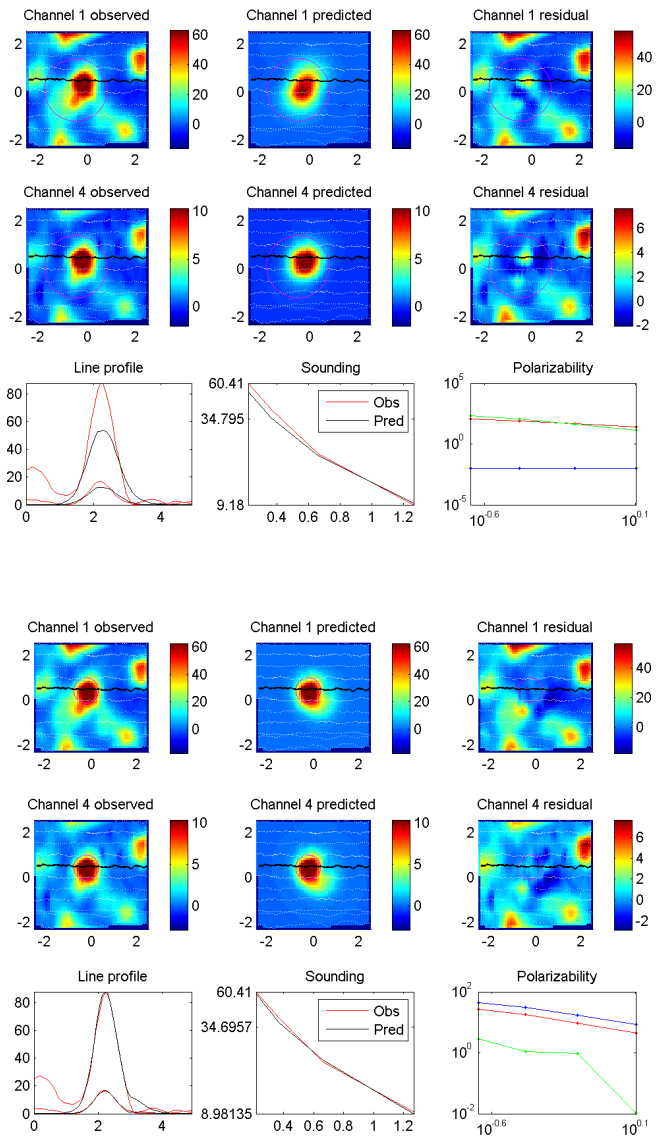


Figure 22. Comparison of fits for Master ID 775. Left: original fit, Right: retrospective fit with tight mask. Plotted line profile is adjacent to the line profile shown in figure 6(b).

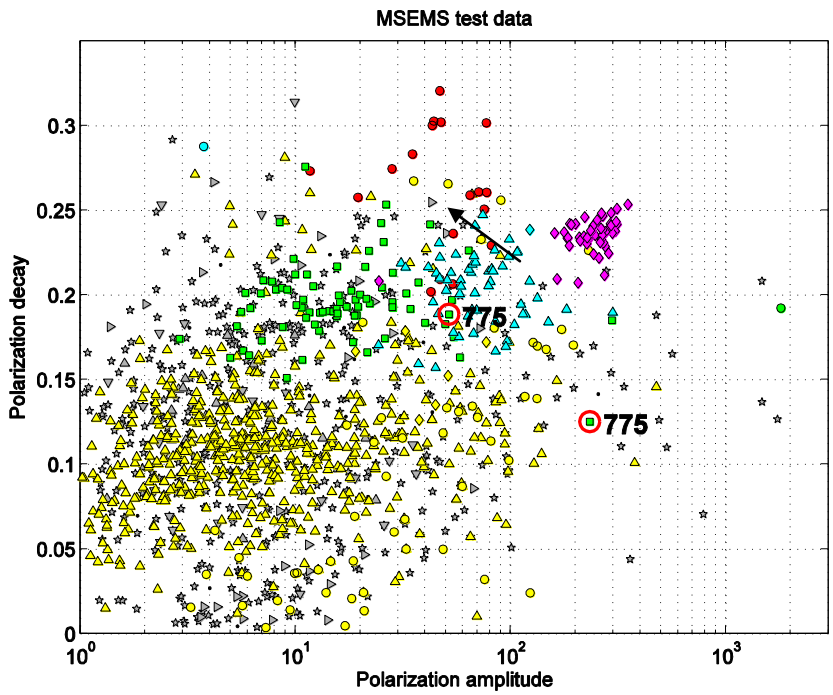
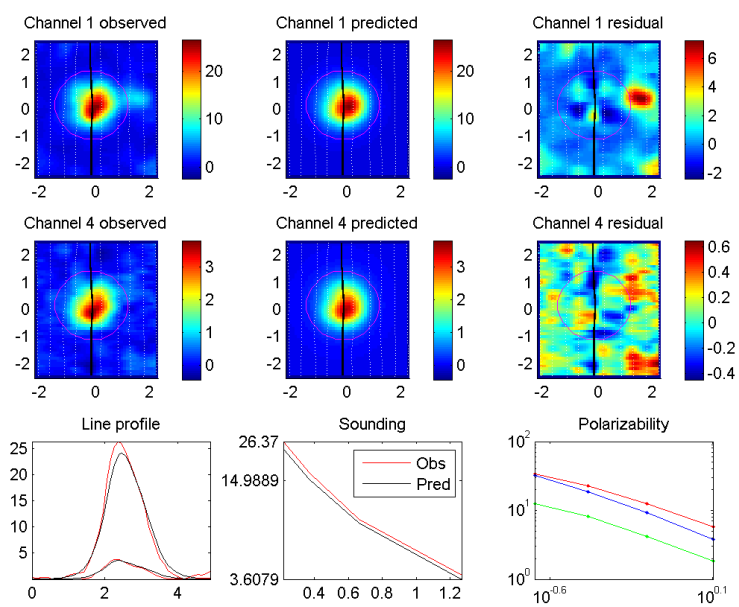
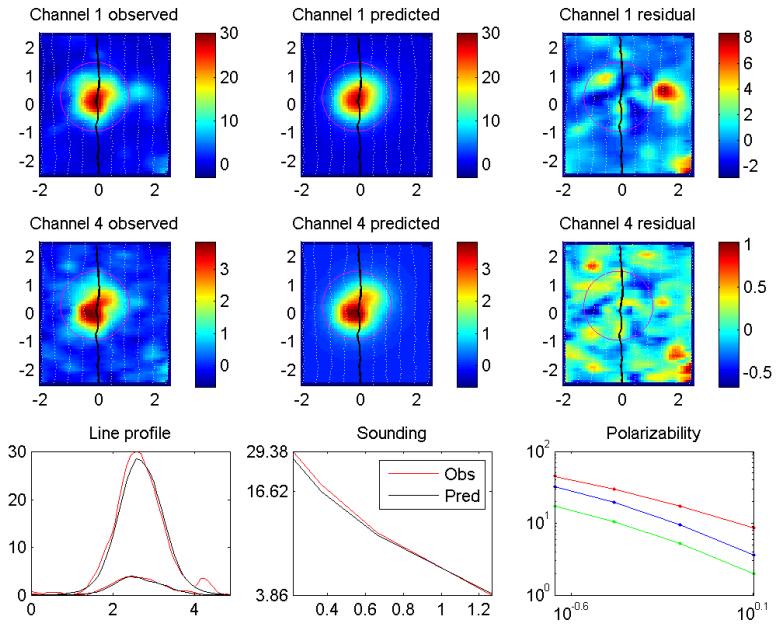


Figure 23. Change in estimated features for Master ID 775 after retrospective inversion with a tight mask.

We next consider Master ID 1444 (Figure 24), which is a false negative in the MTADS data (and something of an outlier in the EM61 cart data). While this target is relatively deep for a 60mm target (32cm) and so has low SNR, there is no obvious fault with any of the fits for this target (

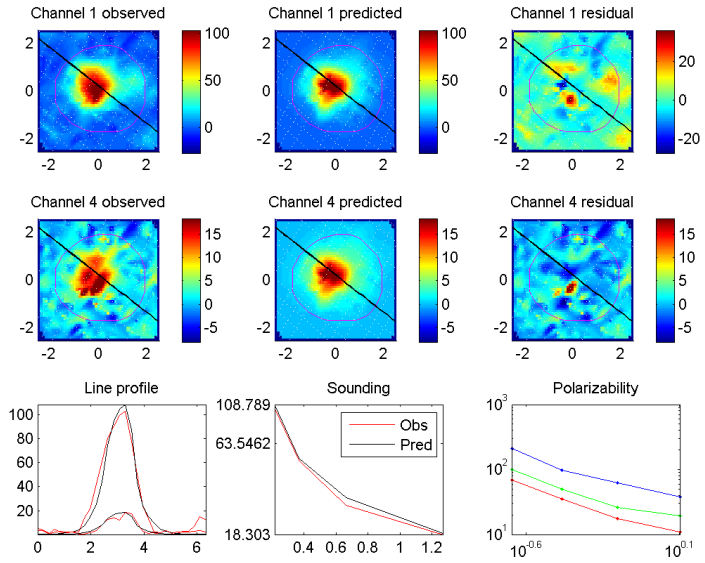


(a) EM61 cart



(b) MSEMS

Figure 25. Comparison of fits over Master ID 1444 (60mm mortar). Line profile is black line in gridded images, showing observed and predicted data at channels 1 and 4.



(c) MTADS EM61

Figure 26). The misclassification of this target in the MTADS data is therefore more an issue of the stop-dig point for this diglist, rather than faulty estimation of features. Master ID 160, which is also highlighted in Figure 15, also appears as something of an outlier in the MTADS test data.

Again there is no obvious problem with the fits, and this target was correctly identified at the specified operating point.

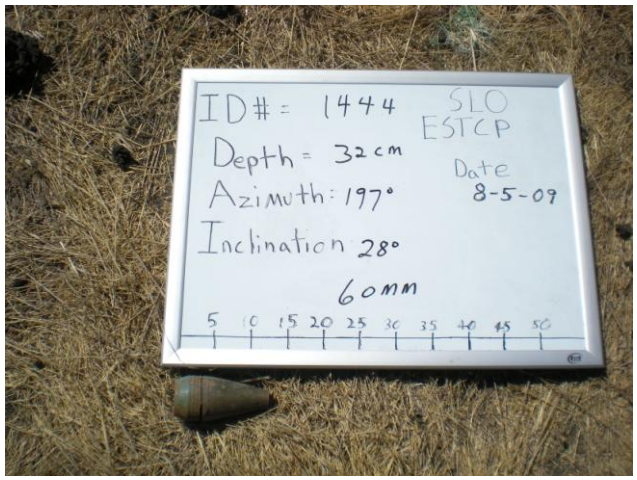
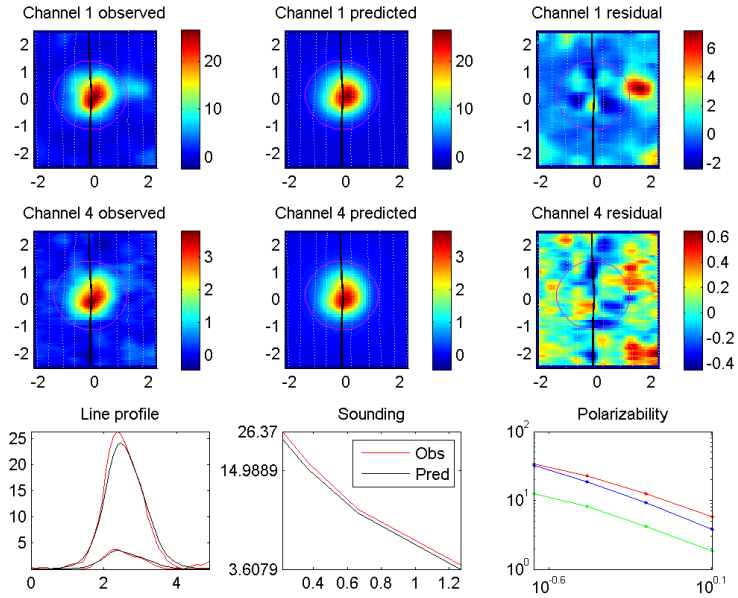
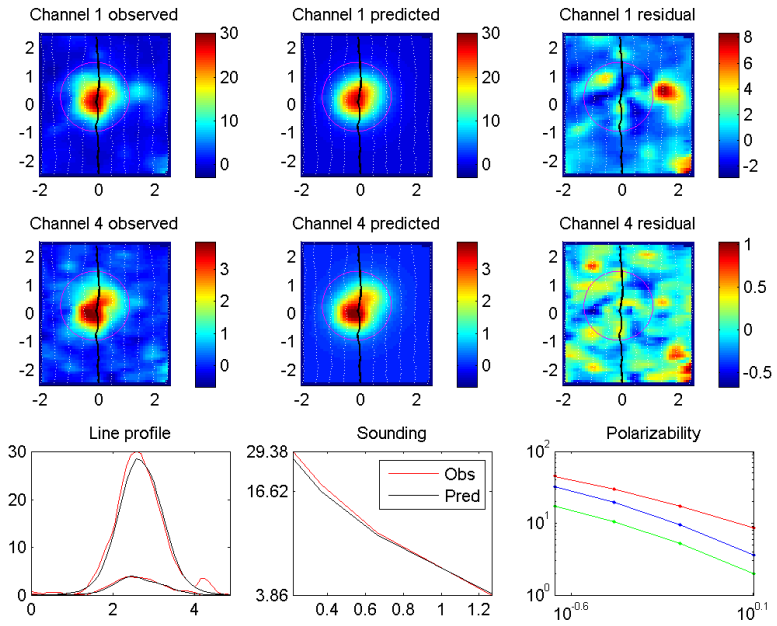


Figure 24. Master ID 1444, a deep 60mm body with unfavorable orientation.

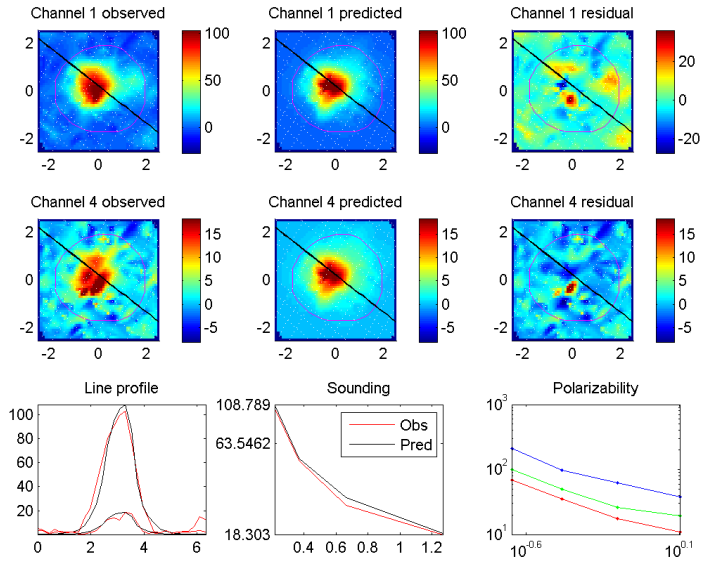


(a) EM61 cart



(b) MSEMS

Figure 25. Comparison of fits over Master ID 1444 (60mm mortar). Line profile is black line in gridded images, showing observed and predicted data at channels 1 and 4.



(c) MTADS EM61

Figure 26. (cont'd) Comparison of fits over Master ID 1444 (60mm mortar). Line profile is black line in gridded images, showing observed and predicted data at channels 1 and 4.

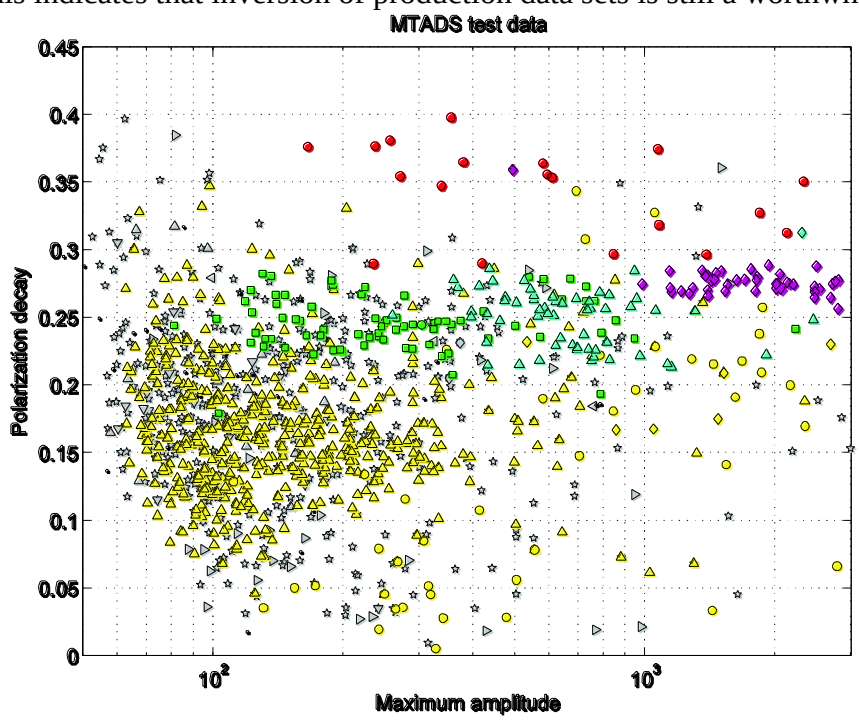
For generation of diglists using polarization decay, we used the following procedure:

1. Identify high SNR anomalies. For MTADS data we used the maximum data value within the mask and defined any anomaly with a peak data >200 milliVolt (mV) to be high SNR. This captures the smallest 81 mm in the training data, but leaves some training 60 mm and 2.36" warheads as low SNR anomalies.
2. Threshold on decay rate for high SNR anomalies down to the mean decay rate for 60 mm.
3. Threshold on decay rate for low SNR anomalies down to the mean decay rate for 60 mm.
4. Threshold on decay rate for the remaining targets, irrespective of SNR

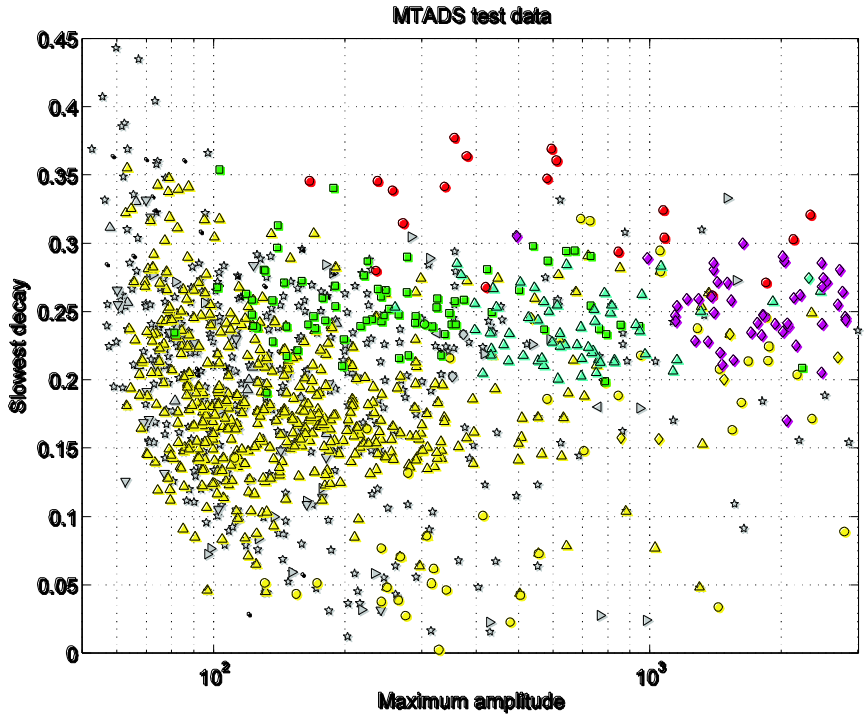
For the MTADS data this procedure produced an obvious kink in the ROC (Figure 17) when we applied step 3. This implies that there were a number of low amplitude, slow decaying non-TOI, as seen in the upper left of Figure 27 (a). For comparison, we also consider a data-based decay feature (slowest decay) in Figure 27 (b). The slowest decay is computed as:

$$\text{Slowest decay} = \max\left(\frac{d(t_4)}{d(t_1)}\right) \tag{20}$$

with the maximum taken over all soundings satisfying $d(t_1) \geq 0.85\max(d(t_1))$. The class distributions in Figure 27 (a) and (b) are qualitatively similar, but the TOI distributions are somewhat tighter for the model-based feature than for the data-based feature, particularly for larger TOI. This indicates that inversion of production data sets is still a worthwhile effort.



(a) Polarization decay versus maximum anomaly amplitude, MTADS test data.



(b) Data features: slowest decay versus maximum anomaly amplitude, MTADS test data.

Figure 27. Comparison of model and data based decay features for MTADS test data.

7.2 ADDITIONAL ANALYSIS OF TEMTADS DATASETS

Figure 28 shows estimated training and test features and accuracy of depth estimation for the TEMTADS data set. Depth recovery for these data is much improved relative to the production data sets. In Figure 28 (b) we highlight a number of targets, including two target classes not seen in the training data (Stokes mortar and 5" rocket) and a number of test 60 mm targets with anomalously fast (i.e. small) primary decay rate. In the following section, we focus on improving parameter estimates for these 60 mm outliers.

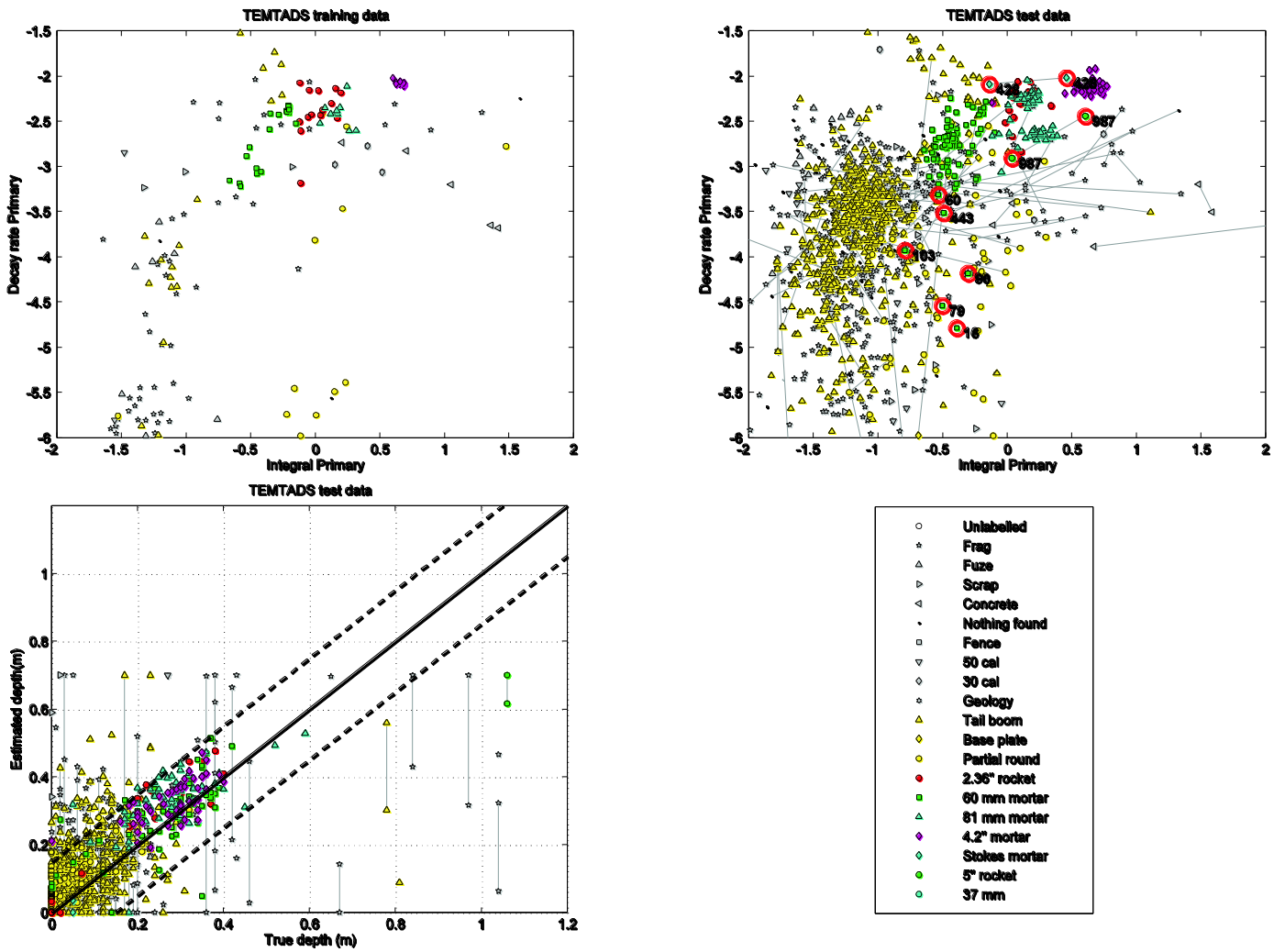


Figure 28. Analysis of TEMTADS data. Feature vectors connected by a line are multi-target inversions.

Four diglists were submitted for this data set:

1. Expert: ranking based upon visual comparison of estimated polarization with library polarizations
2. Library: fingerprinting method with polarizations fixed at library values during inversion. Decision statistic is the maximum correlation coefficient between the observed data and the data predicted by the model.
3. Polarization: match of estimated primary polarization with library polarization. Decision statistic is the minimum misfit between estimated and library primary polarizations
4. Statistical: Non-parametric classifier trained on primary polarization amplitude and decay.

Visual inspection of all fits in the ordered diglists was carried out to ensure that each technique did not miss any obvious TOI. For the statistical classifier, a number (but not all) of the outlying test 60 mm targets were identified in this QC process. Figure 29 shows ROC curves generated

for these classifiers. In this figure we have corrected for one QC mistake: Master ID 16 (a 60 mm) was intended to be included in the "can't decide" category in the submitted diglists but was mistakenly left as "likely clutter." Here we have placed this target at the end of the "can't decide" category (immediately before the first "likely munitions") for expert, polarization and statistical diglists. The library method correctly identified this target as a likely munitions item and so no correction was applied. In Figure 29 we show the location of Master ID 241 in the diglist. As with Master ID 1541 in the production datasets, this target was difficult to identify in the TEMTADS diglists, and the groundtruth photo (Figure 30) raises some doubt as to whether this is actually a TOI.

Table 6. Summary of discrimination performance for TEMTADS classifiers.

Method	# Munitions found	FAR	AUC	False Negatives
TEMTADS Expert	203	279/1080	0.990627	1285 (60 mm) 103 (60 mm)
TEMTADS Library	203	207/1080	0.992301	711 (60 mm)
TEMTADS Polarization	203	455/1080	0.975543	103 (60 mm)
TEMTADS Statistical	203	279/1080	0.985263	103 (60 mm)

Table 6 compares the performance of the TEMTADS classifiers in more detail. The library method achieved the lowest false alarm rate (FAR) and highest AUC. Here the FAR is the number of unnecessary digs required to find all munitions in the ordered diglist (not the number of unnecessary digs at the specified operating point). In this analysis we have treated Master ID 241 as a non-TOI. For all classifiers, false negative targets were 60 mm targets, with Master ID 103 appearing as a false negative in expert, polarization and statistical diglists. Figure 32 shows the fit to the TEMTADS data for this target. It is evident in Figure 32 that the SNR at late times is quite low. To compute a decay parameter for input into statistical classification, we fit a function of the form:

$$L(t) = \sum_{i=1}^{N} A_i \exp(-t/\gamma_i) \tag{21}$$

to the estimated polarizations, with the time constants γ_i spaced logarithmically over the TEMTADS measurement times, so that we only need solve a linear inverse problem for the coefficients A_i. We then used this smoothed estimate to compute the ratio of the primary polarization at t_1=0.042 milliseconds (ms) (channel 1) and t_m=7.856 ms (channel 93). However, if the fitted decay extrapolates below the noise floor of the estimated polarizations, then we obtain a poor estimate of the actual rate of decay. This is seen in Figure 33, which shows the median SNR over all channels versus estimated decay. All fast-decaying 60 mm test targets in Figure 28 have relatively low SNR.

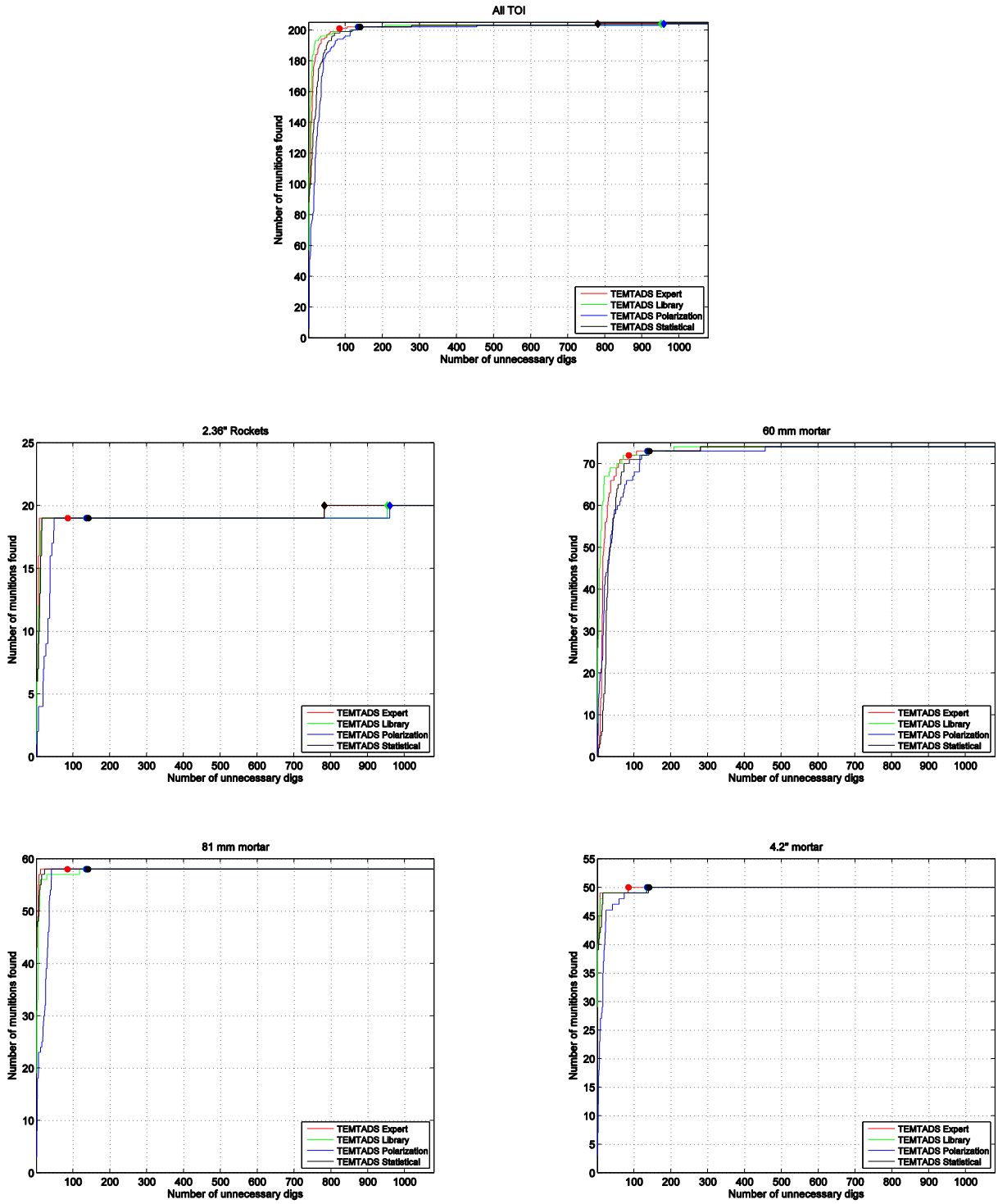


Figure 29. Receiver operating curves for TOI classes, TEMTADS classifiers. Circles are specified operating points, diamonds show the location of Master ID 241 (2.36" rocket) in the diglists.

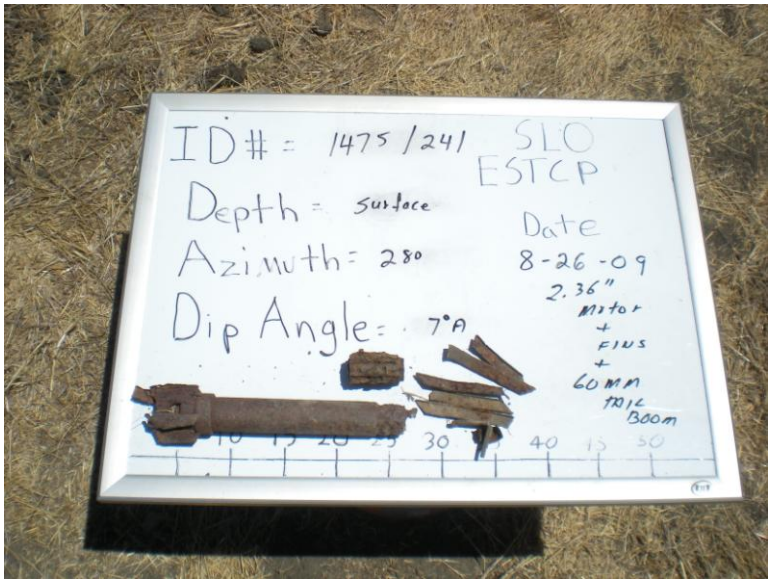


Figure 30. Master ID 241.

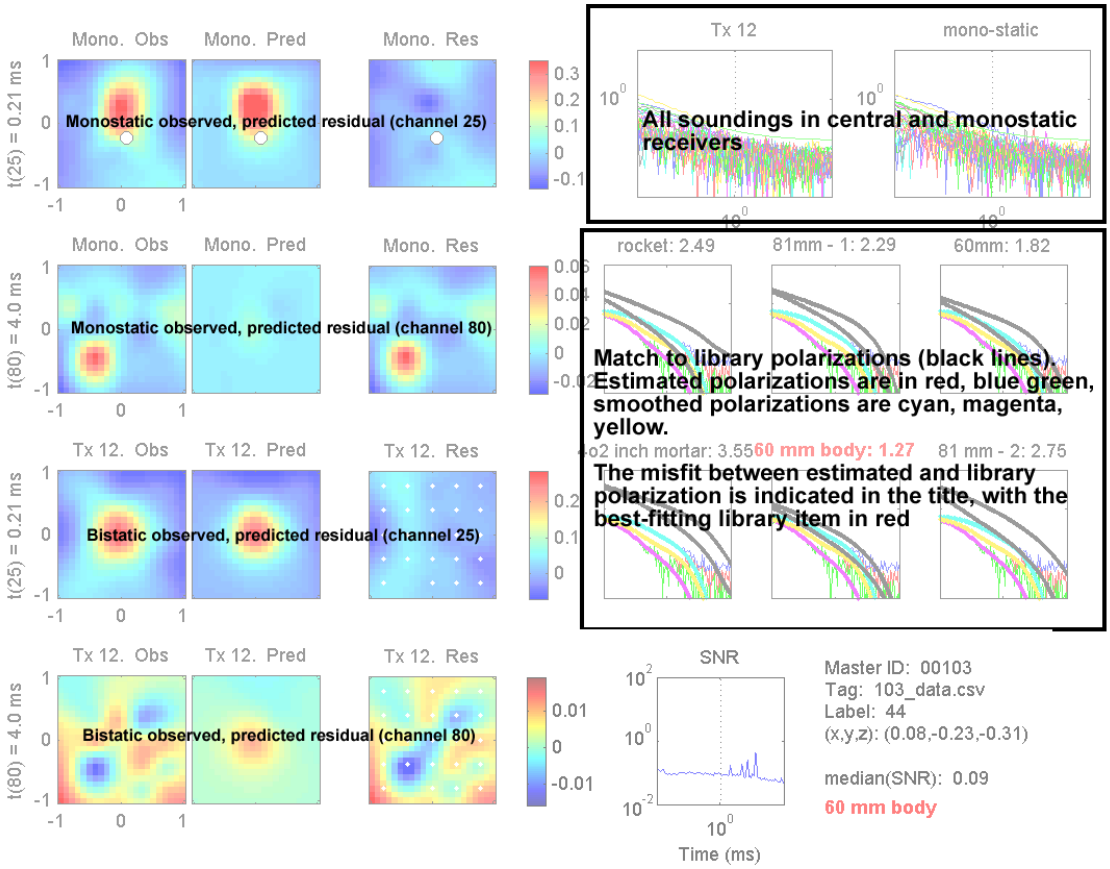


Figure 31. Legend for TEMTADS results plots.

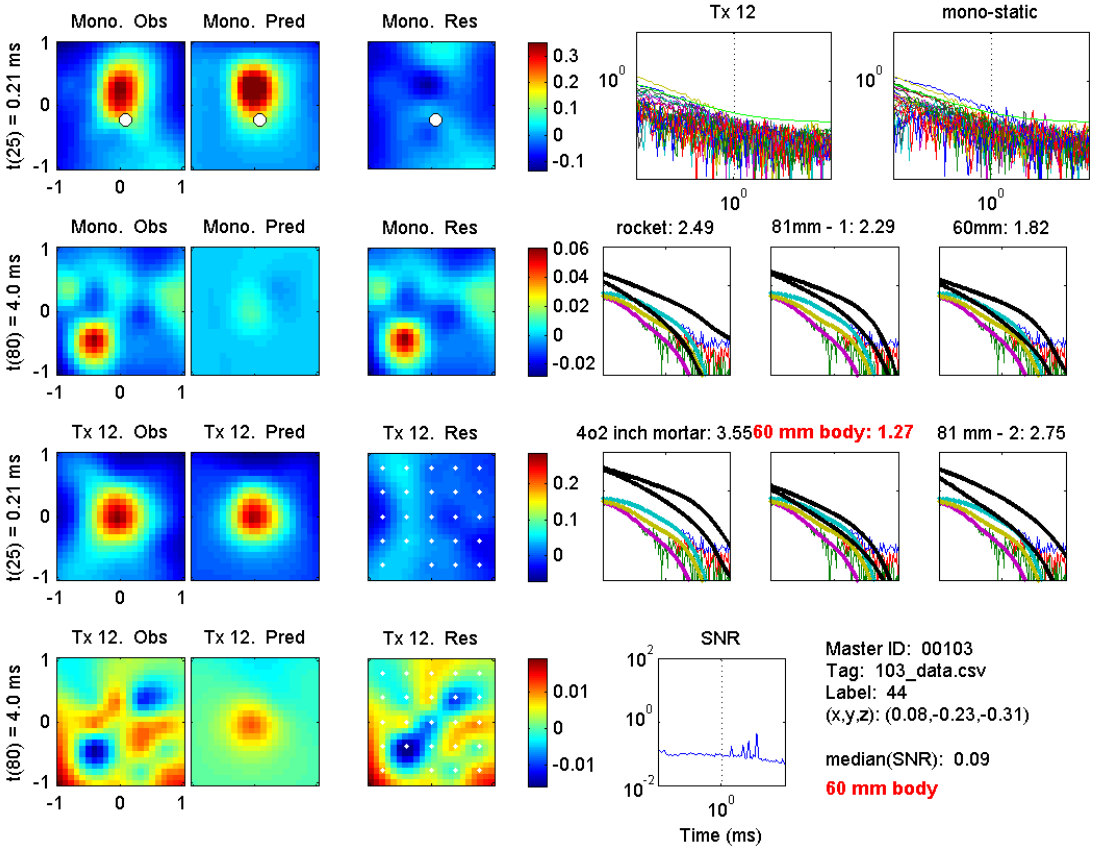


Figure 32. Fit to TEMTADS data for Master ID 103 (60 mm). Legend in Figure 31 explains layout of the inversion results plot.

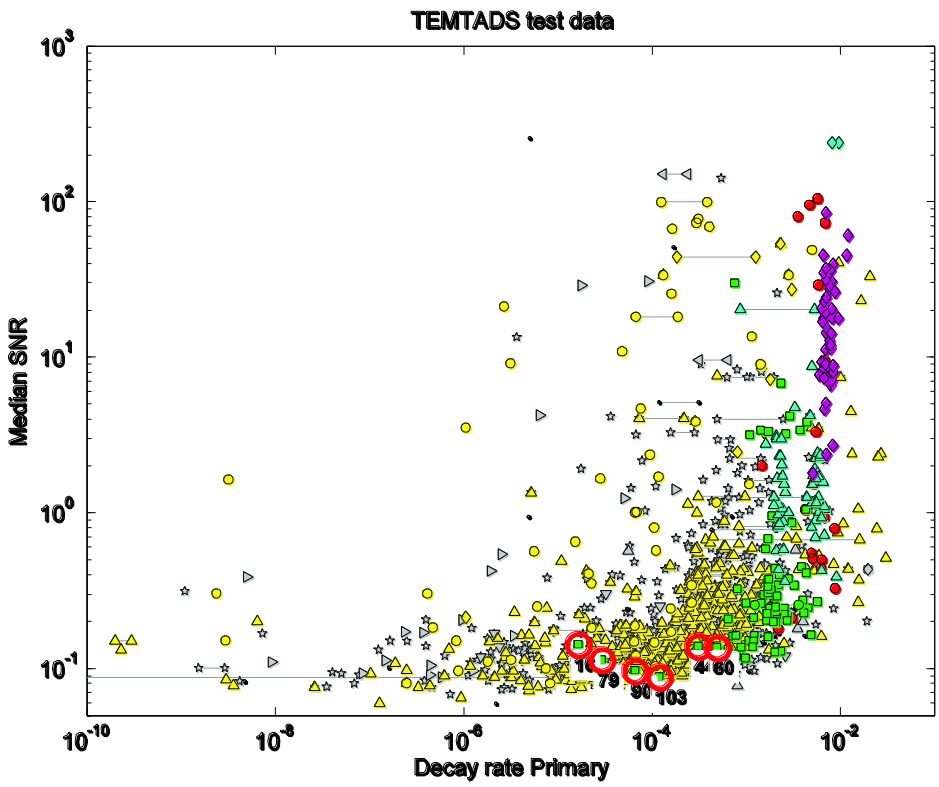


Figure 33. Median SNR versus estimated decay rate of primary polarization for TEMTADS test data. Low SNR targets are circled.

From Figure 33, we conclude that a more robust technique for estimating time decay information is required for input into statistical classification. A first step to improving our parameter estimation is to apply a weighting to the estimated polarizations when we fit decaying exponentials to this function. This is analogous to weighting our data inversions with a percent +floor. Here we compute the weighting on the polarization L at i^{th} time channel as:

$$w_i = \frac{1}{0.01\left|L(t_i)\right| + L_{floor}} \tag{22}$$

with the floor value estimated from the last N channels:

$$L_{floor} = \text{median}\left\{L_i(t_{late} : t_{end})\right\} \tag{23}$$

Figure 34 compares the fit to polarizations using unweighted and weighted least squares. For Master ID 1285 the weighted inversion provides much improved estimates of the polarizations.

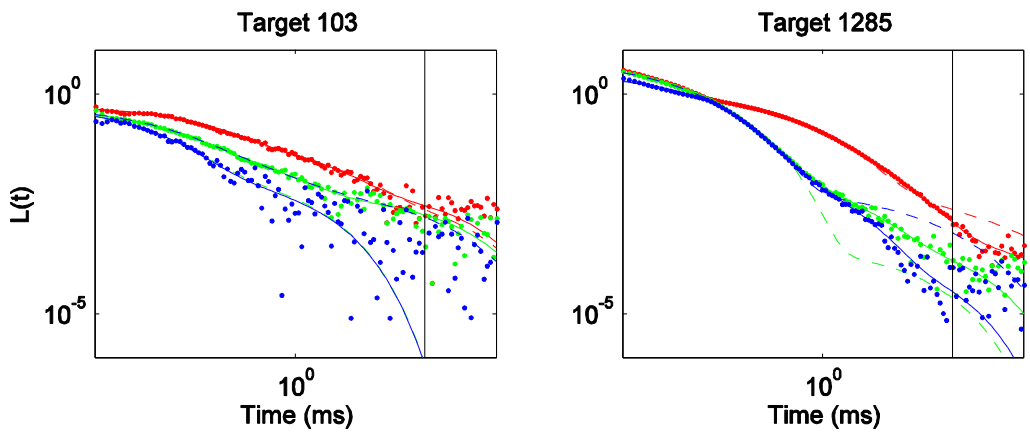


Figure 34. Fits to estimated polarizations (dots) for two targets. Dashed line is unweighted fit, solid line is weighted fit. Vertical line indicates the channel used to compute the polarization decay parameter.

With our improved fit we can then compute the decay parameter as:

$$\text{Polarization decay} = \frac{\max\left(L_i(t_m), L_{floor}\right)}{L_i(t_1)}. \tag{24}$$

With t_m a specified late time channel as before. By taking the maximum of the smoothed polarization and the noise floor in the numerator, we ensure that the decay parameter does not extrapolate below the noise floor. **Error! Reference source not found.** shows the TEMTADS feature space computed in this manner, with t_m taken as channel 93 (as in Figure 28) and the polarization noise floor is estimated over the last 25 TEMTADS time channels. We see that the distributions of 60 mm and non-TOI targets are considerably tightened; all previously small decay rate 60 mm TOI now have larger than average decay rates. There is, however, one 60 mm target (Master ID 1285) which remains a significant outlier to its class after re-computation of the decay parameter.

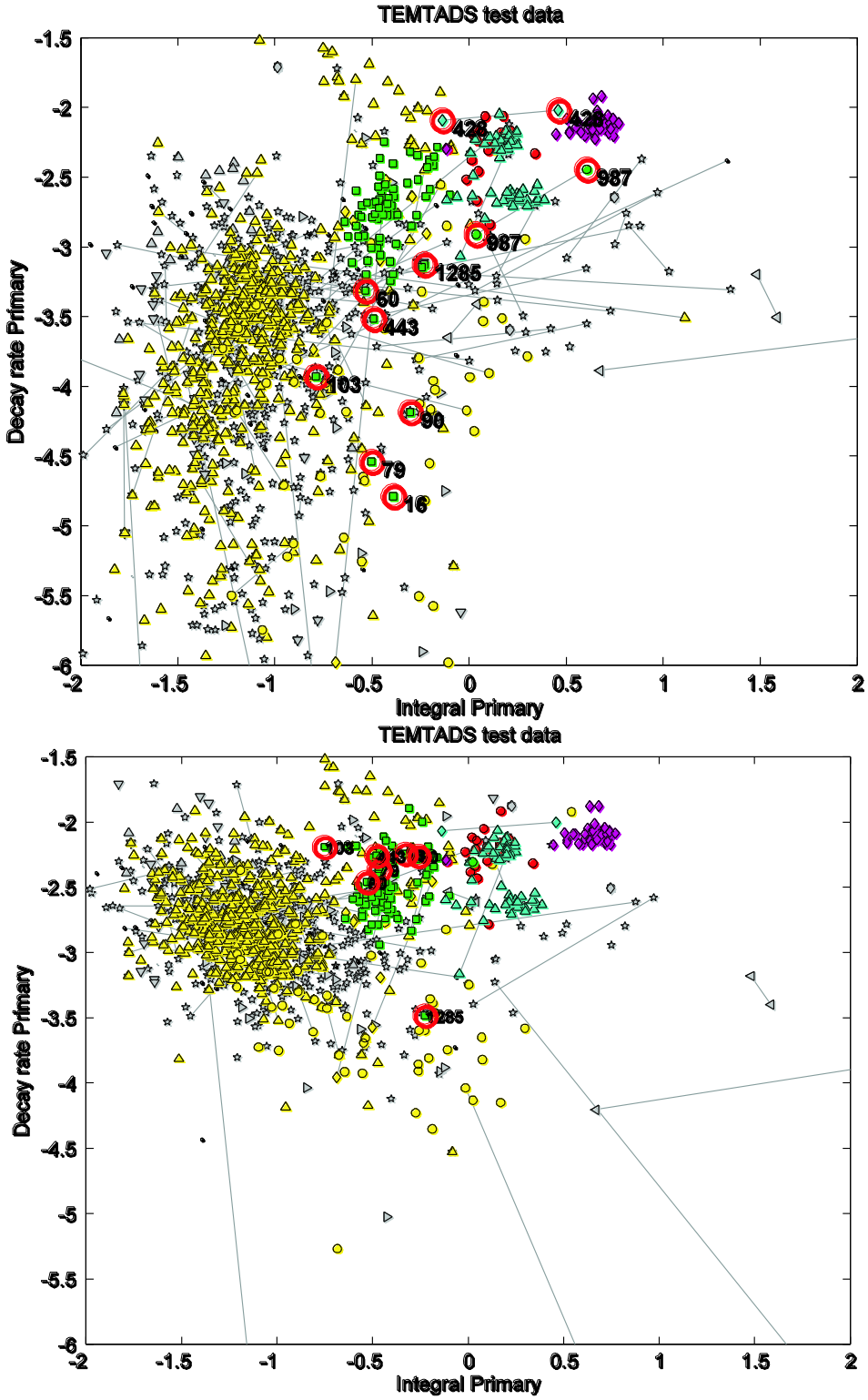


Figure 35. Comparison of primary polarization decay rate estimates without noise floor (left) and with noise floor (right).

As seen in Figure 34, the SNR of recovered polarizations for Master ID 1285 is quite good. However, the decay rate of the primary polarization is much faster than is typically observed for 60 mm targets. A multitarget re-inversion of these data produces an excellent result (Figure 36): we find that the 60 mm at depth is masked by a fast decaying-piece of near surface clutter. While no clutter is seen in the ground truth photo for this target, we speculate that a piece of clutter on the surface was moved during or after surveying.

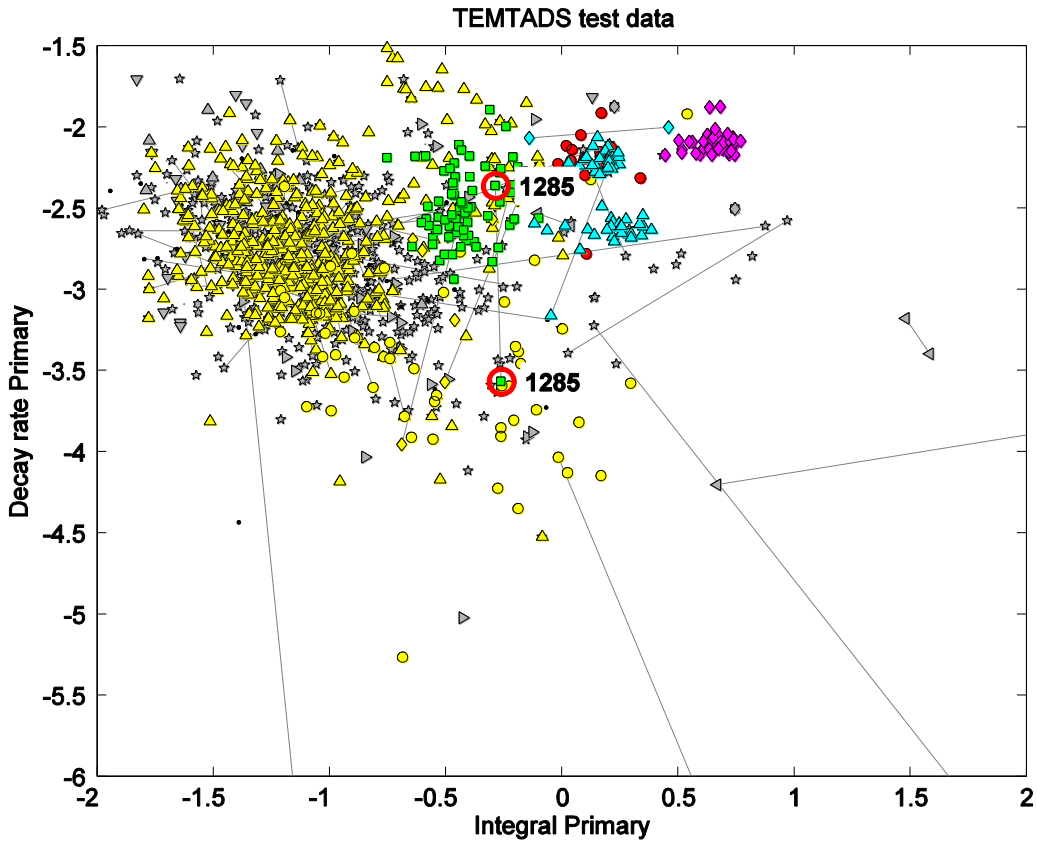


Figure 36. TEMTADS test data with features from retrospective multi-target inversion of Master ID 1285.

Figure 37 shows the decision surface and ROC curve obtained for a PNN classifier trained on retrospective size and decay features. A reduction in TOI outliers comes at the expense of increased overlap between TOI and non-TOI classes, so that the AUC of the retrospective PNN is decreased relative to the ROC for the submitted statistical classification diglist. The submitted library diglist, which achieved the best AUC for all methods applied to the TEMTADS data, is also shown for comparison.

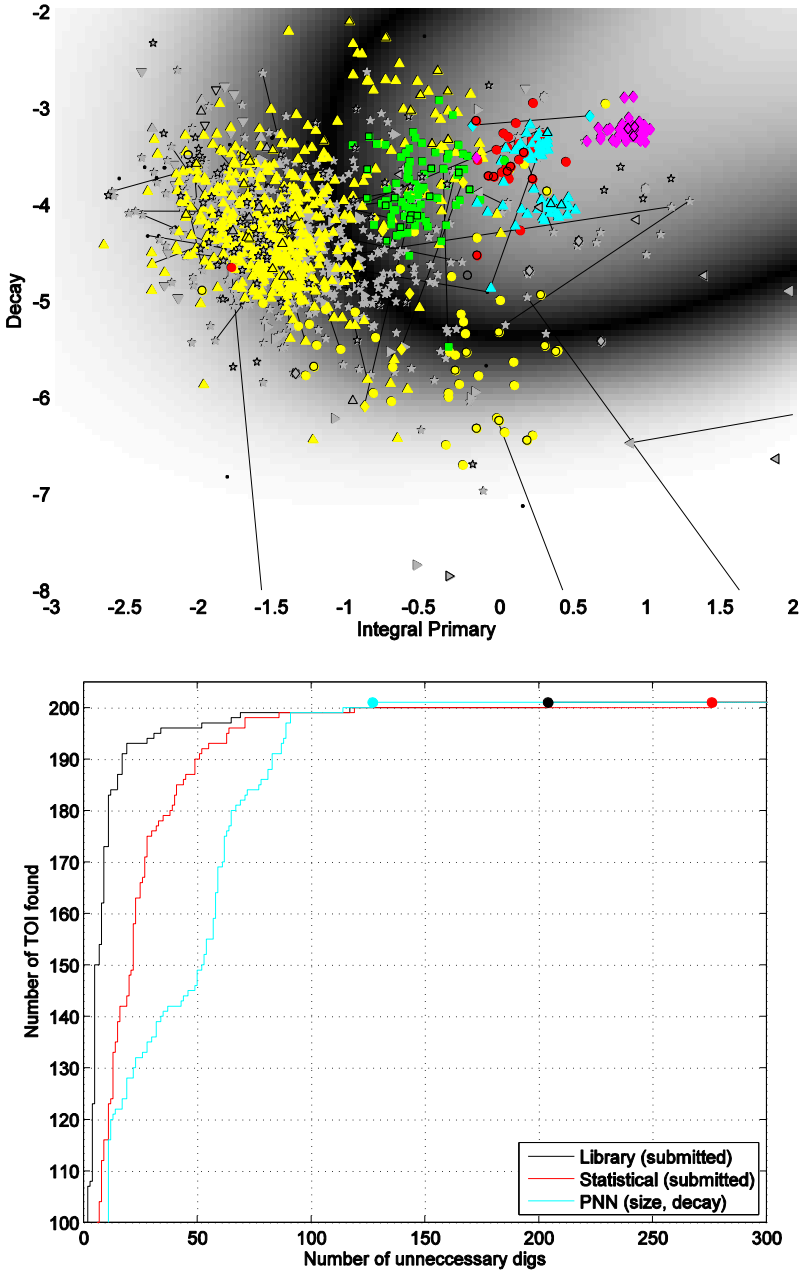


Figure 37. Left: retrospective PNN decision surface with TEMTADS test data. Training feature vectors are outlined in black. Right: ROC curves for library method, statistical classification, and retrospective PNN. Markers indicate the point on each ROC curve at which the last TOI is identified (Master ID 241 is not considered a TOI in this ROC analysis).

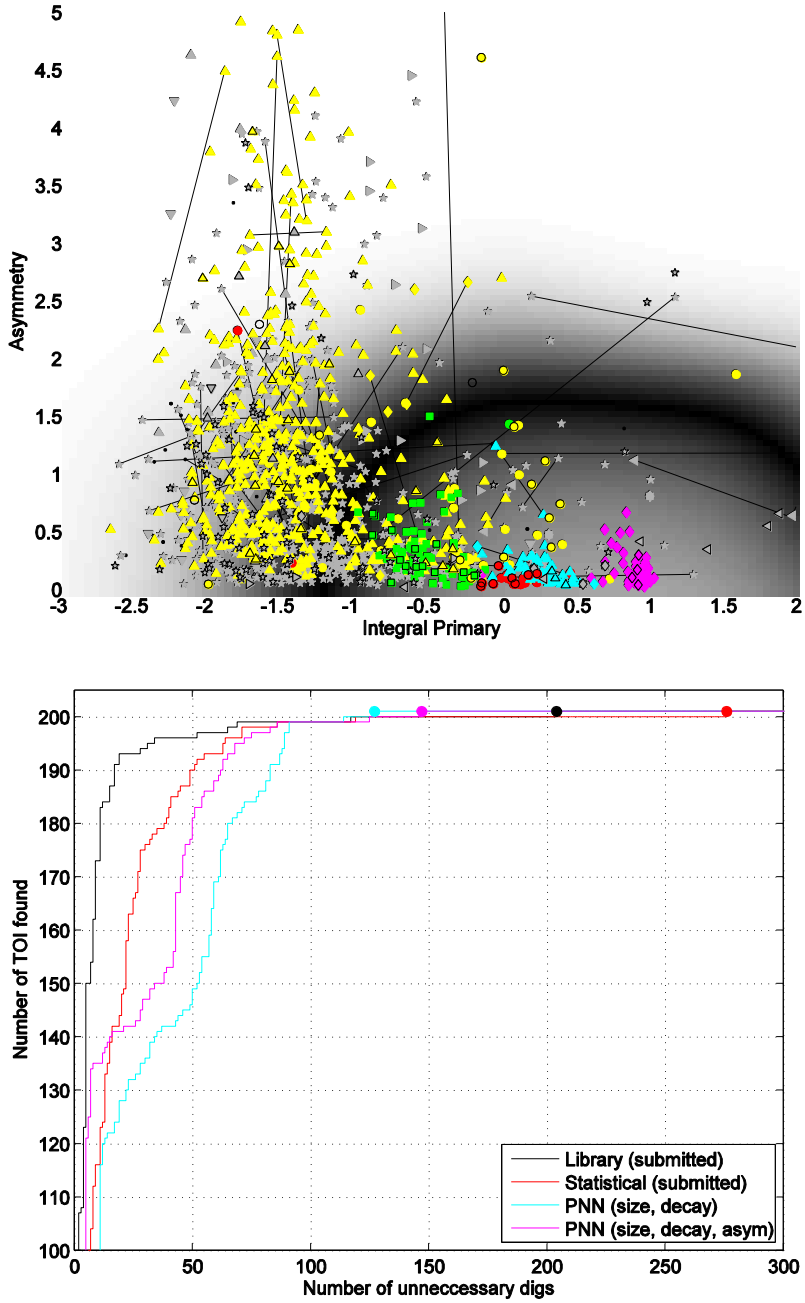


Figure 38. Left: retrospective PNN decision surface with TEMTADS test data. Training feature vectors are outlined in black. Right: ROCs for library method, statistical classification, and retrospective PNNs. Markers indicate the point on each ROC at which the last TOI is identified.

7.2.1 Classification using asymmetry

One way to increase AUC for the retrospective classifier is to include an asymmetry parameter (Figure **38**). The asymmetry parameter is computed as:

$$Asymmetry = \frac{\int |L_2(t) - L_3(t)|\, dt}{\int |L_2(t) - L_1(t)|\, dt} \tag{25}$$

so that an *Asymmetry=0* indicates an axisymmetric target. Here we integrate the estimated polarizations $L_i(t)$ over the time range 0.42-5 ms.

For display purposes, the decision surface in

Figure **38** is computed for a PNN classifier trained in a two-dimensional space whereas the ROC for the retrospective classifier in magenta is trained in a three-dimensional feature space (size, decay and asymmetry). We see that asymmetry provides an initial improvement in discrimination performance, but can be problematic when applied to low SNR targets (i.e 60 mm), resulting in a slight increase in FAR relative to the retrospective PNN trained on size and decay only. For this reason we chose not to use asymmetry in the submitted statistical classification diglist.

7.2.2 Classification using the moments of the polarizabilities

The moments of the polarizabilities have also been suggested as features for discrimination with TEM data. The nth moment of the ith polarizability (polarization) is computed as:

$$M_i^n = \int t^n L_i(t) dt \tag{26}$$

Figure 39 shows log-transformed primary polarization moments 0 and 2. The zeroth moment is equivalent to the primary polarization integral used previously. Higher order moments emphasize late time data and so ideally provide information about late time decay. Here we observe a strong correlation between zeroth and second moments (and similarly for the first moment). A parameter analogous to the polarization decay can be obtained by considering the ratio of second to zeroth moments. A feature space spanned by zeroth moment and moment ratio of the primary polarization appears similar to our usual size/decay feature space. However, classifiers trained on moments have smaller AUC and larger FAR than other retrospective classifiers, primarily due to increased overlap between TOI and non-TOI classes with these features (Figure 40).

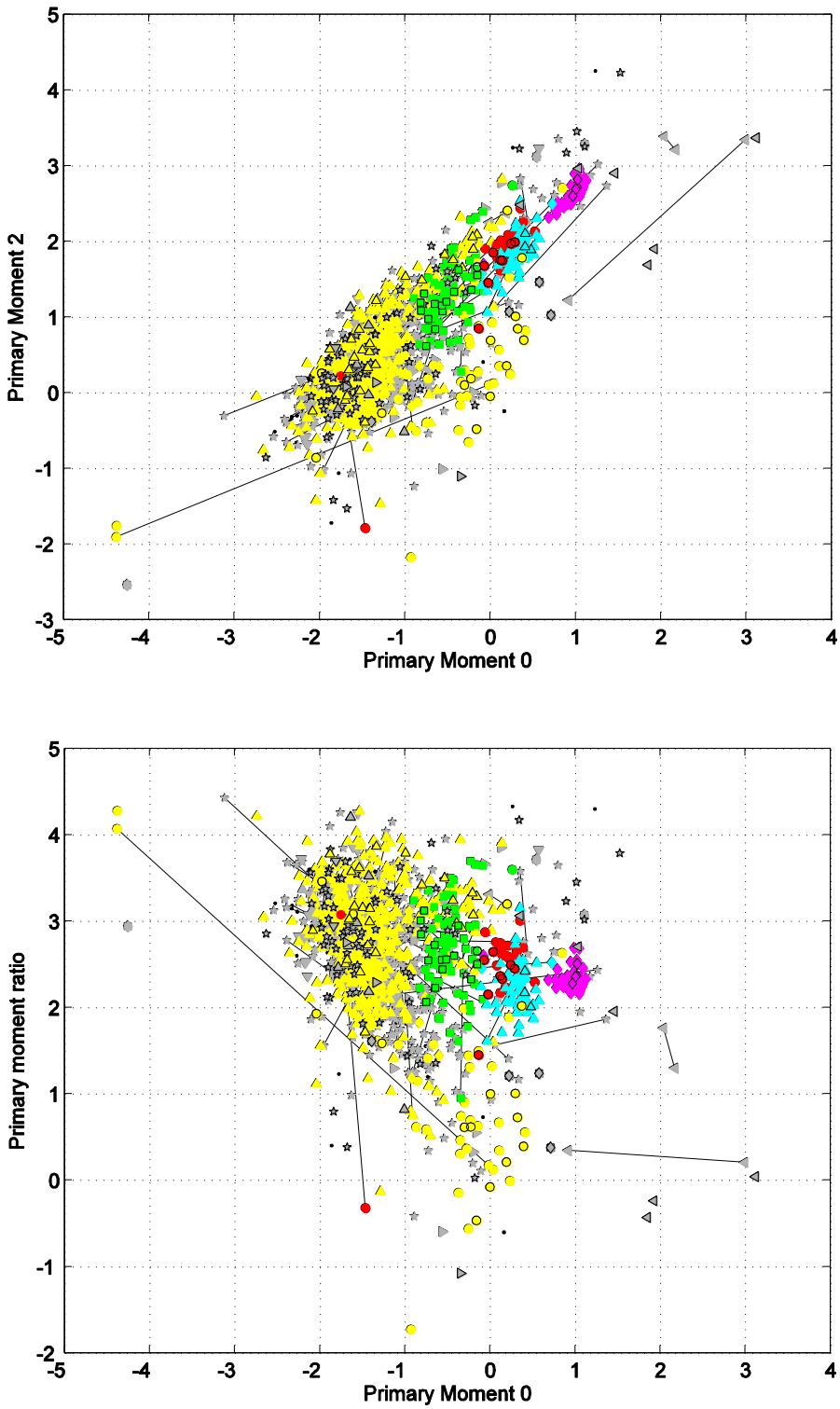


Figure 39. Moments of TEMTADS primary polarization. Left: second versus zeroth moments. Right: moment ratio (second divided by zeroth), versus zeroth moment.

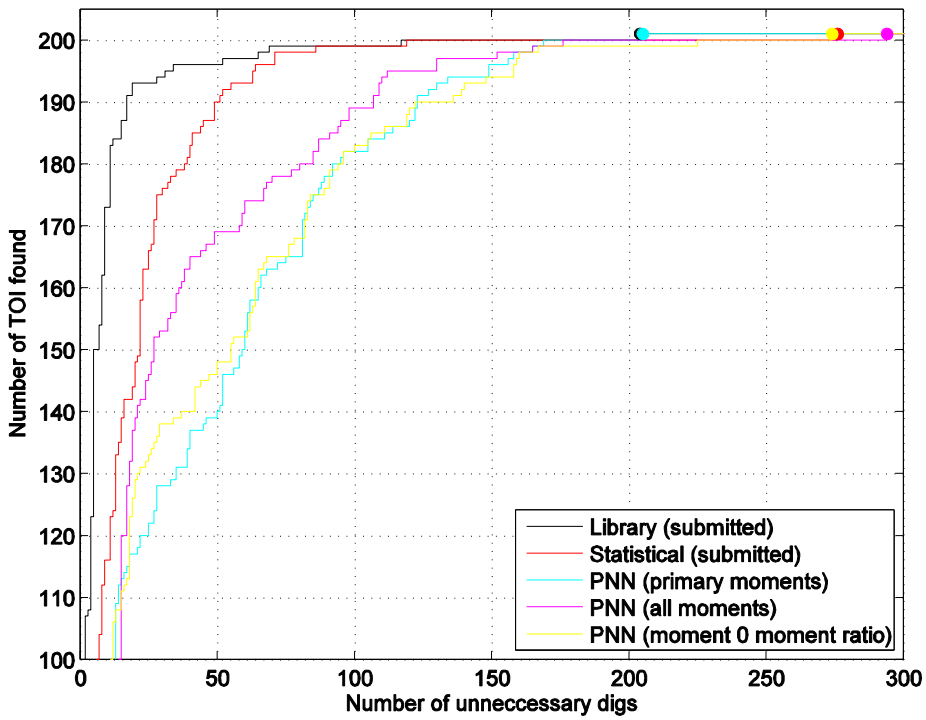


Figure 40. Retrospective PNN classifiers applied to moment-derived features. "primary moments" denotes a PNN trained on zeroth through second moments of the primary polarization, "all moments" is a PNN trained on zeroth through second moments of all polarizations (a nine-dimensional feature space), "moment 0 moment ratio" is a PNN trained on primary zeroth moment and moment ratio.

7.2.3 Automatic feature selection

Our classification strategies are generally restricted to low-dimensional spaces (<10 features), with features selected via experimentation with the training data. This makes data-visualization easy, and may prevent the so-called "curse of dimensionality" (i.e. the number of training data must increase exponentially to ensure accurate estimation of generating distributions). However, TEMTADS data provide a relatively large set of features (115 time channels x 3 polarizations) and it is possible that by restricting our classifiers to low-dimensional features spaces we are omitting information that may improve classification performance. Here we investigate classification and feature selection in high-dimensional spaces using SVMs. SVMs generally perform well in high-dimensional spaces because the decision function is computed as a projection onto a vector **w** comprised of small subset of non-orthogonal training vectors (the support vectors). In addition, redundant or non-discriminative features can easily be identified as small elements in **w**. As a first experiment, we apply the recursive feature elimination (RFE) technique of Guyon et al., 2002 to identify a subset of 20 features from an input feature space spanned by the (log-transformed) smoothed primary polarizations at all 115 time channels. The RFE method works as follows:

1. Train the SVM classifier in a feature space spanned by the set of features **s**; and
2. Eliminate from **s** the feature with the smallest contribution to the weighting **w.** Loop to 1 until the number of elements in **s** attains a prescribed value.

For a linear SVM, the decision function d is computed as a weighted combination of the training feature vectors x_k:

$$w = \sum_k \alpha_k y_k x_k \tag{27}$$

with α_k a Lagrange multiplier determined in training (α_k=0 indicates x_k is not a support vector), y_k the class label (+1,-1) of the training vector. For a nonlinear SVM the weighting cannot be expressed as a linear combination of training vectors, but an analogous criterion for sequentially removing features is derived in Guyon et al., 2002. This corresponds to eliminating features that are "minimally informative" as in the active learning techniques in Zhang et al (or, more generally, for survey design in inverse problems).

Figure 41 shows the estimated primary polarizations in the training data and the channels selected via RFE with a nonlinear (radial basis function) SVM. Interestingly, the feature selection technique chooses three time windows which might qualitatively be described as early, intermediate and late windows. We then train a nonlinear SVM in this feature space. The complexity of the SVM decision boundary for non-separable data depends upon an overlap parameter C and the width of the kernel function σ. We fix the former parameter at C=1 and then estimate σ using five-fold cross validation (a technique related to bootstrapping). Applying the trained classifier to the TEMTADS test data, we obtain a quite favorable result relative to retrospective analyses presented thus far (Figure 42). While the library method still provides the best result in terms of AUC, the SVM RFE method is an improvement in terms of AUC relative to the submitted statistical classifier, and has the lowest FAR obtained in retrospective analysis. This suggests that higher dimensional feature spaces identified by automatic feature selection techniques can improve performance relative to low-dimensional classifiers.

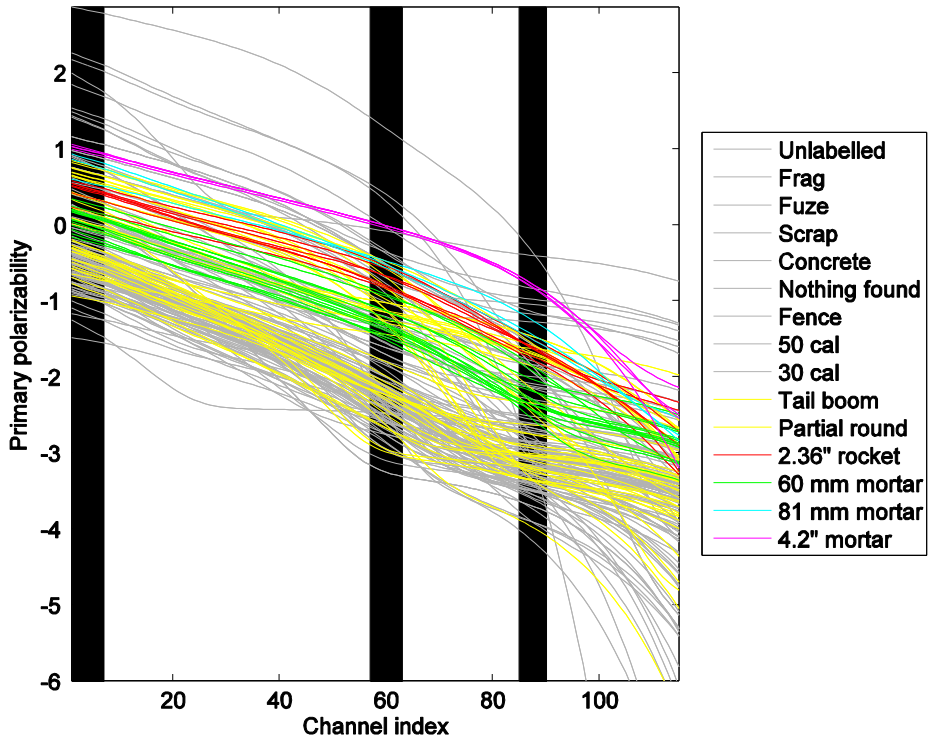


Figure 41. Smoothed polarizabilities for TEMTADS training data. Shaded rectangles indicate 20 time channels selected with recursive feature elimination.

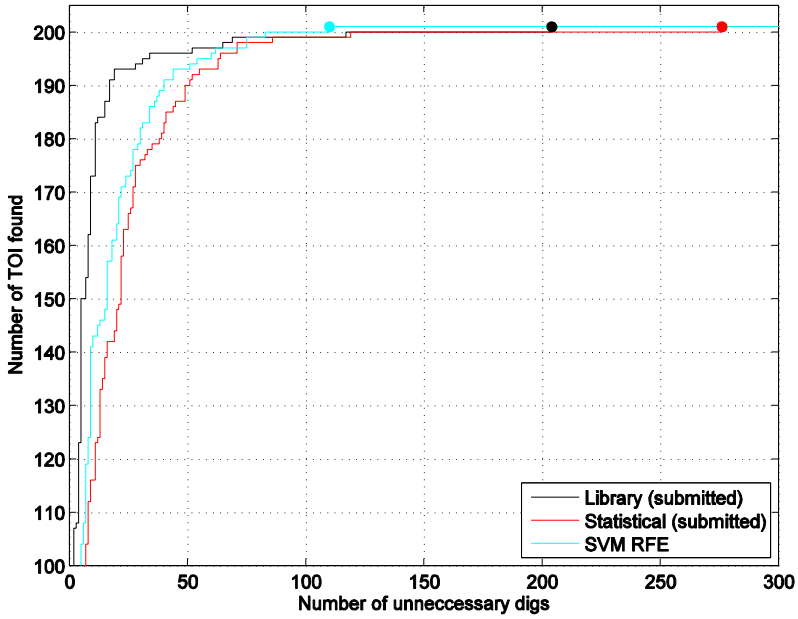


Figure 42. Retrospective analysis of TEMTADS test data using nonlinear SVM in a feature space identified with RFE.

7.2.4 Identifying TOI type

In three of our submitted TEMTADS methods we made (or can infer) predictions of the type of UXO in addition to the usual classification into TOI and non-TOI. For the expert opinion method, the analyst classified certain items as "high-confidence" UXO and for those items he also predicted the UXO type. For the library method, we can infer the UXO type by selecting the item that produces the best-fit to the data. For the statistical classification, the decision surface was constructed using a series of class means and covariances centered on each ordnance class (and clusters of clutter). We can infer the UXO type of a feature vector by assigning it to the closest class cluster. For all three methods the analysts were expecting to encounter 60 mm, 81 mm and 4.2" mortars and 2.36" rockets. TEMTADS had 202 such encounters in the blind-test data (one extra 60 mm and one extra 4.2" mortar were detected during the MTADS screening survey but could not be reached by the TEMTADS).

For the expert-opinion method 216 items were identified by the analyst as high-confidence UXO. 189 of those were TOI and 27 of those were false-negatives, so the analyst was correct 87.5% of the time. 17 of these false negatives occurred in the final 34 items ranked as high-confidence UXO. This means that only 10 false-negatives were encountered in the first 182 items ranked as high-confidence UXO (94.5% success rate). Almost all of the false-negatives were classified by the analyst as 60 mm mortars with three classified as 81 mm mortars and two as 2.36" rockets.

The expert classified 189 of the 202 TOI encountered by TEMTADS as high-confidence TOI. Most of the remaining 13 items (nine 60 mm mortars, three 2.36" rockets and one 81 mm mortar) were still recommended for excavation by the analyst, but no predictions on UXO type were made. Table 7 shows that the analyst made the correct prediction of UXO type on 185 out of 189 items classified as high-confidence UXO: a success rate of over 98%. One rocket (Master ID 791, a damaged rocket) was mislabeled as an 81 mm mortar (polarizations provide slightly better fit to the 81 mm mortar). For the 81 mm mortars, two were mislabeled as 2.36" rockets (Master ID 377 an intact 81 mm, but polarizations better match a rocket, Master ID 907 81 mm mortar without tail, polarizations provide a slightly better match to 2.36" rocket) and one as a 60 mm mortar (Master ID 1386: 81 mm body without tail-fins). The analyst correctly identified all of the 60 mm and 4.2" mortars in the high-confidence UXO category.

For the library method 200 of 202 TOI encountered by TEMTADS were identified as TOI, and 196 of those 200 items were assigned to the correct UXO (Table 8): a 98% success rate. The same two 81 mm mortars mislabeled by the expert opinion method were also mislabeled by the library method. Two 2.36" rockets were mislabeled as 81 mm mortars: Master ID 444 that comprised several rocket motor pieces and which was fit in the statistical method as a multi-object scenario but only as a single object in the library method; and Master ID 1300 which was an intact (but slightly bent) 2.36" rocket at 23 cm, which was slightly better fit as an 81 mm mortar. All 60 mm and 4.2" mortars that were correctly declared TOI were also correctly assigned to the right UXO type.

For the statistical classification method 199 of the 202 TOI encountered were identified as TOI, and 179 of those 199 items were assigned to the correct UXO type (Table 9): a 90% success rate. The lower success rate of the statistical classification method is mostly caused by 60 mm mortars incorrectly identified as 2.36" rockets (six cases) and 2.36" rockets incorrectly identified as 81 mm mortars (10 cases). 95% of the 81 mm mortars and all of the 4.2" mortars were correctly identified.

Table 7. Ability of TEMTADS Expert Interpretation method to correctly label ordnance type. The table includes a column that shows the number of items that were classified as high-confidence UXO by the analyst.

Munition	Number of items	Number HighConf UXO	Number correctly labeled	% correct	Mislabels
60 mm	73	64	64	100	
2.36” rocket	20	17	16	94	81 mm (791)
81 mm mortar	59	58	55	95	2.36” (377, 907), 60 mm (1386)
4.2” mortar	50	50	50	100	
Total	202	189	185	98	

Table 8. Ability of TEMTADS Library method to correctly label ordnance type.

Munition	Number of items	Number classified as TOI	Number correctly labeled	% correct	Mislabels
60 mm	73	72	72	100	
2.36” rocket	20	19	17	89	81 mm (444,1300)
81 mm mortar	59	59	57	97	2.36” (377, 907)
4.2” mortar	50	50	50	100	
Total	202	200	196	98	

Table 9. Ability of TEMTADS Statistical Classification method to correctly label ordnance type.

Munition	Number of items	Number classified as TOI	Number correctly labeled	% correct	Mislabels
60 mm	73	71	65	92	2.36” rocket (192, 302, 344, 353, 441, 1456)
2.36” rocket	20	19	8	42	60 mm (444), 81 mm (123, 188, 413, 448, 464, 547, 791, 1291, 1300, 1420)
81 mm mortar	59	59	56	95	2.36” (340, 467, 907)
4.2” mortar	50	50	50	100	
Total	202	199	179	90	

7.3 ADDITIONAL ANALYSIS OF METALMAPPER DATASETS

Figure 43 shows the ROC curves for the statistical and expert option methods applied the MetalMapper data. For the statistical method there were two false-negatives: Master ID 241 (2.36” rocket motor parts with questionable ground-truth) and Master ID 1502 (only 37mm found). These two items were also false-negatives in the expert opinion method which also contained one additional false-negative: Master ID 775 (multi-object scenario 60 mm mortar with two tail fins nearby). ROC curves computed for individual items are nearly vertical, except for the 60 mm mortars. As expected, these present the biggest difficulty to both methods.

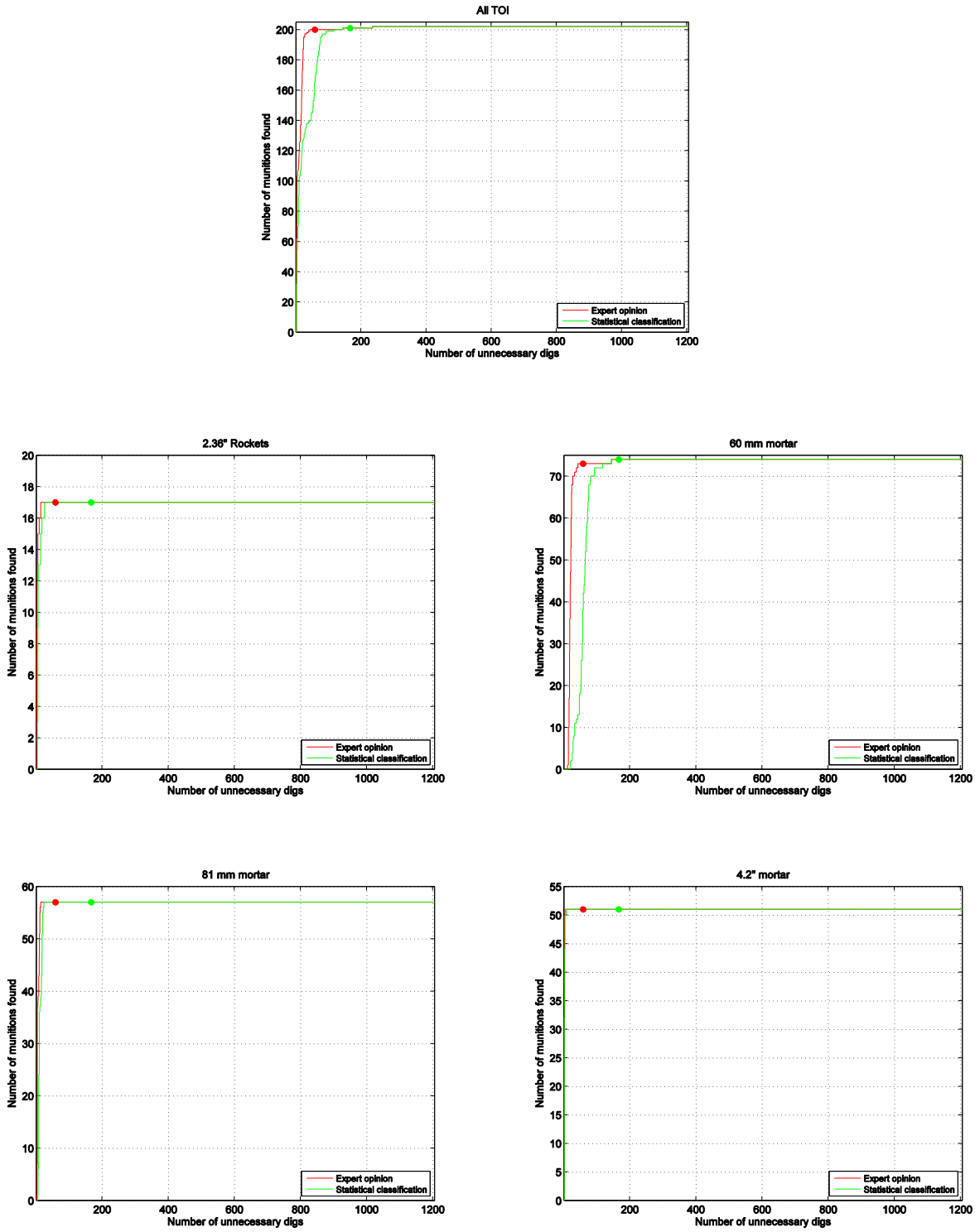


Figure 43. Receiver operating curves for TOI classes, MetalMapper classifiers. Open circles are specified operating points. Master ID 241 was not labeled as a TOI when generating these figures.

The statistical classifier used a feature space consisting of a time-decay and a size based feature vector (Figure 44a). The decision surface was obtained by training a PNN classier on a two class problem (TOI or not-TOI). Ordnance clusters in the blind-test data (Figure 44b) are similar to those in the test data, except there are a number of outliers in the 60 mm class (Master IDs 775 and 1315) and the one obvious outlier in the 2.36" rockets (Master ID 241). When Master ID 775 is refit assuming two objects are in the field of view, the recovered polarizabilities of one of the objects closely matches that from a 60 mm mortar body (Figure 45). Master ID 1315 is a 60 mm rocket motor at 30 cm deep: its decay rate is similar to other 60 mm mortar bodies but its estimated size is too small. Inspection of the MetalMapper data (Figure 46) reveals that the SNR is quite high in all 3 components of each receiver when the vertical axis transmitter is fired, but quite low when the horizontal axis transmitters are fired. The poor SNR from the horizontal axis transmit data probably contributes to underestimate of the size parameter.

The one 37 mm recovered (Master ID 1502) has a smaller size, but a slower decay that the 60 mm mortars. While it was a false-negative in the statistical classification method, its slow decay rate would have allowed it to be identified as suspicious, had we been looking for smaller UXO.

A plot of the predicted versus actual depths shows that MetalMapper has an excellent ability to constrain object depth (Figure 44c).

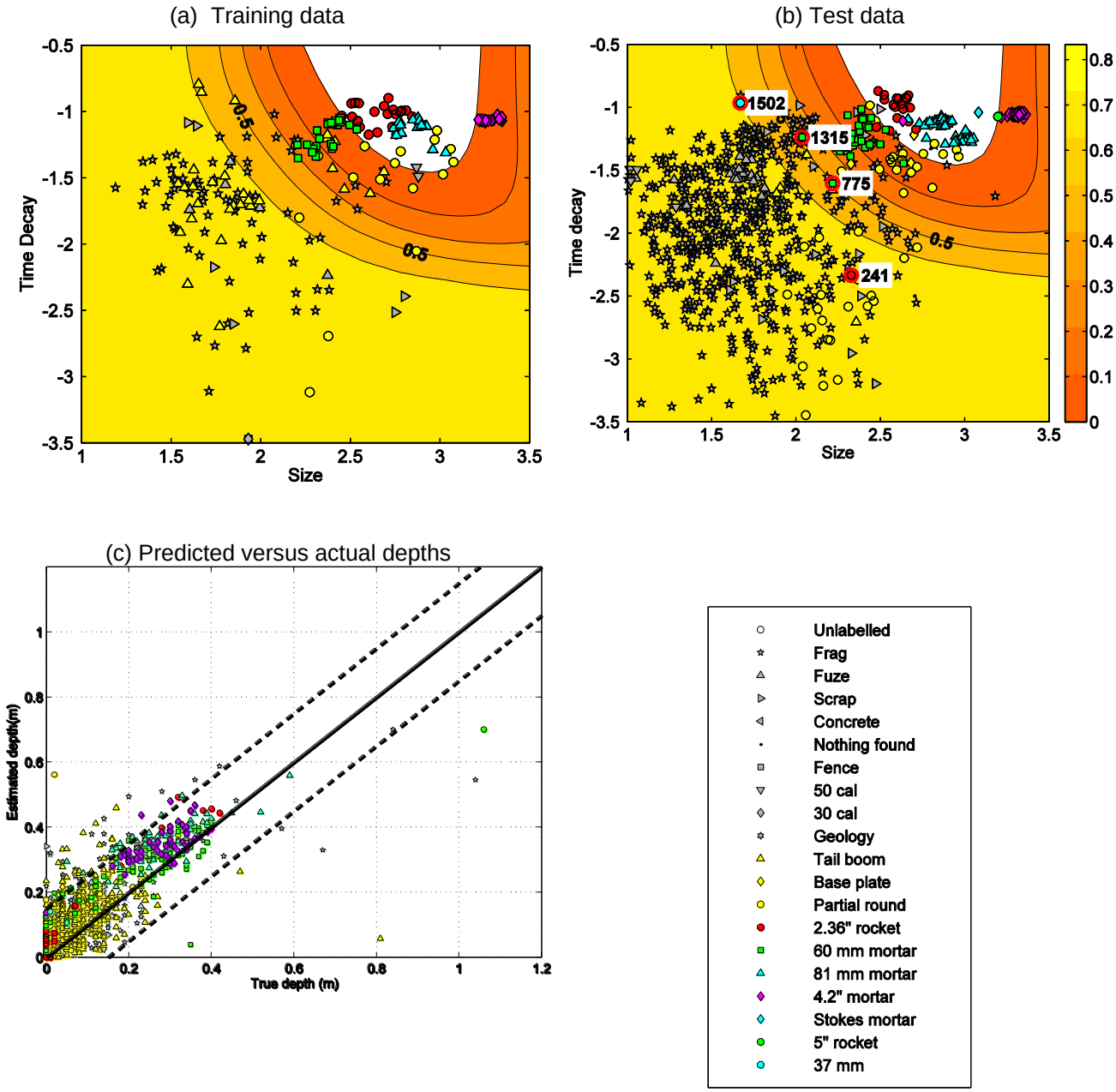


Figure 44. Training (a) and test-data (b) for the MetalMapper statistical classification method, along with a plot of recovered versus actual depths (c).

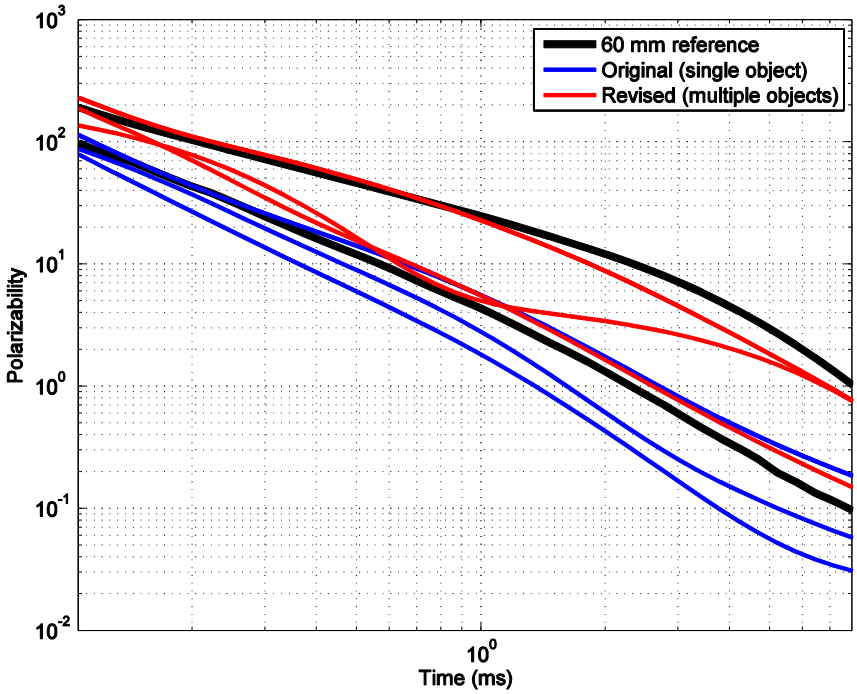


Figure 45. Polarization recovered over Master ID 775 using MetalMapper data inverted as a single object and as a multi-object scenario.

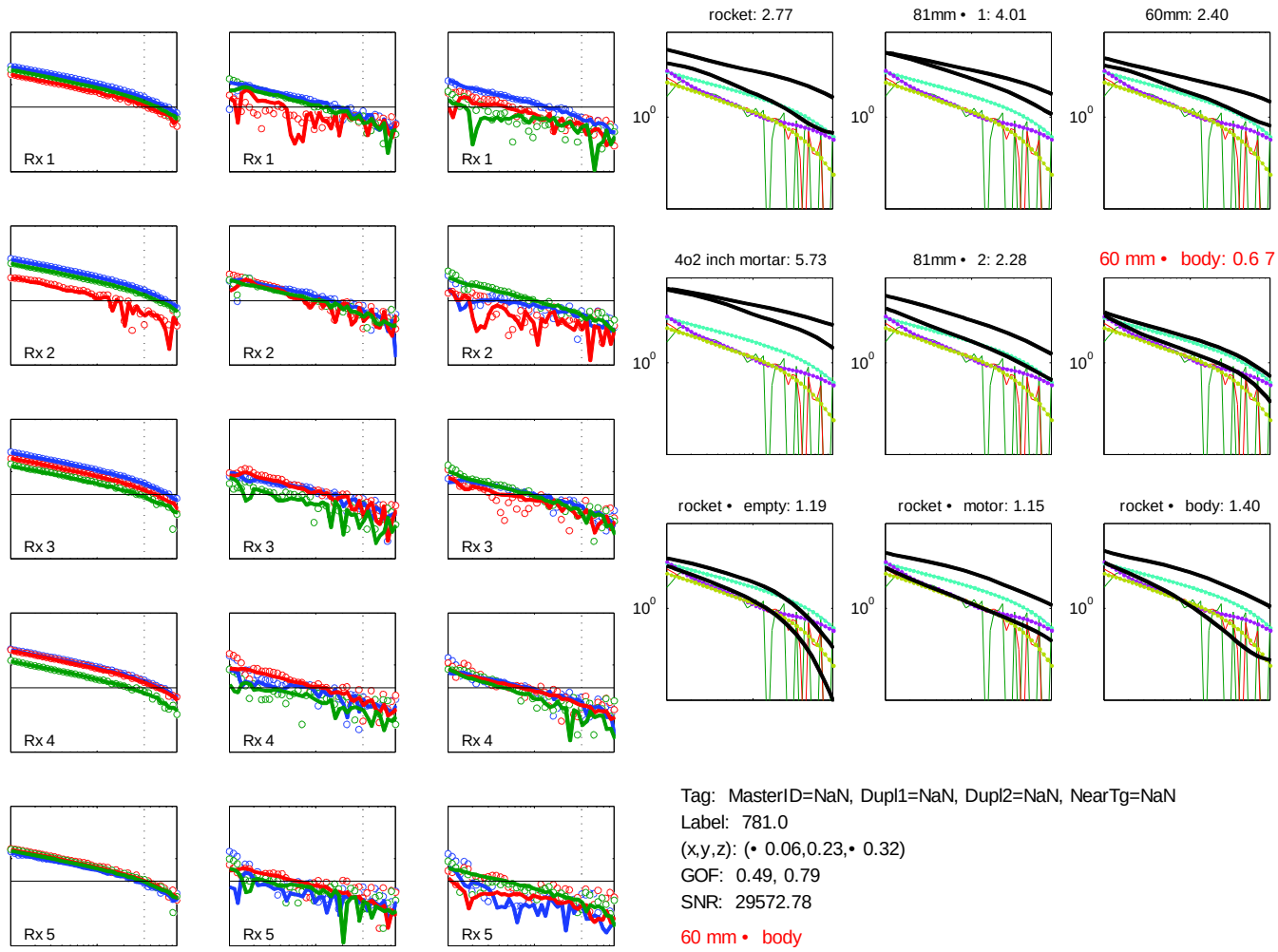


Figure 46. Data fit and polarizabilities extracted from Master ID 1315. Left data plots show results for z-axis excitation, middle for x-axis and right for y-axis. Blue is z-axis receiver, green is x-axis and red is y-axis. Solid line is fitted model, dots are observed data.

7.3.1 More aggressive discrimination strategies

The analysis we conducted and reported in our Training Memo (attached as Appendix A) had lead us to conclude that a statistical classifier based on size, time-decay and asymmetry would constitute an effective discrimination strategy. However, after further consideration we opted to use a more conservative strategy of just the size and time-based feature vectors. We were concerned about the potential failure of the asymmetry feature vector on low SNR data or in cases where there were multiple objects in the field of view. In Figure 47, we show plots of the size and asymmetry feature vectors. The first row of plots were computed using:

$$\text{Asymettry} = \frac{I_{23}}{I_{12}} \quad \text{and} \quad I_{ij} = \int_{t1}^{t2} |L_i(t) - L_j(t)| dt \tag{28}$$

where *t1* and *t2* are the lower and upper time-gates used for the calculation. The numerator is the difference between secondary and tertiary polarizations (should be 0 for axially symmetric objects) and the denominator is the difference between primary and secondary polarizations

(should be large for plate-like objects). Discrimination potential looks promising for the 60 mm mortars, 2.36” rockets and 4.2” mortars but not for the 81 mm mortars. The primary and secondary polarizations of many of the 81 mm mortars in the test-data have similar magnitudes at early times, which makes I_{12} small and which in turn causes the asymmetry measure to be large. This was not a significant problem in the test-data (Figure 47b) although there are several outliers, including two 60 mm mortars and an 81 mm mortar. Master ID 775 has been discussed previously in this report and constitutes a multi-object scenario. Master ID 1285 is a 60 mm mortar body at 35 cm depth, while Master ID 899 is an 81 mm mortar at 17 cm depth. Changing the definition of asymmetry to:

$$\text{Asymettry} = \frac{I_{23}}{I_1} \quad \text{where} \quad I_1 = \int_{t1}^{t2} L_1(t)dt \tag{29}$$

eliminates Master ID 1285 as an outlier, but still results in a relatively large asymmetry metric for Master IDs 775 and 899 (Figure 47d). A plot of polarization size versus inversion misfit (Figure 48) shows that Master IDs 775, 899 and 1285 have atypically large misfits for their size.

The analysis in the last paragraph shows that asymmetry might be an effective feature vector when fit-quality is high, but that considerable care would have to be exercised to avoid trusting the asymmetry measure when fit quality is low (multiple objects, low SNR).

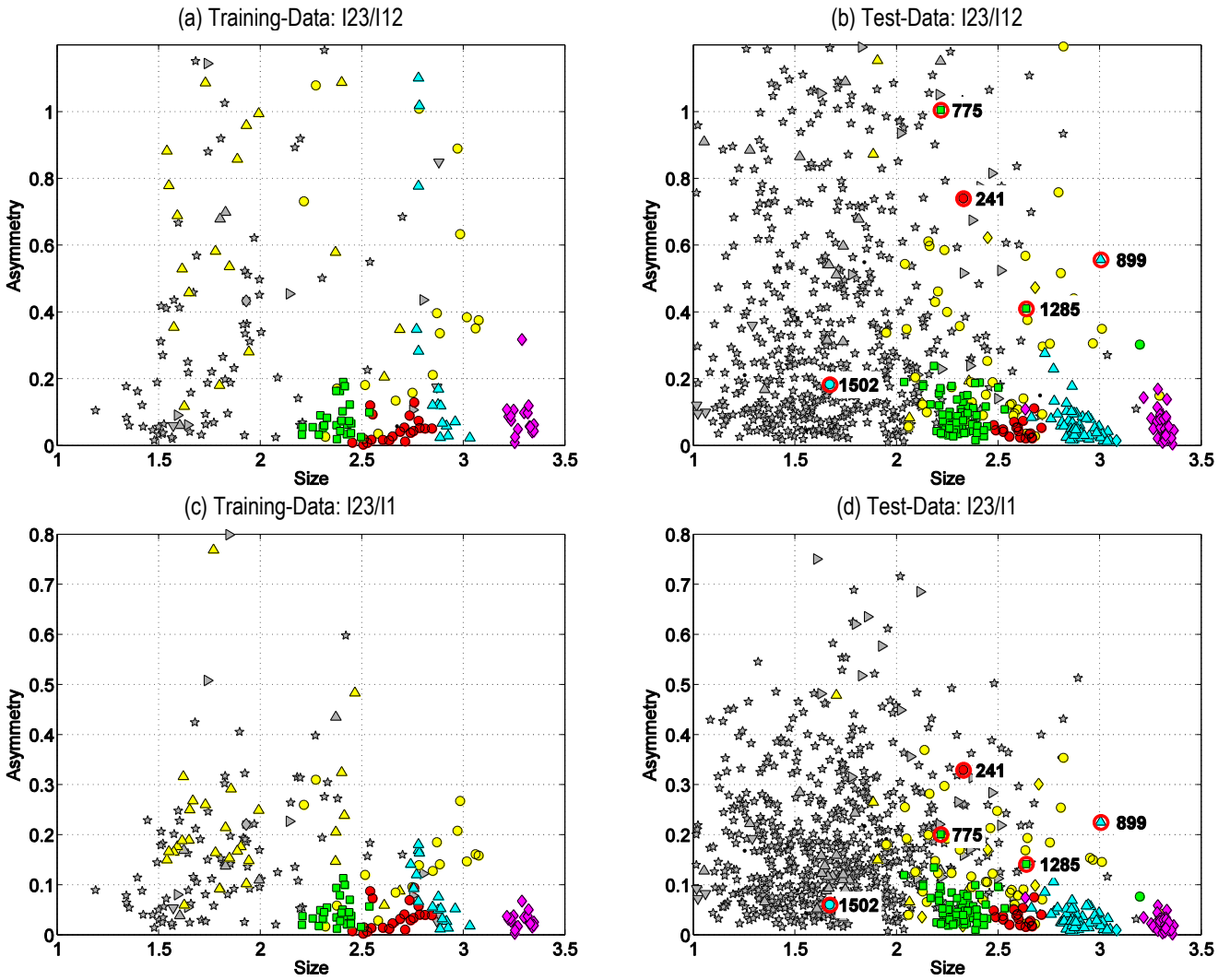


Figure 47. Asymmetry versus size feature space for the MetalMapper. The top-rows uses I23/I12 as the measure of asymmetry while the bottom row uses I23/I1. See the legend in Figure 44 for an explanation of symbols.

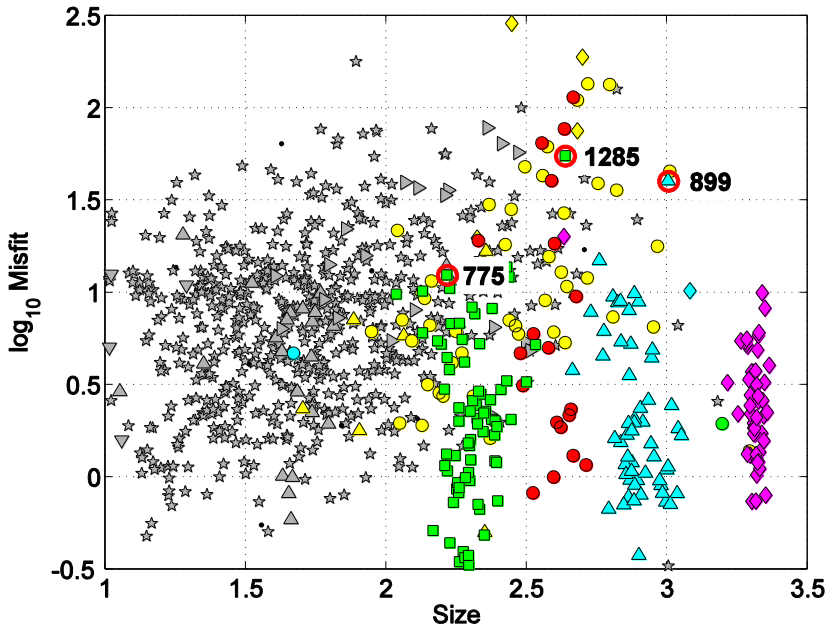


Figure 48. Plot of misfit versus size for MetalMapper feature vectors.

7.3.2 Modified MetalMapper classifier

The ROC curves for the MetalMapper Statistical Method (Figure 43) indicate that the discrimination strategy was very effective on the 81 mm and 4.2" mortars and the 2.36" rockets but not as effective on the 60 mm mortars. The 60 mm mortars are a more difficult object to characterize, but the MetalMapper performance was noticeably inferior to the statistical classifier applied to the TEMTADS. Part of the reason for this is that we used a simple PNN classifier on the MetalMapper data and a more effective Expectation Minimization classifier on the TEMTADS data. Figure 49 shows the training and test-data for an EM classifier applied to the same feature space as was used for the PNN classifier (Figure 44). The revised classification surface is not as smooth as the original PNN classifier and provides a better representation of the distributions of the underlying test and training data. A comparison of ROC curves (Figure 50) indicates that the revised classifier provides a more effective ranking of TOI, although both require almost the same number of excavations to get the last few 60 mm mortars. The last three items dug with the revised classifier are Master IDs 1285 (60 mm body at 35 cm), 775 (multi-object 60 mm discussed previously) and 1315 (60 mm body at a depth of 30 cm).

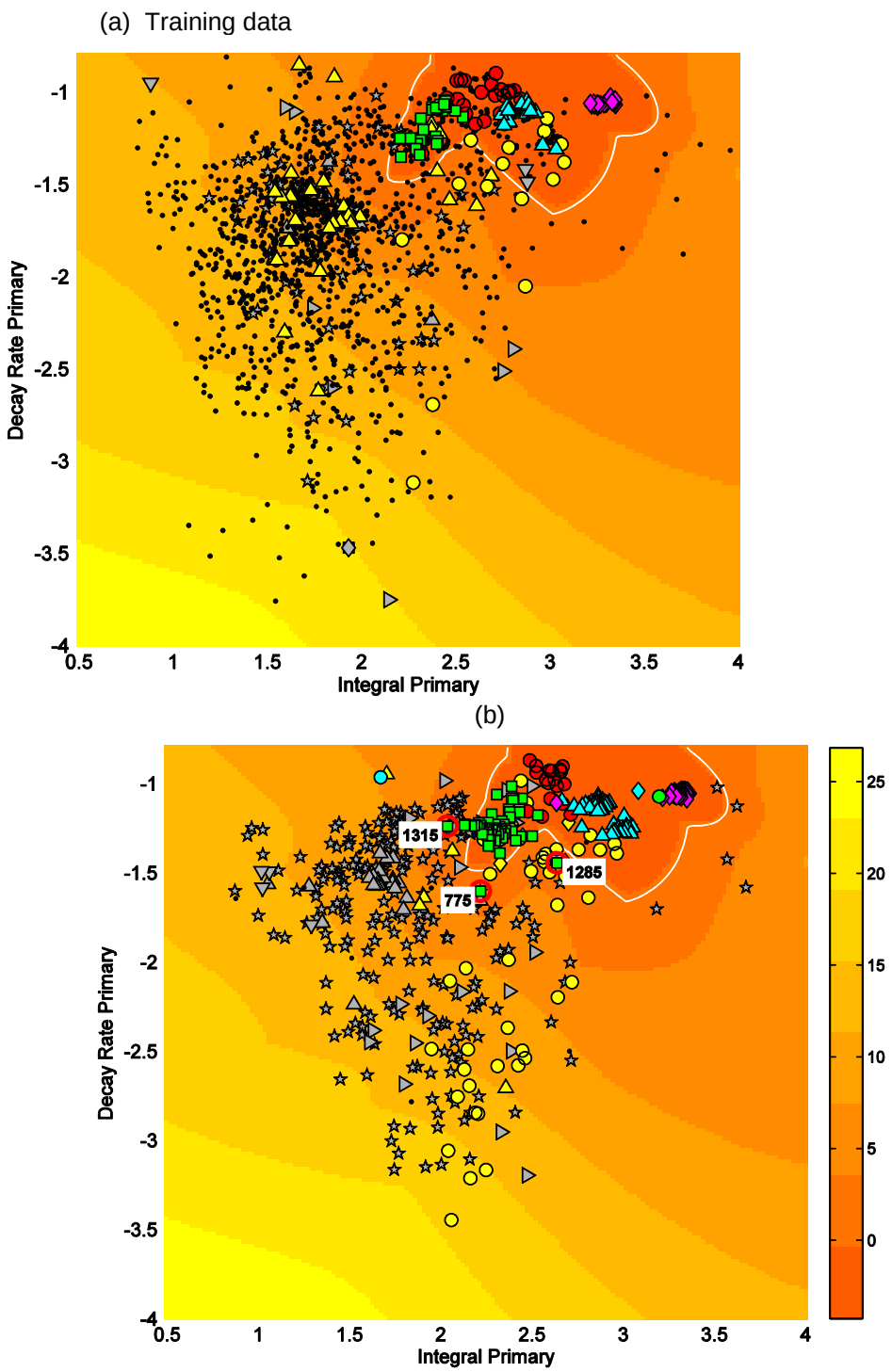


Figure 49. Expectation Minimization classifier applied to the MetalMapper data: (a) training data; and (b) test-data. The white-line is a contour of the decision surface.

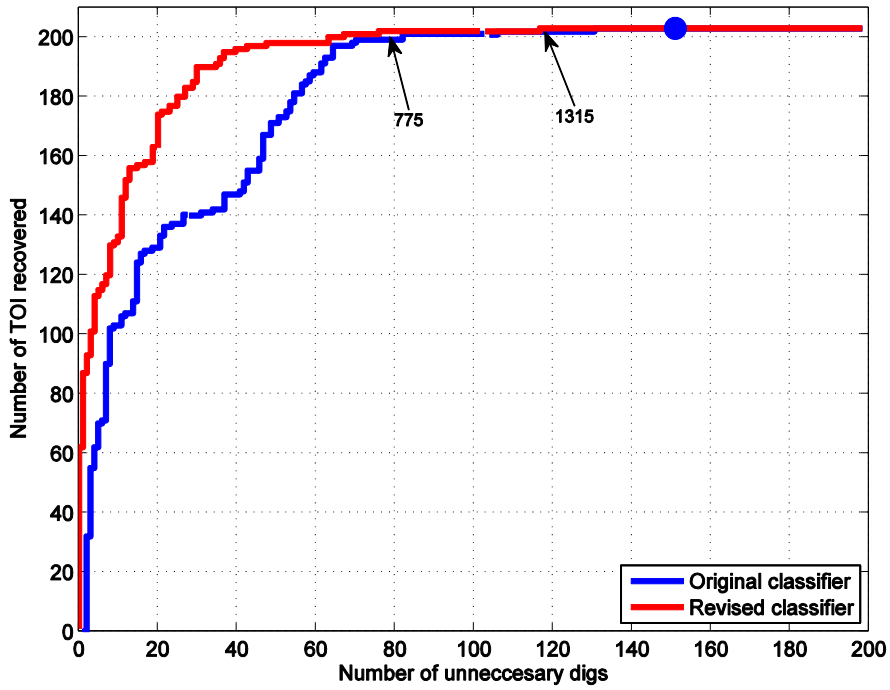


Figure 50. MetalMapper ROC curves for original PNN classifier and a revised classifier.

The EM classifier was constructed using a mean and covariance for each ordnance type. By computing the distance of each feature vector from the class-mean we can predict the UXO type. MetalMapper encountered 201 items that were either mortars of caliber 60 mm, 81 mm or 4.2" or were 2.36" rockets. 200 of these 201 items were correctly identified as TOI (99.5% success) with Master ID 241 (2.36" rocket parts) the only false-negative. The summary in Table 10 indicates that the ordnance type was correctly predicted on 198 of 200 occasions (99% success rate): both miscalls were 2.36" rockets. Master ID1372 was incorrectly labeled as a 60 mm mortar, while Master ID 1300 was mistakenly labeled as an 81 mm mortar.

Table 10. Ability of MetalMapper Statistical method to correctly label ordnance type.

Munition	Number of items	Number identified	Number correctly labeled	% correct	Mislabels
60 mm	74	74	74	100	
2.36” rocket	18	17	15	88	60 mm (1372) 81 mm (1300)
81 mm mortar	58	58	58	100	
4.2” mortar	51	51	51	100	
Total	201	200	198	99	

7.3.3 MetalMapper Library Method after correcting a coding mistake

In the preparation of the library dig-sheet for the MetalMapper we made a coding mistake and sorted the anomalies using the wrong metric (the TEMTADS library diglist was generated using the correct metric). Rectifying the problem resulted in a significant performance improvement as shown by the ROC curve in Figure 51. The revised ranking scheme is significantly better than the original submitted scheme and in fact outperforms the expert method. 196 of 204 TOI are recovered after the recovery of just 19 non-TOI, the next 5 TOI push the total non-TOI count to 51. The final three TOI are found at non-TOI counts of 89 (Master ID 1502: 37 mm projectile), 108 (Master ID 413: 2.36” rocket motor) and 235 (Master ID 241: 2.36” rocket motor, with suspicious ground-truth).

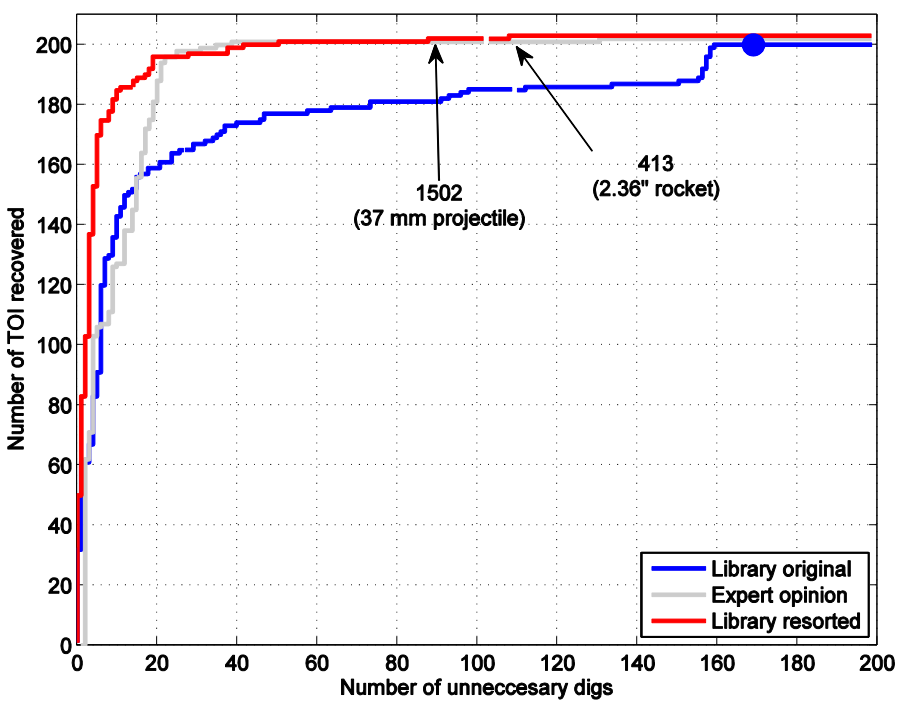


Figure 51. Part of the MetalMapper ROC curve for the submitted library and expert opinion methods, along with a corrected library based method. The blue dot shows the operating point for the original library method.

For the library method, we make a prediction of the item type in conjunction with the estimate of the likelihood the item is a TOI. MetalMapper encountered 201 items that were either mortars of caliber 60 mm 81 mm or 4.2” or 2.36” rockets. 200 of these 201 items were correctly identified as TOI (99.5% success) with Master ID 241 (2.36” rocket parts) the only false-negative. Table 11 provides a summary of how often the ordnance type was correctly identified: from 71% of the time for 2.36” rockets and 86% of the time for the 60 mm mortars to 98% of the time for the 81 mm and 100% of the time for the 4.2” mortars. These are lower success rates than were achieved with the library method applied to the TEMTADS data (Table 8): 98% of TOI assigned to the correct ordnance class). Part of the reason for this, particularly for the 60 mm mortars, is that we included a rocket motor and a rocket warhead in the MetalMapper library but not in the TEMTADS library. These rocket components often provide better fits to the 60 mm mortar bodies than does the polarization used for the 60 mm rocket body.

Table 11. Ability of MetalMapper Library method to correctly label ordnance type.

Munition	Number of items	Number identified	Number correctly labeled	% correct	Mislabels
60 mm	74	74	64	86	2.36” (160, 304, 342, 344, 353, 500, 508, 976, 1285, 1469)
2.36” rocket	18	17	12	71	81 mm (188, 413, 583, 1300, 1420)
81 mm mortar	58	58	57	98	2.36” (899)
4.2” mortar	51	51	51	100	
Total	201	200	184	92	

7.4 PERFORMANCE OF THE BUD

Discrimination with BUD data was carried out in our usual size (integral of the primary polarization) and decay (ratio of the primary polarization at channels 30 (0.942 ms) and 1 (0.145 ms)) feature space. Multiple soundings were acquired over training data targets of interest, and there was some ambiguity as to which soundings best corresponded to the locations of TOI. For this reason we chose to train a PNN classifier using only calibration TOI for which we had high confidence in the estimated polarizations. Although this gave us a very small set of TOI vectors, the decision surface obtained with a PNN classifier in Figure 52a nonetheless seemed quite reasonable. Indeed, the classifier generalizes quite well to the test data in Figure 52b: the PNN does a good job detecting some borderline test 60 mm. There are two outliers in the test data (Master IDs 1444 and 241) which will be difficult to find regardless of the discrimination strategy. Some refinement of the decision surface could provide some improvement in discrimination performance, in particular elimination of a number of test clutter items with large polarization amplitudes (Integral primary > 0 in Figure 52b) and some 60 mm-sized clutter with slow decay (decay rate $\approx$ -4.5 in Figure 52a). Interestingly, adding TOI feature vectors would not provide these improvements; instead, better characterization of the distribution of non-TOI is required. Given the similarity of TEMTADS, Metalmapper, and BUD feature space, regularization of the BUD decision boundary using prior information from other data sets could improve performance. In the simplest case, we could simply refine the decision boundary by manually introducing kernels into the distribution of non-TOI.

The ROC curve for the BUD discrimination strategy compared to the EM61 cart deployed over the same area (Figure 53) demonstrates that performance on this site was comparable. The EM61 rises faster to the 95% detection level, but requires the excavation of a larger number of false-positives to find the last 60 mm mortar. For the BUD, there were a significant number of can't decide anomalies (9, including Master ID 1444) that pushed the ROC curve up to close to the 100% level after 81 false-positives (the one missed item is Master ID 241 which is non-hazardous and likely not TOI). Appendix B provides additional discussion about the BUD data processing approach.

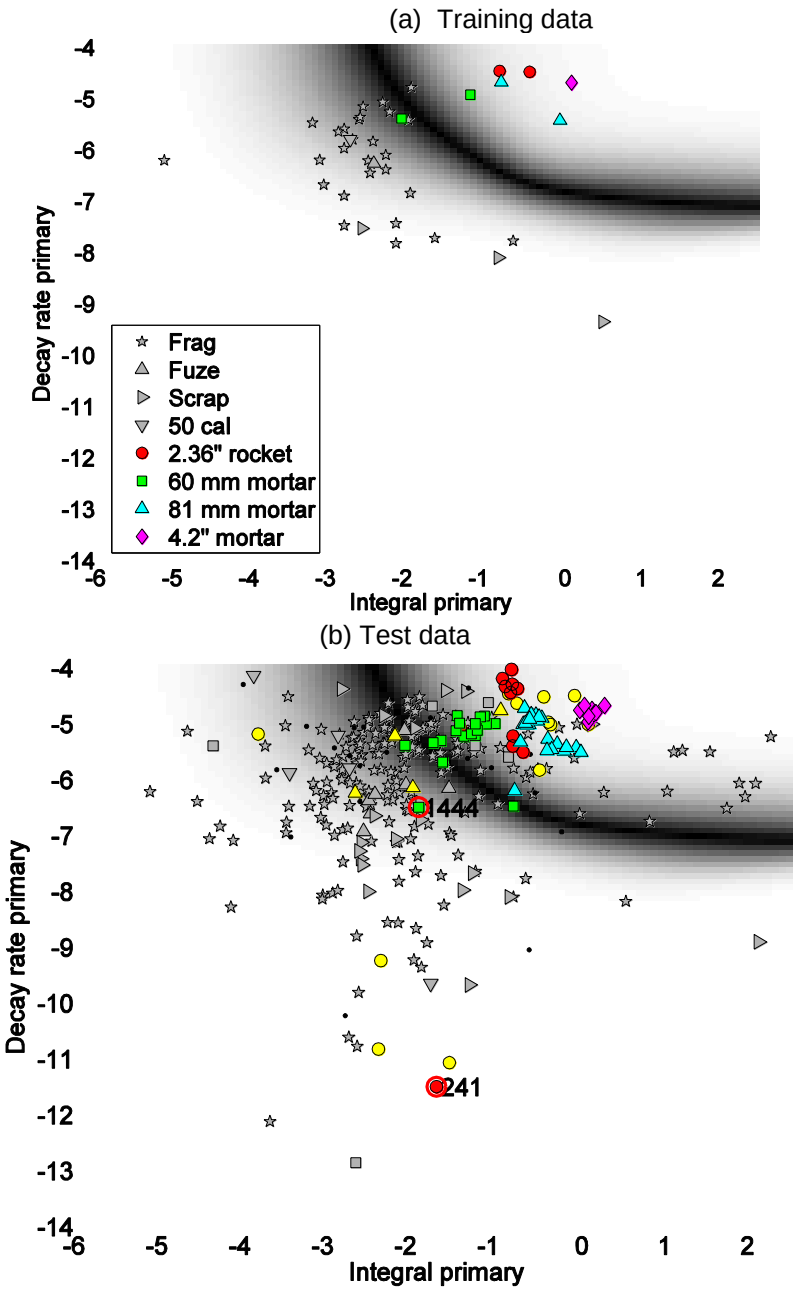


Figure 52. Feature vectors and classification boundary for BUD training (a) and test (b) data.

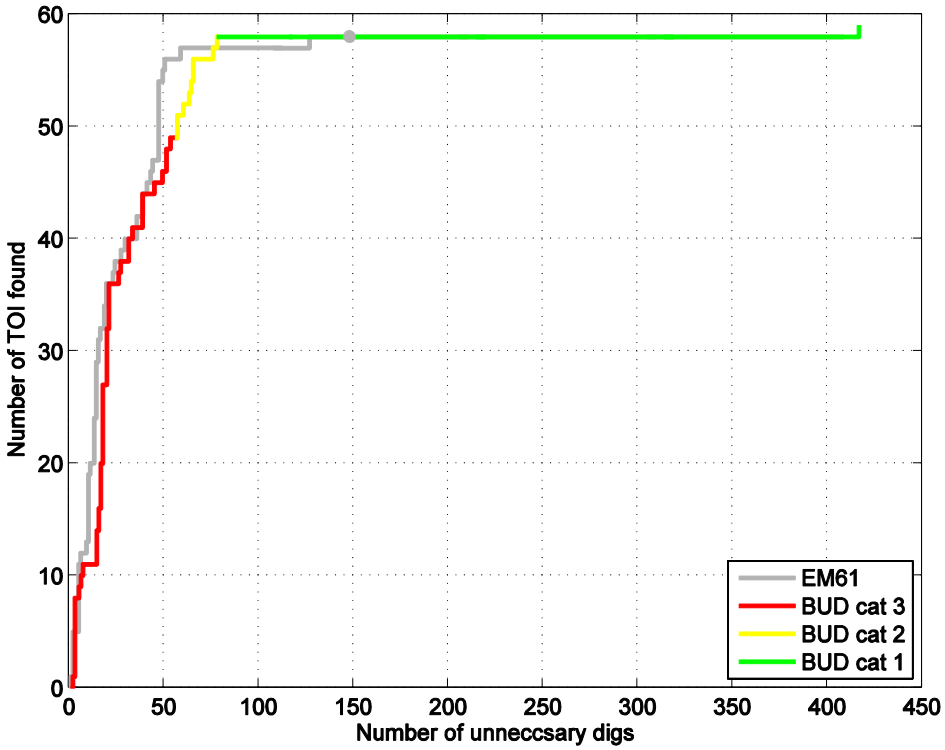


Figure 53. ROC curve for the BUD compared to the EM61 deployed over the same area.

7.5 SUMMARY OF EACH METHOD

Below we provide a summary of the performance of each method.

7.5.1 MTADS magnetometer array: size-based ranking

The detection threshold used for the MTADS magnetometer array was exceeded on over 5000 occasions so magnetic detections were not used in this study. Instead, dipole models were fit to MTADS magnetic data in all locations where the MTADS EM-61 array flagged a detection. Dig-sheet ranking was based on the size of the dipole moment and 460 excavations were required at the operating point: 196 of 200 TOI were recovered. The 60 mm mortars presented the greatest difficulties. Overall, magnetic discrimination was a failure at the SLO site.

7.5.2 EM61 cart: time-decay ranking

A total of 1276 items were detected by the EM61 in the blind-test area including 208 targets of interest. The size of the polarization tensor was not a useful discrimination metric for the 60 mm mortars. Consequently dig-sheet ranking was based on a time-decay rate estimated from polarization tensors fit to the data, with 586 excavations required at the operating point. All 208 TOI were recovered at this point with 378 of 1068 non-hazardous items also excavated (35% of the maximum number of false-positives).

7.5.3 MSEMS cart: time-decay ranking

A total of 1284 items were detected by the MSEMS in the blind-test area including 205 targets of interest. Dig-sheet ranking used the same method as the EM-61 cart with 599 excavations required at the operating point. All 205 TOI were recovered at this point with 394 of 1279 non-hazardous items also excavated (37% of the maximum number of false-positives).

7.5.4 MTADS EM61 array: time-decay ranking

A total of 1187 items were detected by the MTADS EM61 array in the blind-test area including 206 targets of interest. Dig-sheet ranking used the same method as the EM-61 cart with 685 excavations required at the operating point. All but one TOI (a 60 mm mortar) were recovered at this point with 480 of 981 non-hazardous items also excavated (49% of the maximum number of false-positives).

7.5.5 TEMTADS cued-interrogation

A total of 1282 items were included in the blind-test data for the TEMTADS, with 206 targets of interest. Four different methods for dig-sheet ranking were used:

i. Statistical classification: Digsheet ranking was based on using an Expectation Minimization Classifier applied to a 2-D feature space comprising a size and a time-decay feature. 340 excavations were required at the operating point and 203 TOI were recovered along with 137 of 1076 clutter items (12.7% of maximum number of false-positives). The three false negatives comprised two 60 mm mortars and one 2.36” rocket. One 60 mm mortar (Master ID 16) was identified by QC as a TOI but was mistakenly left

off the "can't decide" list. The other 60 mm mortar (Master ID 103) had very low SNR and the recovered polarization tensor model underestimated both the size and time-decay. The 2.36" rocket "false-negative" (Master ID 241) comprised the same rocket motor pieces that generated a false-negative on all next-generation datasets.

ii. Library method: Digsheet ranking was based on comparing the unconstrained polarization tensor fits to polarization tensors constrained by a library of expected ordnance items. At the operating point 335 excavations were required and resulted in the recovery of 204 TOI and 131 of 1076 clutter items (12.2% of maximum number of false-positives). The false negatives comprised the 2.36" rocket motor parts (Master ID 241) and a 60 mm mortar (Master ID 711). This false negative occurred due to confusion regarding which anomaly to fit of several possible options in the field of view of the sensor.

iii. Expert opinion: The initial digsheet ranking was based on the statistical classification method, but an "expert" analyst manually removed items in the TOI list that were thought to be non-TOI. A total of 283 excavations were required at the operating point and 202 TOI were recovered along with 81 of 1076 clutter items (7.5%). The method produced the same three false-negatives as the statistical classification method along with one additional item (Master ID 1285: a multi-object scenario).

iv. Polarization tensor match: Digsheet ranking was based on the match between the recovered polarization tensor and pre-stored polarizations representing the expected ordnance types. A total of 335 excavations were required and 203 TOI were recovered along with 132 of 1076 clutter items (12.3%). False negatives were the same as the statistical classification method.

Ordnance type was predicted by three of the methods. The correct ordnance type was predicted in 179 of 199 cases (90% success rate) for the statistical classification method, for 196 of 200 cases (98% success rate) for the Library method and 185 of 189 cases (98% success rate) for the Expert opinion. All methods had 100% success rate on the 4.2" mortars and only the statistical classifier couldn't achieve 100% success with the 60 mm mortars. The 2.36" rockets and 81 mm mortars were more difficult to distinguish and were occasionally incorrect assigned to the wrong ordnance type.

7.5.6 MetalMapper cued interrogation:

A total of 1409 items were included in the blind-test data for the MetalMapper, with 204 targets of interest. Three different methods for dig-sheet ranking were used:

i. Statistical classification: Digsheet ranking was based on using a PNN classifier applied to a 2-D feature space comprising a size and a time-decay feature. 368 excavations were required at the operating point and 202 TOI were recovered along with 166 of 1205 clutter items (13.8%). The two false negatives comprised Master ID 241 (rocket motor pieces) and Master ID 1502, an unexpected 37 mm projectile.

ii. Library method: Digsheet ranking was based on comparing the unconstrained polarization tensor fits to polarization tensors constrained by a library of expected ordnance items. At the operating point (and after correctly for an initial coding mistake) 378 excavations were required and resulted in the recovery of 203 TOI and 175 of 1205

clutter items (14.5%). The false negative comprised the 2.36” rocket motor parts (Master ID 241).

iii. Expert opinion: The initial digsheet ranking was based on the statistical classification method, but an “expert” analyst manually removed items in the TOI list that were thought to be non-TOI. A total of 258 excavations were required at the operating point and 201 TOI were recovered along with 57 of 1205 clutter items (4.7%). The method produced the same two false-negatives as the statistical classification method along with one additional item (Master ID 775: a multi-object scenario).

Ordnance type was predicted by two of the methods. The correct ordnance type was predicted in 198 of 200 cases (99% success rate) for the statistical classification method, and 184 of 200 cases (92% success rate) for the Library method. The statistical classifier was correct on all 60 mm, 81 mm and 4.2” mortars and wrong with two of the 2.36” rockets. The Library method predicted the correct ordnance type of all 4.2” mortars and all but one 81 mm mortar. Some difficulties were experienced distinguishing the 60 mm mortars and 2.36” rockets, partly because polarizations representing 2.36” rocket warheads and motors were included in the library.

7.5.7 BUD cued interrogation

A total of 473 items were included in the blind-test data for the BUD sensor, including 59 TOI. Digsheet ranking was based on a Probabilistic Neural Network (PNN) classifier applied to a feature space comprising size and time-decay features estimated from the recovered polarizabilities. At the operating point, 139 excavations were required and 58 TOI were recovered along with 81 of 414 non-hazardous items (19.6%). The one false-negative was Master ID 241: the rocket motor pieces.

8.0 COST ASSESSMENT

The demonstration costs for each of the different sensor technologies and cooperative methods were tracked throughout the demonstration. The effort required to perform each element of the processing, modeling, classification, and discrimination was tracked for 9 ranked dig sheets representing different sensor technologies and cooperative inversion approaches. Additionally, preliminary work to adapt UXOLab to analyze the TEMTADS, BUD, and MetalMapper data were performed, and in general improve the efficiency of the entire process. Table 12 presents the detailed breakdown showing labor hours and total costs for each sensor.

Geophysicists at UBC-GIF provided significant support in addition to the SKY staff. Their labor hours for inversion, classification and QC were accurately tracked, and costs assigned using equivalent labor categories. Time-spent on preparatory activities were not tracked by UBC-GIF, thus the estimates shown in Table 12 under that category are probably at least a factor of 2 or 3 lower than actual costs.

The cooperative inversion costs reflect the additional cost to perform the cooperative inversion after single inversions had been completed.

Discrimination costs are shown for the statistical classification method for each of the advanced datasets. Costs for the library method would be slightly different (lower) while that for the expert analysis would be higher (have to include time for an experienced interpreter to analyze each anomaly).

Table 12. Cost Breakdown for the San Luis Obispo Discrimination Study.

Category	Prep		Inversion		Classification		QC		Total	
	Hours	Cost	Hours	Cost	Hours	Cost	Hours	Cost	Hours	Cost
EM61 Cart	12	$1,231	79	$8,126	7	$1,274	14	$2,548	112	$13,180
MSEMS EM61	11	$1,131	93	$10,675	5	$511	19	$2,932	128	$15,249
MSEMS cooperative			29.5	$3,034			14	$2,548	48.5	$6,093
MTADS Magnetometer	2.5	$255	115	$9,117	3	$546	16	$2,913	136.5	$12,831
MTADS EM61	9	$1,059	111.5	$11,422	5	$511	4	$728	129.5	$13,720
MTADS cooperative	4	$728	2	$206			2	$206	8	$1,140
TEMTADS	71	$7,456	283	$29,031	39	$4,623	79	$9,726	472	$50,836
MetalMapper	63	$2,571	97	$6,588	80	$8,491	49	$5,964	289	$23,614
BUD	40	$4,137	40	4,086	8	$1,456	3	$366	51	$5,959
General Activities									72	$9,455
Total	**213**	**$18,568**	**850**	**$82,245**	**147**	**$17,412**	**200**	**$27,931**	**1,482**	**$155,652**

9.0 MANAGEMENT AND STAFFING

A flow chart showing the managerial hierarchy and the relationship between the principal investigator (PI) and other personnel is shown in Figure 54.

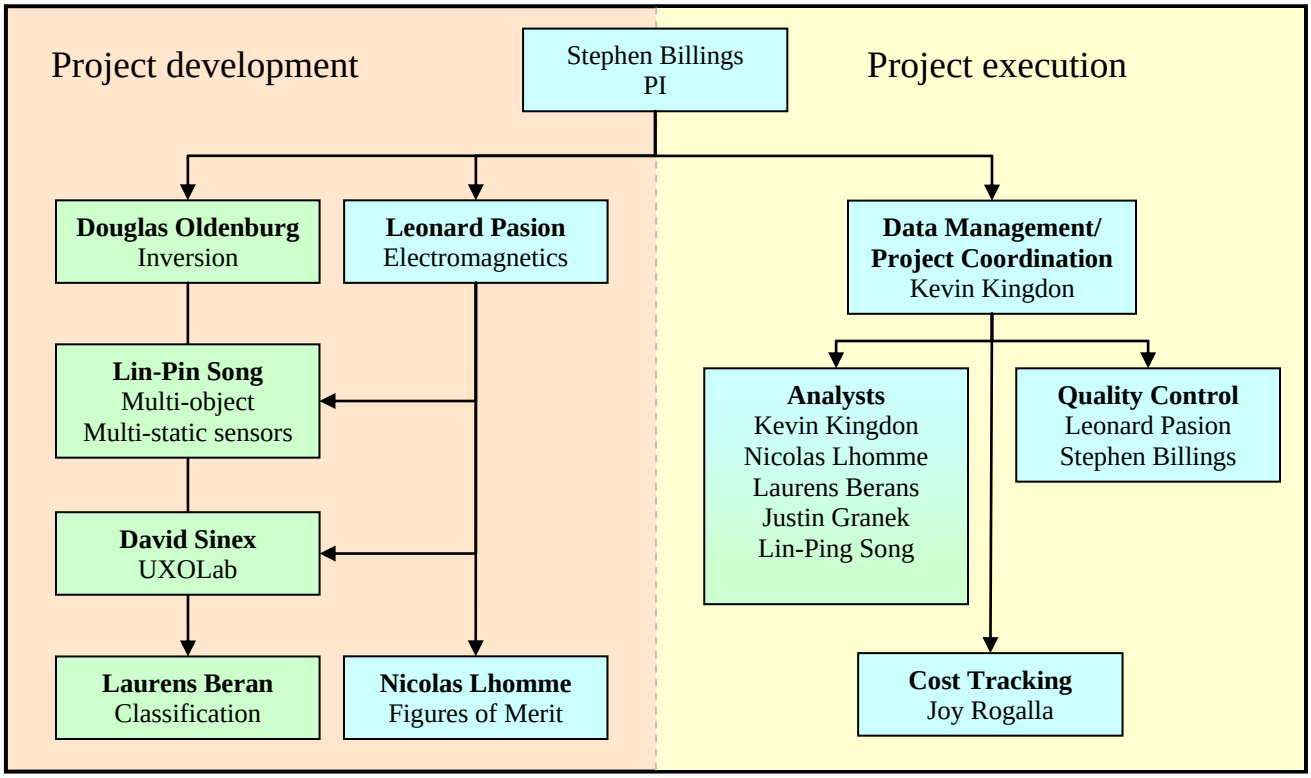


Figure 54. Project management hierarchy showing Sky Research personnel in blue and UBC-GIF personnel in green. The hierarchy is split between the development and execution components.

10.0 REFERENCES

Bell, T. H., B. J. Barrow, and J. T. Miller, 2001, Subsurface discrimination using an electromagnetic induction sensor, *IEEE Transactions on Geoscience and Remote Sensing*, 39, 1286 – 1293.

Billings, S. D., L. R. Pasion, and D. W. Oldenburg, 2002, Discrimination and identification of UXO by geophysical inversion. Phase II: Inversion of total-field magnetics, USACE Engineer Research and Development Center, Tech. Rep. ERDC/GSL TR-02-16.

Billings, S.D., 2004, Discrimination and classification of buried unexploded ordnance using magnetometry: *IEEE Transactions of Geoscience and Remote Sensing*, 42, 1241 - 1251.

Billings, S.D., 2008, Data Modeling, Feature Extraction, and Classification of Magnetic and EMI Data, ESTCP Discrimination Study, Camp Sibert, AL, Project 200504: Practical Discrimination Strategies for Application to Live Sites, Demonstration Report. ESTCP Program Office, September.

Carin, L., Zhang, Y. & Liao, X., 2004, Detection of Buried UXO via Active Selection of Labeled Data, proceedings from the UXO Forum, St. Louis, March 9-12, 2004.

Collins, L., Zhang, Y., Li, J., Wang, H., Carin, L., Hart, S., Rose-Phersson, S., Nelson H. and McDonald, J. R., 2001, A comparison of the performance of statistical and fuzzy algorithms for unexploded ordnance detection: *IEEE Transactions on Fuzzy Systems*, 9, 17-30.

Guyon I., J. Weston, S. Barnhill, V. Vapnik, Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning, 46, 2002.

Hart, S.J. et al., 2001, Using Physics Based Modeler Outputs to Train Probabilistic Neural Networks for Unexploded Ordnance (UXO) Classification in Magnetometry Surveys. *IEEE Trans. Geosci. Remote Sensing* 39, 797-804.

Pasion, L. & Oldenburg, D. 2001, A Discrimination Algorithm for UXO Using Time Domain Electromagnetics: *Journal of Engineering and Environmental Geophysics*, 28, 91-102.

Pasion, L. R., Billings, S. D. & Oldenburg, D. W., 2003, Joint and Cooperative Inversion of Magnetics and Electromagnetic Data for the Characterization of UXO Discrimination Problems: *The Symposium on the Application Geophysics to Engineering and Environmental Problems (SAGEEP)*, San Antonio, Texas, U.S.A., Feb 6-10.

Smith and Lee. The moments of the impulse response: A new paradigm for the interpretation of transient electromagnetic data. Geophysics, 2002.

Wax, M. & Kailath, T., 1985**,** Detection of Signals by Information Theoretic Criteria: IEEE Transactions On Acoustics, Speech, And Signal Processing, ASSP-33, NO. 2, 387-392.

Zhang, Y., Collins, L. M., & Carin, L., 2003a, Model-based statistical signal processing for UXO discrimination: Performance results from the JPG-V demonstration: *Proceedings of SPIE Volume 5089*, 1116-1126.

Zhang, Y., Collins, L., Yu, H., Baum, C. E. & Carin, L., 2003b, Sensing of Unexploded Ordnance with Magnetometer and Induction Data: Theory and Signal Processing. *IEEE Trans. Geosci. Remote Sensing*, 41, 1005-1015.

Zhang, Y., E. Dura, L. Carin, 2004, Detection of Buried Targets Via Active Selection of Labeled Data: Application to Sensing Subsurface UXO. IEEE Trans Geosci. Remote Sensing. 42.

APPENDICES

Appendix A: Training Memo

Training Memorandum

Feature Extraction and Classification of Magnetic and EMI Data, San Luis Obispo, CA

ESTCP MM-0504 Practical Discrimination Strategies for Application to Live Sites

Sky Research, Inc. & The University of British Columbia

September 10, 2009
Draft
Version 2.0

1. INTRODUCTION

As part of the ESTCP UXO Discrimination Study, Sky Research and UBC-GIF will submit the following 7 dig-sheets:

a) Magnetics, size-based: Production of a dig-sheet ranked according to dipole moment;
b) EM-61, statistical (Contractor): Statistical classification of features derived from the Contractor EM-61 data and the production of a ranked dig-sheet;
c) EM-61, statistical (MTADS): Statistical classification of features derived from the MTADS EM-61 data and the production of a ranked dig-sheet;
d) EM-61 and magnetics, statistical: As per b) but with EM-61 fits constrained by the magnetics data and with the addition of the features from the magnetometer data (remanence, moment etc); and
e) Man-Portable Simultaneous EMI and Magnetometer System (MSEMS): cooperative inversion of the MSEMS EM61 data using depth constraints from the MSEMS magnetometer;
f) Time Domain Electromagnetic Towed Array Detection System (TEMTADS) cued interrogation array data: Statistical classification of features derived from the TEMTADS;
g) TEMTADS library: Library based discrimination applied to TEMTADS data;
h) MetalMapper cued interrogation data: Statistical classification of features derived from the MetalMapper data.
i) MetalMapper library: Library based discrimination applied to MetalMapper data;

We had intended to process and analyze data from the Berkeley UXO Discriminator, but that was not delivered to us on time.

This is an updated version of an earlier draft of this document which includes the discrimination strategies for the TEMTADS and MetalMapper datasets.

This document describes the fitting parameters used for each data type, and discusses the ranking strategy for each method.

2. MTADS MAGNETOMETER DATA

a. Analysis of Test-Pit data

Magnetometer data were collected over 60 and 81 mm mortars, 2.36” rockets and 4.2” mortars which were placed in a test-pit at a range of different orientations and depths. The following parameters were used to invert the magnetometer data from the test-pit.

- Earth’s magenetic field: Inclination = 59.53^{o}, Declination = 13.59^{o}, Magnitude = 48000 nT;
- Noise-floor = 1.0 nT;
- Default circular masks of 3 m diameter.

All dipole model fits were found to be acceptable over all orientations of all items. The recovered moments in directions parallel and perpendicular to the Earth’s magnetic field are plotted in Figure 1 along with “dipole feasibility curves” for each of the ordnance items.

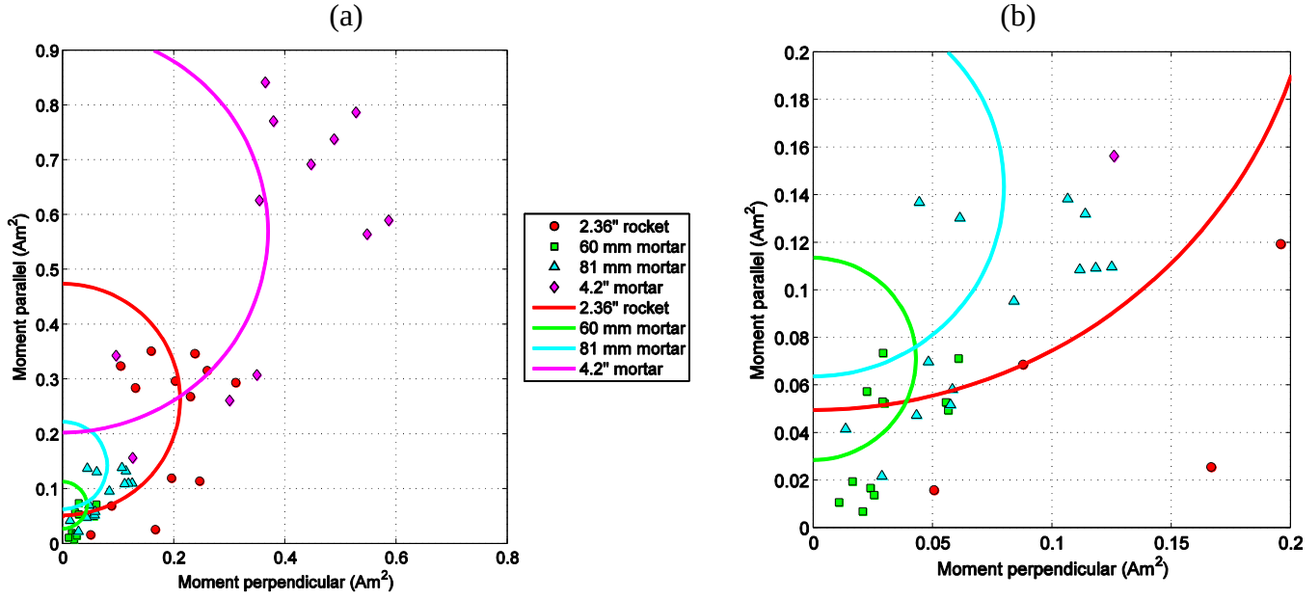


Figure 1: (a) Fitted magnetic moments from the testpit with dipole feasibility curves overlain. The plot in (b) is a zoomed in version of the plot in (a).

Each of the items tends to cluster around its respective dipole feasibility curve. We not that both the 2.36” rocket and 60 mm mortar appear to have considerable remanence as evidenced by the fits with large angles relative to the Earth’s magnetic field. These would be ranked low by the apparent remanence metric.

If the sensors are assumed to be 25 cm off the ground (rather than the nominal 30 cm in the demonstration plan), then there is good agreement between predicted and actual depths (Figure 2). All predicted depths are within 21 cm, and all but five are within 10 cm of the actual depths. The depths of the deeper items appear to be slightly under predicted.

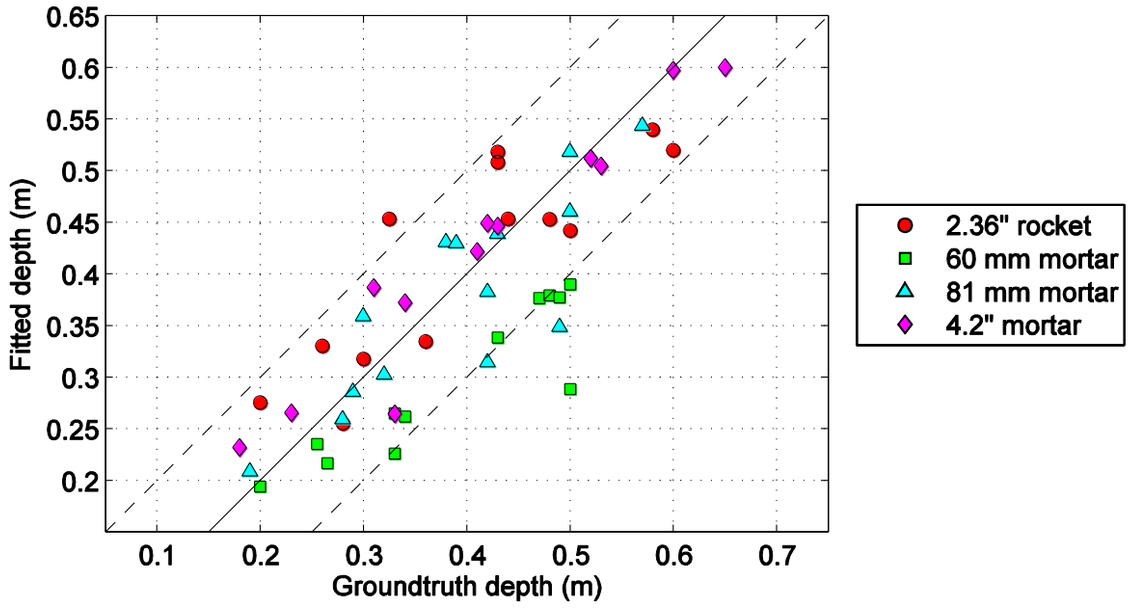


Figure 2: (a) Predicted versus actual depths for the MTADS magnetometer data on the test-pit.

b. Analysis of ground-truth data

The training data were inverted using the same procedures as the testpit. There were 142 inverted anomalies, with 103 having acceptable dipole fit and 39 having an unacceptable fit. The dipole parameters and ground-truth information are summarized in Figure 3a and b. A number of the TOI in the training data have large angles relative to the Earth's magnetic field, either due to magnetic remanence or error in the fitting process. There are several 60 mm projectiles with very small moments, the smallest of which is just over 0.01 Am^2. Many of the small moments come from 60 mm mortars that just comprise the main-section of the projectile, without the tail-boom attached. Items with moments as low as 0.01 Am^2 will need to be excavated as potential UXO indicating that discrimination is not likely to be very efficient at this site.

Figure 4 shows that the recovered depths agree reasonably well with the reported ground-truth depths. Most items are within 20 cm (with the majority within 10 cm) but there is a tendency to push some of the small shallow items a bit deep.

c. Discrimination strategy

With the potential for significant remanence in the TOI, we will opt to use the moment to prioritize digging order. We take the viewpoint that if the moment is very small that the item cannot possibly be an TOI. Figure 5 compares the size of the moment against the apparent magnetic remanence. The TOI tend to have large moments and small apparent remanence, but there are exceptions in both cases. Figure 6 shows the ROC curves that arise when ranking by either quanitity. Apparent remanence outperforms the moment in the middle section of the curve, but both require that most of the training data be excavated to recover all TOI. We would not use magnetic discrimination at this site, but for comparative purposes, we will opt to submit a digsheet with a moment of 0.01 Am^2 used to delineate the dig/no-dig threshold.

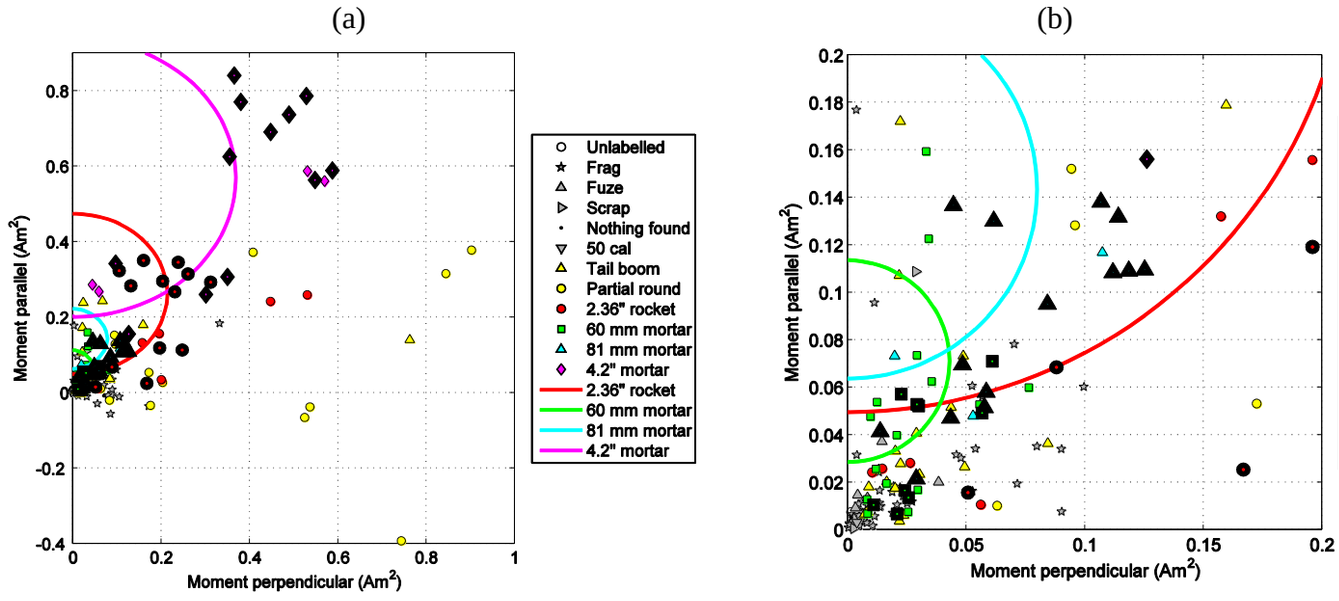


Figure 3: Dipole model fits to ground-truth items at SLO: (a) moments perpendicular and parallel to Earth's field; and (b) same as (a) but with reduced range.

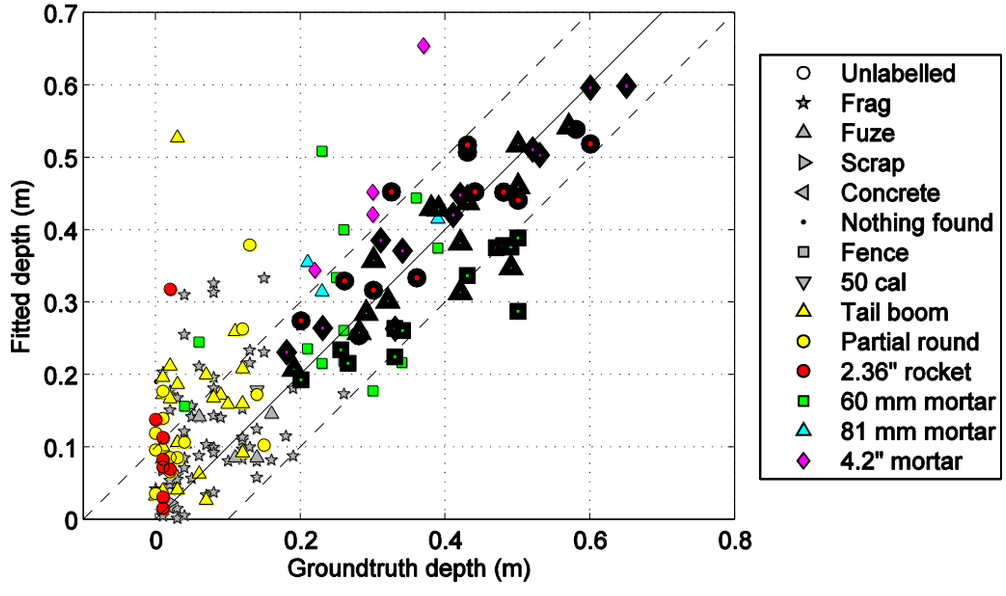


Figure 4: Predicted versus actual depths for the MTADS magnetometer data for the training and test-pit data (with the latter delineated by bold symbols).

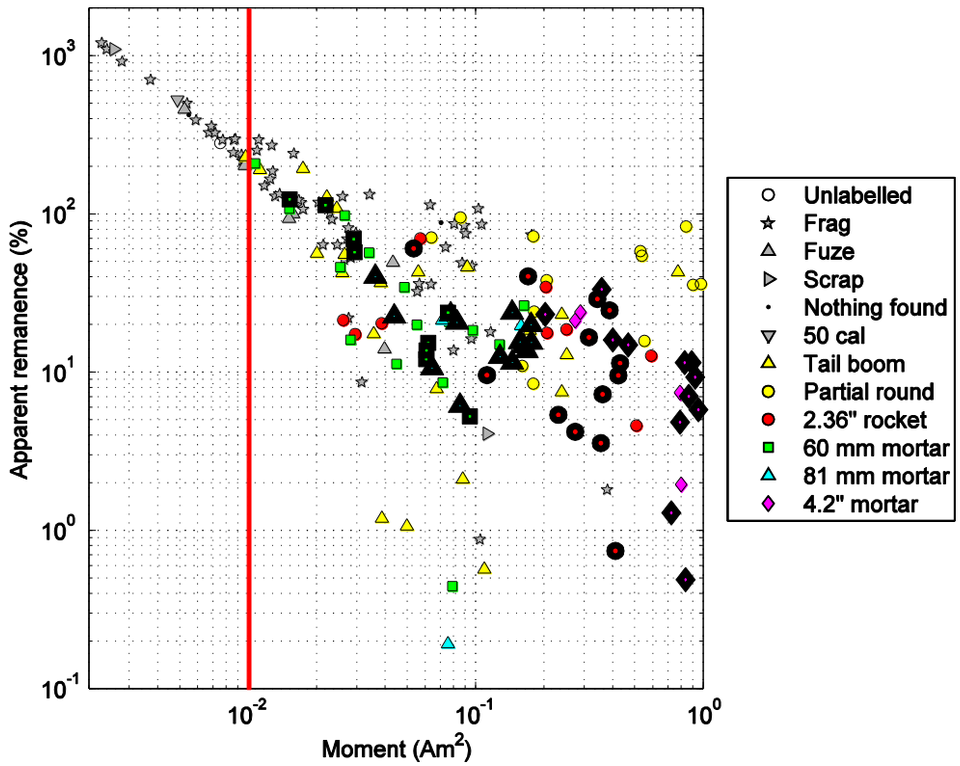


Figure 5: Moments and remanence of seeded UXO and unlabelled items, along with cutoff values for digging prioritized by moment and by remanence.

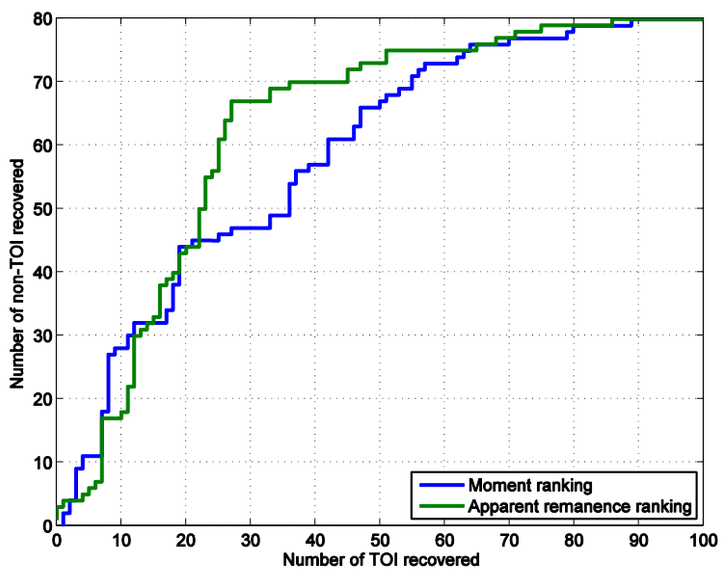


Figure 6: Moments and remanence of seeded UXO and unlabelled items, along with cutoff values for digging prioritized by moment and by remanence.

3. EM-61 CART DATA

a. Analysis of Test-pit results

Data were collected over the four primary targets of interest (TOIs) in horizontal, 45° nose down, and vertical nose down and nose up orientations with a Geonics EM-61 cart recording at four time gates. The data were processed within UXOLab using the following workflow:

- *Error estimation:* Estimates of data errors consisted of two components:
  - Floor noise was estimated from measurements of the testpit area made without targets present.
  - For each datum, we calculated an additional error equal to 10 percent of the datum amplitude.
- *Masking*: The spatial extent of data used in the inversions was determined using the elliptical masking technique described in the Demonstration Plan.
- *Model*: The data were fit using 2 unique polarizations for the dipole tensor. The amplitude of each polarization was estimated at each of the four time channels. The 13 element model vector is

  $m = [X, Y, Z, \varphi, \theta, L_1(t_1), L_1(t_2), L_1(t_3), L_1(t_4), L_2(t_1), L_2(t_2), L_2(t_3), L_2(t_4)]$

  where (X,Y,Z) is the location, (φ,θ) are the orientation angles and $L_i(t_j)$ is the ith polarization at the jth time channel.

All 25 data sets acquired over test-pit targets were successfully inverted using this procedure. Sensor height was assumed to be 0.4 m for all inversions. Figure 7 compares estimated and actual depths obtained for these targets. There is poor agreement between fitted and actual depths for these inversions.

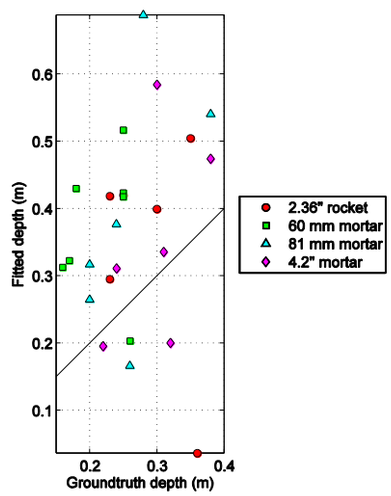


Figure 7. Fitted vs. groundtruth depths for testpit items

Figure 8 shows the fit to observed test-pit data acquired over a horizontal 81 mm mortar, with the misfit versus target depth for this inversion shown in Figure 9. The fit to the observed data is quite good and the target depth is reasonably well constrained, and yet the recovered depth (58 cm) errs significantly from the reported ground truth depth (38 cm).

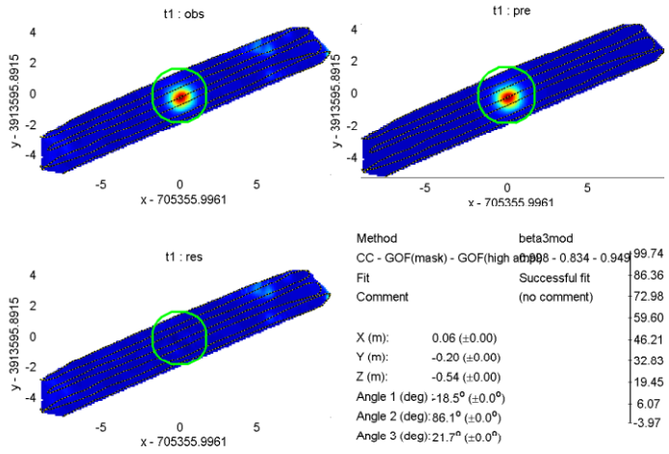


Figure 8. Observed, predicted and residual data for channel 1 of EM61 cart data acquired over a horizontal 81 mm target.

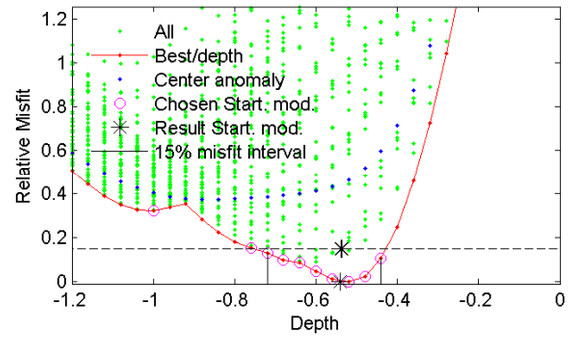


Figure 9. Misfit versus depth curve. Estimated target depth is -0.58 m, true depth is -0.38 m.

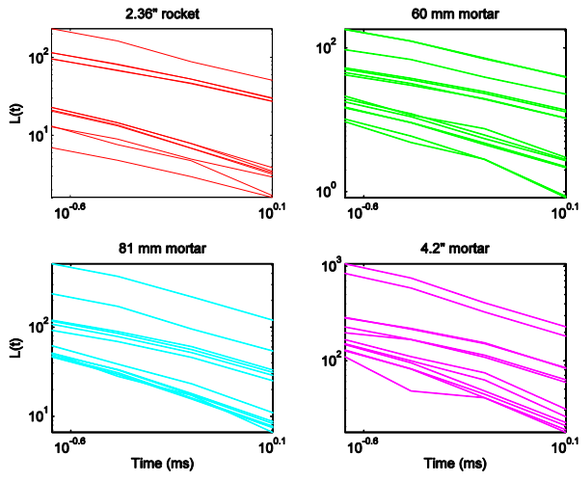


Figure 10. Estimated 2 dipole polarizations from EM61 cart testpit data

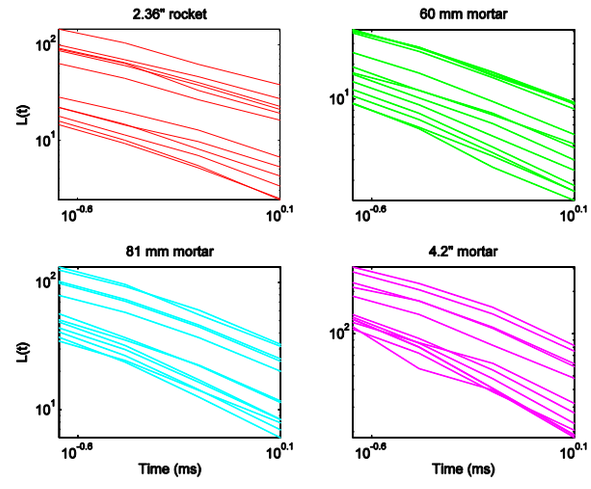


Figure 11. Estimated 2 dipole polarizations from EM61 cart test-pit data, with estimated depth constrained by ground truth depths

Figure 10 shows the estimated polarizations (both $L_1(t_j)$ and $L_2(t_j)$) from all targets in the test-pit data. While the rate of decay of polarizations is generally consistent, the polarizations do not cluster in amplitude because we are unable to sufficiently constrain target depth for these data. Figure 11 shows the same fits, but with target depth constrained to lie within +/- 5 cm of the ground truth depth. There is a somewhat improved grouping of the polarizations.

b. Analysis of training data

EM-61 cart training data were inverted with a three dipole instantaneous amplitude model. We use a three-dipole model (rather than a two-dipole model as for the test-pit data) because many targets in the training data are expected to be non-axisymmetric. Of the 198 EM-61 training data targets, 68 were judged by the data analyst to have passable fits to the observed data. Figure 12 and Figure 13 compare inversions for passed and failed targets in this data set. Figure 12 is representative of many of the failed inversions in that the gridded data image shows lobes on the target anomaly. These were due to inadequate lag correction of the data and were subsequently relagged and reanalyzed, which reduced the number of failed anomalies by 18.

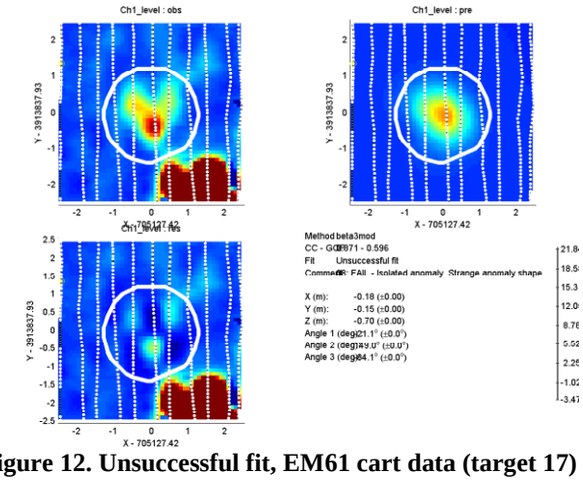


Figure 12. Unsuccessful fit, EM61 cart data (target 17)

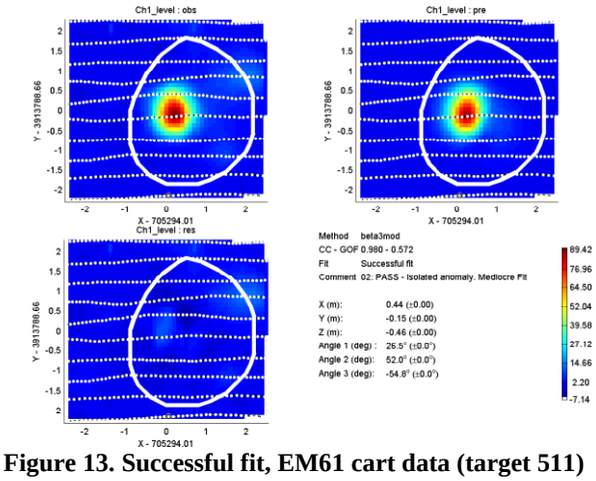


Figure 13. Successful fit, EM61 cart data (target 511)

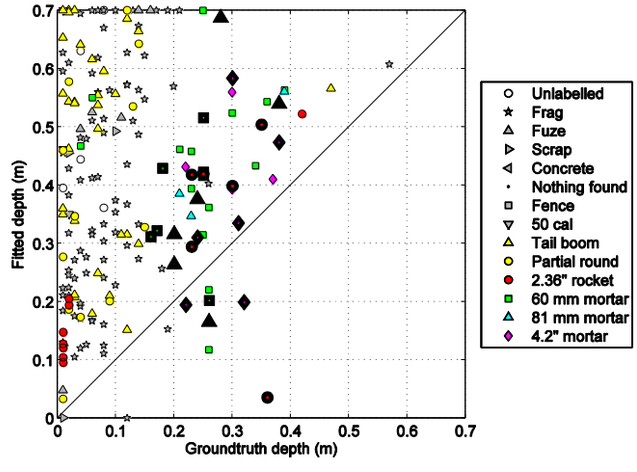


Figure 14. Fitted vs. ground truth depths for training and test-pit (bold symbols) targets

As seen with the test-pit targets, estimates of target depth are generally deeper than ground-truth depths (Figure 14). All estimated depths are shown here, regardless of the pass/fail status of the inversion. This is not inconsistent with the results of previous demonstrations, where we found there was considerable ambiguity in the optimal target depth owing to local minima of the misfit function. One option to reduce the

sensitivity of the result is to use multiple models corresponding to local minima of the misfit function. However, we find that for many cases in the training data that local minima near the correct target depths are not present and the inversion drives the solution to the lower depth bound (Figure 15).

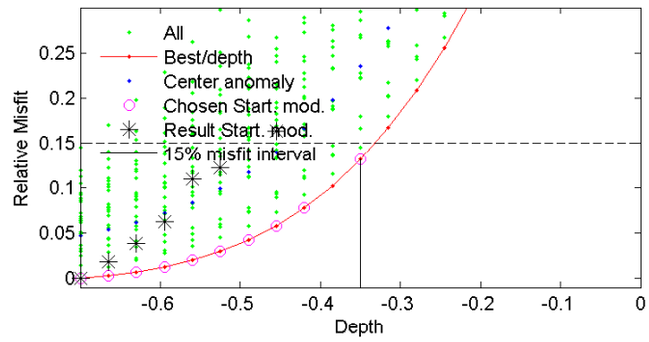


Figure 15. Misfit versus depth curve for target 410, EM61 data. True target depth is 25 cm.

We also computed the following data features for both the training and test-pit data

- Mean decay: the mean of $d_i(t_4)/d_i(t_1)$ for all soundings within the target mask, restricted to only include values which fall within [0,1].
- Data energy: the energy $\Sigma d_i^2(t_1)/N$ within the target mask at the first time channel (N=number of data in the mask)

Figure 16 shows data features for the test-pit and training data sets. There is good correspondence between decay rates from training and test-pit data features for larger targets of interest (4.2", 2.36", 81 mm). The 60 mm training items tend to have lower data energy than the test pit items, likely because the training data 60 mm often do not have an intact tail. We note also one fast-decaying (small mean decay) 2.36" rocket warhead which is an outlier to the overall distribution of TOIs. The ground truth photo of this target (#1260, Figure 19) suggests that this item is cracked, and this defect might greatly increase the rate of decay of induced currents.

Figure 17 shows inversion-based features extracted from observed EM-61 training data with a three-dipole instantaneous amplitude model. All estimated features are shown here, regardless of the pass/fail status of the inversion. The calculated features are

- Polarization amplitude = $(\Sigma L_i(t_1)^2)^{1/2}$
- Polarization decay = $(\Sigma L_i(t_4)^2)^{1/2} / (\Sigma L_i(t_1)^2)^{1/2}$

The polarization amplitude is proportional to the amplitude of the induced moment in the presence of a unit primary field along each of the target's principle axes. The polarization decay is the ratio of the polarization amplitude at the fourth and first time channels. These features are analogous to those previously used for discrimination in the Camp Sibert demonstration, where we used the amplitude and decay of the primary polarization. Using the total polarization amplitude here simplifies the computation somewhat by eliminating the requirement to identify the primary polarization. The polarization amplitude is strongly correlated with target depth and so is not a particularly useful feature for discrimination with these data.

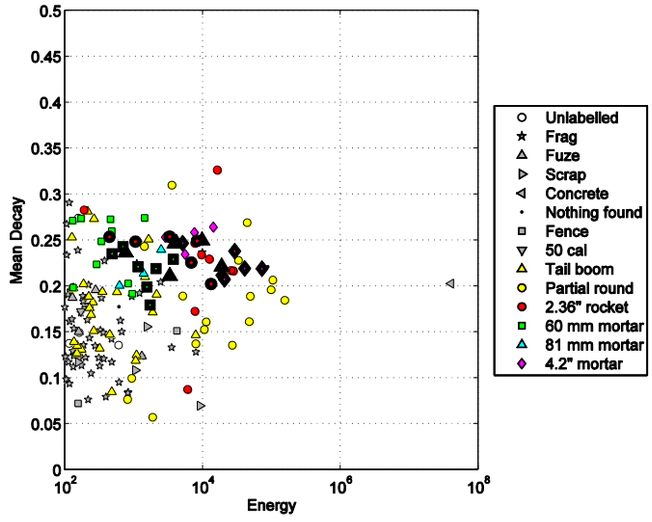


Figure 16. Data features extracted from EM61 cart data. Features in bold are test-pit items

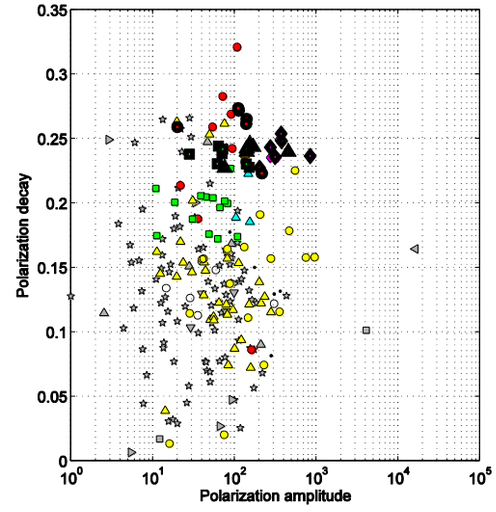


Figure 17. Model features extracted from EM61 cart data. Features in bold are test-pit items.

In Figure 17 we find that the polarization decay is a useful parameter for discrimination between targets of interest and clutter. Again there is an outlying TOI corresponding to the cracked 2.36" warhead in Figure 13. The polarization decay seems relatively insensitive to errors in depth estimation: Figure 18 shows a misfit versus depth curve for a test-pit target of interest (81 mm), with a true depth of 0.27 m. The solution at depth greatly overestimates polarization amplitude, but the shallow and deep solutions have similar polarization decays of 0.23 and 0.24, respectively.

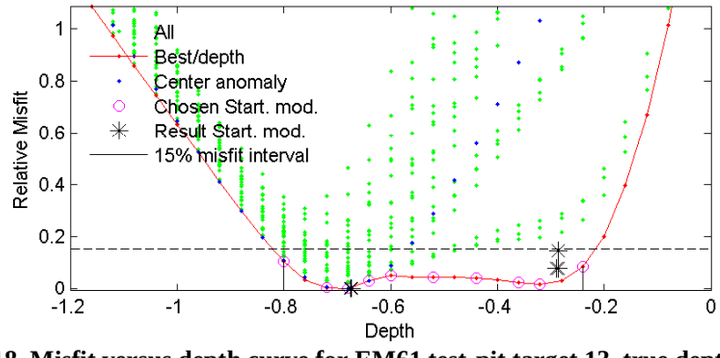


Figure 18. Misfit versus depth curve for EM61 test-pit target 13, true depth is 0.27 m.

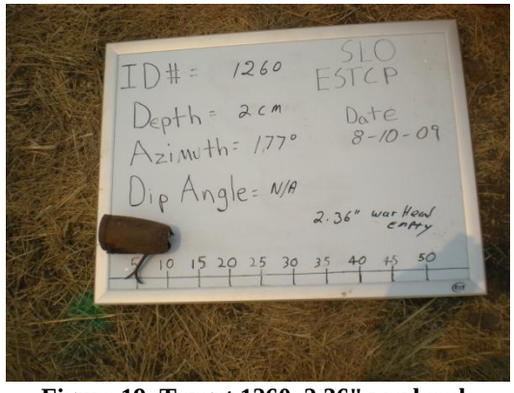


Figure 19. Target 1260, 2.36" warhead.

c. Creation of a dig-list for the EM-61 cart

Canonical analysis of the data and inversion features in Figure 16 and Figure 17 confirms that the model-based polarization decay feature provides the best separation between targets of interest and clutter items (Figure 20).

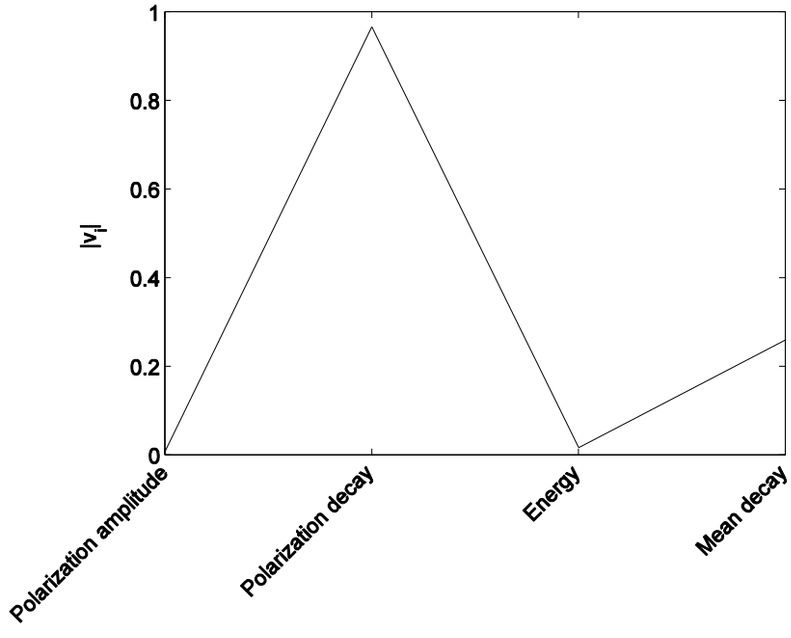


Figure 20. Components of the first canonical eigenvector applied to data and model features.

Based upon this analysis, we propose to generate a dig-list for the EM-61 cart data using the polarization decay parameter extracted with an inversion. Because a single parameter will be used to discriminate between targets of interest and clutter, no statistical classifier is required: we can simply threshold on polarization decay, starting with large values.

Figure 21 compares the bootstrap performance of this approach with various statistical classifiers. Based on these simulations, mean expected performance of a threshold on polarization decay is comparable to, or slightly better than, statistical classification with both model-based features.

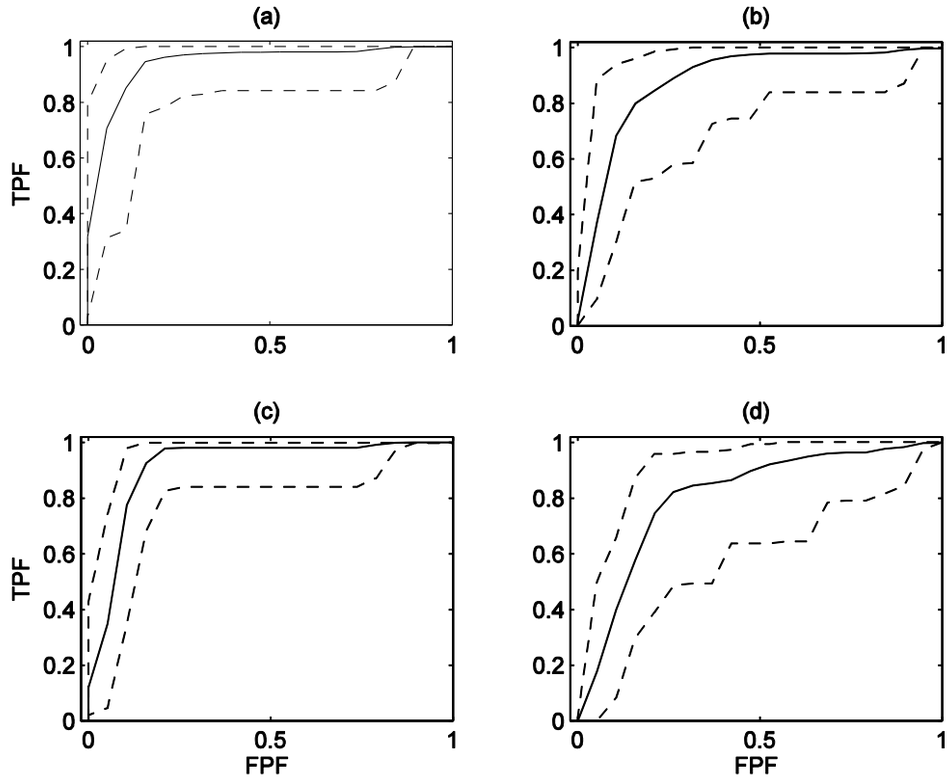


Figure 21. 0.632 bootstrap comparison of classification strategies for EM-61 data. Solid line is mean performance and dashed line shows maximum and minimum bounds over all bootstrap resamples. (a) Quadratic discriminant analysis (QDA) applied to two-dimensional feature space spanned by polarization amplitude and decay (model features). (b) QDA applied to two-dimensional feature space spanned by energy and mean decay (data features). (c) Threshold on polarization decay. (d) Threshold on energy.

Anomalies with high signal energy at the first time channel (greater than 10^6, as shown in Figure 16) that fall after our predetermined cut-off value will be classified as can't analyze to ensure they are labelled. We determine a cut-off value the for polarization decay parameter based upon the following analysis (bootstrapping cannot be employed here because there is no statistical classifier training). We assume that, based upon the training data, targets of interest comprise 25 percent of the EM61 target picks. We also assume that the distribution of the polarization decay for ordnance targets is normal, with mean and variance also estimated from the training data (the outlying 2.36" rocket is not included in this analysis). An upper bound on the last occurrence of TOIs can then be computed by integrating the distribution of TOI polarization decay up to a critical value corresponding to the probability P=1/N, where N is the expected number of TOIs in the test data. This produces a cutoff value of 0.1456 for the polarization decay (Figure 22).

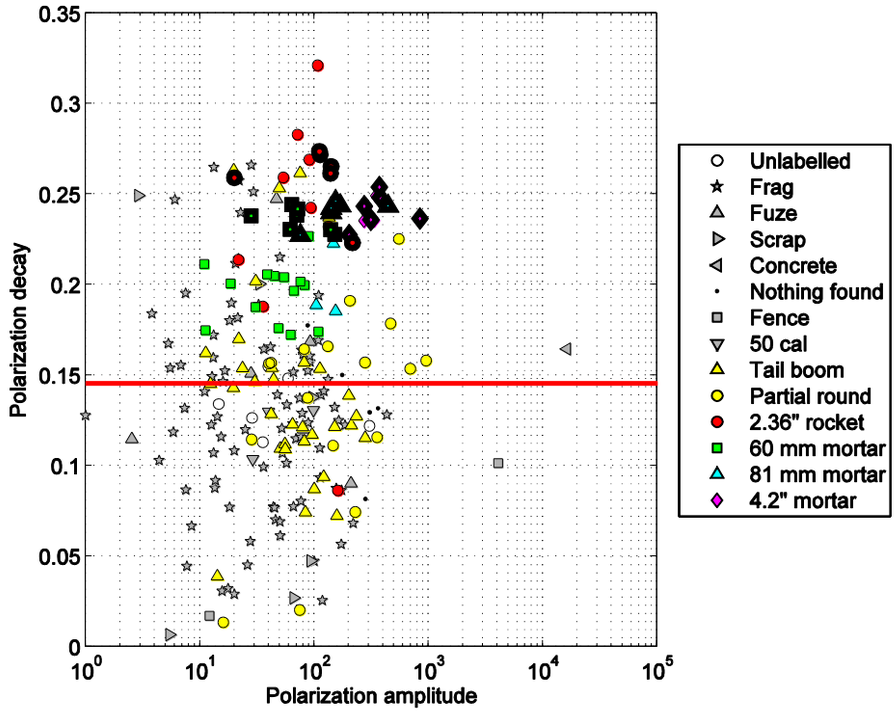


Figure 22. EM-61 model features from test-pit and training data showing estimated cut-off in polarization decay for test data.

4. MTADS EM-61 ARRAY DATA

a. Analysis of Test-pit results

Data were collected over the four primary targets of interest (TOIs) in horizontal, 45° nose down, and vertical nose down and nose up orientations with an MTADS EM-61 array recording at four time gates. The data were processed within UXOLab using the following workflow:

- *Error estimation:* Estimates of data errors consisted of two components:
  - Floor noise was estimated from measurements of the testpit area made without targets present.
  - For each datum, we calculated an additional error equal to 10 percent of the datum amplitude.
- *Masking*: The spatial extent of data used in the inversions was determined using the elliptical masking technique described in the Demonstration Plan.
- *Model*: The data were fit using 2 unique polarizations for the dipole tensor. The amplitude of each polarization was estimated at each of the four time channels. The 13 element model vector is

  $m = [X, Y, Z, \varphi, \theta, L_1(t_1), L_1(t_2), L_1(t_3), L_1(t_4), L_2(t_1), L_2(t_2), L_2(t_3), L_2(t_4)]$

  where (X,Y,Z) is the location, (φ,θ) are the orientation angles and $L_i(t_j)$ is the ith polarization at the jth time channel.

All 53 data sets acquired over test-pit targets were successfully inverted using this procedure. Sensor height was assumed to be 0.335 m for all inversions. Figure 23 compares estimated and actual depths obtained for these targets. Depth recovery is somewhat improved relative to the EM-61 cart. Poor depth estimation for some targets is likely attributable to poor data coverage: test-pit measurements were made with a single pass over the target.

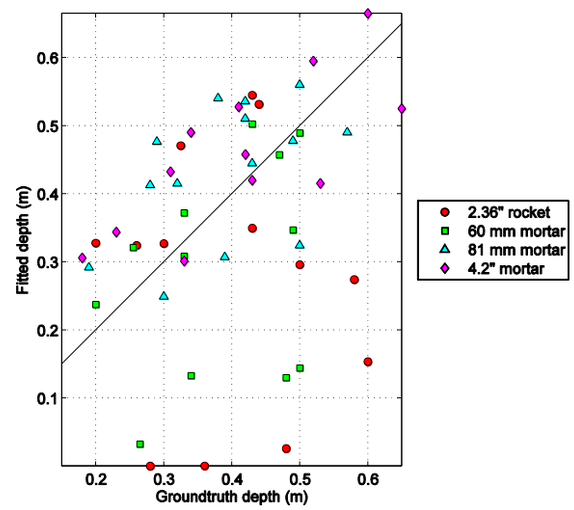


Figure 23. Fitted vs. groundtruth depths for testpit items, MTADS EM-61 array.

b. Analysis of training data

MTADS EM-61 training data were inverted with a three dipole instantaneous amplitude model. Floor noise was estimated by windowing several anomaly-free areas of the survey and computing data standard deviations within these windows. Figure 24 shows noise standard deviations for the three sensors in the MTADS array for all windows. Sensor 2 has, on average, a slightly elevated noise level relative to the neighboring sensors, but the difference is not sufficiently large to motivate using a separate noise floor for each sensor. Based upon this analysis, we used a noise floor of 6 mV plus a 10 percent error when inverting the MTADS EM-61 training data

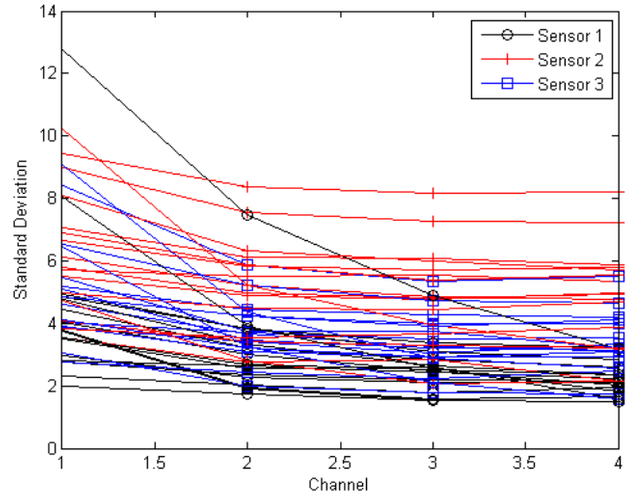


Figure 24. Background noise estimation for MTADS EM-61 data.

Of the 182 EM-61 training data targets, 169 were judged by the data analyst to have passable fits to the observed data. Figure 25 compares actual and estimated depths for these data.

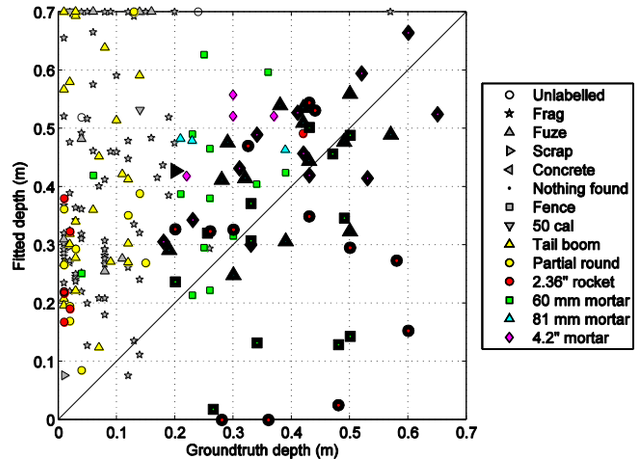


Figure 25. Fitted vs. groundtruth depths for training items, MTADS EM-61 array.

The majority of the training targets are deeper than their reported ground truth depths. A large number of non-TOI items are estimated to be at the maximum depth constraint of 0.7 m. Many of these targets are associated with low SNR, diffuse anomalies which can only be reproduced by placing the dipole source at depth. For example, Figure 26 and Figure 27 show inversion results and misfit versus depth curves for a low SNR anomaly.

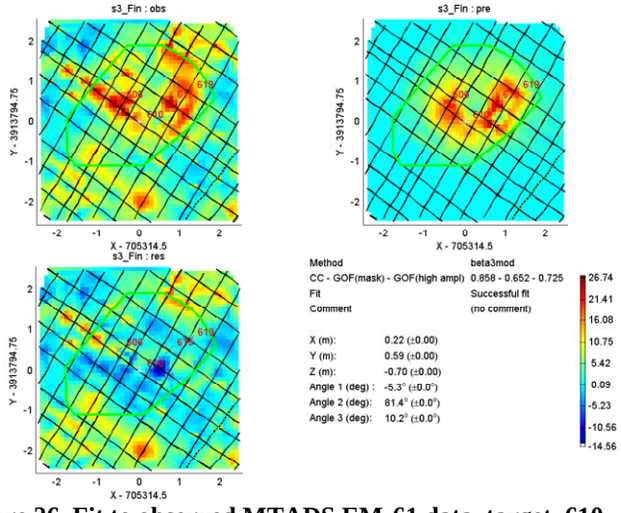


Figure 26. Fit to observed MTADS EM-61 data, target 610.

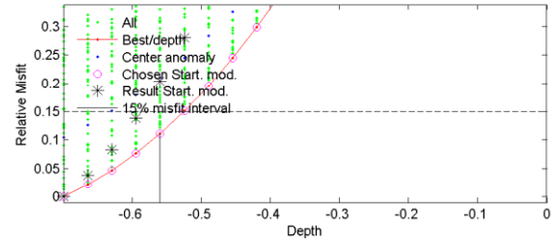


Figure 27. Misfit versus depth curve for target 610,.

Although several targets have been picked on this anomaly, we can fit the data reasonably well with a single target at 0.7 m. Figure 28 shows the ground truth at this location: the anomaly is produced by many small pieces of frag. Other targets within the mask of Figure 4 occur at shallower depths.

Figure 28. Ground truth for target 610.

Figure 29 shows estimated model-based features for the MTADS EM-61 data, with both passed and failed inversions included. These features are the same as those described for the EM-61 cart in this memo. At first inspection the separation between TOIs and clutter on the basis of the decay parameter appears much worse than for the EM-61 cart, with many fast-decaying ordnance items which appear as outliers to the distribution of TOIs. One of these is the cracked 2.36" warhead previously encountered in the EM-61 cart training data. All outlying test-pit items correspond to horizontal cross-track measurements. Because test-pit

measurements were made with a single pass over the target, these data only excite the fast-decaying, transverse polarization of the target, and so the estimated decay parameter is much smaller because there is no contribution from the axial polarization.

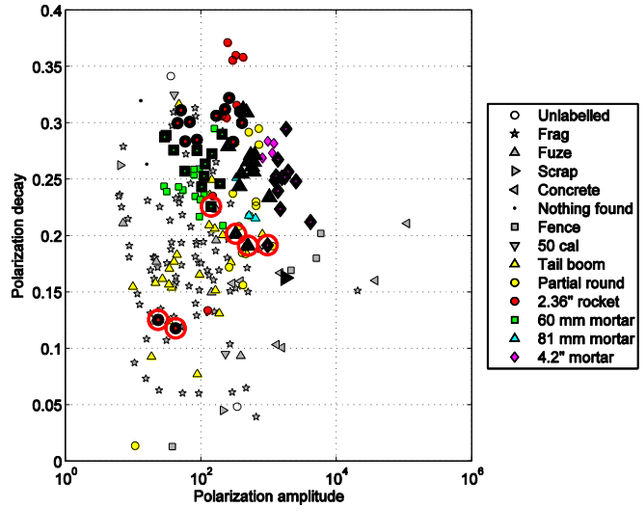


Figure 29. Model features from inversion of MTADS EM-61 data. Features in bold are test-pit items, and circled features are horizontal crosstrack test-pit items.

Some of the test-pit TOIs with polarizations in the range [0.15 0.2] are angled horizontal cross-track. In these cases the estimated polarization decay is also likely faster than a survey with multiple passes. These fast decays are not likely to be encountered in the test data, where perpendicular passes have been made over targets.

c. Creation of a dig-list for the MTADS EM-61

Based upon the preceding discussion, we select the same-decay based strategy for the MTADS as for the EM-61 cart. We note, however, that the MTADS EM-61 has generally slower decay rates than observed for the EM-61 cart. Again this is because the MTADS excites the target from multiple passes and so has a better chance of exciting the slow-decaying axial polarization. This discrepancy between the two surveys requires that we specify a different cut-off threshold for the MTADS data. Repeating the analysis used for the EM-61 data, we select a cut-off in polarization decay at 0.1813 (**Figure 30**). As explained in the preceding discussion, this analysis excludes outlying test-pit and training targets.

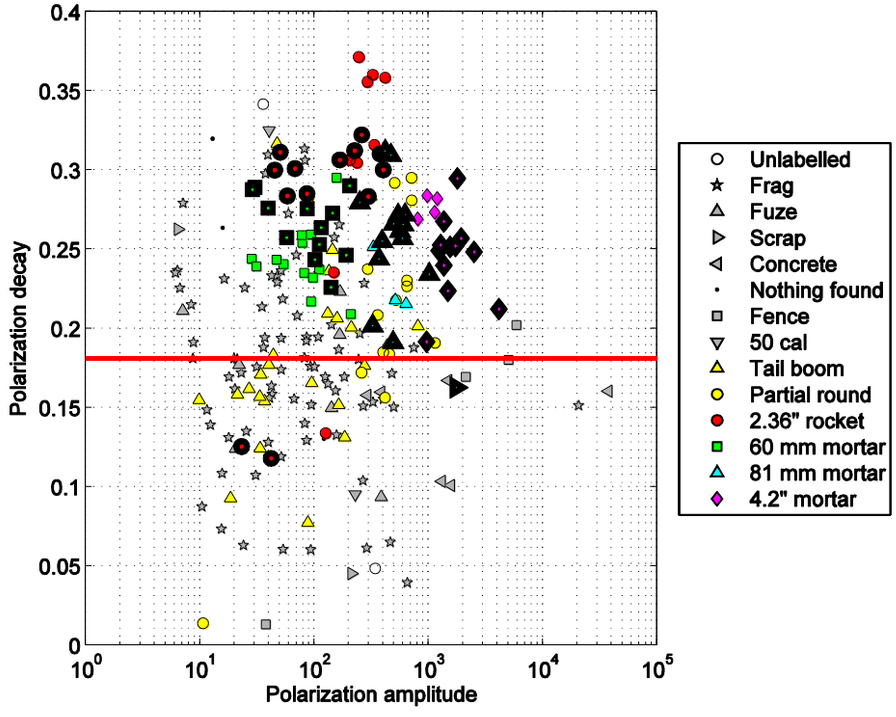


Figure 30. MTADS EM-61 model features from test-pit and training data showing estimated cut-off in polarization decay for test data.

5. MTADS EM-61 ARRAY DATA COOPERATIVE INVERSION

Figure 31 compares estimated and ground truth depths for MTADS EM-61 inversions with and without cooperative constraints. Cooperative inversion provides much improved depth estimation. There is a commensurate improvement in the clustering of TOIs by size (i.e. polarization amplitude), particularly for larger TOIs (Figure 32). However, for the smaller TOIs (60 mm), there is no noticeable improvement in the clustering of feature vectors by size.

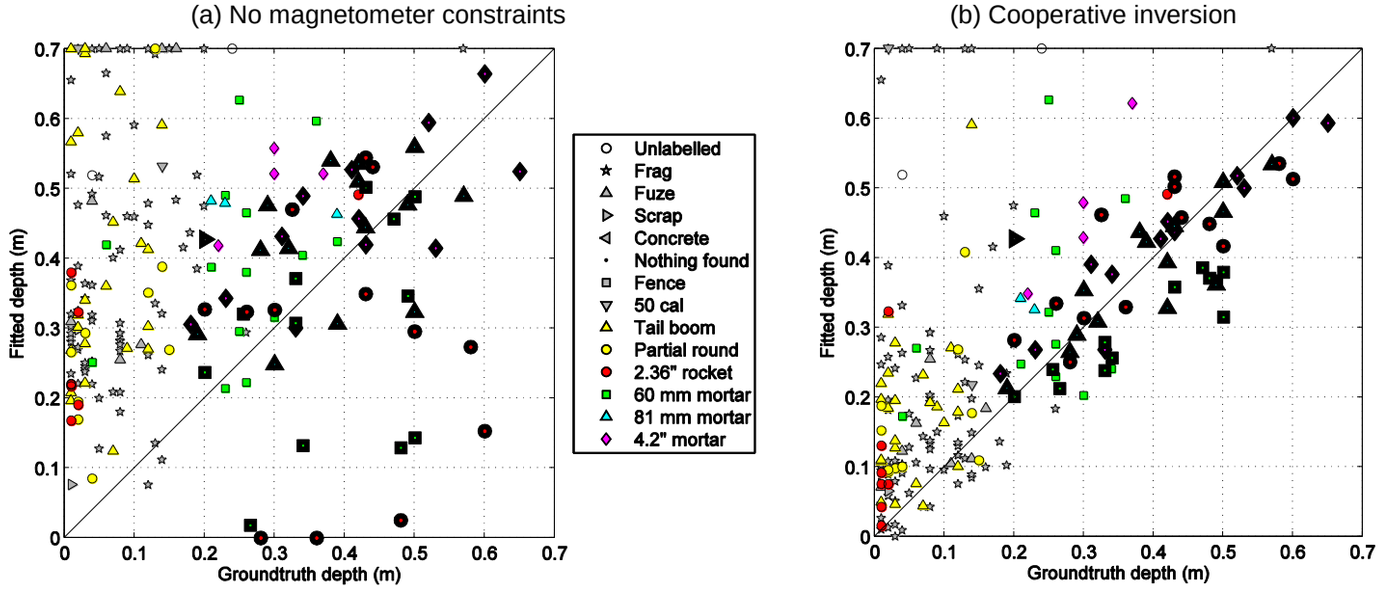


Figure 31. Fitted vs. groundtruth depths for training items, MTADS EM-61 array.

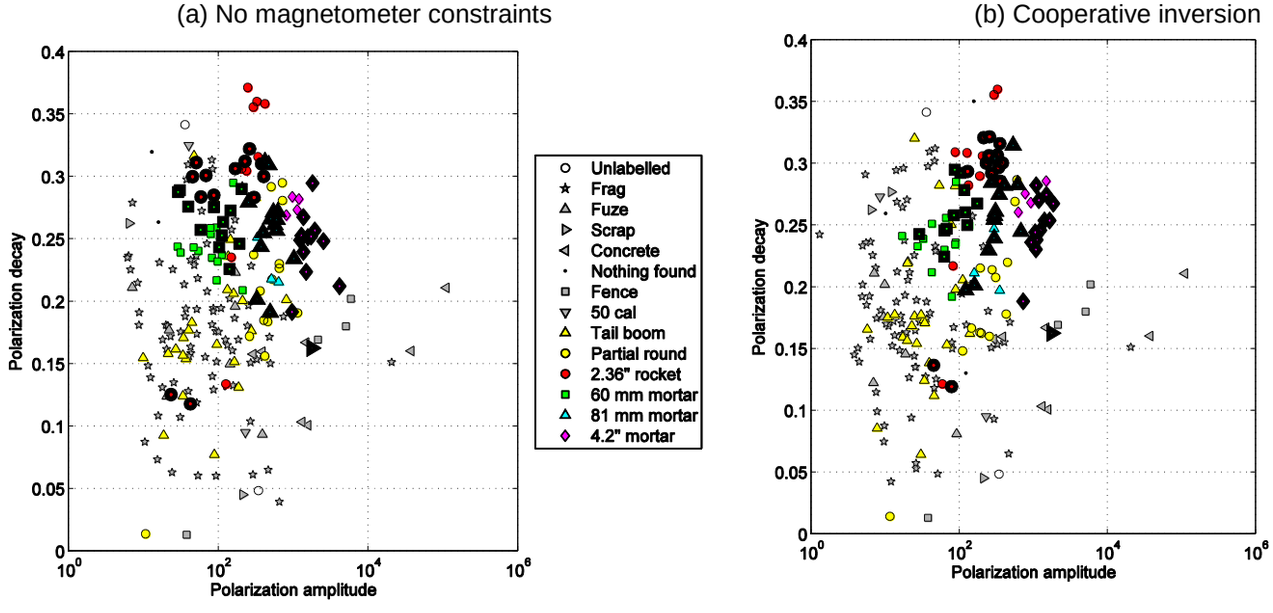


Figure 32. Estimated model features for MTADS EM-61 array.

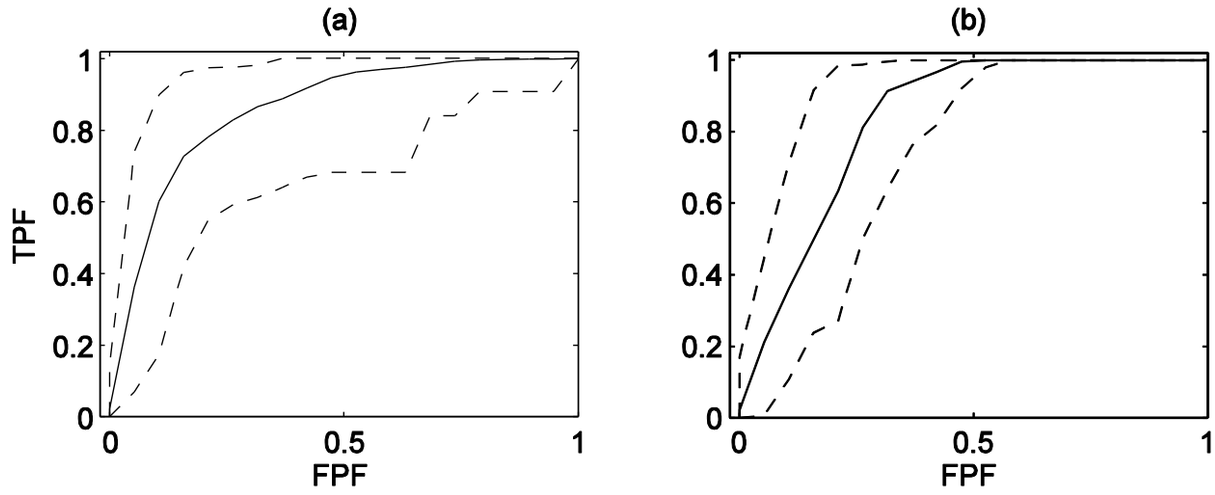


Figure 33. Bootstrap analysis of MTADS EM-61 data. Solid line is mean ROC, dashed lines are upper and lower performance bounds. (a) Neural network classifier trained on polarization amplitude and decay. (b) Threshold on polarization decay.

Figure 33 shows a bootstrap ROC analysis with features extracted by cooperative inversion from the MTADS EM-61 data. We compare a probabilistic neural network classifier (PNN) trained on polarization amplitude and decay with thresholding on polarization decay alone. The performance of the PNN ROC is initially improved relative to decay thresholding. This is because the cooperative inversion provides an improved grouping of larger TOIs. However, for smaller TOIs the polarization amplitude parameter is less useful as a discriminant (even with cooperative inversion constraints) and so the PNN ROC has difficulty detecting these targets. Based upon this analysis we choose to threshold on polarization decay alone for the MTADS EM-61 data and **will not produce a diglist based upon cooperative inversions**.

6. MSEMS EM-61 AND MAGNETOMETER COOPERATIVE INVERSION

a. Analysis of Test-pit results

Data were collected over the four primary targets of interest (TOIs) in horizontal, 45° nose down, and vertical nose down and nose up orientations with the Man-Portable Simultaneous EMI and Magnetometer System (MSEMS) which consists of an EM-61 and magnetometer mounted on a cart. The MSEMS data set were cooperatively inverted by using dipole location estimates from MSEMS magnetics data as *a priori* information. Upper and lower constraints on the location are defined to be twice the estimated variances of the estimated location parameters, i.e.:

$$X^{mag} - 2\sigma_X^{mag} < X < X^{mag} + 2\sigma_X^{mag}$$

$$Y^{mag} - 2\sigma_Y^{mag} < Y < Y^{mag} + 2\sigma_Y^{mag}$$

$$Z^{mag} - 2\sigma_Z^{mag} < Z < Z^{mag} + 2\sigma_Z^{mag}$$

where the estimated location from the inversion of magnetics data is (X^{mag}, Y^{mag}, Z^{mag}) and their estimated standard deviations are $\left(\sigma_X^{mag}, \sigma_Y^{mag}, \sigma_Z^{mag}\right)$. The noise and mask definitions are the same as for the non-cooperatively inverted data.

The cooperative instantaneous amplitude, three-polarization model fits were found to be acceptable for all 25 cooperatively inverted test-pit items. Sensor height of the EM-61was assumed to be 0.4 m and 0.55m for the magnetometer for all inversions. Figure 34 compares estimated and actual depths obtained for these targets. The recovered depths are improved through incorporation of the magnetometer constraints.

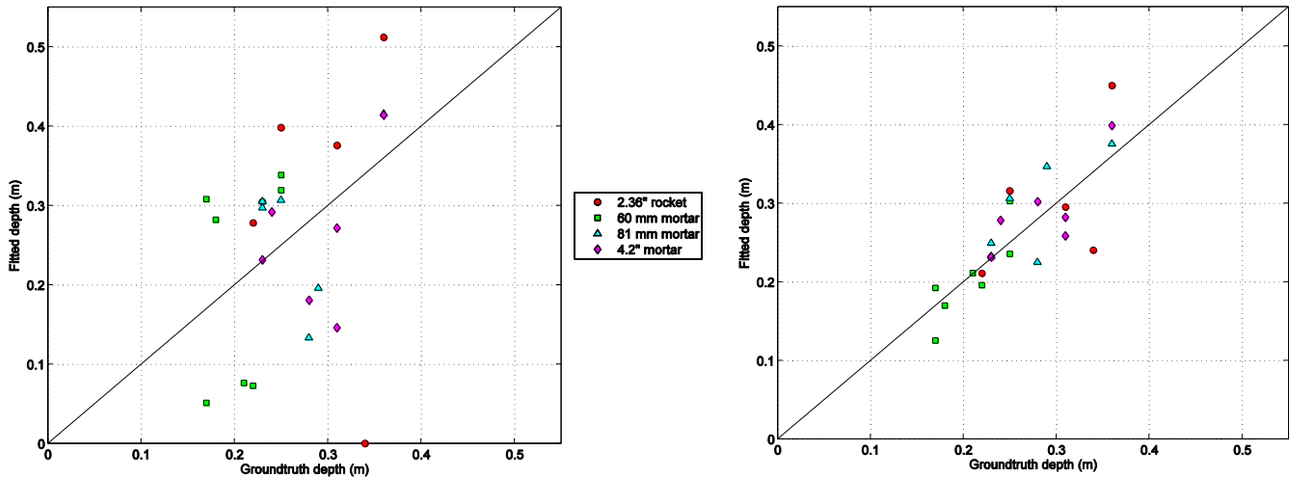


Figure 34. Fitted vs. groundtruth depths for testpit items. Instantaneous amplitude three-polarization inversion results without cooperative constraints are indicated in the left image while the right image illustrates improvements in recovered depths obtained through incorporating depth constraints from the magnetic data.

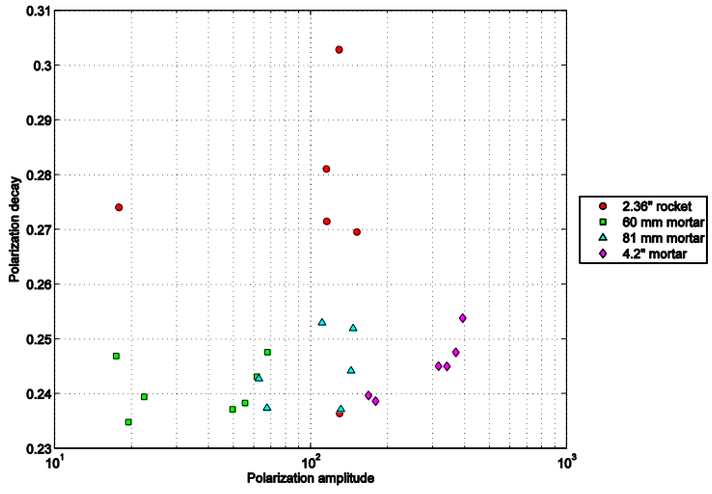


Figure 35 Model features extracted from MSEMS data. Instantaneous amplitude three-polarization inversion results without cooperative constraints are shown.

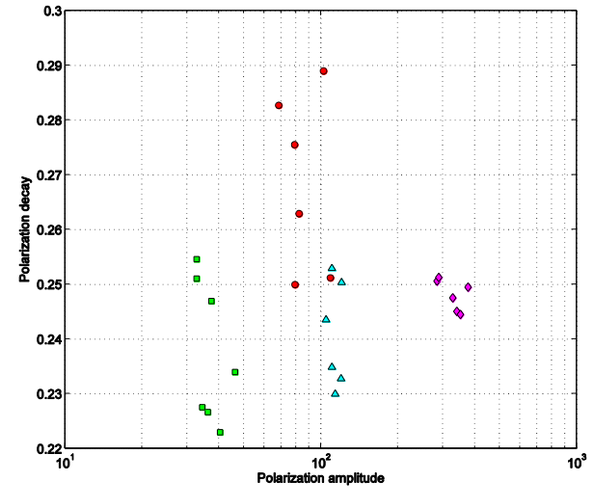


Figure 36 Model features extracted from MSEMS data. Instantaneous amplitude three-polarization inversion results using cooperative constraints illustrates the improvements in clustering of model features for a given target type.

The amplitude and time-decay based parameters of the test-pit targets are more tightly clustered for the cooperatively inverted data compared to the noncooperatively inverted data (Figure 35 and Figure 36). The improved estimates of parameters, particularly the polarization's tighter clusters in amplitude are due to the more accurate location and depth estimates returned by the cooperative inversion process.

b. Analysis of training data

EM-61 cart training data were cooperatively inverted with a three dipole instantaneous amplitude model using magnetic constraints. We use a three-dipole model because many targets in the training data are expected to be non-axisymmetric. Of the 195 training data targets, 107 were judged by the data analyst to have passable fits to the observed data. Figure 37 and Figure 38 compare inversions for passed and failed targets in this data set. Figure 37 is representative of many of the failed inversions caused by overlapping target responses.

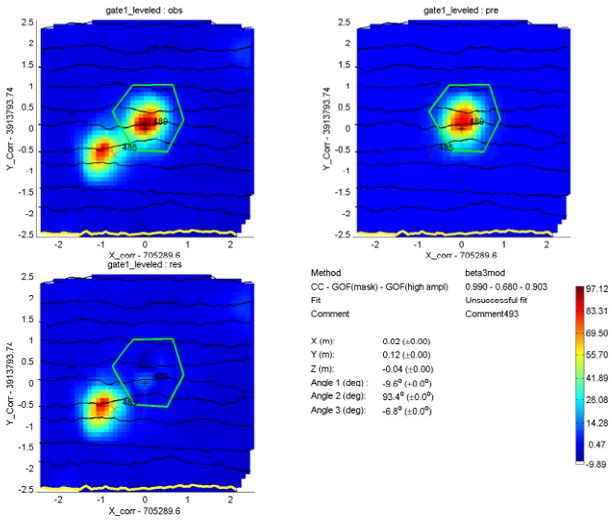


Figure 37 Unsuccessful fit, MSEMS data (target 489)

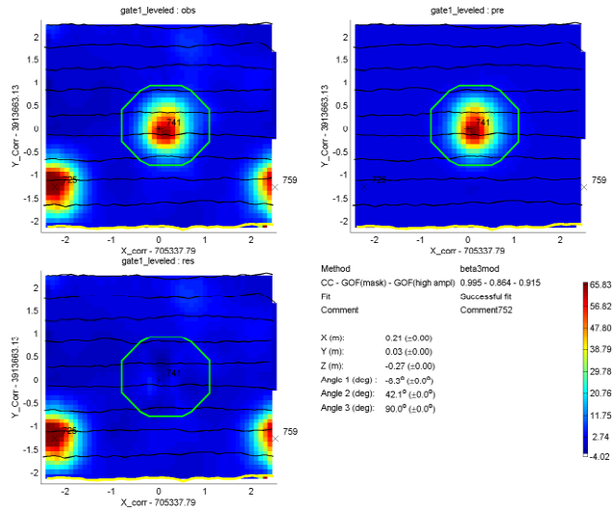


Figure 38 Successful fit, MSEMS data (target 741)

Closely spaced anomalies, as depicted in Figure 37, can be problematic for the cooperative inversions which rely on a depth constraint obtained from magnetometer data. Consider the image of the magnetometer data shown in Figure 38. While there are two closely spaced yet distinct targets evident in the EM61 cart data of Figure 2, The magnetic response of Figure 39 appears to rather represent a combined response of the targets 485 and 489. Performing a cooperative inversion using the mag constraint obtained for target 485 results in a recovered depth that is too deep, placing the target at a depth of 20cm. Ground truth information indicates that this target is a partial round found at a depth of 4cm. In this case, using the mag constraint pushes the target deeper because the mag response is actually a combined response from both targets 485 and 489. Inverting target 485 without using the magnetic depth constraint results in a recovered depth of 11cm, closer to the true depth. Overlapping anomalies are a common occurrence in the SLO training data and require careful attention to masking details during the QC process.

As observed with the test-pit targets, estimates of target depth are improved with the incorporation of cooperative constraints from the magnetic data (Figure 40). All estimated depths are shown here, regardless of the pass/fail status of the inversion.

Figure 41 shows inversion-based features extracted from observed EM-61 training data with a three-dipole instantaneous amplitude model. All estimated features are shown here, regardless of the pass/fail status of the inversion. The calculated features are

- Polarization amplitude = $(\Sigma L_i(t_1)^2)^{1/2}$
- Polarization decay = $(\Sigma L_i(t_4)^2)^{1/2} / (\Sigma L_i(t_1)^2)^{1/2}$

The polarization amplitude is proportional to the amplitude of the induced moment in the presence of a unit primary field along each of the target's principle axes. The polarization decay is the ratio of the polarization amplitude at the fourth and first time channels. These features are analogous to those previously used for discrimination in the Camp Sibert demonstration, where we used the amplitude and decay of the primary polarization. Using the total polarization amplitude here simplifies the computation somewhat by eliminating the requirement to identify the primary polarization. The polarization amplitude is strongly correlated with target depth and so is not a particularly useful feature for discrimination with these data.

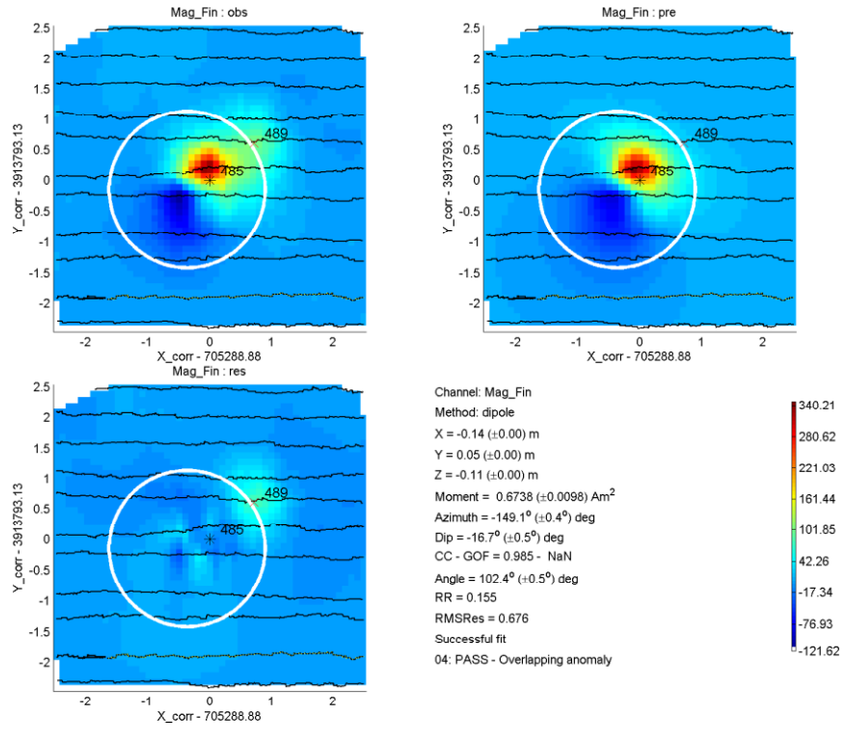


Figure 39 Closely spaced targets produce a combined magnetic response, MSEMS magnetics data (target 485)

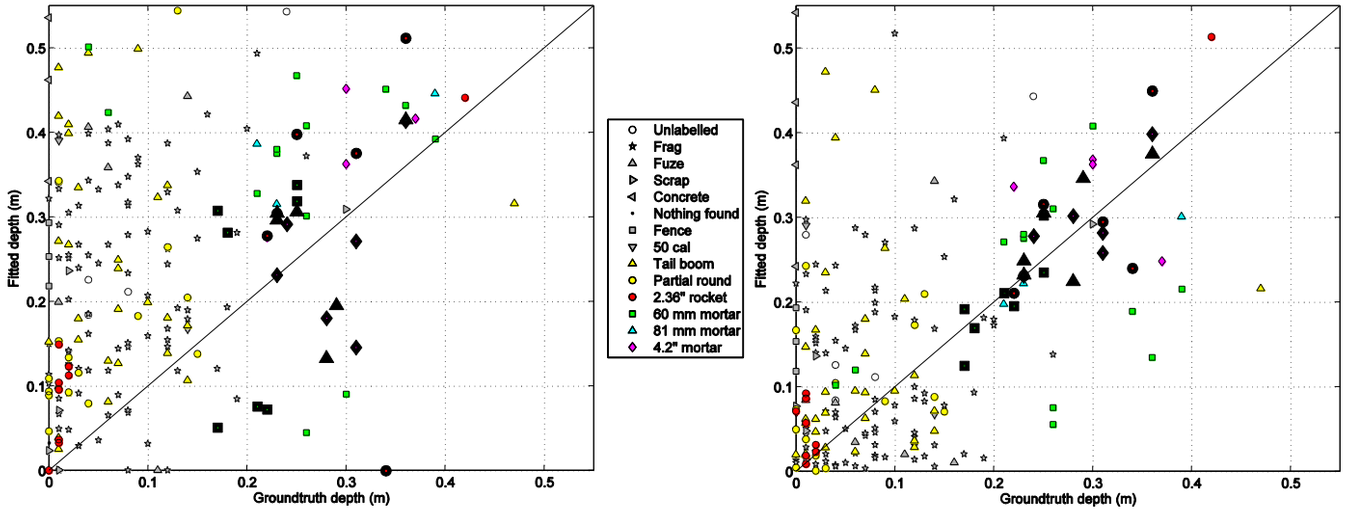


Figure 40. Fitted vs. ground truth depths for training and test-pit (bold symbols) targets. Instantaneous amplitude three-polarization inversion results without cooperative constraints are indicated in the top image while the bottom image illustrates improvements in recovered depths obtained through incorporating depth constraints from the magnetic data.

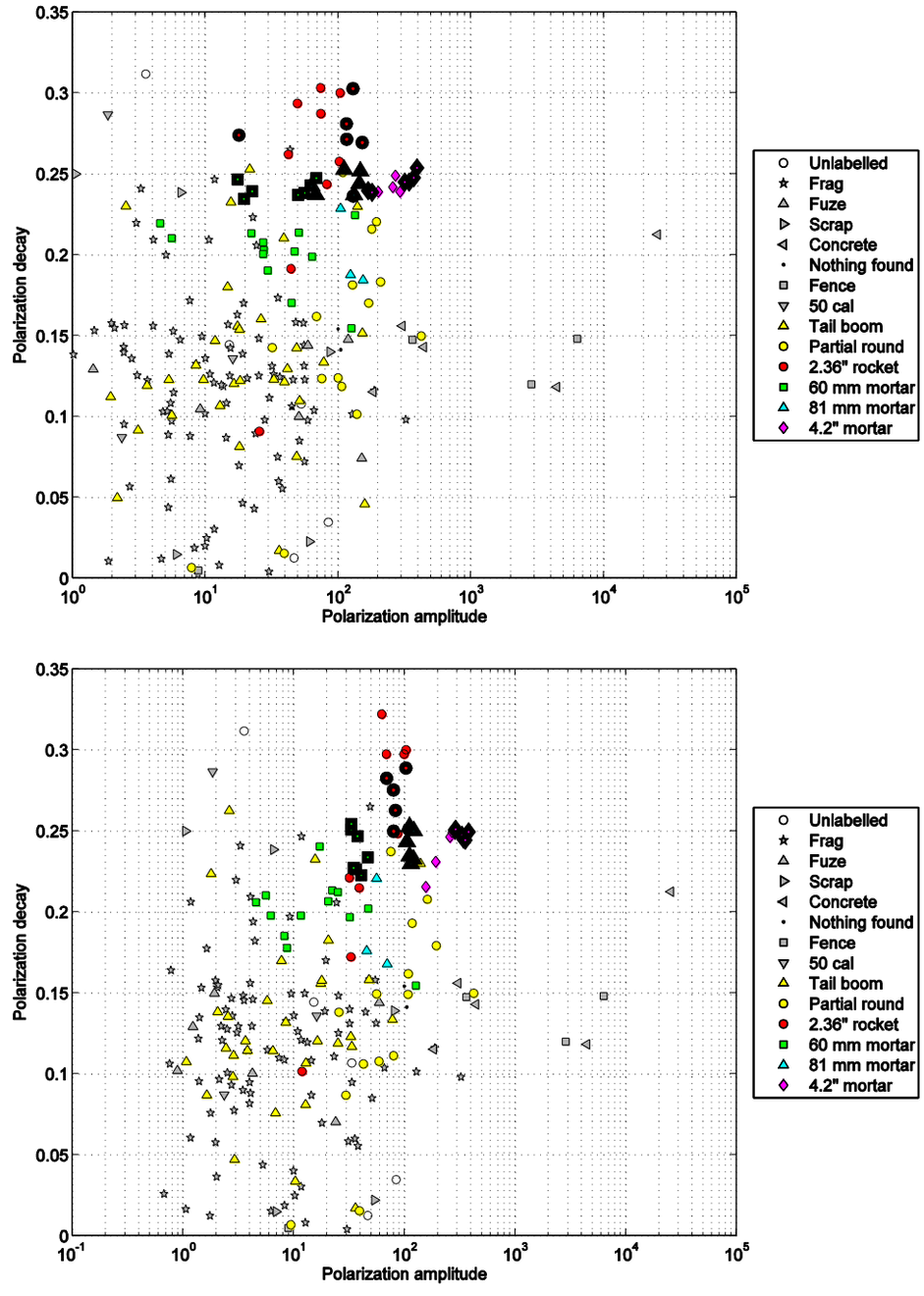


Figure 41 Data features extracted from MSEMS data. Features in bold are test-pit items. The top plot is without cooperative constraint applied while the bottom plot includes magnetic constraints.

In Figure 41 we find that the polarization decay is a useful parameter for discrimination between targets of interest and clutter. In both Figure 40 and Figure 41 there is an outlying TOI corresponding to the cracked 2.36" warhead.

c. Creation of a dig-list for MSEMS

In designing a discrimination strategy for the MSEMS, we follow the same bootstrap analysis as was employed for the other detection mode surveys. Figure 42 compares the bootstrap performance of discrimination with a neural network versus thresholding on polarization decay. The neural network provides an initial improvement in discrimination capability by finding large, slow-decaying TOIs, thereby increasing the area under the ROC (AUC) metric. However, it provides no significant advantage over the decay rate threshold in terms of final false alarm rate.

Again, we conclude that thresholding on polarization decay is an effective (and simple) strategy for discrimination of TOIs with MSEMS data. We select a cut-off value of the polarization threshold of 0.1394 based upon the statistics of the training data (Figure 43).

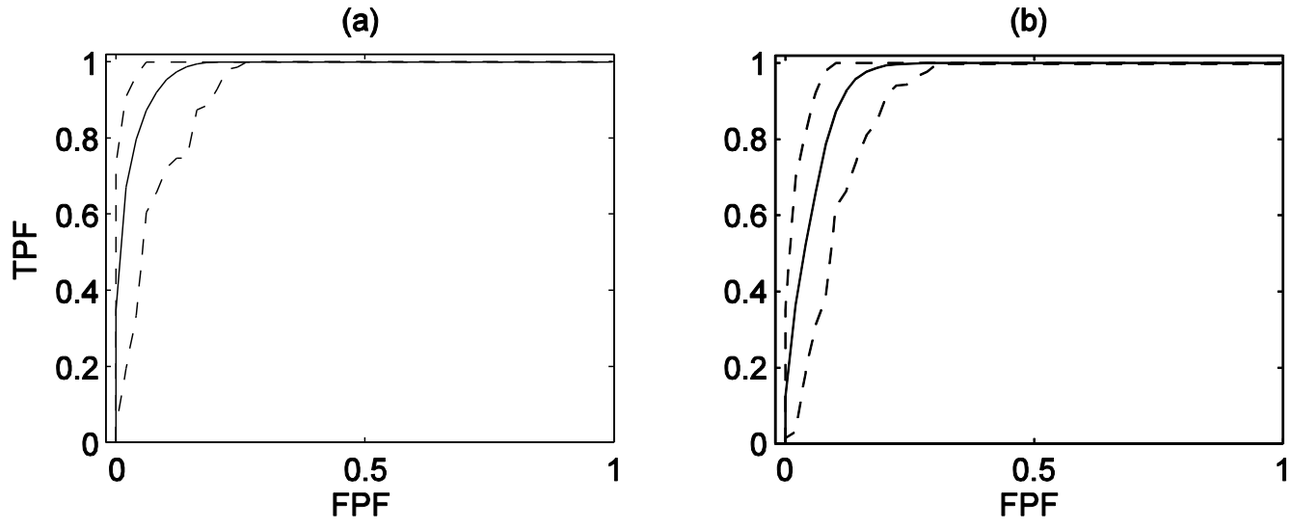


Figure 42. Bootstrap analysis of MSEMS data. Solid line is mean ROC, dashed lines are upper and lower performance bounds. (a) Neural network classifier trained on polarization amplitude and decay. (b) Threshold on polarization decay.

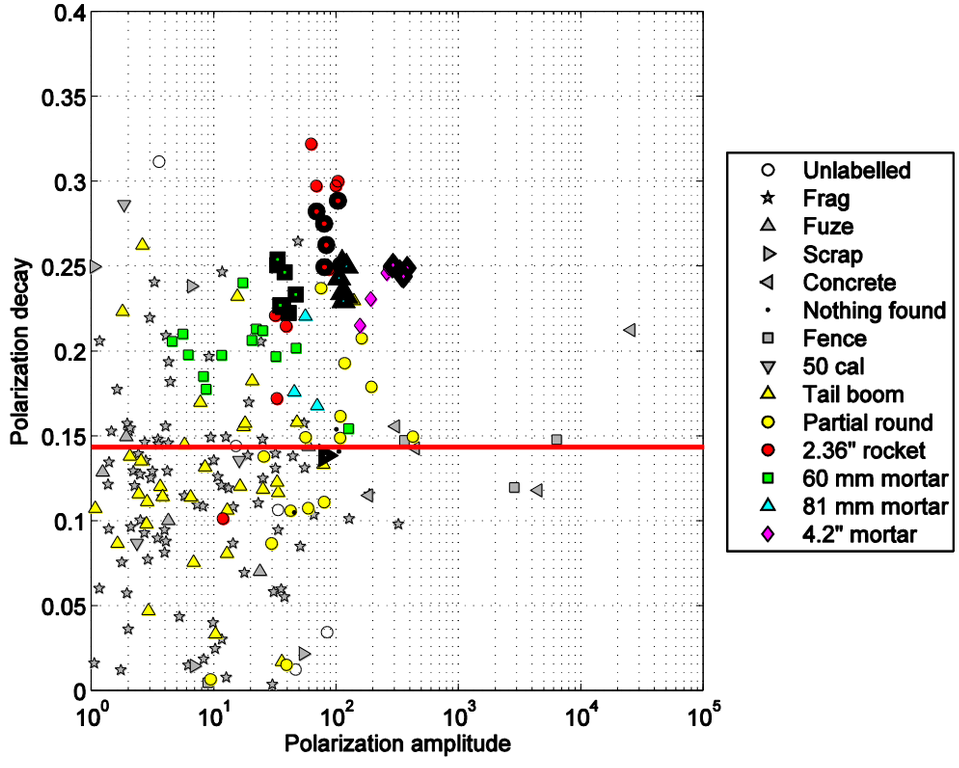


Figure 43. Operating point cut-off for MSEMS discrimination.

7. TEMTADS CUED INTERROGATION: STATISTICAL

TEMTADS has an array consisting of 5 x 5 transmitters and receivers and has 115 logarithmically spaced gates between 0.042 ms and 24.35 ms. It is deployed in a cued-interrogation mode and for each transmit excitation, TEMTADS records the response at all of the receivers. Thus it has spatial-temporal data of size 625 x 155 for points.

The test-pit and training data we are using were all pre-processed by the data collection demonstrator: including time gate correction, normalization by transmitter current, and background subtraction. It was reported that sensor 21 in the array did not work properly in an intermittent manner (personal communication with James B. Kingkon). We checked the data and do see bad decay behavior in Rx-21 for some transmitters but the measured data associated with Tx-21 look physically reasonable. In our processing, we remove any data from Rx-21.

In the inversion processing of TEMTADS data, the following parameters were used:

- **Estimate of Data error**:
  - Background noise was estimated from the test-pit background measurements. We estimated a standard deviation of 1 mV to 0.02 mV across time channels for all receivers.
  - We assumed 5 percent noise on each data point.
- **Masking**: Data above a SNR of 0.02 were used in the inversions. Optionally, spatial masking was applied to anomalies in terms of mono-static images. For suspected overlapping cases, no spatial masking was used, and instead a multi-object inversion was undertaken.
- **Model:** The data were inverted using a 3-dipole model. For n = 115 time gates, we have 351 unknowns to be determined:

$$m = [X, Y, Z, \varphi, \theta, \gamma, L_1(t_1), L_2(t_1), L_3(t_1), \ldots, L_1(t_n), L_2(t_n), L_3(t_n)]$$

  where (X,Y,Z) is the location, (φ,θ,γ) are the orientation angles and $L_i(t_j)$ is the i-th polarization at the j-th time channel.

a. Analysis of Test-pit results

The test-pit training data were collected with respect to four ordnance items: 60 mm mortar, 81 mm mortar, 2.36 in rocket and 4.2 in mortar, and one symmetric item (a shotput). TEMTADS data were collected over an open pit at a sensor height of 16.5 cm from the ground. For each ordnance item, there were 6 or 7 measurements corresponding to different positions and orientations such as horizontal, vertical, inclined, and nose up and down.

Figure 44 shows the comparison of the predicted and actual ground-truth depths. For the two vertical cases (nose up and down) of 2.36 in rocket, the depths are underestimated probably because the top part of the object dominates the EMI responses. Overall there is excellent agreement between actual and predicted depths.

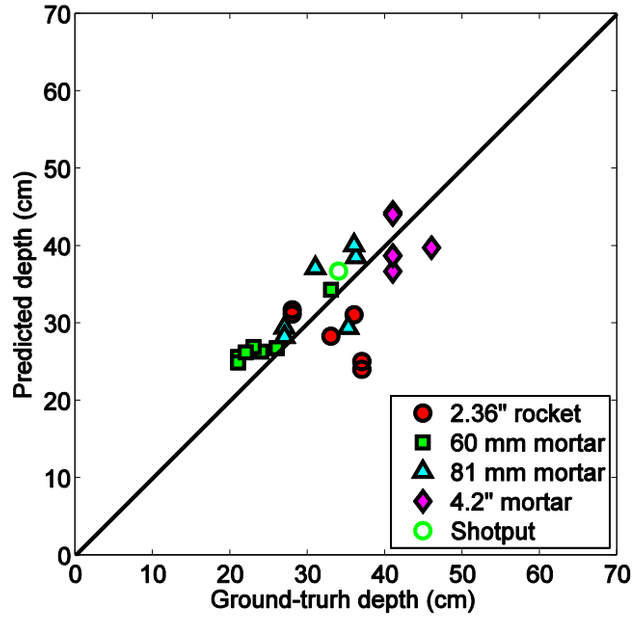


Figure 44. Test-pit TEMTADS data. Recovered depths versus the ground-truth ones.

Figure 45 show the recovered polarizations over each of the test-pit items. In each of the plots, parameterized polarizations are represented in green, recovered polarizations for the same item that were configured (depth, orientation) differently are in other colors. Overall, the polarizations for each item are almost invariant with respect to object orientation and thus provide a unique target signature. For the 2.36 inch rocket, the secondary polarizations are not always equal (Figure 45b): specifically when in the noise down position, where the sensor data is most sensitive to the twisted tail part of the item. For all items, the decay characteristics of the primary polarizations are well behaved and smooth to at least 10 ms. Secondary polarizations for 2.36 in rocket and 60 mm mortar behave smoothly to 1 - 2 ms and for 81 mm mortar and 4.2 in mortar to 5 - 8 ms.

b. Analysis of training data

The TEMTADS training data set contains 178 anomalies. According to the training list, there are fourteen 60 mm mortars, three 81 mm mortars, four 2.36 in rockets and four 4.2 in mortars. Two non-UXO items (1301 and 1373) were treated as target of interest (TOI). Figure 4 is the recovered depths against the measured depths from the site. For 4.2 in mortars, the predicted depths are around 5-15 cm deeper than the ground-truth ones, but the polarizations from the associated training data were well recovered and almost identical those from test-pit data (not shown here). Errors in depth estimates could be due to the topography at SLO, as TEMTADS depth estimates are derived assuming the sensors are 16.5 cm above flat ground.

For the 2.36” rocket the depths of the four shallow cases (1253, 1289, 1301 and 1373) and the deepest one at 42 cm (1019) are well predicted. In the other two shallow cases (565 and 33), the depths are in error by 7.3 and 9.6 cm. Generally, the recovered polarizations of intact 2.36 in rockets match well those derived from the test-pit data. For the TOI like 2.36 rocket motor (1301 or 1373), their polarizations resemble those of the intact ones, as shown in Figure 47.

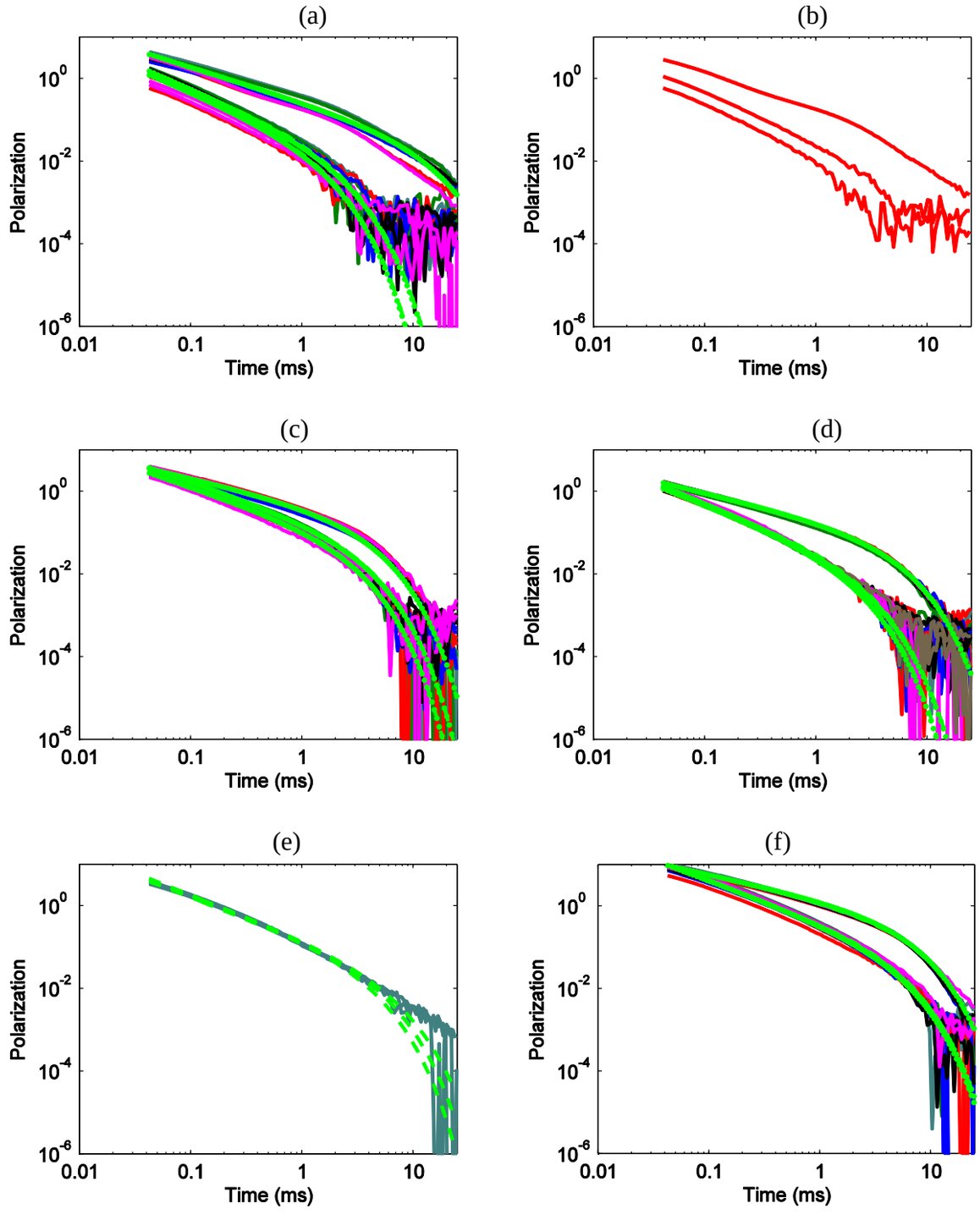


Figure 45. Recovered polarizations from the test-pit TEMTADAS data. (a) 2.36 in rocket. (b) 2.36 in rocket, inclined, nose down. (c) 81 mm mortar. (d) 60 mm mortar. (e) shotput. (f) 4.2 in mortar. In (a) and (c)-(f) except green curves that are parameterized ones, all other curves in each plot represent recovered polarizations when the same item was positioned at various depths and orientations (horizontal, vertical, inclined).

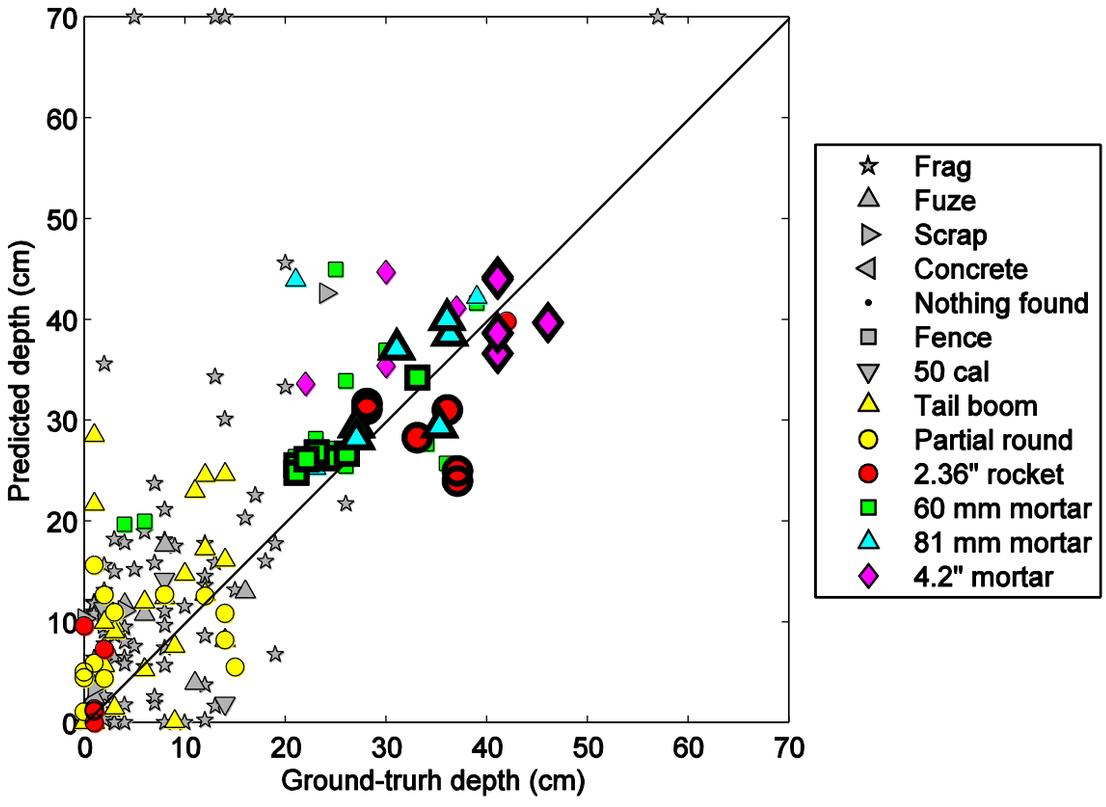


Figure 46. Predicted depths v.s. ground-truth ones for TEMTADS test-pit and training data.

For 81 mm mortar, the depth corresponding to 1339 is 20 cm deeper than the groundtruth. For the other two cases (1081 and 1342), the predicted depths of 25 and 42 cm are close to the ground-truth of 23 and 39 cm respectively. The polarizations extracted from 1342 agree with those from the test-pit data. The polarizations from 1081 and 1339 are almost identical but show different decay characteristics than those from the test-pit data and have larger amplitudes, as shown in Figure 48. These two mortars are a different type than the one measured in the test-pit.

For 60 mm mortars, there are 3 cases (410, 522 and 831) where the predicted depths (45, 20 and 20 cm) are deeper than the ground-truth (25 cm, 6 cm, and 4 cm). For the latter two cases, a similar discrepancy was noticed in the MetalMapper data. The depths for the other 8 cases are accurately predicted. For this small UXO, the recovered polarizations show some variations as compared to the polarizations obtained from the test-pit data. Figure 49 presents the polarizations extracted from anomaly 522. Its primary polarization decay is fast when approaching 1 ms and later. Similar decay characteristics are observed in a number of other anomalies (e.g., 36, 410 and 1309). Inspection of the ground-truth indicates that these 60 mm mortars are missing their tails, so that the faster decay rate is expected. For the intact 60 mm mortars (111, 511 and 831), the recovered polarizations are in a very good agreement with those of the test-pit data. For several of the 60 mm projectiles, signals are too weak to allow accurate recovery of secondary polarizations. For example anomaly 1023 (Figure 50) has a maximum response in the first time-channel of 1.57 mV. As discussed in the next section, besides the SNR issue, there are some suspected overlapping anomalies that also can make single-object inversion inaccurate.

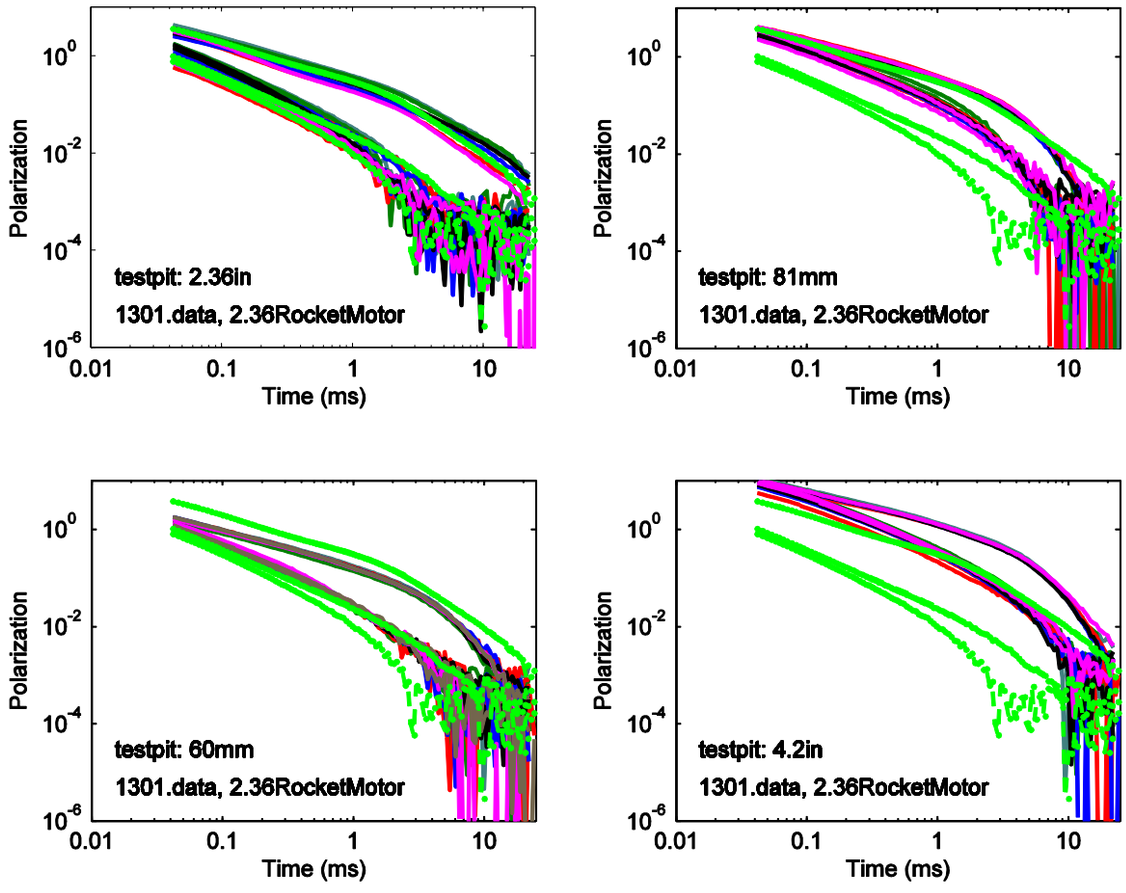


Figure 47. Anomaly 1301. 2.36 Rocket Motor. The recovered polarizations (in green) from 1301 are plotted against the polarizations of four configurations obtained from the test-pit data.

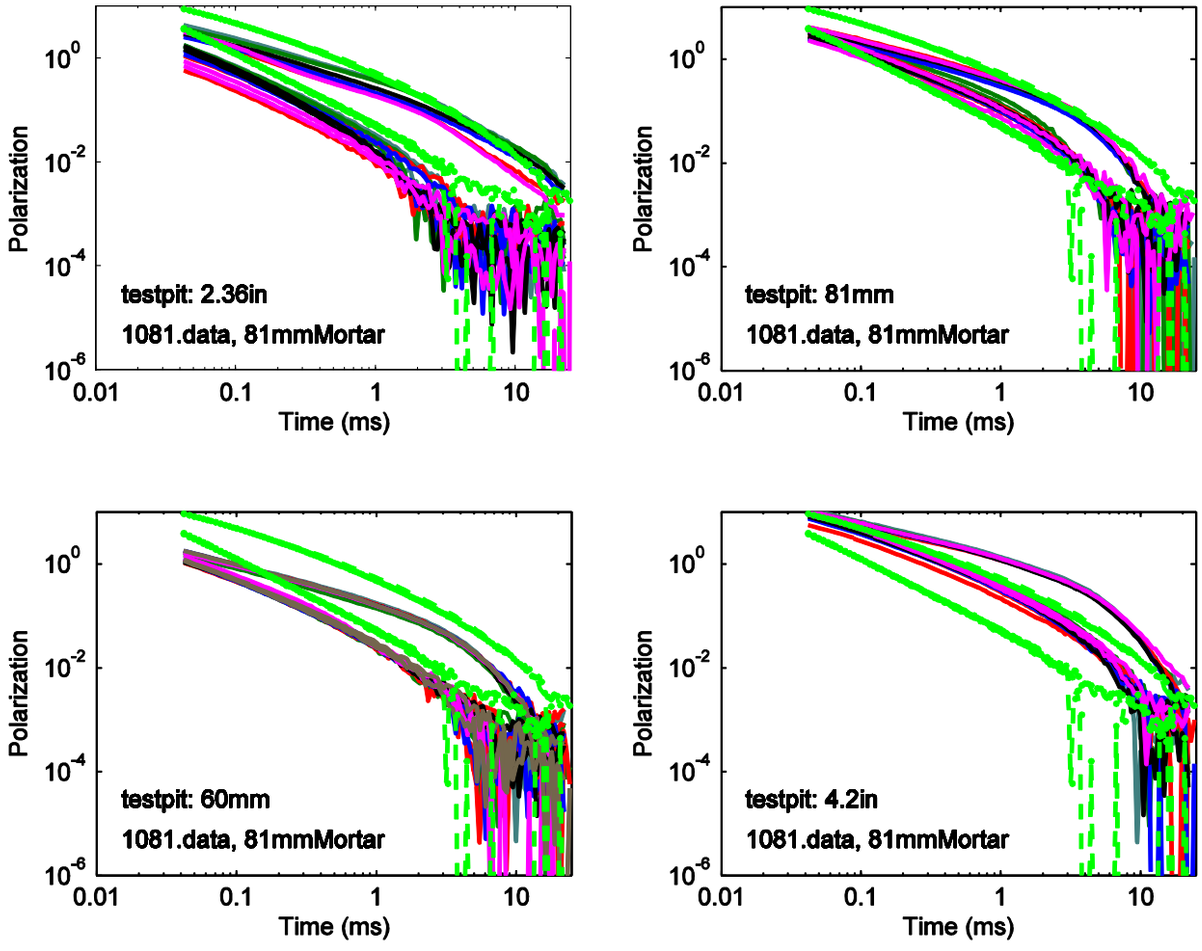


Figure 48. Anomaly 1081, 81 mm mortar. The recovered polarizations (in green) are plotted against the polarizations obtained over four configurations in the test-pit data.

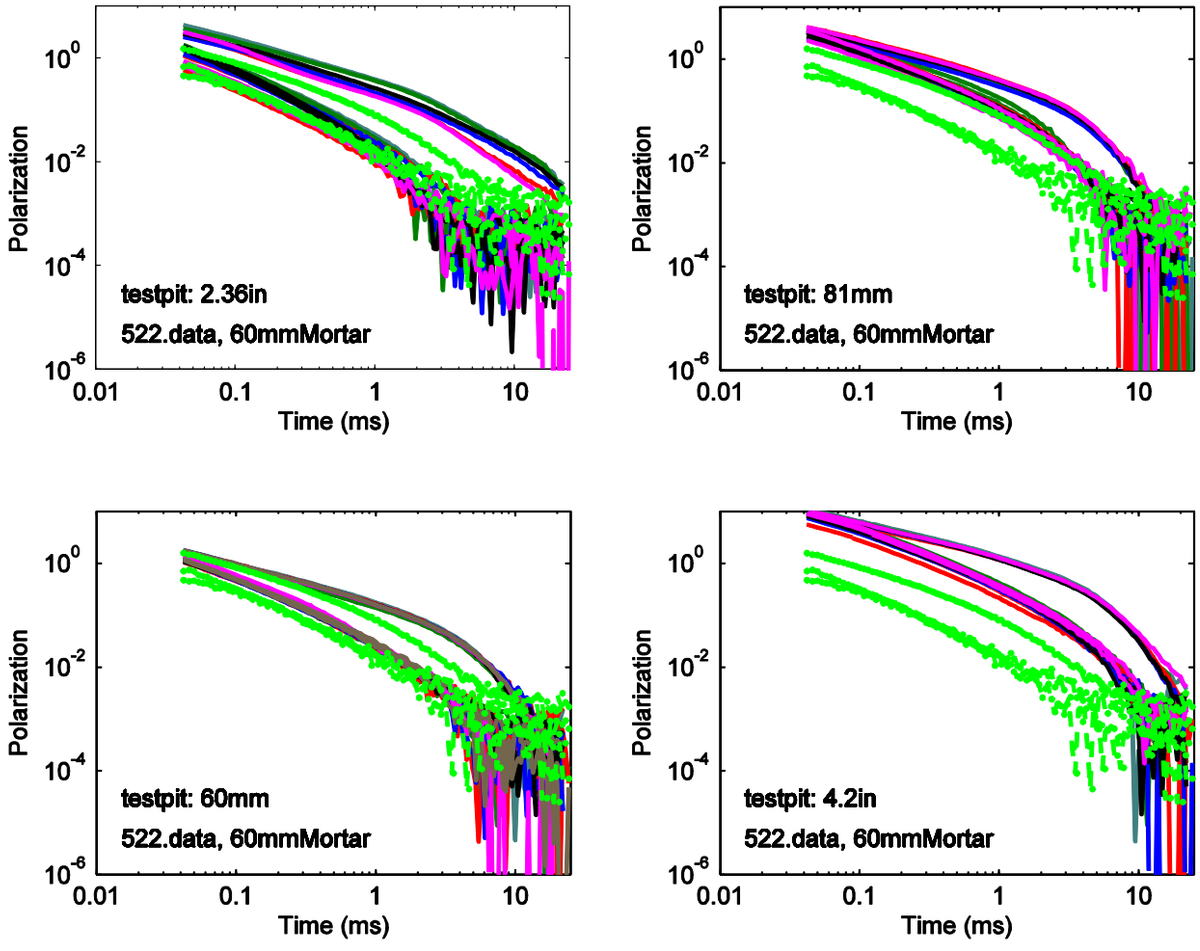


Figure 49. Anomaly 522, 60 mm mortar. The recovered polarizations (in green) from 522 are plotted against the polarizations of four items measured over the test-pit.

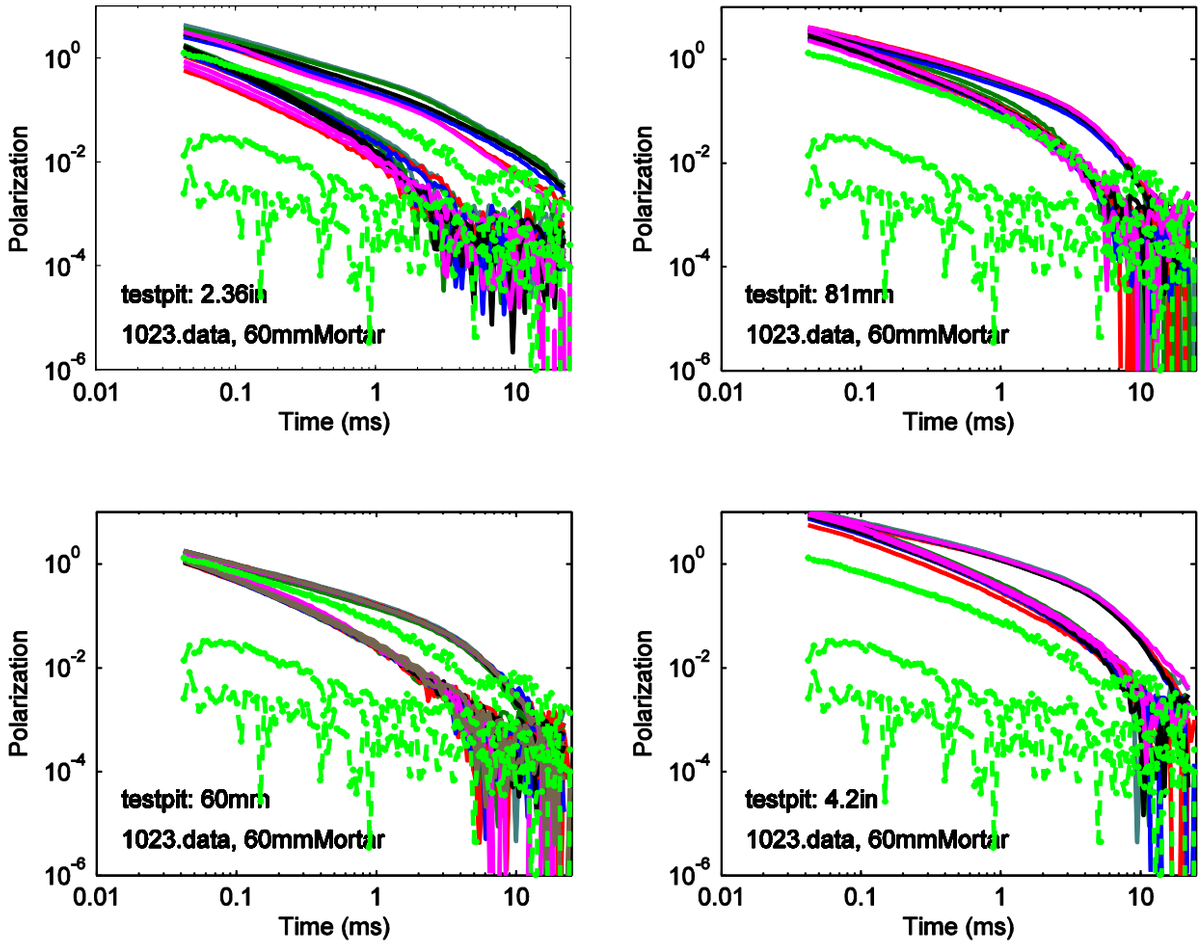


Figure 50. Anomaly 1023, 60 mm mortar. The recovered polarizations (in green) from 1023 are plotted against four polarizations obtained from the test-pit.

c. Approach for dealing with overlapping objects

Through visual review of single object inversion results, we consider the following 43 anomalies as suspected multi-object cases:

1107 1111 1146 1260 1292 1312 1340 1378 22 31 386 389 410 424 461 46 471 489 510 535 557 580 587 593 603 608 610 624 627 634 635 653 654 667 678 693 695 770 792 804 866 878 904.

Some of these are listed as multiple objects in identification column of the training list.

We use information theoretic criteria (ITC) to automatically estimate the number of objects. The ITC are composed of a data-based log likelihood function for a given model and a penalty function that counterbalances model complexity (for details see the demonstration plan for this project). The ITC can be implemented either in detection-only or joint detection and estimation modes. In detection-only mode, the number of objects is determined separately and estimation of model parameters (e.g., locations and dipolar polarizations of objects) is then followed. In the joint detection and estimation, both groups of unknowns are

determined simultaneously by executing inversions for possible models. In both modes, the model with minimum ITC value is selected as the model we use for our analysis. Under Gaussian statistics, the detection-only ITC can be implemented rapidly with a closed-form expression for a sequence of assumed sources.

Figure 51 shows the histogram by applying the detection-only ITC to those subsets of TEMTADS training data: three cases (535, 627, 635) are detected as single object cases, three cases (46, 471, 1378) are predicted to have more than two objects (this is likely overestimated), while for most two objects are predicted.

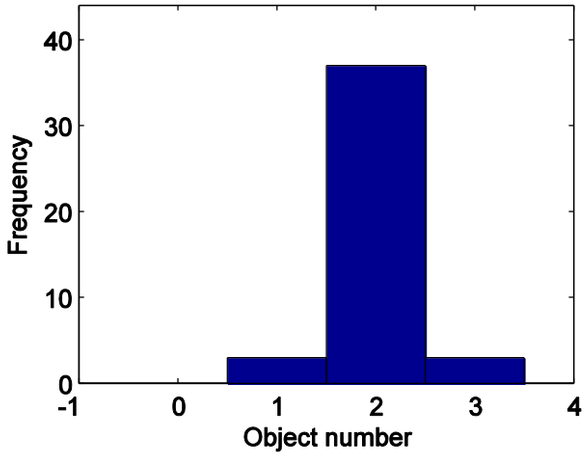


Figure 51. Statistics for MDL detection of object number among 43 suspected anomalies.

Figure 52. The picture of the item for anomaly 489.

According to the ground-truth information anomaly 489 (picture shown in Figure 52) was from a 60 mm mortar that had to be blown-in-place. We were unable to obtain satisfactory fits to the data (Figure 53 Figure 54) using a single-object model and the corresponding recovered polarizations (Figure 55) are larger and display the characteristics of a non-UXO item.

Figure 56 plots Minimum Description Length (MDL) against the number of equivalent dipole sources for anomaly 489. The MDL value of 11.7 at three-polarization, is reduced to a minimum of 5.5 at five-polarizations. This indicates that the anomaly is caused by two objects, one asymmetric with three polarizations and one axially symmetric with two polarizations. When executed in joint detection and estimation mode the MDL value for a single object is 8004 compared to 1809 for a two-object model.

With the two-object inversion, the predicted data agree very well with the observed data (Figure 57 and Figure 58). At 1 ms, the anomaly in the center stands out and the strong anomaly in the corner is gone. Figure 59 shows the recovered polarizations after two-object inversion, in which green and black dots represent the two different models and the other color curves show the polarizations of 60 mm mortar inverted from the test-pit data. One polarization closely matches the polarizations extracted over the 60 mm mortars in the test-pit. The second object, with two major polarizations that are almost equal, is mostly likely a piece of clutter whose polarization response is strong at early times but which quickly diminishes it amplitude at times approaching 1 ms.

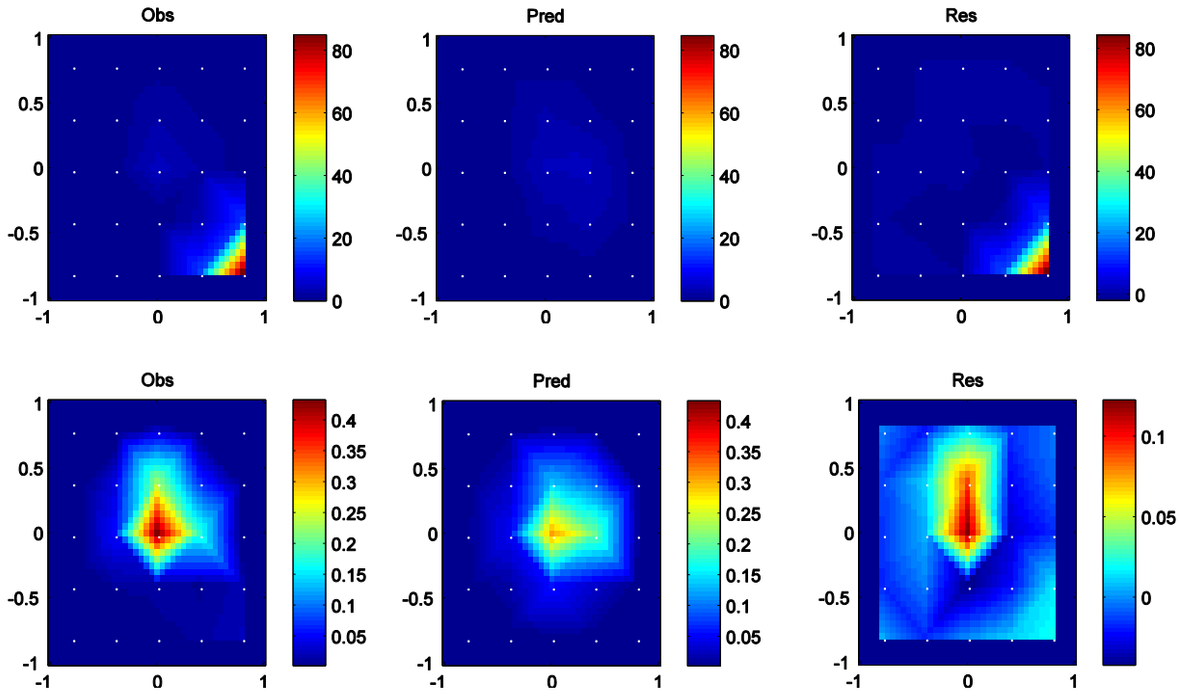


Figure 53. Observed data for anomaly 489 and the predicted data and residuals after single-object inversion. The first row is at t1=0.042 ms and the second row is at t53=0.99 ms.

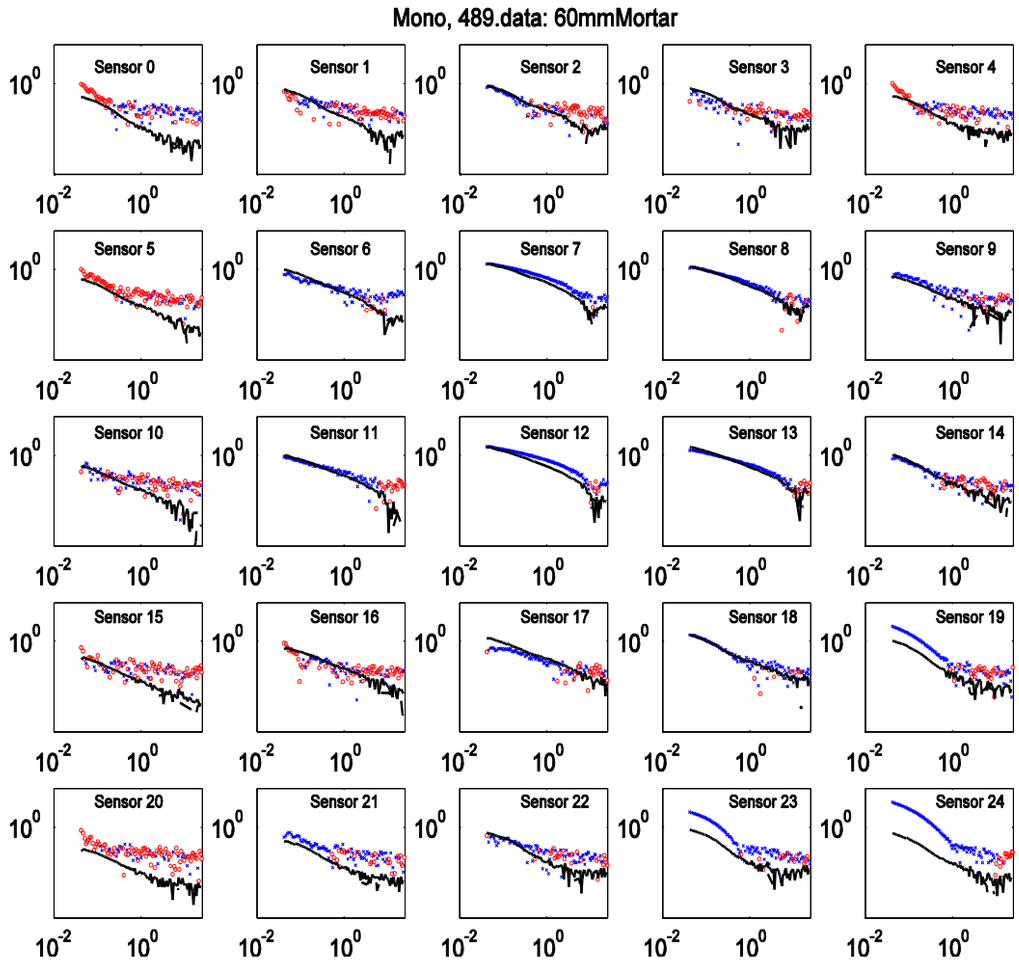


Figure 54. For anomaly 489, the observed (red negative and blue positive) and predicted data (green negative and black positive) from a single-object inversion when each sensor itself transmits and measures signals.

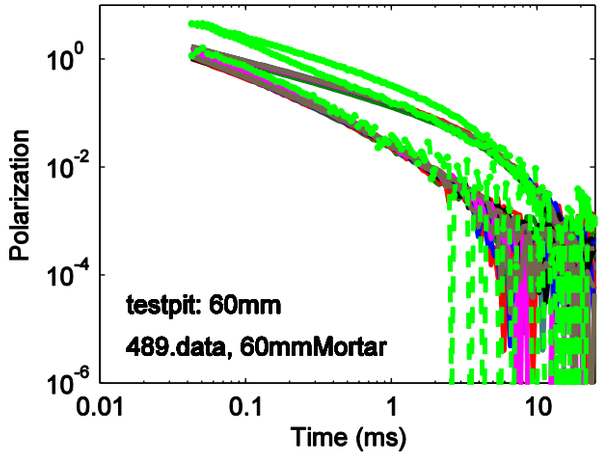


Figure 55. Recovered polarizations (green thick dots) after single-object inversion of anomaly 489. Other color curves represent recovered polarizations of 60 mm mortar from the test-pit data.

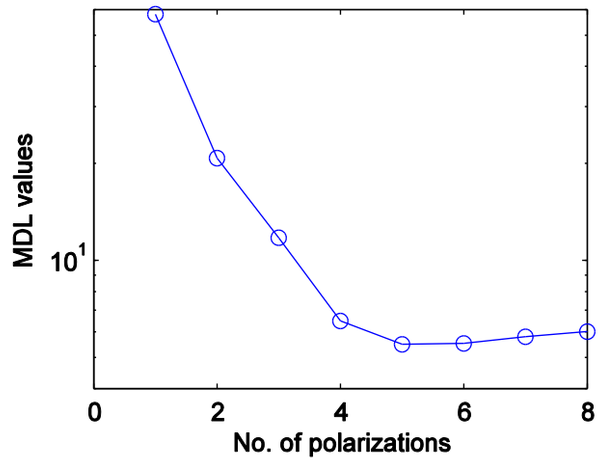
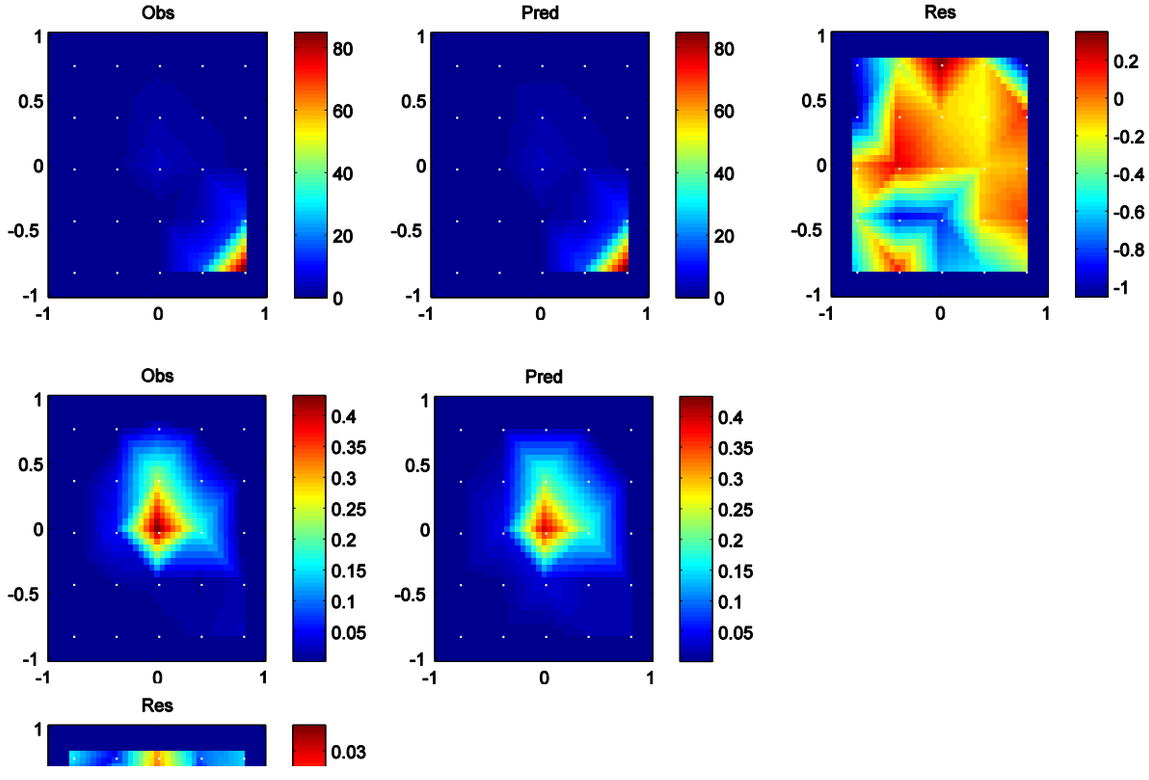


Figure 56. MDL value versus the number of polarizations for anomaly 489.



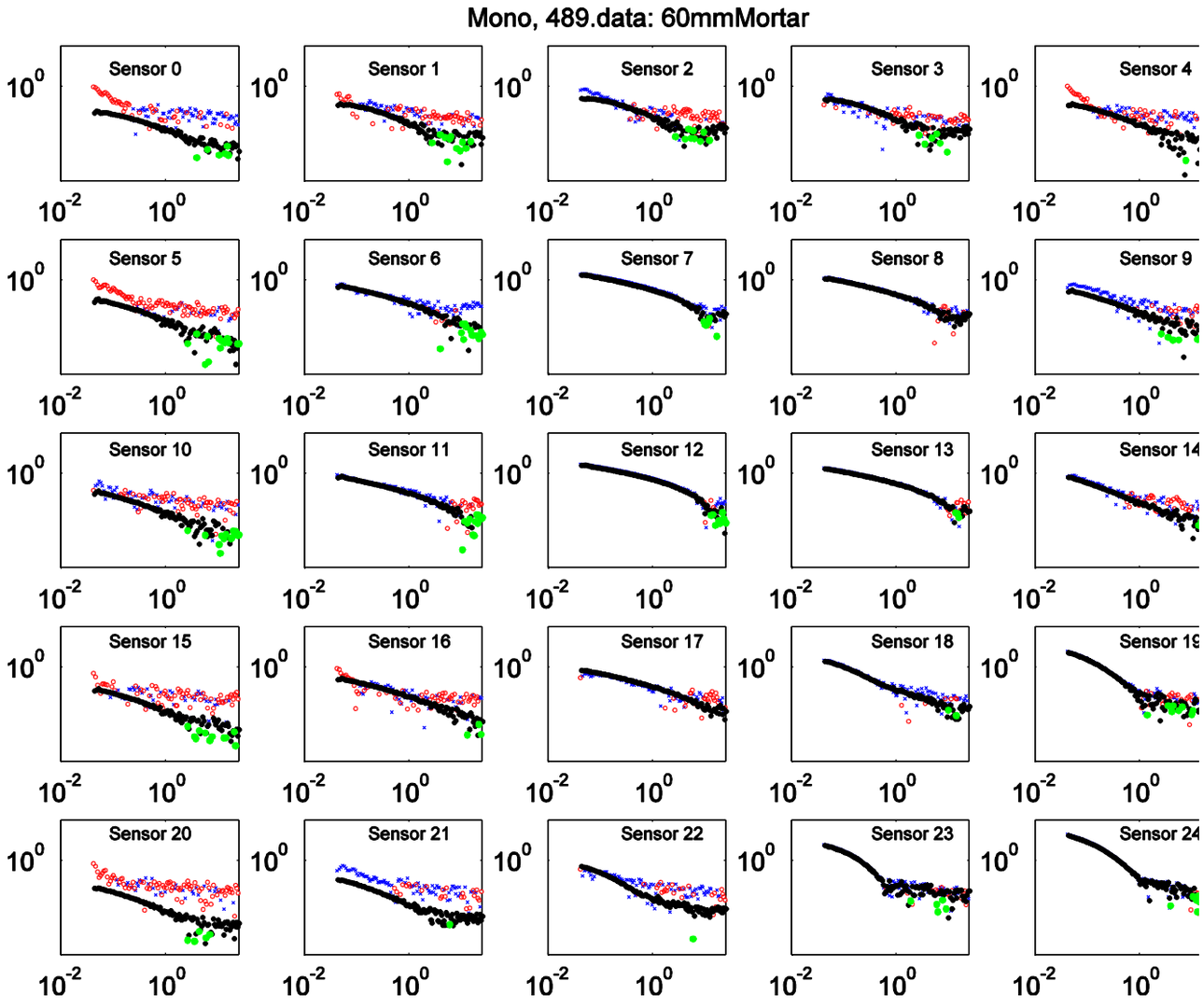


Figure 58. For anomaly 489 observed (red negative and blue positive) and predicted data (green negative and black positive) from two-object inversion, when each sensor itself transmits and measures signals.

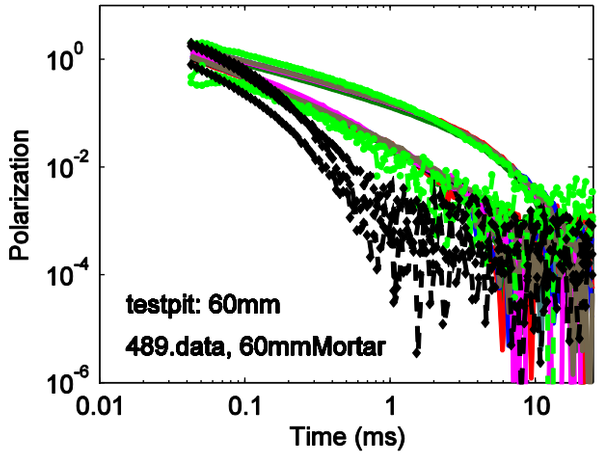


Figure 59. Recovered polarizations for anomaly 489 after two-object inversion (green and black thick dots). The other color curves in the plots represent the recovered polarizations of 60 mm Mortar from the test-pit data.

d. Feature vectors with discrimination potential

Polarizations for each item were obtained by inverting for the diagonal components of the polarizability tensors. Figure 60 plots an example of inverting data over a 60 mm mortar body. We would expect that the recovered polarizabilities should be smoothly decreasing. However, we solve for the amplitude of the polarizability at each time channel using noisy data. In order to "de-noise" the recovered polarizabilities we parameterize the curve as the sum of exponentials

$$d(t) = \sum_{n=1}^{N} a_n \exp\left(-\frac{t}{\tau_n}\right)$$

where τ_n are a set of log-spaced time constants that span the measurement window of the sensor. The coefficients a_n are solved via a linear inversion. Figure 60 compares the recovered polarizabilities (in red, green, and blue lines) with smoothed versions of the polarizabilities (in black). The smooth version of the secondary polarizabilities is obtained by averaging the secondary polarizabilities then smoothing. Note that we used this smoothing process instead of Pasion-Oldenburg because that parameterization couldn't match the recovered decay curves over the entire time-range spanned by the TEMTADS (this likely occurs because the transmitter waveform of the TEMTADS does not approximate a true step-off response).

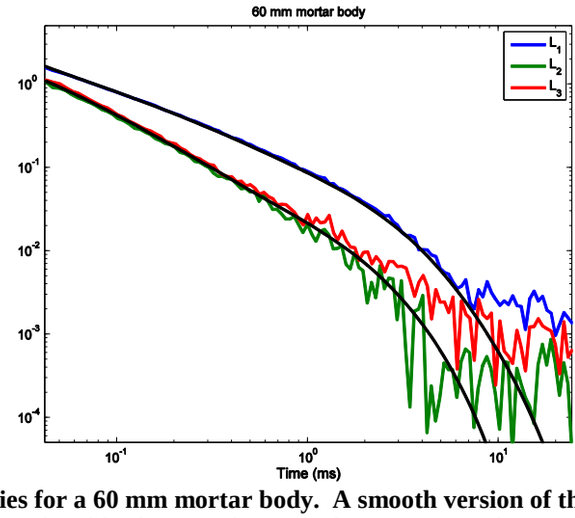
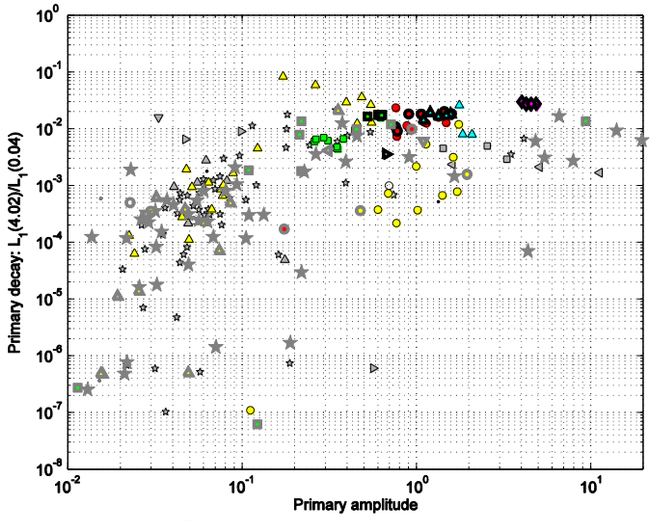


Figure 60. Recovered polarizabilities for a 60 mm mortar body. A smooth version of the estimated polarizabilities are plotted in black. The secondary polarizability in black is determined by averaging the secondary polarizabilities then smoothing.

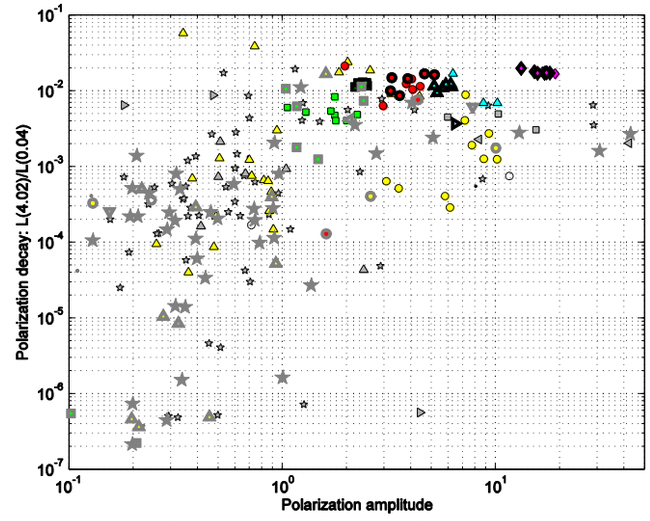
Figure 61 plots a number of feature parameters derived from the smoothed polarization fits to the training and test-pit data. In the plots, test-pit items are shown with a solid black outline, while multi-object items are plotted with a grey outline. For the multi-object fits, no effort has been made to determine which of the two models is most likely ordnance. In each case, visual examination of the plots reveals that one of the multi-object models lies near one of the UXO clusters in each feature space. The features plotted are:

- Primary polarization size (evaluated as the integral of $L_1(t)$ from t = 0.04 to 5 ms);
- Secondary polarization size (evaluated as the integral of $L_2(t)$ from t = 0.04 to 5 ms);

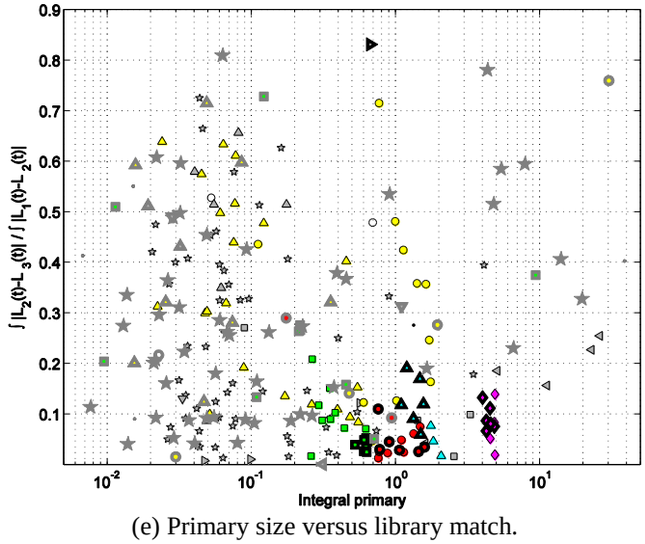
(a) Primary size versus primary time-decay



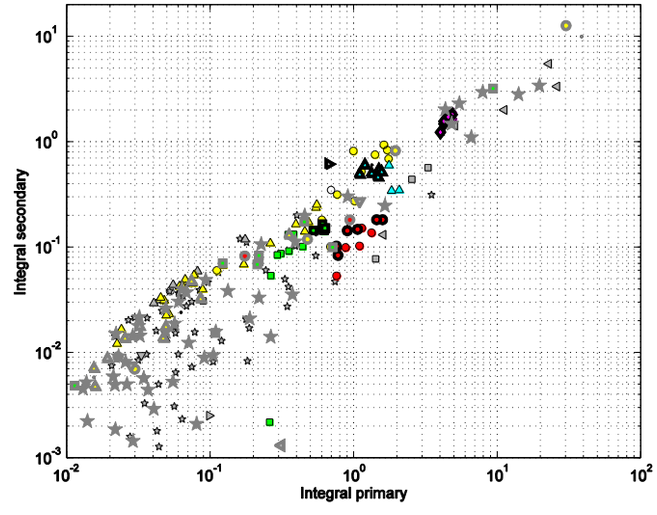
(b) Total polarizability size verus time-decay



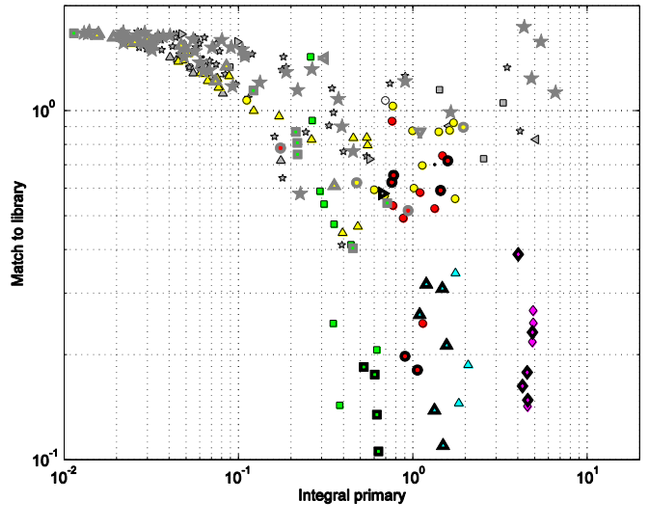
(c) Primary size versus asymmetry



(d) Primary versus secondary size



(e) Primary size versus library match.



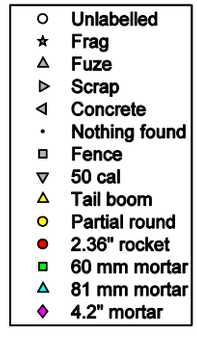


Figure 61. Feature vectors extracted from TEMTADS polarization fits to test-pit (dark outline) and training data. Multi-objet anomalies are indentified by a grey outline.

- Primary time-decay evaluated as $L_1(t=4\text{ ms})/L_1(t=0.04\text{ ms})$. We found that if we went much later in time that the decay parameter for the 60 mm bodies became very small;
- Polarization amplitude evaluated as the sum of squares of each polarization at t=0.04 ms;
- Polarization decay, as for primary decay but using the sum of squares of the three polarizations.
- Asymmetry evaluated as $\int |L_3(t) - L_2(t)| dt \;/\; \int |L_1(t) - L_2(t)| dt$ where the polarizations are ranked from largest (L_1) to smallest (L_3) and the integrals are evaluated from t = 0.04 to 5 ms.
- Match to library, evaluated as $\min_{k=1:6} \left\{ \sqrt{\sum_{i=1}^{3} \left[\int |L_i(t) - J_i^k(t)| dt \right]^2} \right\}$ where $J_i^k(t)$ is the i-th polarization of the k-th item in the ordnance library (see the next section on library methods) and the integrals are evaluated from t = 0.04 to 5 ms.

As with the previous generation sensors, a combination of object size and decay rate provides good separation between the UXO and a lot of the clutter (Figure 61a & b). These two parameters are most effective when evaluated using just the primary polarization (Figure 61a), and not the sum of the polarizations (Figure 61b).

With this next-generation sensor technology we anticipate being able to constrain the secondary polarizations so that shape should provide a useful diagnostic. A plot of the primary and secondary size reveals a fairly tight clustering of the UXO classes (Figure 61d). However, the secondary polarization does not appear to add much in terms of discrimination information.

motivation behind the "asymmetry" parameter is that $\int |L_3(t) - L_2(t)| dt$ will be small for rod-like objects (ideally it will be equal to zero) and large for plate-like objects, with the opposite behavior expected for $\int |L_1(t) - L_2(t)| dt$. Thus asymmetry will be large for plates and asymmetric items and small for rod-like items with an axis of symmetry: and indeed the training and test-pit data appear to support this assertion (Figure 61c). Surprisingly, some of the 81 mm mortars in the test-pit have the largest asymmetry values amongst the UXO class. The item with largest value is anomaly 504, a 60 mm body at 23 cm, which has very low signal-to-noise ratio.

The "match to library" parameter will be small when the recovered polarizations closely match one of the items in the library (Figure 61e). However, even when the 60 mm bodies are included in the library (see the section on the library method), there are still some items that don't match the library very well. The worst outlier is anomaly 1023, a 60 mm body at 36 cm depth. Another 60 mm body doesn't match the library well (anomaly 504 at 23 cm depth).

The feature space plots indicate that some combination of primary size, primary decay and asymmetry would make the most robust feature set. We have some concerns regarding the accuracy of the asymmetry parameter on the smaller items, particularly the 60 mm bodies. Thus some hybrid strategy may work best and we estimate performance of different methods in the following section.

e. Discrimination strategy

In Figure 62 and Figure 63 we compare 0.632 bootstrap performance on TEMTADS training data for various choices of feature space and class definitions. Each figure shows the bootstrap performance of QDA classifiers trained on the features indicated in each respective subplot. The figure captions indicate the TOI

class used in training. For example, the 2.36” TOI class indicates that the QDA classifier is trained on two classes: 2.36” rockets and everything that is not a 2.36” rocket. When generating the bootstrapped ROC for this classifier, however, we do not include other TOIs (4.2”, 81 mm, etc.) as false positives. Some caution is required when interpreting these plots. In particular, the “merged all” (using all TOI classes in training) bootstrap ROCs appear quite favorable (Figure 64b). However, since different numbers and types of TOIs are used to generate each figure, direct comparisons between figures are not necessarily valid.

For larger TOIs (4.2, 2.36”, 81 mm), the feature space spanned by primary amplitude, primary decay and asymmetry has good average performance. This classifier seems comparable to our usual two-dimensional amplitude vs. decay classifiers, but has the added potential to identify partial rounds which may appear similar to medium-sized ordnance in terms of amplitude and decay alone. Increasing the dimensionality of the feature space necessitates an increase in the number of features required for training, and so for medium ordnance (2.36” and 81 mm), we will merge these classes together as a single TOI class (Figure 63a). For small ordnance, the asymmetry parameter seems less beneficial.

These results suggest that it might be advantageous to apply separate classifiers trained to find individual target classes, and then to merge these classifiers together to generate a single diglist for the test data. We considered several strategies with the following classifiers:

1. QDA trained on 4.2” mortars as TOI class with primary amplitude, primary decay and asymmetry.
2. QDA trained on 2.36” rockets and 81 mm mortars as TOI class with primary amplitude, primary decay and asymmetry.
3. QDA trained on 60 mm as TOI class with primary amplitude and primary decay (but not asymmetry).
4. QDA trained on all ordnance classes (4.2”, 2.36” rockets, 81 mm mortars and 60 mm) as TOI class with primary amplitude and primary decay (standard approach, no asymmetry information)

Figure 64 and Figure 65 illustrate the benefits of this approach by comparing the following combinations of classifiers

1. **Size,decay, asymmetry:** the maximum TOI probability from classifiers 1-3.
2. **Size, decay:** classifier 4.

Figure 65 shows a clear improvement in median classification performance with incorporation of asymmetry information as described in (a), and so we use this approach for discrimination of TEMTADS data. While the worst case performance from the bootstrap analysis is quite poor (Figure 64 and Figure 65), this likely represents very “unlucky” realizations of bootstrapped training and test data. We expect that the performance obtained with this approach on the actual data will be closer to the median or best case realizations. A stop dig point will be determined for this classifier using a bootstrap analysis.

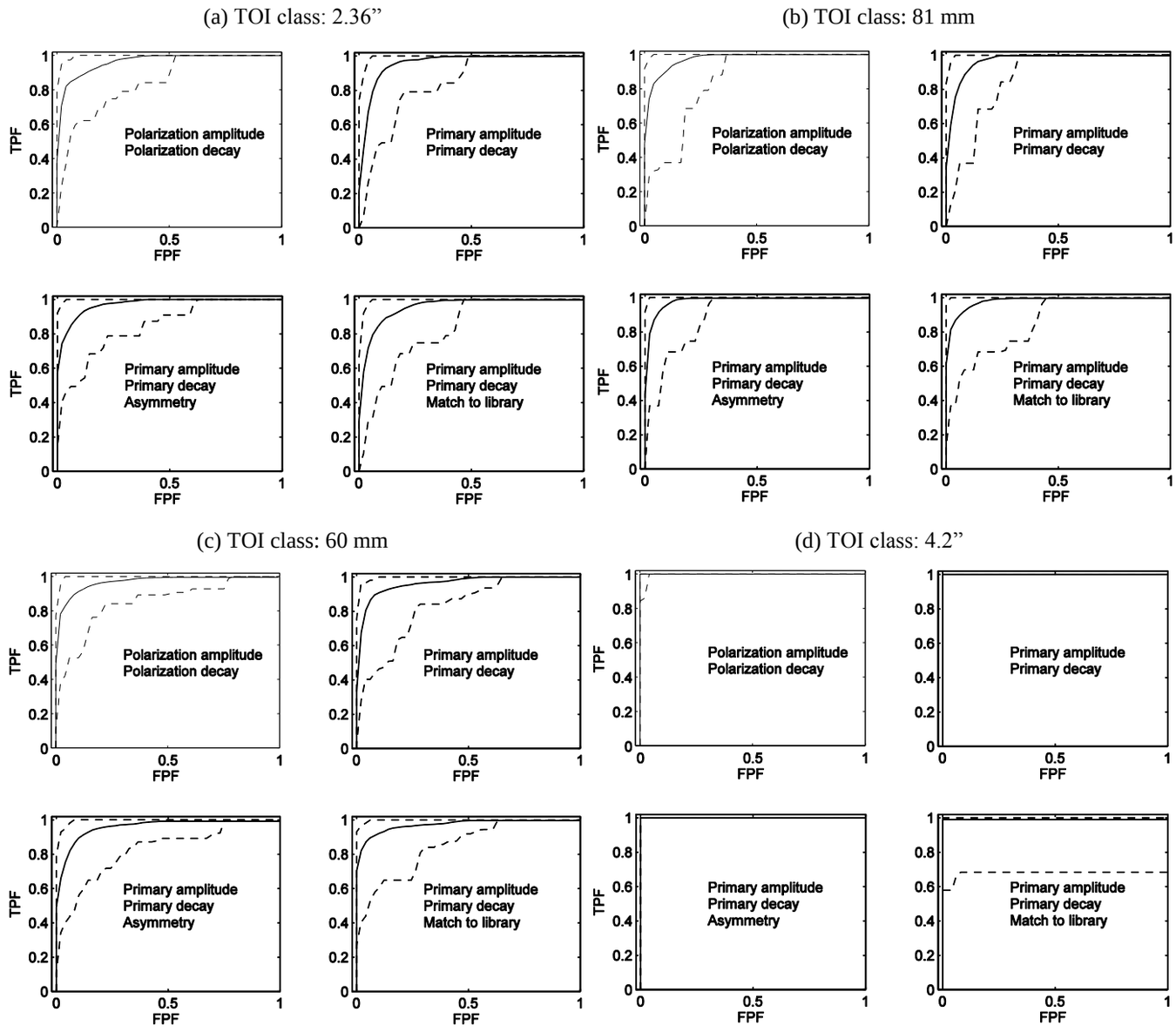


Figure 62. Bootstrap analysis of classification performance for each ordnance class when using different features in a quadratic discriminant classifier.

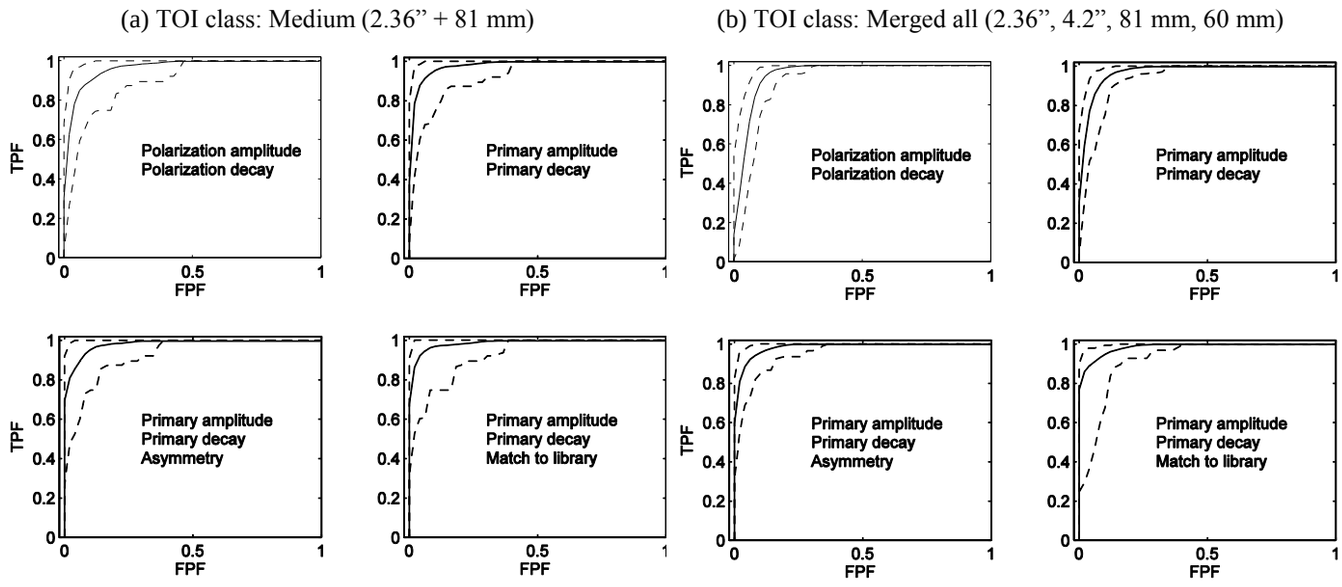


Figure 63. Bootstrap analysis of classification performance combined ordnance classes when using different features in a quadratic discriminant classifier.

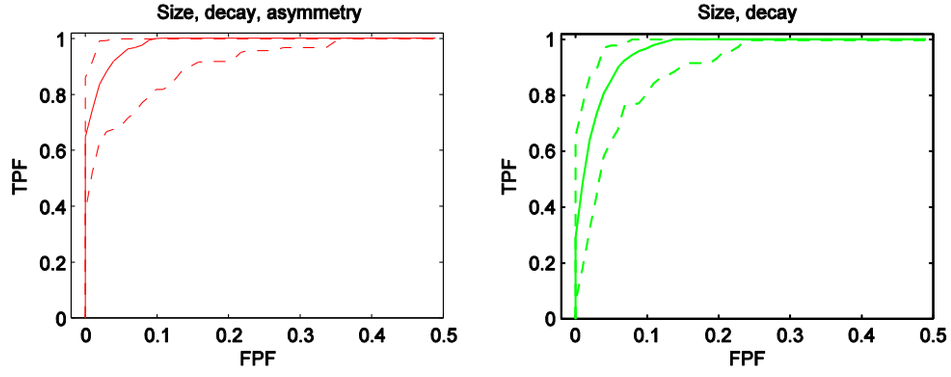


Figure 64. Bootstrap comparison of classification strategies for TEMTADS data.

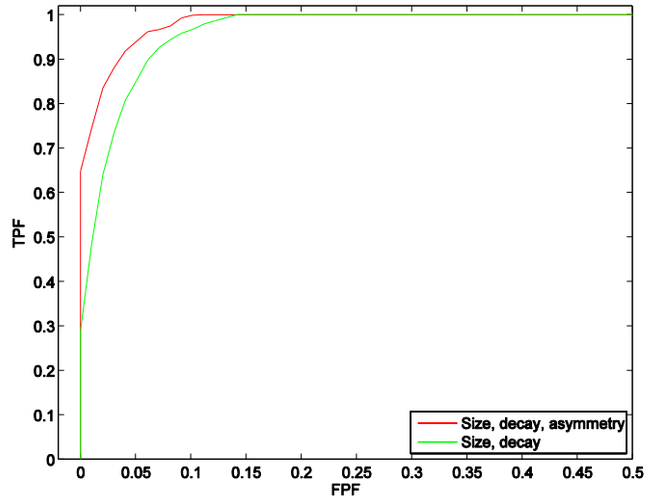


Figure 65. Median bootstrap ROCs for TEMTADS classification strategies.

8. TEMTADS CUED INTERROGATION: LIBRARY METHOD

A library of polarizations was originally generated by inverting test pit data acquired over a 2.36 inch Rocket, 60 mm mortar, 81 mm mortar, and 4.2 inch mortar. Data were acquired at a number of different target depths and orientations. For the library we assume that each target has axial symmetry (i.e. secondary polarizatibilities are equal). The polarizaibilities for these different anomalies were then averaged to produce a single polarizability for each target type. Polarizabilities were then smoothed using the method outlined in Section **XX**. Analysis of the training data suggested a pair of Targets of Interest (TOI) should be added to the library to improve discrimination performance. The additional items were (1) a 60 mm mortar body and (2) an additional 81 mm mortar with decay characteristics significantly different than the 81 mm mortar used in the test pit. Figure 66 plots the different polarizabilities in the library.

a. Library based discrimination method

We consider two different approaches that utilize the library of polarizabilities.

Method 1 "Library Inversion": We first implement the library based method of Pasion et al. (2007). In this method we determine which target in the library most likely produced the anomaly. For each target in our library a non-linear inverse problem is solved for the position and orientation that minimizes the least-squares difference between the observed data anomaly and the data predicted from each target. To determine if the anomaly is likely generated by one of the targets we can either find the predicted data with the maximum correlation to the observed data.

Once we have determined which target in the library most likely produced each data anomaly, two approaches are considered for constructing a dig-list . In the first approach we simply sort all the anomalies according to correlation coefficient between the observed data and the data predicted by the best fit target. In the second approach we compare the data predicted by the best fit target of the library with the data predicted by the unconstrained inversion (i.e. the inversion for 3 polarizabilities). In the fit quality is much better with 3 polarizabilities than with the item in the library, then we label that anomaly as being less likely to belong to the library.

Method 2 "Polarization Match": The second application of the polarizability library involves inverting the data for the polarizability tensor, then comparing the estimated polarizabilities with those in the library. This method is essentially an automated way of comparing of recovered polarizabilities with polarizabilities of expected targets. We rank likelihood of a target by the norm of the difference between the recovered polarizability and candidate target polarizability in the library. A dig-list is generated by comparing the misfits of all the anomalies, i.e. those anomalies whose estimated polarizabilities have the closest match to a polarizability in the library are given the highest priority dig.

a. Library based discrimination results

Figure 67 compares the performance of both methods when using the six member library. Figure 67(a) summarizes the performance when applying Method 1 by comparing the data fits of the unconstrained inversion to the library based inversion. That is, we consider an anomaly to be more likely to be from a target in the library if the data fit of the unconstrained inversion is approximately the same as the best fit inversion using a polarizability from the library. Figure 67(b) summarizes the performance of Method 2. For both methods target 1023 and 489 are problematic.

Figure 68 demonstrates why anomalies 1023 and 489 are problematic. In both cases the data were unable to accurately constrain the secondary polarizabilities. The principle polarizability is more accurately estimated than the secondary polarizabilities.

Figure 67(c) summarizes the performance of Method 2 when comparing the sum of polarizabilities (i.e. $\Sigma L_i(t)$). We have learned from previous experience that the sum of polarizabilities can be a more robust parameter to estimate than the individual polarizabilities. However, the sum of polarizabilities loses some of the shape information that can be derived by looking at the relative magnitudes of the individual polarizabilities. Figure 67(c) demonstrates that the total polarizability allows for a lower FAR. However, the algorithm is less efficient, i.e. more scrap is dug at the start of the digging process. Nonetheless, this is the algorithm that we will use for our library based discrimination of TEMTADS data.

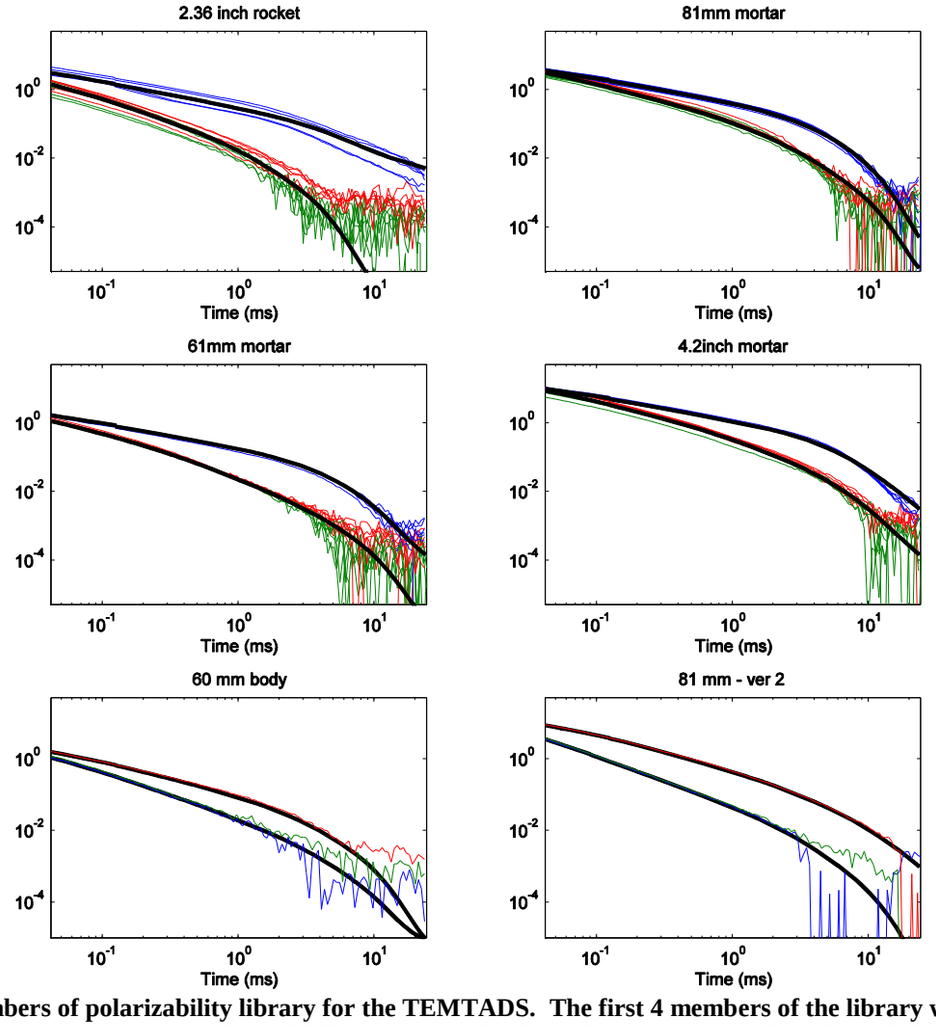
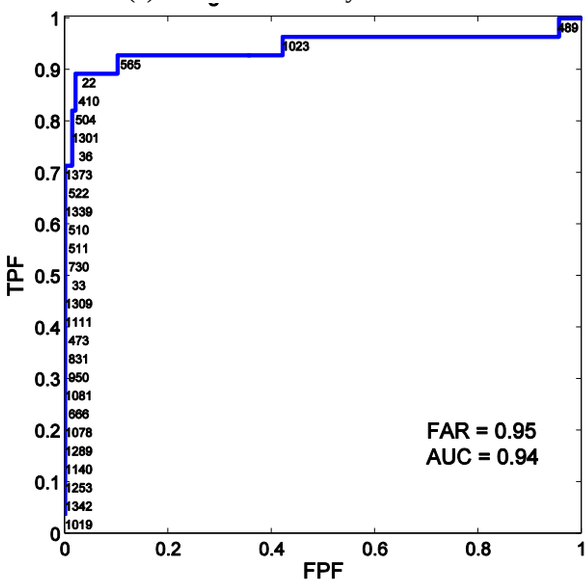
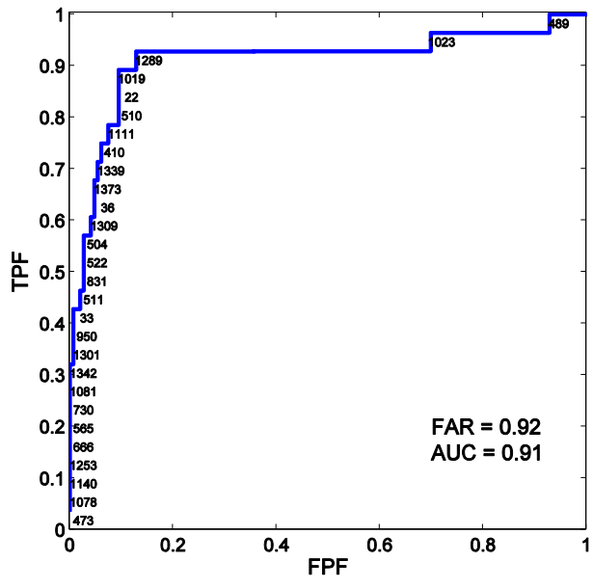


Figure 66. Members of polarizability library for the TEMTADS. The first 4 members of the library were determined from test pit data acquired at a number of different target depths and orientations. The polarizabilities for the other items were obtained from the training data.

(a) Method 1: Library Inversion



(b) Method 2. Polarization Match



(c) Method 2. Polarization Match using total polarizability (i.e. ΣL_i)

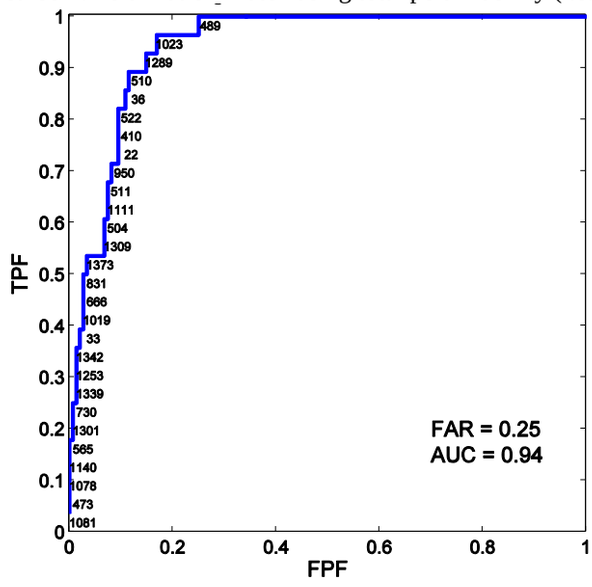


Figure 67 Comparison of library based discrimination methods. The inability of the algorithm to identify targets 1023 and 489 as high priority items lead to high false alarm rates. A more robust (i.e. lower FAR) approach is to use the total polarizability (c).

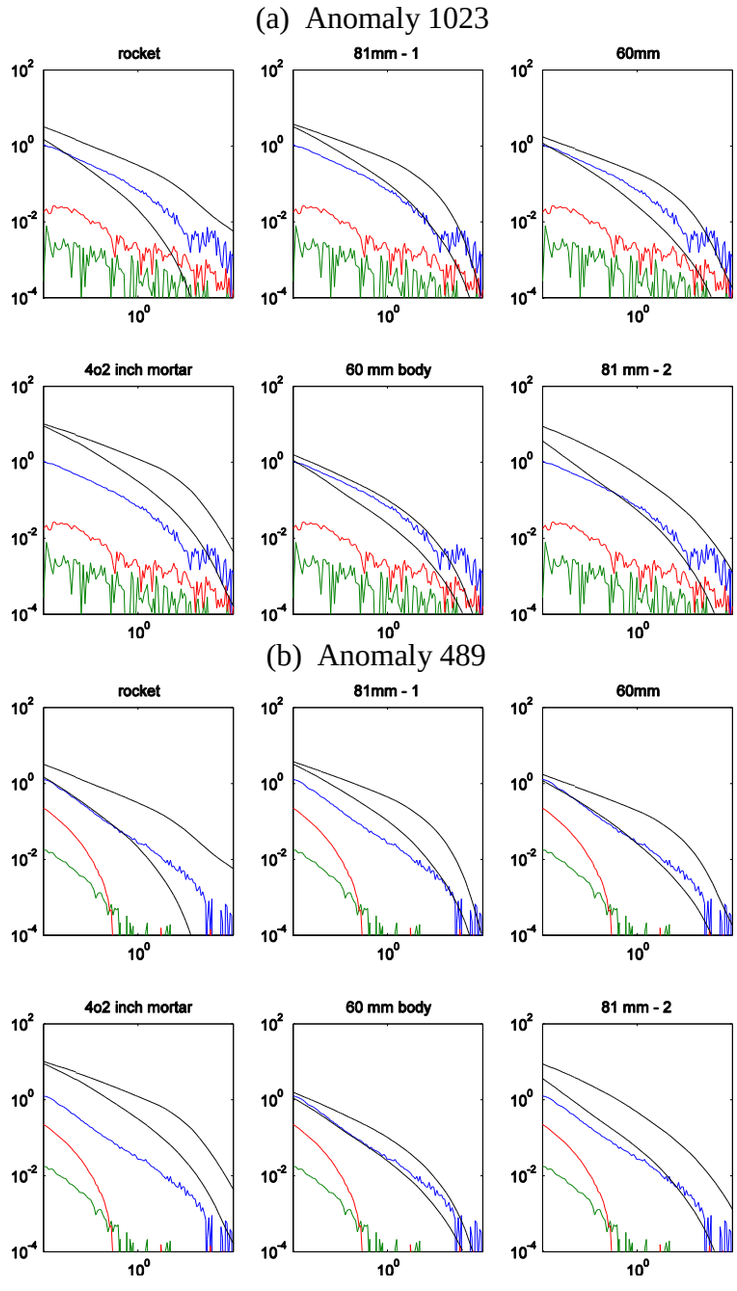


Figure 68 Comparison of recovered polarizabilities of anomalies 1023 and 489 with members of the library. In both cases, the target of interest is a 60 mm mortar body. The recovered polarizabilities in both cases have the characteristics of inversion of low signal to noise anomalies, i.e. the primary polarizability is somewhat constrained, while the secondary polarizatibilities are not well constrained.

9. METAL MAPPER CUED INTERROGATION: STATISTICAL CLASSIFICATION

a. Analysis of Test-pit results

Data were collected over the four primary targets of interest (TOIs) in the test pit (33 datasets) and in "free air" (34) over 50 time gates ranging from 0.024 ms to 7.912 ms. Since early time channels can be contaminated with sensor related noise, we omit the first 8 channels during our analysis. The earliest time channel we consider is at 0.106 ms. All data anomalies were processed within using the following workflow:

- *Error estimation:* Estimates of data errors consisted of two components:
  - Floor noise of 50 mV across all time channels for all receivers was estimated from measurements of the test pit area made without targets present.
  - For each datum, we calculated an additional error equal to 2 percent of the datum amplitude.
- *Masking*: No masking was applied, i.e. all soundings were used for inversion.
- *Model*: The data were fit using 3 unique polarizations for the dipole tensor. The amplitude of each polarization was estimated at each of the forty two time channels. The 108-element-model vector is

$$m = [X, Y, Z, \varphi, \theta, \gamma, L_1(t_1), L_1(t_2), L_1(t_3), \ldots, , L_2(t_1), L_2(t_2), L_2(t_3), \ldots, L_3(t_1), L_3(t_2), L_3(t_3), \ldots]$$

where (X,Y,Z) is the location, (φ,θ,γ) are the orientation angles and $L_i(t_j)$ is the ith polarization at the jth time channel.

Figure 69 shows the recovered polarizations from each of the different orientations measured in the test-pit or in air. Polarization values at later times are often noisy, and we utilize the same exponential smoothing strategy that was applied to the TEMTADS to produce smoothed versions of the polarizabilities.

Figure 70 compares estimated and actual depths obtained for these targets. When calculating the depth, we assume a sensor height (or stand off) of 0.21 m. Depth is accurately predicted within 0.05 m. The accuracy with which depth was estimated suggests accurate characterization of the dipole polarizabilities for the four test pit targets. As a sensitivity analysis we re-inverted the data with 100 mV noise level and obtained the same results (i.e. predicted depth within 0.01 cm of the 50 mV result for all inversions).

b. Analysis of training data

Training data was obtained for 171 anomalies detected in the field. The same procedure was applied to invert those anomalies. All inversions yielded satisfactory results without any need to adjust parameters used when processing the test pit data. We also found that inversions of MetalMapper data were typically unaffected by the proximity of neighboring target, owing to the small sensor footprint.

Depth is generally well recovered for the majority of UXO (Figure 71). A number of noticeable outliers are predicted 0.10-0.20 m too deep:

a. Three 60 mm mortars. Upon comparing the recovered polarizations with those derived from 60 mm mortars in the test pit, we noticed that

   i. Target 510 shows a faster than normal time decay and depth error Δd= 0.10 m.

   ii. Target 522 has lower amplitude and faster time decay and Δd= 0.15 m.

   iii. Target 831 has the expected polarization characteristics and Δd = 0.15 m.

b. One 81 mm mortar: Target 1339 with Δd =0.16 m.

c. One 4.2” mortar: Target 730 with Δd =0.18 m.

Inversion diagnostics for item 522 are shown in Figure 72. The item is predicted at 0.21 cm instead of 0.05 cm despite an almost-perfect fit (with correlation coefficient close to 1). The depth-misfit curve shows a sharp, localized minimum. This well constrained solution appears to be characteristic of MetalMapper inversions.

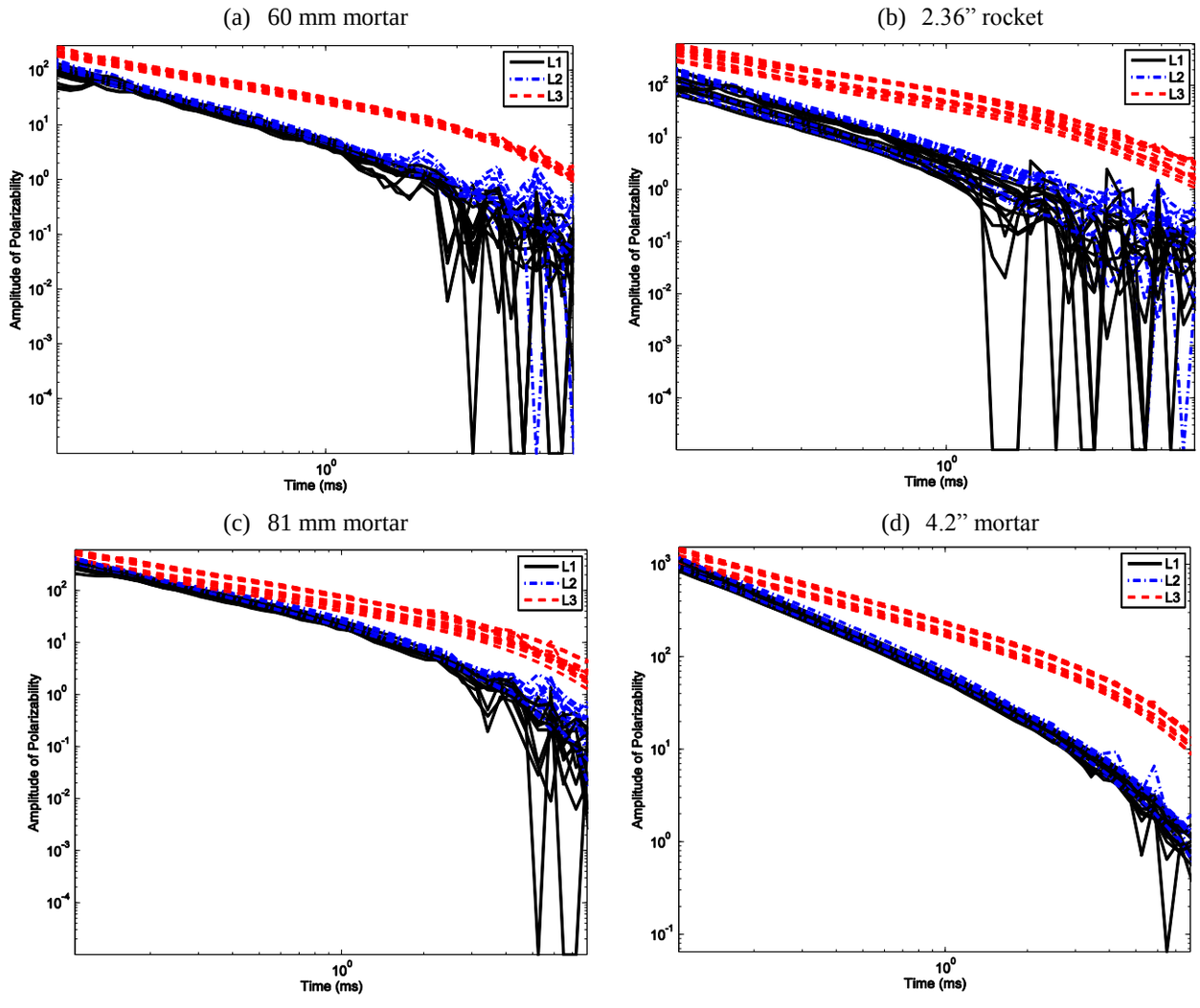


Figure 69. Variability of polarizability curves and late time instability for different types of UXO.

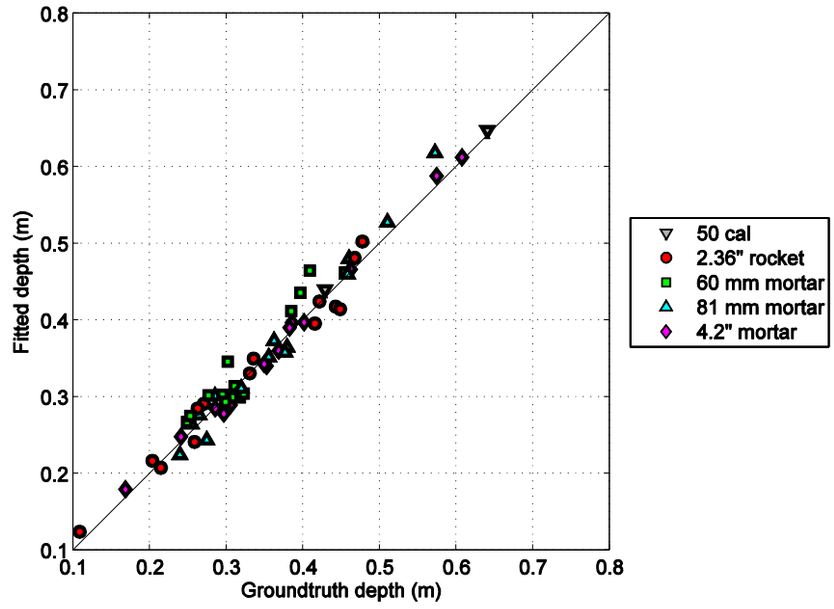


Figure 70. Fitted vs. ground-truth depth for test pit and free air measurements with MetalMapper sensor.

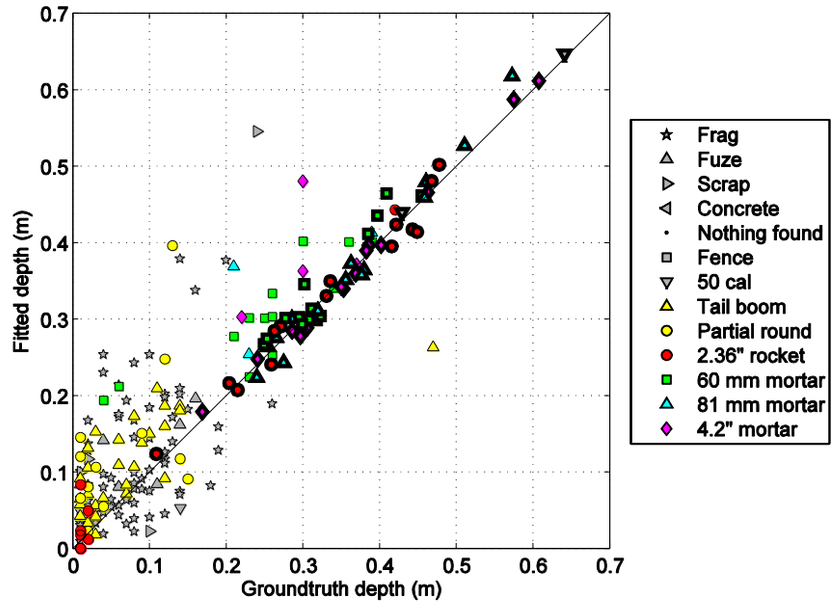


Figure 71. Fitted vs. ground-truth depth for test pit and free air items (MetalMapper survey in cued static data collection mode).

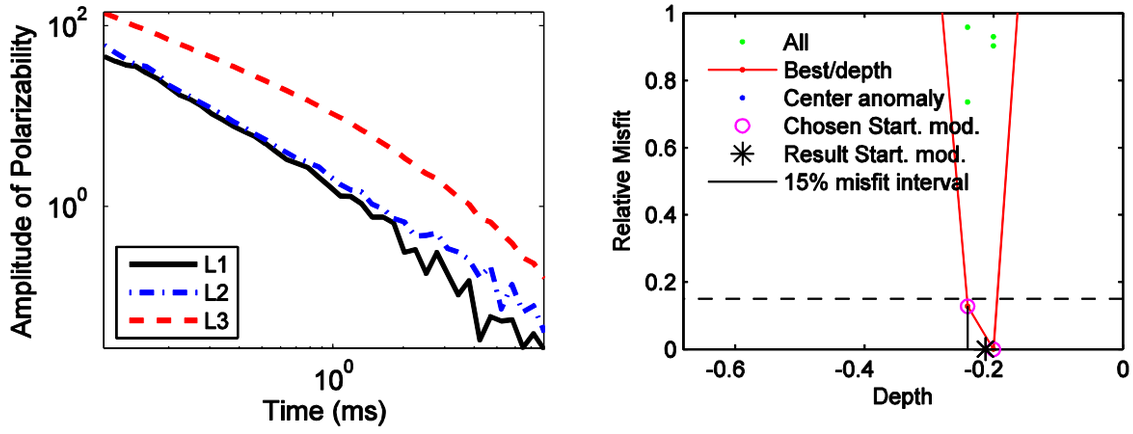


Figure 72. Inversion diagnostic features for 60 mm mortar with 0.15 m depth error (Target 522). Left: Recovered polarizability curves. Right: Depth-Misfit curve.

Comparison of the recovered polarizations of the 60 mm item with polarizabilities inferred from the test pit and free air measurements shows that the amplitude and time decay of this item differ from all typical UXO (Figure 73). The ground truth photo seems to indicate a small body at a deeper depth than reported, which could explain the discrepancies.

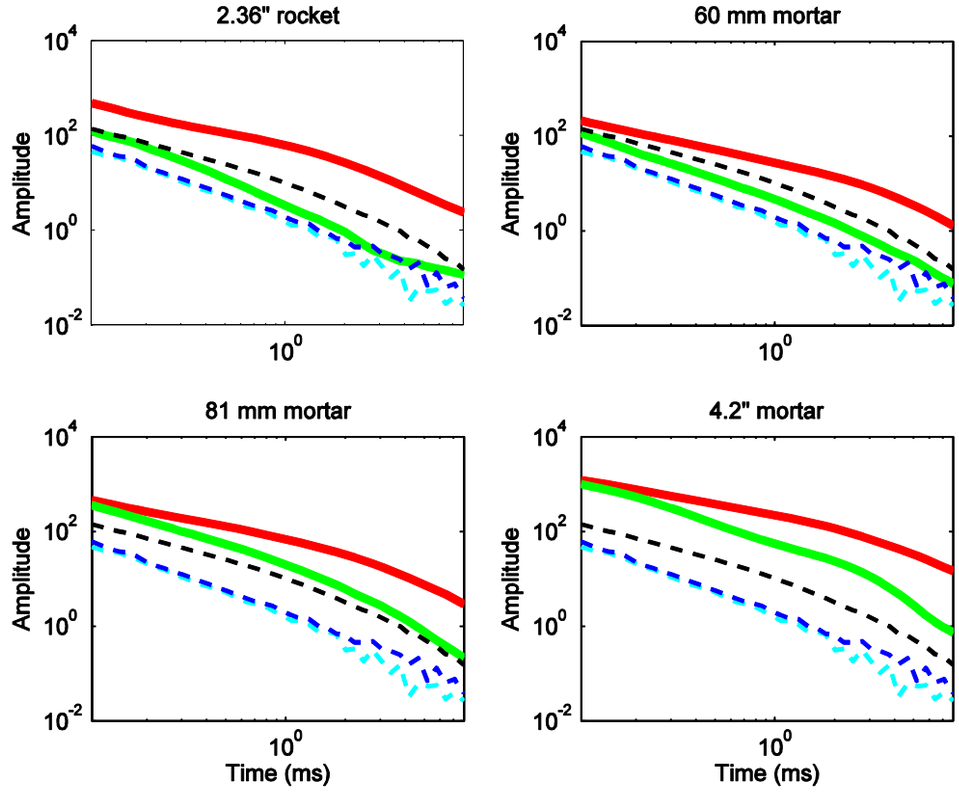


Figure 73. Polarizability curves for all four types of targets of interest (red and green curves for primary and secondary polarizations, respectively) compared to 60 mm mortar with poor depth estimate (Target 522, dashed curves).

c. Feature vectors with discrimination potential

We propose to base the classification mostly on physical parameters that can be inferred from the dipole model. The different features to test are:

1. Amplitude of the polarizability (square root of the sum of the square polarizability components at first time channel);
2. Time decay rate of the polarizability amplitude between two given times;
3. Asymmetry computed as the ratio of the primary polarizability and the mean of the secondary and tertiary polarizabilities at a given time channel, (A1);
4. Asymmetry of the target taken as the ratio of the secondary and tertiary polarizabilities at a given time channel, (A2);
5. Asymmetry as ratio of integrals of the differences between smoothed polarizations (secondary minus tertiary and primary minus secondary), (A3);
6. A fit quality parameter: the Correlation Coefficient (CC) between the inverted observed and predicted data.

Motivated by the classification strategy adopted for the previous sensors, we utilized the polarizability amplitude at the first time channel (norm([L1 L2 L3](t1))) and its decay rate as features to train a probabilistic neural net classifier. Decay rate is defined as the ratio between the polarizability amplitude at early and late times. Figure 74 and Figure 75 show the distribution of features for the different types of test and training items with a time decay computed at 0.27 ms (10th channel) and 1.31 ms (25th channel). Classification with the early decay feature gives the impression that parameters are well clustered in model space, which would generally be the most desirable option, while the later decay shows larger variability within each UXO class and larger separation with clutter items. Signal at later times is often too weak to extract robust polarizabilities. Recovered polarizabilities can be smoothed by fitting a sum of exponentials to approximate the late time decay (as was done with the TEMTADS data).

Close examination of Figure 76 and Figure 78 reveal that the inferred time decay rates for two of the three field 81 mm mortars are faster than their test pit counterparts (Targets 1081 and 1339, with associated depth errors of 0.02 and 0.16 m, respectively). These are different mortars compared to the ones measured in the test-pit.

The quality of data from monostatic sensors acquiring in a dynamic mode is generally not sufficient to warrant using target shape information. However, the ability of the MetalMapper to accurately estimate polarizabilties should allow us to use an asymmetry feature. Figure 77 through Figure 79 contain feature space plots defined by asymmetry and principle polarizability amplitude. The two 81 mm with fast time decay do not appear as outliers when utilizing the asymmetry measures of Figure 77 through Figure 79, which suggests that polarizability amplitude, decay rate and asymmetry can complement each other to separate UXO from clutter.

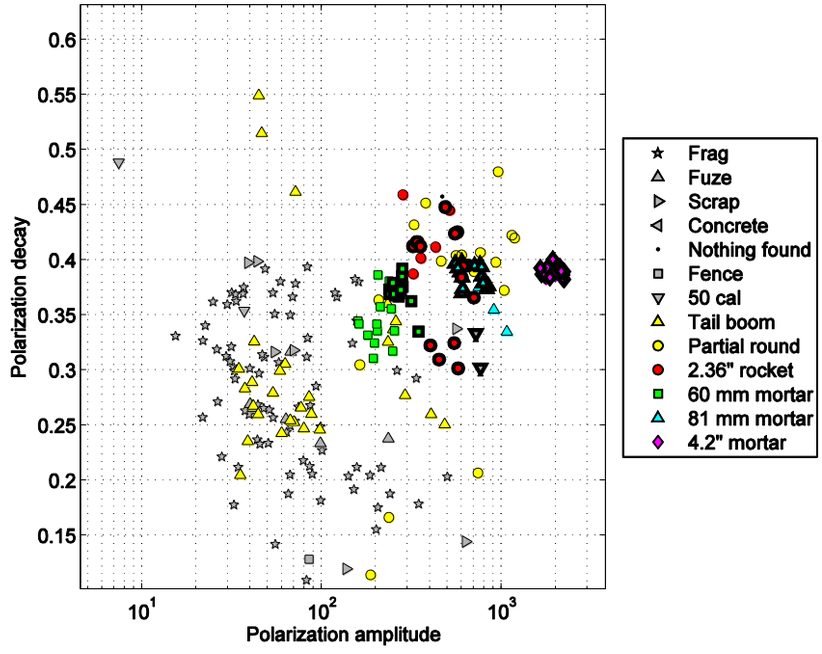


Figure 74. Feature plot with time decay obtained from polarizability at 0.27 ms.

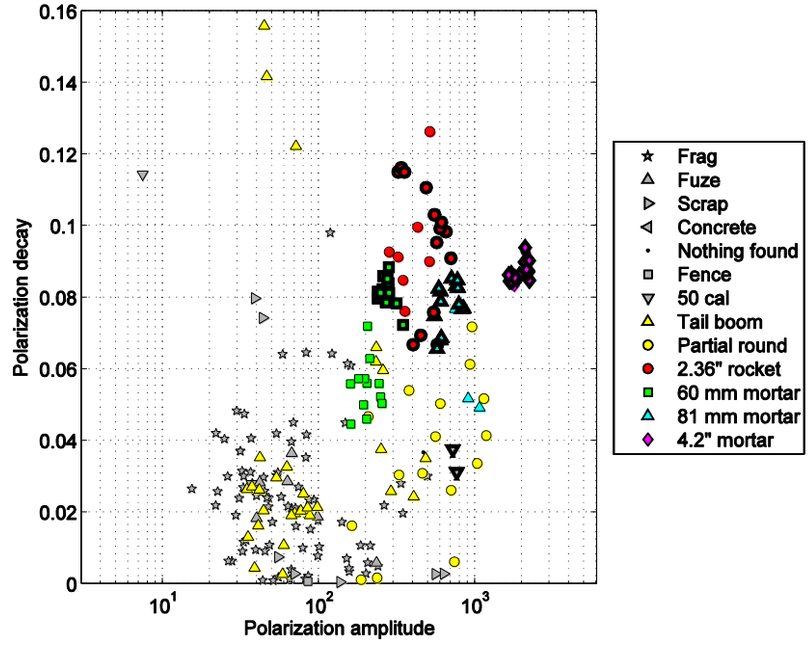


Figure 75. Feature plot with time decay obtained from polarizability at 1.314 ms (no smoothing of late polarizabilities).

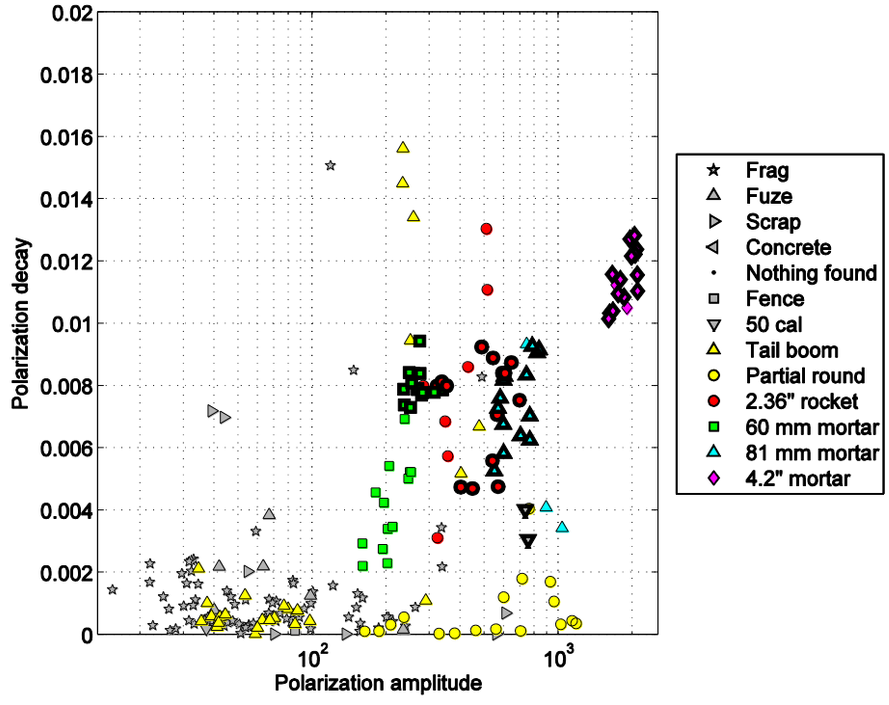


Figure 76. Feature plot with time decay obtained from polarizability at 6.1 ms (with smoothing of late polarizabilities).

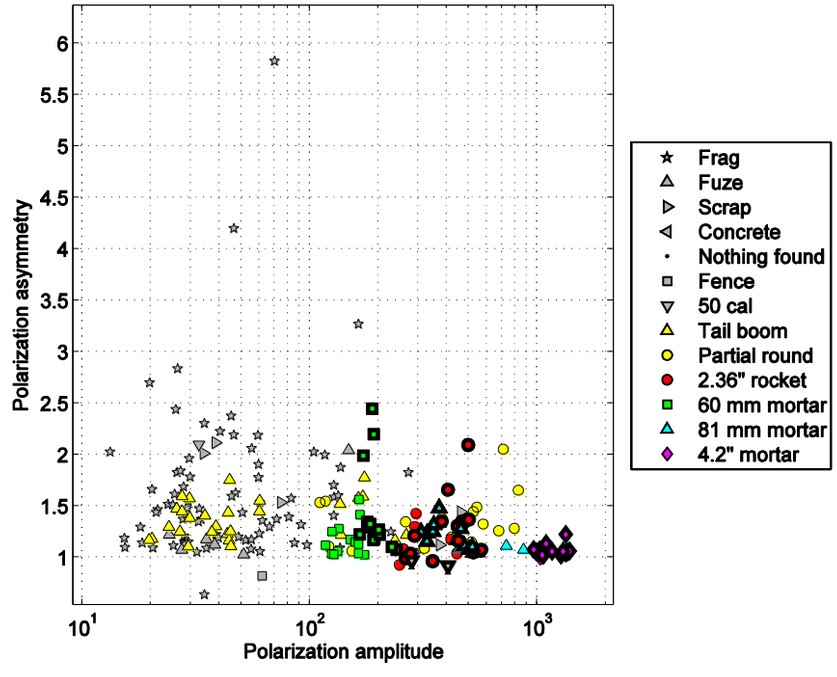


Figure 77. Feature plot with polarization amplitude (principal polarization) and polarization asymmetry (ratio of secondary and tertiary polarizations) at 0.118 ms.

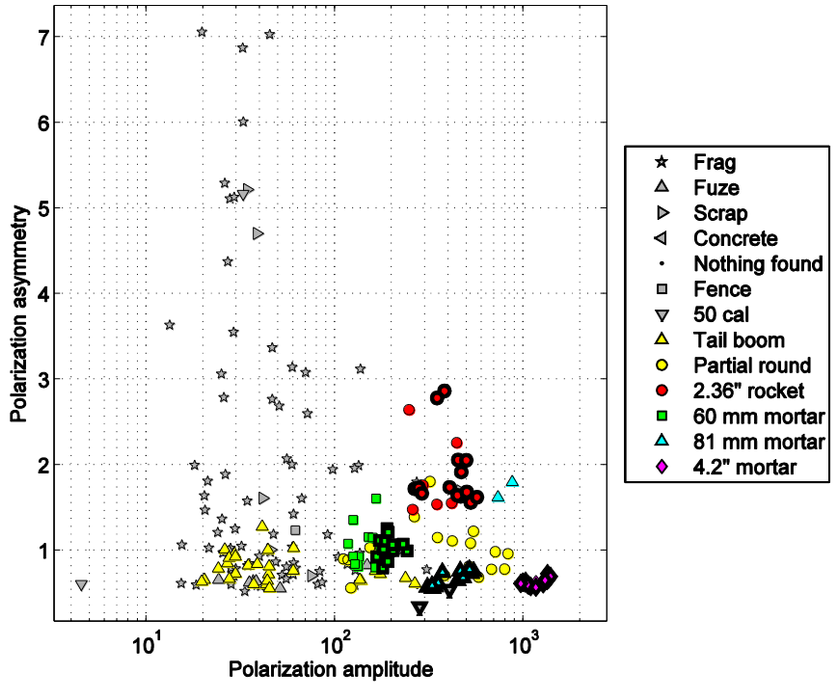


Figure 78. Feature plot with polarization amplitude (principal polarization) and polarization asymmetry (ratio of primary and mean of secondary and tertiary polarizations) at 0.118 ms.

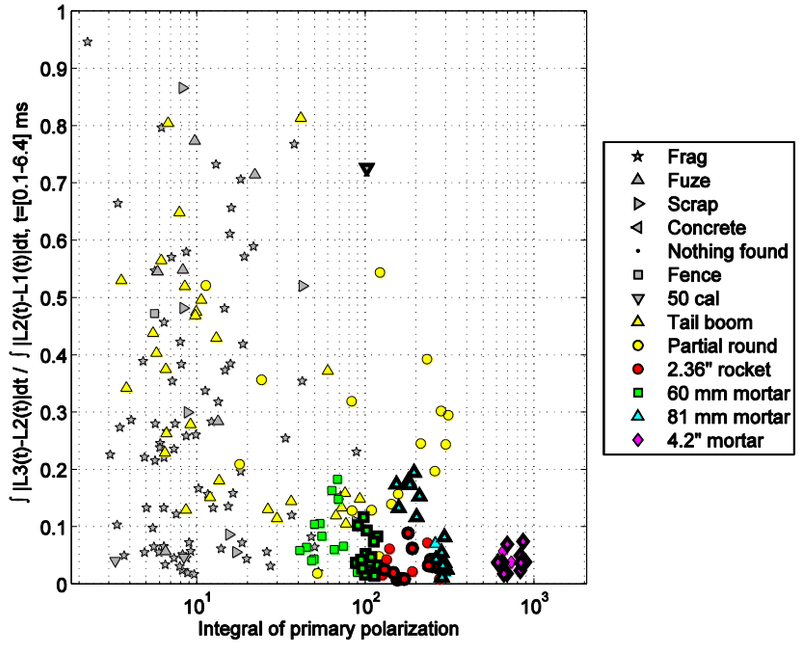


Figure 79. Feature plot with polarization amplitude (here the integral of the smoothed primary polarization, from 0.1 to 6.1 ms) and polarization asymmetry (ratio of integrals of differences between smoothed polarizations: secondary minus tertiary and primary minus secondary).

Defining a Classifier

d. Classification strategy

Efficient discriminating features can be identified by building a ROC curve to assess the performance of each method in leaving non-UXO items in the ground. Similarly to the other sensor data, we use a statistical classifier to discriminate UXO from clutter. In the following we mainly apply a Probabilistic Neural Network (PNN) to distinguish between two classes, UXO and non-UXO, utilizing a given set of features (physical parameters, fit), training on a training set (TS) and being validated on a validation set (VS). Training is possible with ground truth obtained from 67 anomalies from the test pit (TP) and 171 anomalies in the field training (FT) set.

To assess the effectiveness of different classification strategies and avoid building a classifier that is too sensitive to particular training items we perform a **bootstrap analysis** by training on different subsets to classify FT. For each bootstrap iteration, the training set is obtaining by random sampling and replacement of the TP+FT ground-truth sets. Each training subset is built to be a quarter of the size of TP+FT to reproduce an experiment in which a small training set is used to classify a larger validation set. Sampling is done on a target type basis so that the relative distribution of target types (all four UXO + general clutter) is similar to that of TP+FT.

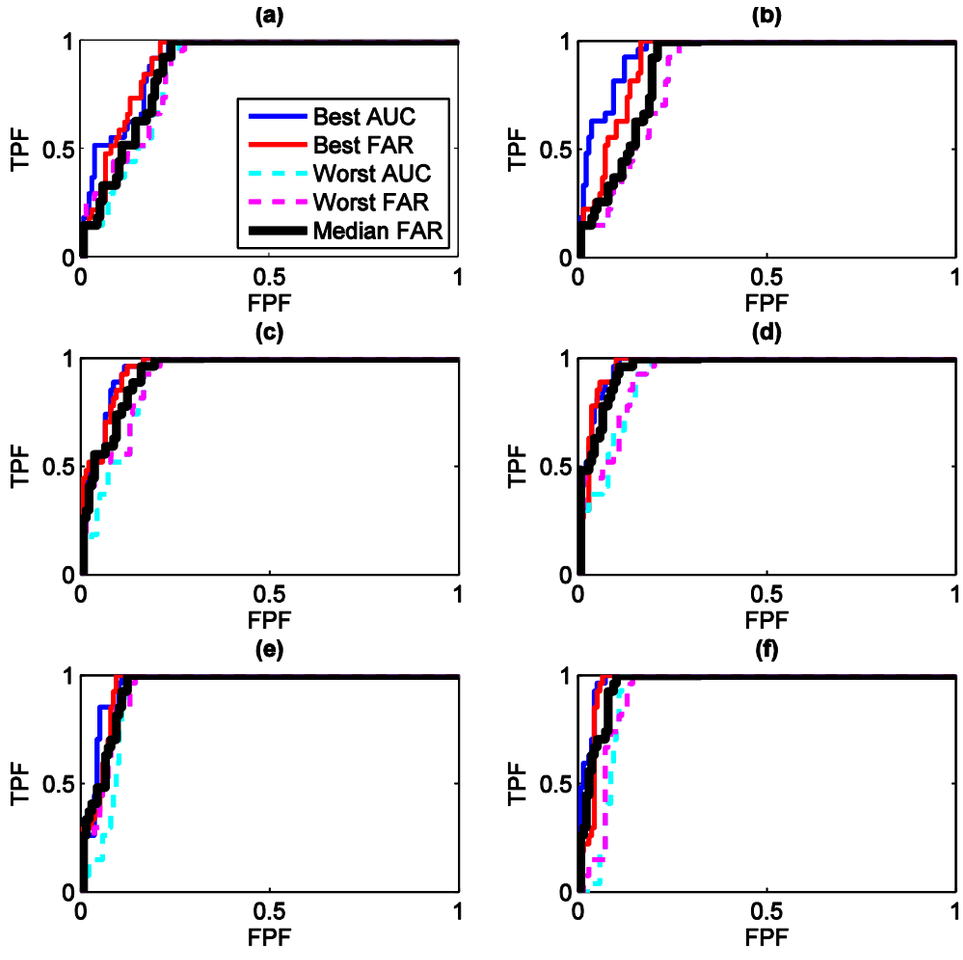


Figure 80. ROC curve comparison for amplitude of polarizability (norm of L(t1)) versus time decay (ratio of norm(L) at t1 and tn). (a) tn=0.27 ms. (b) Same plus CC. (c) tn=1.3 ms. (d) Same plus CC. (e) tn=6.1 ms. (f) Same plus CC. The ROC curve efficiency can be measured with the False Alarm Rate (FAR) when all UXO are recovered or the Area Under the Curve (AUC).

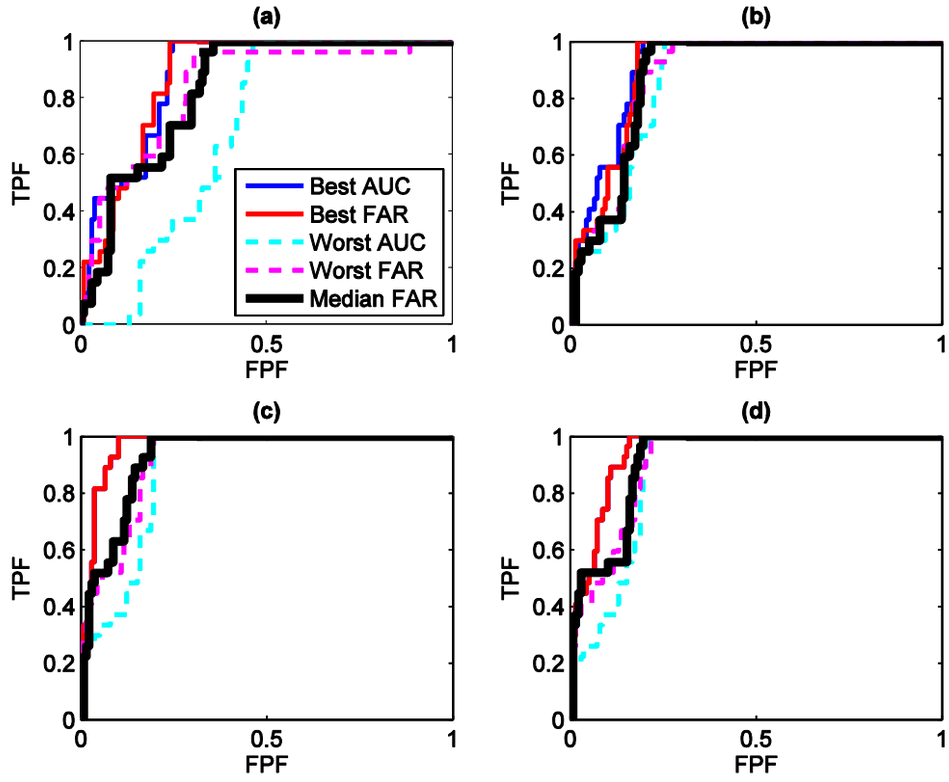


Figure 81. ROC curve comparison for amplitude of principal polarizability versus asymmetry. (a) Asymmetry A1 as the ratio of the primary and mean of secondary and tertiary polarizabilities. (b) Asymmetry A2 as the ratio of the secondary and tertiary polarizabilities. (c) Asymmetry A3 as ratio of integrals from 0.1 to 2.3 ms of differences between smoothed polarizations (secondary minus tertiary and primary minus secondary). (d) Same for integral from 0.1 to 6.1 ms.

Figure 80 shows that classification is improved when utilizing late time decay, from 6.1 ms to 1.3 ms and 0.27 ms (see polarization plots in Figure 74 through Figure 76). Time decay at 6.1 ms is obtained by smoothing the recovered polarizabilities, which could be unstable and lead to errors. In contrast, performance at 1.3 ms is acceptable, and incidentally corresponds to a similar time range as that of a Geonics EM-61 (latest channel at 1.266 ms). Best performance is achieved when including the correlation coefficient as a third feature.

Asymmetry of the target can help distinguish a body of revolution (likely a UXO) from an asymmetric piece of shrapnel. The best measure of asymmetry is obtained when taking the ratio of the secondary minus tertiary and primary minus secondary (A3), as shown in Figure 81. Combining amplitude, time decay, asymmetry and, optionally, correlation coefficient leads to great efficiency at recovering all UXO without dig out significant amount of clutter (Figure 82). This result applies whether using time decay up to 6.1 ms or stopping at 1.3 ms, where signal is stronger and parameter recovery is more robust.

We propose to use amplitude, correlation coefficient time decay and asymmetry (A3) in the 0.1-1.3 ms time range to classify the field anomalies with MetalMapper data.

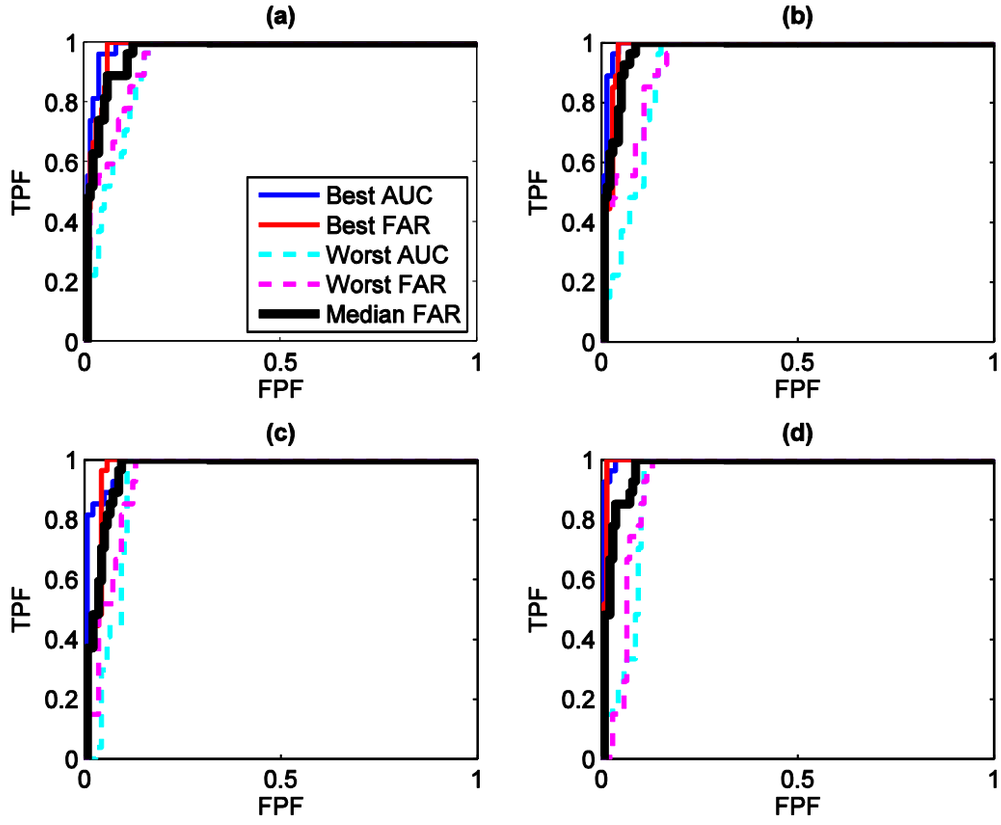


Figure 82. ROC curve comparison for amplitude of polarizability versus time decay versus asymmetry (ratio of integral of secondary and tertiary polarizabilities and integral of primary minus secondary). (a) Amplitude, decay and asymmetry A3 at t=1.3 ms. (b) Same plus CC. (c) Amplitude, decay and asymmetry A3 at t=6.1 ms. (d) Same plus CC.

10. METAL MAPPER CUED INTERROGATION: LIBRARY METHOD

The objective of library based methods is to determine which member of a library of potential targets most likely produced an observed data anomaly.

a. Library Generation

A library of polarizations was originally generated by inverting test pit data acquired over a 2.36 inch Rocket, 60 mm mortar, 81 mm mortar, and 4.2 inch mortar. Data were acquired at a number of different target depths and orientations. For this library we assume that each target has axial symmetry (i.e. secondary polarizatibilities are equal). The polarizaibilities for these different anomalies were then averaged to produce a single polarizability for each target type.

Analysis of the training data suggested a number of Targets of Interest (TOI) should be added to the library to improve discrimination performance. The additional items were (1) a 60 mm mortar body, (2) an additional 81 mm mortar with decay characteristics significantly different than the 81 mm mortar used in the test pit, (3) a 2.36 inch rocket body, (4) 2.36 inch rocket motor, and (5) an empty 2.36 inch rocket. We show the different polarizabilities in the library in Figure 83.

b. Library based discrimination method

We used the same methods as those used for the TEMTADS library based method.

c. Library based discrimination results

We test the different methods on the training data. Test pit data were not included in the analysis. Figure 84 compares the performance of both library methods when using a six member library. The six members are the first 6 targets in Figure 1 (2.36 rocket, 4.2 inch mortar, 60 mm mortar with and without fins, and 2 types of 81 mm mortar). Figure 84(a) plots contains the ROC curve when applying Method 1 with a correlation coefficient to prioritize the dig list. Figure 84(b) summarizes the performance when applying Method 1 by comparing the data fits of the unconstrained inversion to the library based inversion. That is, we consider an anomaly to be more likely to be from a target in the library if the data fit of the unconstrained inversion is approximately the same as the best fit inversion using a polarizability from the library. In this case, ID 33 and ID 1373 are left too late in the dig list since they are not similar to any member of the library. Figure 85 contains photos of targets 33 and 1373. Figure 84(c) summarizes the performance of Method 2.

The two methods are repeated using a 9 member library that also contains target 33 and 1373 (Figure 86). There is little change when prioritizing digs according to correlation coefficient alone. As expected performance of the remaining methods is improved. This improvement is quantified by the lower FAR and high AUC measures of performance. We note that targets 33 and 1373 are significantly different from scrap such that the efficiency of either algorithm (quantified by the area under the curve (AUC)) is not reduced. Based on these results, we will utilize method 1 with 9 items in the library.

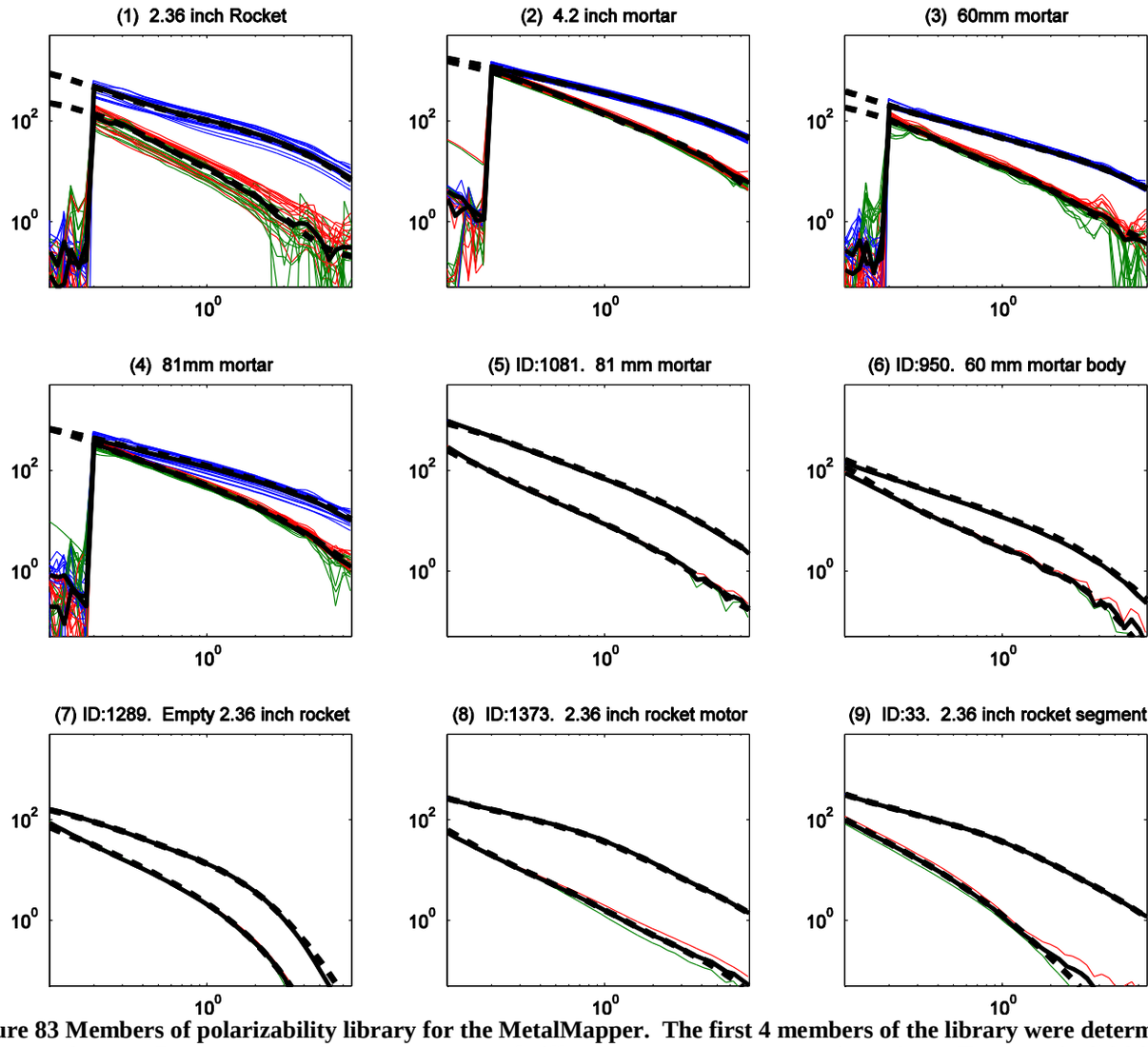
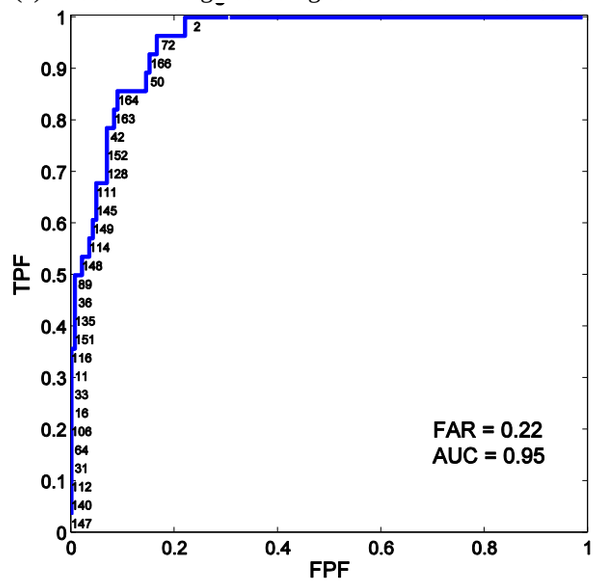
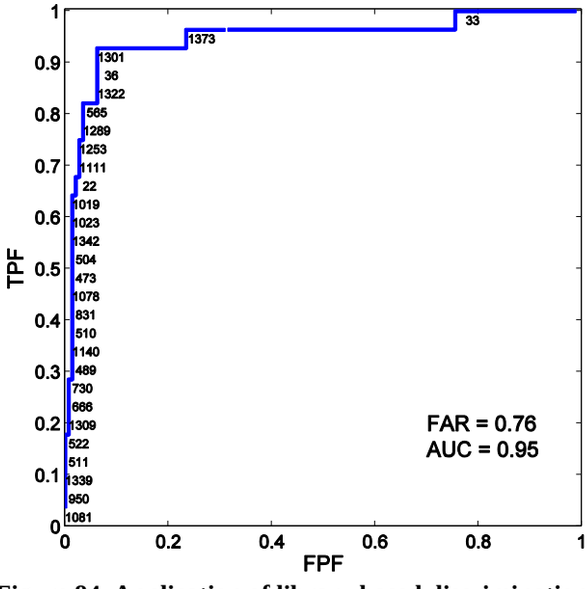


Figure 83 Members of polarizability library for the MetalMapper. The first 4 members of the library were determined from test pit data. Data were acquired at a number of different target depths and orientations. The polarizaibilities for these different anomalies were then averaged to produce a single polarizability for each target type. The remaining polarizabilities (i.e. 5 through 9) were added once a number of Target of Interests (TOI) not included in the test pit data were found in the training data. These targets include (5) ID 1081: an additional 81 mm mortar, (6) ID 950: A 60 mm mortar body, (7) ID 1289: Empty 2.36 inch rocket, (8) ID 1373: 2.36 inch rocket motor, and (9) ID 33: segment of a 2.36 inch rocket.

(a) Method 1: Dig according to correlation Coefficient



(b) Method 1: Library Inversion.



(c) Method 2: Polarization Match

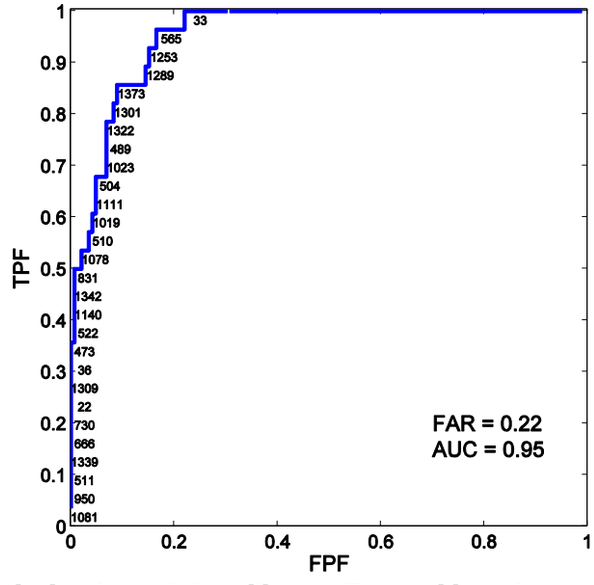


Figure 84 Application of library based discrimination methods using a six item library. Two problematic anomalies that greatly affects the false alarm rates are anomalies 33 and 1373.

(a) Target ID 33

(b) Target ID 1373

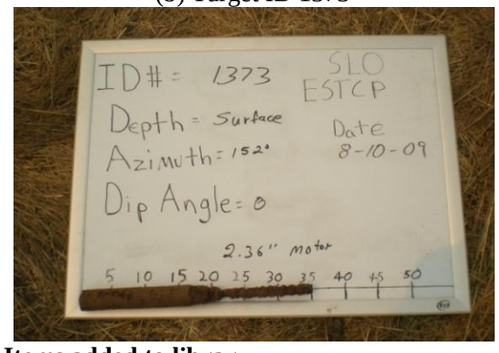
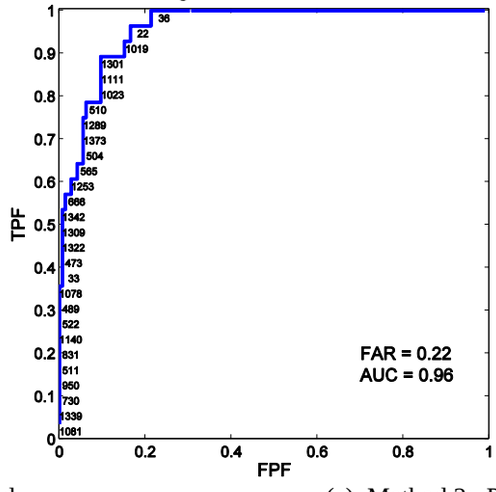
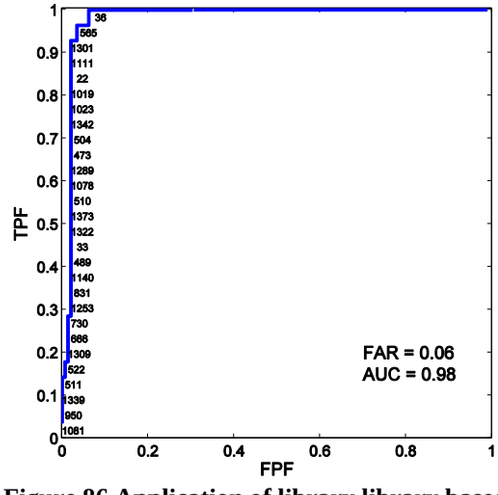


Figure 85 Items added to library

(a) Method 1: Dig according to correlation Coefficient



(b) Method 1: Library Method



(c) Method 2: Polarization Match

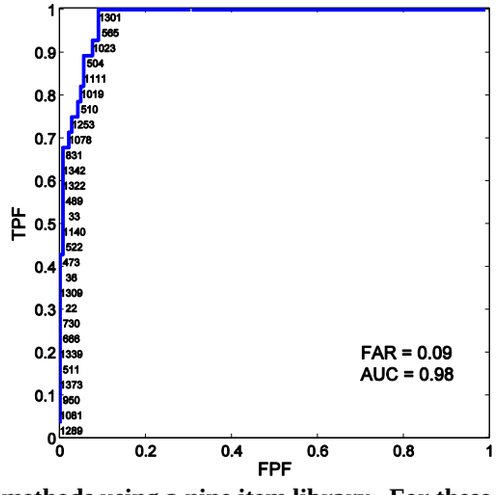


Figure 86 Application of library library based discrimination methods using a nine item library. For these results Targets 33 and 1373 were added to the library. Targets 33 and 1373 are significantly different from scrap such that the efficiency of either algorithm (quantified by the area under the curve (AUC)) is not reduced.

Appendix B: Processing of BUD data

The BUD sensor system consists of three orthogonal transmitters and eight pairs of differenced receivers. The useable time window is between 0.14 ms and 1.4 ms and comprises 35 logarithmically spaced time-gates.

The BUD data we received were all pre-processed: including normalization by transmitter current, background subtraction, and so on. The estimates of data errors in all receivers across time channels are provided in the BUD data and were used as standard deviations in our processing. Based on our processing experience of using TEMTADS and MetalMapper we assumed 3 percent noise on each data point. We processed the BUD data only for one location although BUD generally acquired around 10-location measurements for each Master ID anomaly.

Figure B1 is an inversion example of SLO_302-0316 data, i.e, Master ID is 302 and the BUD measurement location is numbered as 0316. The observed and predicted data are shown in Figure B1 and agree closely for most transmitter/receiver combinations. Figure B2 is the recovered polarizations for this anomaly versus the polarizabilities extracted from the BUD calibration data and shows that the item is most likely a 60-mm mortar with tail.

In April, 2009, we set up the BUD sensor in the UXOLab and tested our inversion using the BUD data collected at the former Camp Sibert.

In October, 2009, we received the BUD data collected at SLO. During this time, we were not aware of that the BUD sensor was re-configured. The processing of BUD calibration data showed that recovered polarizabilities for each of the interested UXOs were inconsistent and the fits to the observed data were generally quite poor. We spent a significant amount of time testing our algorithms to make sure there were no mistakes in our software. We contacted Dr. Erika Gasperikova to inquire about any possible changes in the BUD system found that the sensor had been reconfigured for the SLO study. Around the middle of December, we received sphere and spheroid test-stand data from Dr. Erika Gasperikova and using these test-stand data we finally corrected the sign setup in one transmitter that previously worked for the Sibert data.

Details of the classification method used are presented in section 7.4.

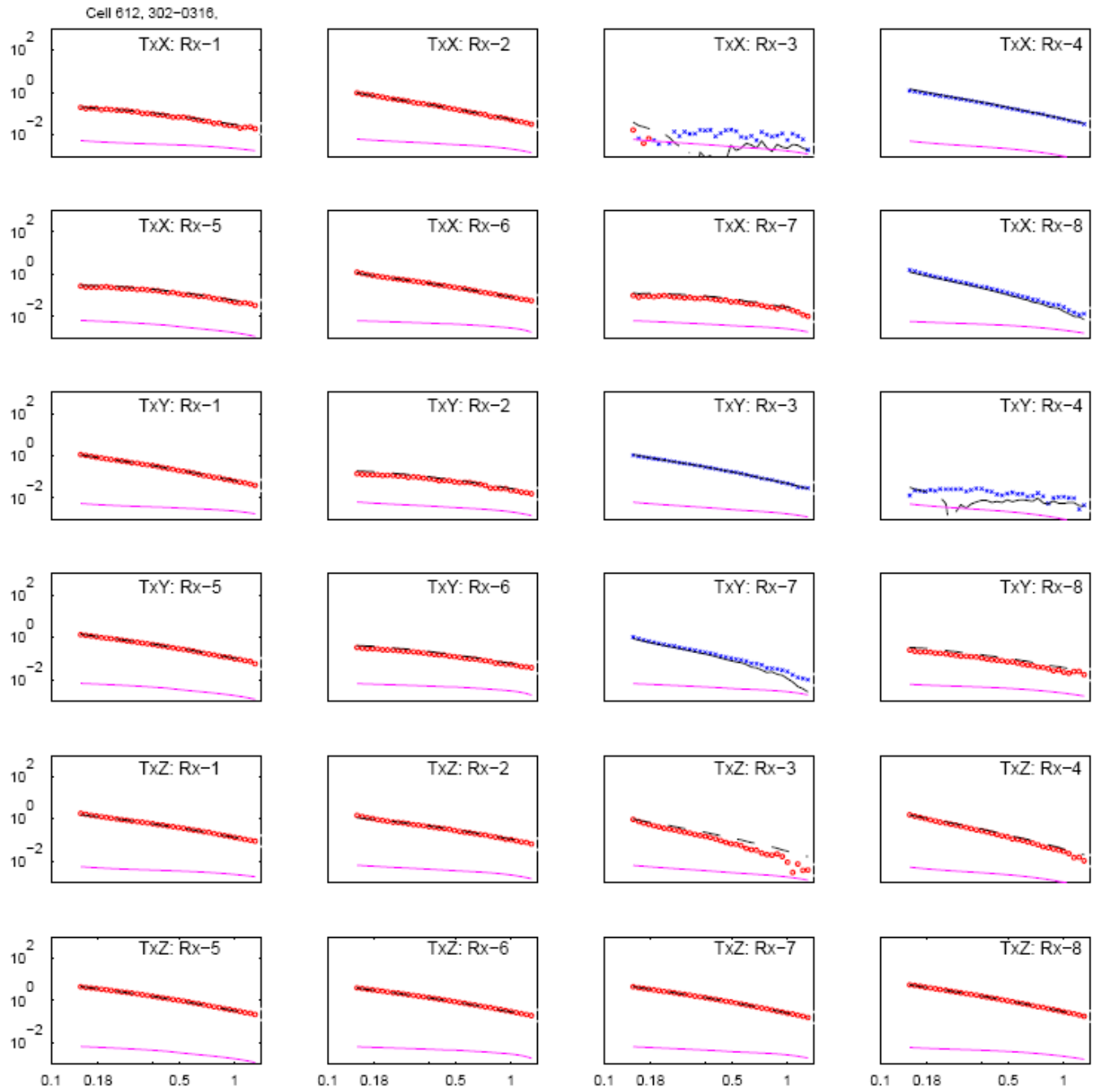


Figure B1. BUD inversion of SLO_302-316.data. The observed (red negative, blue positive) and predicted (black, dash negative, solid positive). The magenta curve in each subplot represents the estimated data error.

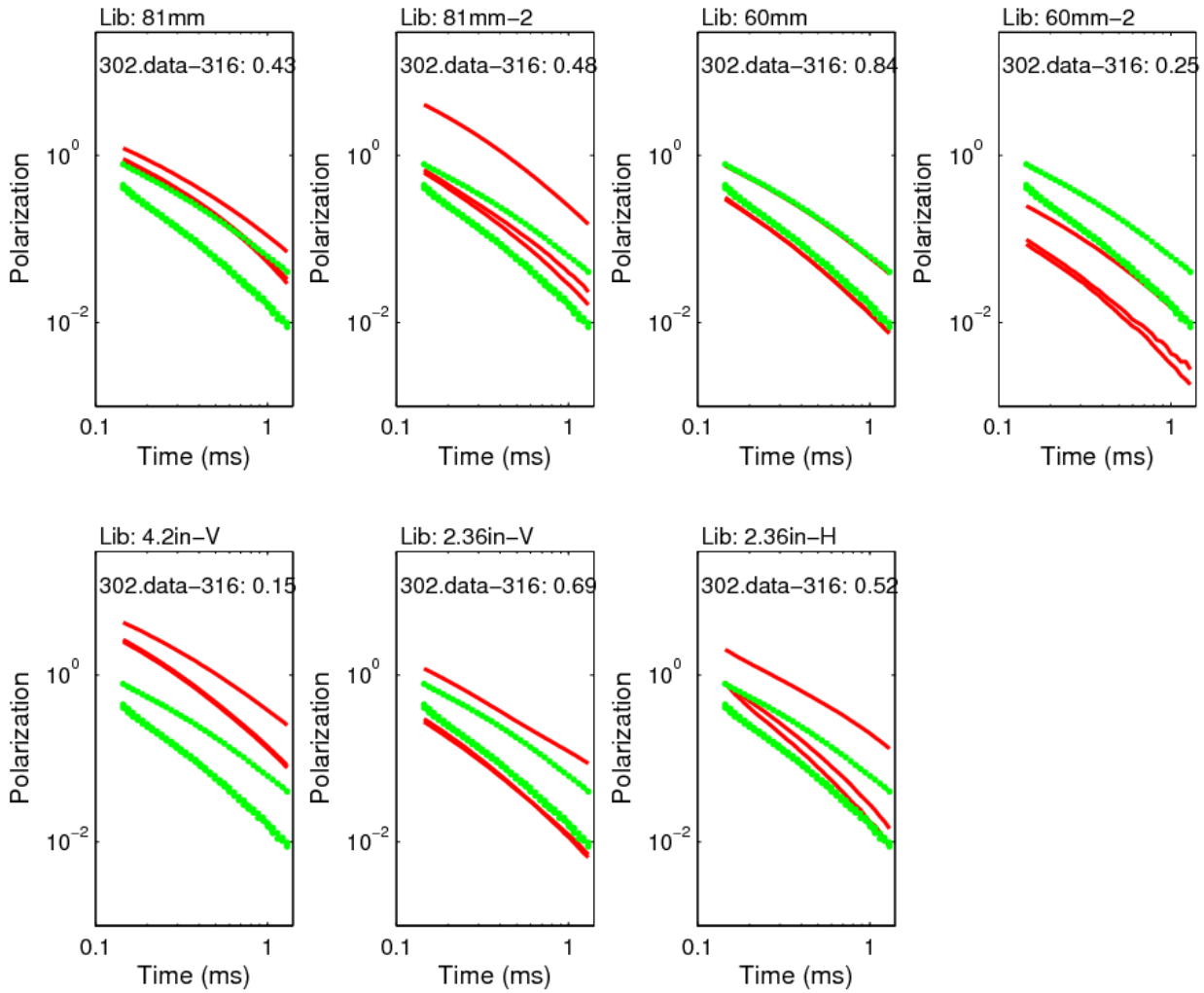


Figure B2. The recovered polarizations (green curves) from SLO_302-316.data against the 6 sets of polarizabilities extracted from BUD SLO calibration and training data.

Appendix C: Points of Contact

POINT OF CONTACT Name	ORGANIZATION Name Address	Phone Fax E-mail	Role in Project
Stephen Billings	Sky Research Inc, 112A/2386 East Mall Vancouver, BC, V6T-1Z3	541 552 5185 stephen.billings@skyresearch.com	Principal Investigator
Leonard Pasion	Sky Research Inc, 112A/2386 East Mall Vancouver, BC, V6T-1Z3	541 552 5186 leonard.pasion@skyresearch.com	Quality Control officer
Kevin Kingdon	Sky Research Inc, 112A/2386 East Mall Vancouver, BC, V6T-1Z3	541 552 5187 kevin.kingdon@skyresearch.com	Project management and personnel coordination
Joy Rogalla	Sky Research 445 Dead Indian Memorial Rd Ashland, OR	541 552 5104 joy.rogalla@skyresearch.com	Cost tracking