FINAL REPORT

Decision Support Tools for Munitions Response Performance Prediction and Risk Assessment

SERDP Project MR-2226

SEPTEMBER 2016

Laurens Beran
Barry Zelt
David Lutes
Black Tusk Geophysics Inc.

Distribution Statement A
This document has been cleared for public release

SERDP
DOD ▪ EPA ▪ DOE

Page Intentionally Left Blank

This report was prepared under contract to the Department of Defense Strategic Environmental Research and Development Program (SERDP). The publication of this report does not indicate endorsement by the Department of Defense, nor should the contents be construed as reflecting the official policy or position of the Department of Defense. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the Department of Defense.

Page Intentionally Left Blank

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 Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. 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 ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
09/01/2016	Final Report	April 2012 - September 2016

4. TITLE AND SUBTITLE
Decision Support Tools for Munitions Response Performance Prediction and Risk Assessment

5a. CONTRACT NUMBER
W912HQ-12-C-0011

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. Laurens Beran
Dr. Barry Zelt
Dr. David Lutes

5d. PROJECT NUMBER
MR-2226

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Black Tusk Geophysics, Inc.
401 / 1755 West Broadway
Vancouver, BC V6J 4S5

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Strategic Environmental Research and Development Program
4800 Mark Center Drive
Suite 17D03
Alexandria, VA 22350-3605

10. SPONSOR/MONITOR'S ACRONYM(S)
SERDP

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

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

13. SUPPLEMENTARY NOTES
N/A

14. ABSTRACT
As advanced UXO classification is transitioned from the research community to industry, there is an increasing need to develop practical tools to aid in the application of this technology to live-site munitions response projects. This project has aimed to anticipate the requirements of industry by developing algorithms and software to aid in critical decisions that must be made by geophysical data analysts and project managers during the course of a munitions response project.

15. SUBJECT TERMS
Unexploded ordnance (UXO), electromagnetic (EM), sensors, electromagnetic induction (EMI), data acquisition, feature extraction, classification.

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. Laurens Beran
U	U	U	UU	128	19b. TELEPHONE NUMBER *(Include area code)* 604-428-3382

Page Intentionally Left Blank

Abstract

Objective. New technologies for detection and classification of buried unexploded ordnance (UXO) have the potential to significantly reduce the cost of munitions response projects. In particular, electromagnetic (EM) sensors developed specifically for this problem can reliably discriminate between ordnance and non-hazardous metallic clutter. The classification process involves fitting a physical model to observed sensor data and then using the parameters of this model to make inferences about the physical properties of a detected target.

ESTCP demonstration projects have shown that advanced classification with next generation EM sensors consistently outperforms commercial standard systems. To further the use of advanced classification in munitions response projects, decision support tools are required to help project geophysicists, managers, and stakeholders understand:

(1) how to best deploy available technologies for a particular remediation problem;
(2) how to ensure with high confidence that all targets of interest are identified following remediation efforts.

While significant advances have been made in the acquisition and processing of geophysical data for classification of buried munitions, the success of any classification strategy strongly depends on the site characteristics, including range of munitions types and clutter, geological background, topography, and vegetation. The objective of this project is to develop and validate the components of a decision support system (DSS) that will help individual site-managers and teams design surveys and data processing strategies to achieve optimal discrimination performance at the lowest attainable cost for a given site.

Technical Approach. Our technical approach to this project considered two major research topics:

(1) *Performance prediction*. We developed techniques to model the performance of advanced sensors at a site. We considered both physical modelling of target response thresholds and developed statistical models to assess the feasibility of classification under site-specific conditions.
(2) *Risk assessment*. We developed statistical models to assess the posterior probability that targets of interest (TOI) remain in the ground following remediation efforts.

Results. This report describes work on the following topics:

(1) *Optimizing detection with multistatic sensors*. As monostatic sensors are replaced with multistatic, multi-component sensors for dynamic detection surveys, a threshold analysis tool is required to determine the minimum anomaly amplitude expected for a target of interest at a specified maximum clearance depth. We describe analysis tools that we have developed for threshold analysis with MetalMapper and TEMTADS2x2 sensors. We also develop an algorithm for objectively selecting time channels and receiver components for target picking with dynamic multistatic data. Our approach defines a detection channel that is a linear combination of received channels. The weightings of received channels comprising the optimized detection channel are estimated by maximizing the expected signal to noise ratio for a target of interest at a specified maximum clearance depth. Finally, we consider delineation of regions in dynamic detection data where classification cannot be applied. We show that a singular value analysis of the received data can be used to filter out isolated anomalies and reduce the size of areas designated for "mag and dig" operations.

(2) *Performance prediction*. We first develop efficient Monte Carlo (MC) methods for predicting the variability of estimated polarizabilities under site-specific conditions (e.g. sensor noise, target

density, etc.). This approach provides a rigorous means to simulate the probability of correct classification of specified UXO and non-UXO.

Once EMI sensor data have been collected and inverted, initial predictions of classification performance can be updated with site-specific information. We develop a number of data and model quality metrics to assess the overall difficulty of the classification task. These include: mean polarizability misfit with respect to library items, signal to noise ratio, and a metric that uses the point-to-point variability of soundings or polarizabilities to determine the number of channels that can be used for classification. We combine these metrics into a "Dataset Degree of Difficulty" (DDD) that categorizes the classification difficulty at the site. This approach gives the data analyst an objective measure with which to assess the feasibility of classification at a site using available information. However, retrospective analysis of cued MetalMapper data sets indicates that the DDD is predictive of average classification performance and cannot reliably predict false alarm rates at the point where all TOI are identified. We address this difficulty with a regularized linear regression algorithm that directly learns the relationship between performance metrics and the observed false alarm rate. Finally, we extend the regression model to generate predictions at an early stage of a project (prior to data collection) by imputing missing metrics with values drawn from past sites.

(3) *Risk assessment*. A random compliance sampling approach has been suggested for UXO risk assessment, and we extend this approach to account for the bias in prioritized digging, thereby reducing the number of excavations required to test for outlying UXO. We then discuss and compare methods for identification of outliers to the distribution of UXO via generative models of the receiver operating characteristic (ROC). Next, we consider how seeded items emplaced for quality control can be used to increase confidence in the classification process, and we model this process by constraining the ROC model. Finally, we briefly turn to the problem of identifying novel, or unique, UXO with prioritized validation digs. We propose a metric that combines features of the geophysical model estimated for each detected target to identify novel UXO. The metric requires no prior information about the UXO present at a site.

Benefits. The methods developed under this project will aid managers in designing a cost-effective remediation effort prior to deployment and in adjusting and optimizing survey design and data processing as more information becomes available. These tools will help users and stakeholders understand the potential benefits and limitations of advanced classification.

Contents

Abstract i
List of Figures iv
List of Tables vi
List of Acronyms vi
Acknowledgements vii
1. Introduction 1
2. Technical background 1
2.1. Data acquisition 1
2.2. Feature extraction 2
2.3. Classification 5
3. Detection with multistatic sensors 7
3.1. TEMTADS2x2 detection analysis 8
3.2. MetalMapper detection analysis: effect of target azimuth and polarizability ratio 12
3.3. Optimizing the detection channel 16
3.4. Detection modeller software 18
3.5. Delineating "mag and dig" regions in dynamic detection data 20
3.5.1. Target density estimation 21
3.5.2. Singular value decomposition of dynamic data 21
3.5.3. Multi-object inversion of dynamic data 26
4. Performance prediction 32
4.1. Dataset degree of difficulty 32
4.2. Regularized regression for performance prediction 36
4.3. *A priori* performance prediction 39
5. Risk assessment 41
5.1. Compliance sampling 41
5.2. Biased compliance sampling 43
5.3. Generative models for risk assessment 47
5.3.1. Estimating probabilities with the binormal model 49
5.4. Comparison of risk assessment methods 50
5.5. Risk assessment with seed items 51
5.6. Identifying novel TOI 55
5.6.1. Matching with a comprehensive library 55
5.6.2. Model metrics 56
5.7. Software for risk assessment 58
5.8. Conclusions: risk assessment 59
6. Conclusions and Implications for Future Research/Implementation 61
References 62
Appendix A. Project publications and presentations 64
Appendix B. Uncertainty analysis and performance prediction for unexploded ordnance classification 65
Appendix C. Retrospective analyses of cued MetalMapper data sets 81
Appendix D. Detecion modeller user manual 107

List of Figures

1 Electromagnetic induction (EMI) survey 2
2 Sequential inversion approach 4
3 Computing the ROC from the score distributions of TOI and non-TOI 6
4 Comparison of EM-61 and TEMTADS2x2 detection maps at Camp Ellis. 7
5 TEMTADS2x2 detection profile 9
6 Effect of TEMTADS2x2 orientation on detection profile 10
7 TEMTADS2x2 detection threshold as a function of line-spacing 11
8 Predicted clearance depths for selected TOI 11
9 Dependence of MetalMapper along track profile on target azimuth 13
10 Dependence of target azimuth producing a minimal z-component response on target location 14
11 Dependence of MetalMapper detection profile on polarizability ratio 15
12 Optimal channel weights for target detection with the MetalMapper 17
13 Detection modeller software display 19
14 Manual mag and dig region for Camp Hale TEMTADS2x2 data 20
15 Application of pixel-based region delineation algorithm to Camp Hale TEMTADS2x2 data 22
16 Dipole filter applied to synthetic multi-object scenarios in dynamic TEMTADS2x2 data 23
17 SVD applied to synthetic multi-object scenarios in TEMTADS2x2 data 25
18 Effect of moving sensor on SVD analysis for synthetic three-object scenarios 26
19 Comparison of singular value and dipole filter analysis of Camp Hale dynamic TEMTADS2x2 data 28
20 High density regions delineated by thresholding on images of monostatic data, SV4 and dipole filter 29
21 Multi-object inversion of multistatic dynamic detection data 30
22 Comparison of high density regions delineated by threshold on images of monostatic data, SV4 and multi-object inversion residuals 31
23 False alarm rate and Area under the Curve vs. Dataset Degree of Difficulty (DDD) for classification of ESTCP MetalMapper datasets 35
24 Accounting for spatial target density in performance predictions 36
25 Elastic net regression of false alarm rate for classification of MetalMapper data 38
26 Assumed size/decay features for *a priori* performance prediction 40
27 *A priori* performance predictions 40
28 Receiver operating characteristic generated by classification of MetalMapper data collected at Camp Beale, CA 42
29 Polarizabilities and representative photos of targets of interest at Camp Beale 42
30 Illustration of unbiased and biased compliance sampling 44
31 Area under the receiver operating characteristic curve versus dataset degree of difficulty for ESTCP MetalMapper data sets 45
32 Comparison of power as a function of sample size for unbiased and biased compliance sampling 46

33 Application of biased compliance sampling to three ROC curves from classification of cued MetalMapper data 46
34 Fitting the ROC with a binormal model 48
35 Binormal model fits to example ROC curves 48
36 Binormal model fits to example ROC curves showing low correlation coefficients 49
37 Probability that i^{th} dig is a TOI, for classification of Pole Mountain MetalMapper data 51
38 Comparison of risk assessment methods applied to Camp Beale classification diglist 52
39 ROC curves for ordnance classes at Camp Beale 53
40 Unconstrained and constrained binormal fits for classification of ISO seed items at Camp Beale 54
41 Confidence that all TOI in each class have been found at the point in the Camp Beale dig list where all ISO seeds are found 55
42 Number of matches to each class of item in MetalMapper polarizability library 56
43 ROC for τ metric applied to Camp Beale MetalMapper data 59

List of Tables

1 EM sensors used for UXO classification 3
2 Metrics for identifying novel targets of interest 57
3 Summary statistics for retrospective analyses of cued MetalMapper data sets 81

List of Acronyms

APP	Apparent Principal Polarizabilities
AUC	Area Under the ROC Curve
BTG	Black Tusk Geophysics
CCR	Combined Classifier Ranking
cm	Centimeter
DDD	Dataset Degree of Difficulty
DSS	Decision Support System
EM	Electromagnetic
EMI	Electromagnetic induction
ESTCP	Environmental Security Technology Certification Program
FAR	False Alarm Rate
FOM	Figure of Merit
FPF	False Positive Fraction
ISO	Industry Standard Object
m	Meter
MC	Monte Carlo
mm	Millimeter
MMR	Massachusetts Military Reservation
MPV	Man-Portable Vector Sensor
MRM	Multi-static Response Matrix
MSNR	Model Signal to Noise Ratio
ms	Millisececond
mV	MilliVolt
QC	Quality Control
ROC	Receiver Operating Characteristic
SERDP	Strategic Environmental Research and Development Program
SNR	Signal to Noise Ratio
SWPG	Southwest Proving Ground
SVD	Singular Value Decomposition
SV4	Fourth Singular Value
TEM	Time-domain electromagnetic
TEMTADS	Time Domain Electromagnetic Towed Array Detection System
TOI	Target of Interest
TPF	True Positive Fraction
UXO	Unexploded Ordnance
USACE	United States Army Corps of Engineers

Acknowledgements

This work was funded by the Strategic Environmental Research and Development Program (SERDP).

1. Introduction

As advanced UXO classification is transitioned from the research community to industry, there is an increasing need to develop practical tools to aid in the application of this technology to live-site munitions response projects. This project has aimed to anticipate the requirements of industry by developing algorithms and software to aid in critical decisions that must be made by geophysical data analysts and project managers during the course of a munitions response project.

This report summarizes progress towards these goals and is organized as follows. In section 2 we provide a brief background on advanced classification with electromagnetic sensors. We then examine detection with multistatic sensors in section 3, including a discussion of detection threshold analysis and optimal channel selection. In section 4 we describe work on performance prediction. We investigate how metrics of data and model quality can be used to assess classification difficulty and to directly predict false alarm rates. Finally, in section 5 we develop and compare methods for risk assessment. Specifically, we consider how an objective confidence can be used to determine a final stop dig point, and how seed items can be quantitatively used to increase confidence that all TOI have been found.

2. Technical background

Advanced classification of buried munitions requires a number of steps:

(1) Data acquisition: detection of buried targets with a geophysical sensor.
(2) Feature extraction: characterization of each target with features estimated through inversion with a parameterized physics-based forward model.
(3) Classification: prioritization of targets for digging using estimated features.

The ultimate goal of this processing is to identify all targets of interest with a minimal number of false alarms.

2.1. **Data acquisition.** In the data acquisition stage, a geophysical sensor is deployed at the site. Time-domain electromagnetic (TEM) sensors are most commonly used for detection of buried metallic targets. These instruments actively transmit a time-varying primary magnetic field which illuminates the earth. The variation of the primary field induces currents in buried targets and these currents in turn produce a secondary field which can be measured by a receiver at the surface (figure 1)

In a detection mode survey, a TEM sensor passes over an area in nominally straight, parallel lines, with line spacing and instrument height dictated by instrument geometry and detection considerations. Subsequent cued interrogations may revisit previously-identified targets and acquire high signal-to-noise ratio (SNR) data in a small area about the target. Recently developed systems for cued-interrogation illuminate the target with multiple transmitters and receivers (a multistatic configuration) from a single observation location and thereby avoid the requirement for accurate positioning of moving sensors. Table 1 compares the industry-standard EM-61 sensor with newer multistatic systems developed specifically for UXO detection and classification.

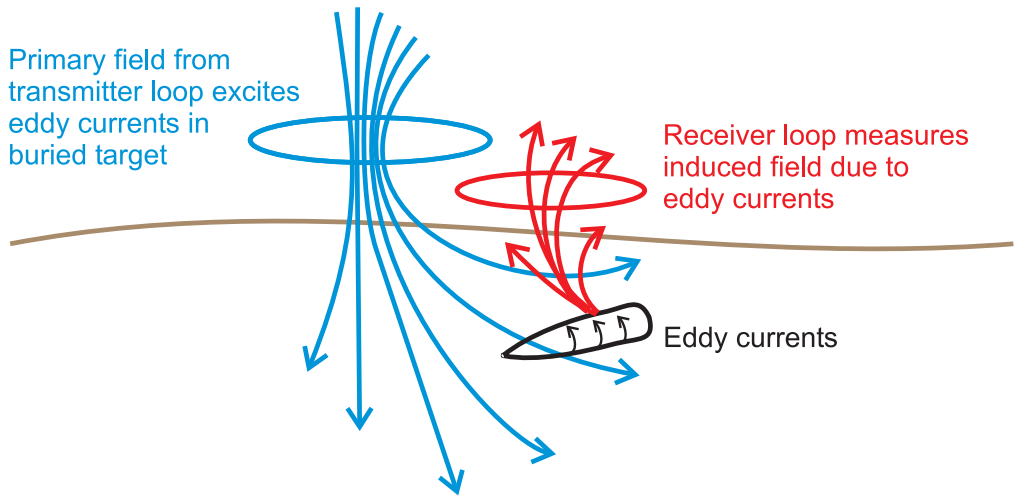


FIGURE 1. Electromagnetic induction (EMI) survey. Eddy currents are induced in a buried target by a time-varying primary field. Decaying secondary fields radiated by the target are then measured by a receiver at the surface.

2.2. **Feature extraction.** Once target anomalies have been identified in the observed geophysical data, we can characterize each anomaly by estimating features which will subsequently allow a classification algorithm to discern targets of interest (TOI) from non-hazardous clutter (non-TOI). These features may be directly related to the observed data (e.g. anomaly amplitude at the first time channel), or they may be the parameters of a physical model. Advanced classification relies upon physics-based modeling, with the observed magnetic field $\mathbf{B}(t)$ radiated by a buried target usually represented as a time-varying dipole

$$\frac{\partial \mathbf{B_s}}{\partial t}(\mathbf{r},t) = \frac{p(t)}{r^3}\left(3(\hat{\mathbf{p}}(t)\cdot\hat{\mathbf{r}})\hat{\mathbf{r}} - \hat{\mathbf{p}}(t)\right) \tag{1}$$

with $\mathbf{r} = r\hat{\mathbf{r}}$ the separation between target and observation location, and $\mathbf{p}(t) = p(t)\hat{\mathbf{p}}(t)$ a time-varying dipole moment

$$\mathbf{p}(t) = \frac{1}{\mu_o}\mathbf{P}(t)\cdot\mathbf{B_o}. \tag{2}$$

The induced dipole is the projection of the primary field $\mathbf{B_o}$ onto the target's polarizability tensor $\mathbf{P}(t)$ (Bell et al., 2001). Here the elements of the polarizability tensor ($P_{ij}(t)$) represent the convolution of the target's B-field impulse response ($\tilde{\mathbf{P}}(t)$) with the transmitter waveform $i(t)$ (Wait, 1982)

$$P_{ij}(t) = \frac{\partial}{\partial t}\int_{-\infty}^{\infty} \tilde{P}_{ij}(t'-t)i(t')dt'. \tag{3}$$

The polarizability tensor is assumed to be symmetric and positive definite and so can be decomposed as

$$\mathbf{P}(t) = \mathbf{A}^T\mathbf{L}(t)\mathbf{A} \tag{4}$$

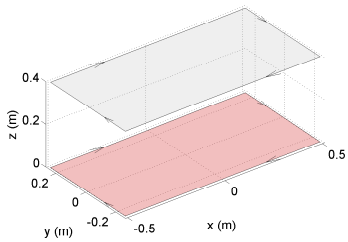
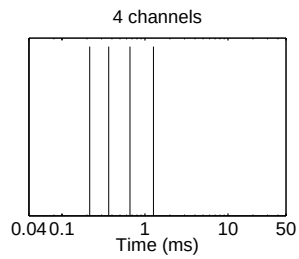
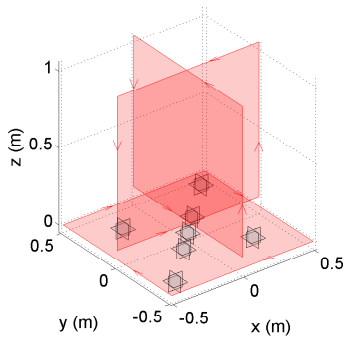
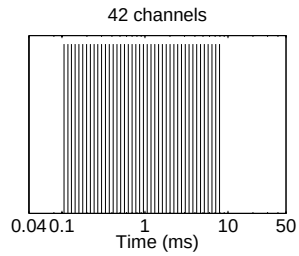
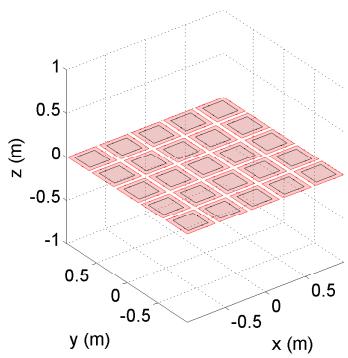
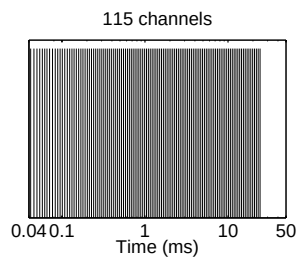
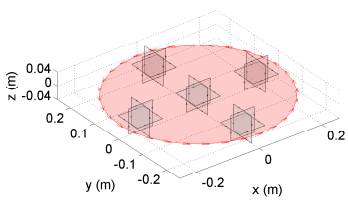
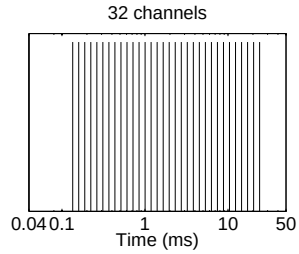
Sensor	Geometry	Channels
EM61 MKII		
MetalMapper		
TEMTADS		
MPV		

TABLE 1. Electromagnetic sensors used for UXO classification. In geometry plots transmitters and receivers are red and black, respectively. Time gates are representative of typical settings for a cued survey mode.

with $\mathbf{A}$ an orthogonal matrix which rotates the coordinate system from geographic coordinates to a local, body centered coordinate system. The diagonal eigenvalue matrix $\mathbf{L}(t)$ contains the principal polarizabilities $L_i(t)$ $(i = 1, 2, 3)$, which are assumed to be independent of target orientation and location.

From a set of observations of the electromagnetic field, the inverse problem is then to find the set of model parameters (location, orientation, and polarizabilities) that best fits the data. The model vector $\mathbf{m}$ can be estimated by minimizing a norm (e.g. least squares) quantifying the misfit between observed ($\mathbf{d}^{obs}$) and predicted ($\mathbf{d}^{pred}$) data. For the TEM dipole model, the least squares estimate must generally be obtained iteratively, owing to the nonlinear relationship between model parameters and predicted data in equation 1. However, if the location of the target ($\mathbf{r}$) is assumed known, then the forward modeling becomes linear, so that

$$\mathbf{d}^{pred} = \mathbf{G}\mathbf{m} \tag{5}$$

with $\mathbf{G}$, the forward modeling matrix, implicitly dependent on target location. The least squares model estimate is then given by

$$\begin{aligned} \hat{\mathbf{m}} &= (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}^{obs} \\ &= \mathbf{G}^\dagger\mathbf{d}^{obs} \end{aligned} \tag{6}$$

with

$$\mathbf{G}^\dagger = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T \tag{7}$$

denoting the pseudo-inverse. In this formulation, the model vector at location $\mathbf{r}$ is parameterized in terms of the six unique elements of the polarizability tensor $\mathbf{P}$

$$\mathbf{m} = [P_{xx},\ P_{xy},\ P_{xz},\ P_{yy},\ P_{yz},\ P_{zz}]^T\,. \tag{8}$$

In practice, the vector $\mathbf{m}$ is estimated at each time channel in a sequential, or two stage, inversion strategy (Song et al., 2011). As illustrated in figure 2, we first solve a nonlinear inverse problem for target location. We then solve a linear problem for the polarizability tensor elements (equation 8) at our fixed location estimate. Decoupling the location and polarizability parameters in this way

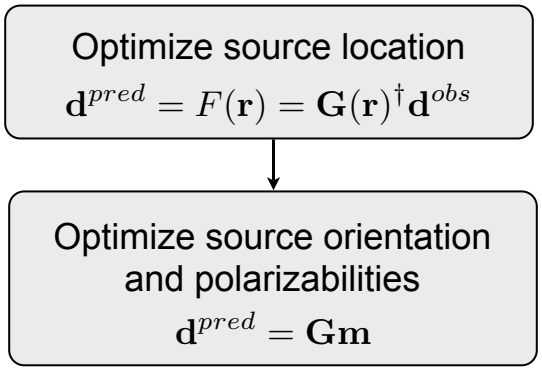


FIGURE 2. Sequential inversion approach for estimation of dipole model parameters. We first estimate target location $\mathbf{r}$; the predicted data in this case are a nonlinear functional $F(\mathbf{r})$. We then estimate target orientation and polarizabilities. At a fixed location and orientation, the predicted data are related to the model via a linear forward operator $\mathbf{G}$.

allows for efficient parallel solution of the linear problem at all time channels. Target orientation

and principal polarizabilities are subsequently estimated via joint diagonalization (Cardoso, 1996). This algorithm returns a single eigenvector matrix $\mathbf{A}$ for all channels, corresponding to a fixed target orientation. The eigenvalues at each time channel are then an estimate of principal polarizabilities.

2.3. **Classification.** To rank detected targets for digging, we use the information in our observed geophysical data. Features of the observed data, estimated without resorting to inversion with a physics-based model, can sometimes suffice as criteria to discriminate between ordnance and non-ordnance targets (e.g. Williams et al. (2007)). However, because dipole model parameters can be related to intrinsic properties such as target size and shape, features derived from the estimated parameters are often more reliable for discriminating between TOI and non-TOI.

Classification with TEM data is most often performed by comparing estimated polarizabilities with library responses and then ranking a target based on some measure of closeness between observed and expected responses. Care must be taken here to use parameters which can be reliably estimated: late time polarizabilities are more susceptible to noise and poor estimates may unduly affect the discrimination decision. Pasion et al. (2007) solve this problem with a fingerprinting algorithm that inverts for target location and orientation while holding principal polarizabilities fixed at their library values. Reducing the model's degrees of freedom in this way makes the inversion less susceptible to fitting the noise. Targets are then dug based upon the proposed library item which produces the best fit to the observed data.

The output of any automated classification algorithm is a decision statistic (y), or score, that is used to rank detected targets from likely TOI to likely non-TOI. For example, a library classifier uses the misfit of estimated polarizabilities with library polarizabilities as the decision statistic. As shown in Figure 3, the receiver operating characteristic (ROC) is then a plot of the true positive fraction (TPF) versus the false positive fraction (FPF), which are defined as the cumulative score distributions for TOI and non-TOI

$$
\begin{aligned}
TPF(x) &= \int^{x} p(y|TOI)dy \\
FPF(x) &= \int^{x} p(y|non-TOI)dy.
\end{aligned}
\tag{9}
$$

In the context of munitions response projects, the false alarm rate (FAR) at which all ordnance are detected on the ROC (i.e. $TPF = 1$) is the crucial metric by which site managers evaluate the efficacy of remediation efforts. An advanced technique that results in good initial detection of TOI but fails to find outlying TOI until late in the dig order will be judged unsuccessful.

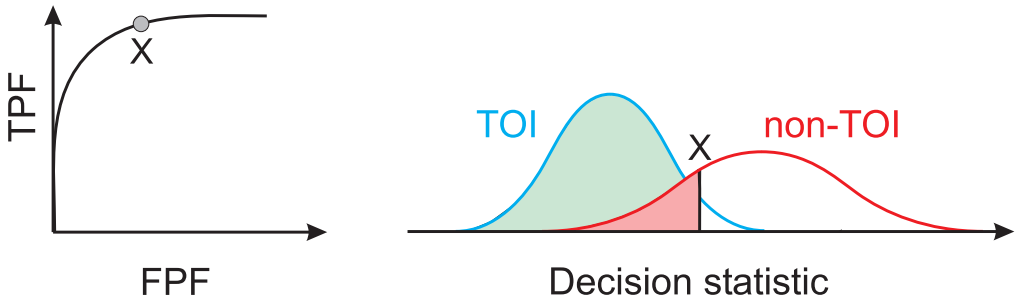


FIGURE 3. Left: the receiver operating characteristic (ROC) shows the true positive fraction (TPF) as a function of the false positive fraction (FPF). Right: the ROC at point x can be modeled as the integral of the distributions of TOI and non-TOI (true and false positives, respectively) with respect to the decision statistic.

3. Detection with multistatic sensors

Detection with multistatic sensors has obvious benefits compared to conventional EM-61 arrays, including:

(1) *Improved resolution.* The smaller area of cube receivers translates to less smearing of detected anomalies and improves isolation of discrete targets in highly cluttered areas. This is illustrated in figure 4, which compares detection surveys at Camp Ellis.

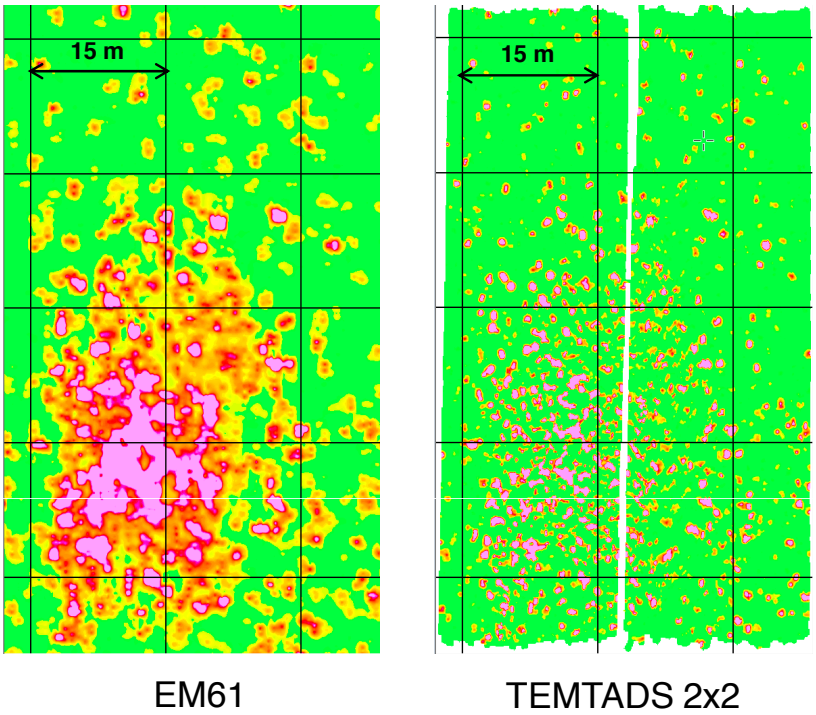


Figure 4. Comparison of EM-61 and TEMTADS2x2 detection maps at Camp Ellis.

(2) *Longer off-time.* Extending off-time measurements beyond the last EM-61 channel (2.1 ms) may reduce total anomaly counts for cued interrogation by eliminating fast-decaying clutter.

(3) *Immunity to soil response.* At sites where there are viscous magnetic soils, measurements orthogonal to a halfspace soil response can be used to identify target anomalies that would otherwise be undetectable. multistatic soil compensation techniques have recently been applied to MPV and MetalMapper datasets acquired at Waikoloa, HI.

(4) *One-pass detection and classification.* It is often possible to reliably classify high SNR targets detected in a dynamic, multistatic survey. This reduces the number of targets that must be revisited for cued interrogation.

As advanced classification becomes mandatory for munitions response projects, target detection will increasingly be done with multistatic sensors. Bell and Barrow (2014) have developed a dipole filter that fits dynamic detection data with dipole sources. The correlation of observed and predicted data at any location can then be used as a criterion to pick targets. The dipole filter simplifies target picking by converting a threshold on observed data to a threshold on the amplitude of recovered

sources. An appropriate dipole filter source threshold can be determined by Monte Carlo simulations that characterize the dependence of estimated polarizabilities on target location, orientation and sensor noise.

Similarly, a dipole fitting approach can be used to predict equivalent monostatic data for dynamic MetalMapper data. This produces intuitive detection maps similar to figure 4 where isolated peaks in the gridded image correspond to discrete targets. Finally, Shubitidze et al. (2016) develop a dipole clustering approach that clusters target locations estimated by inversion of individual dynamic soundings acquired with multistatic sensors.

These advanced target selection algorithms simplify the target picking process, but independent quality control of the resulting picks is still required. Targets picked directly off the observed data can be used for this QC step. When targets are picked directly off of observed multistatic data, a minimum anomaly amplitude must be calculated for expected targets of interest. If picking off a gridded image, the threshold calculation should consider the target response over all receivers used in the gridding. Alternatively, if targets are initially picked using the data acquired along a line by individual receivers, separate thresholds can be determined for each receiver.

To facilitate QC target picking, we have extended modelling of worst case detection thresholds to multistatic, multicomponent EMI arrays. In the remainder of this section we discuss modelling of detection threholds for TEMTADS2x2 and MetalMapper arrays. We then present a method for selecting a detection channel by maximizing worst case SNR. Beran and Billings (2016) provides more detail on threshold modeling and channel selection. We conclude this section with an overview of the Detection Modeler software developed under this project for threshold modeling.

3.1. **TEMTADS2x2 detection analysis.** To illustrate some of the considerations that go into modelling detection thresholds with multistatic sensors, we consider detection of 37 mm projectiles with the TEMTADS 2x2 for a 30 cm clearance depth. Figure 5 shows the resulting detection profile for the TEMTADS2x2 sensor. As with the EM-61, the minimum amplitude response corresponds to a horizontally-oriented target. In this orientation, the primary field is minimially coupled with the target's primary polarizability. For the 2x2, the minimum amplitude response within the sensor footprint is directly between the transmitters (i.e. cross-track location $x = 0$).

In Billings and Beran (2016), we consider how arrays of EM-61 coils can be assembled to optimize detection criteria. Analogous to the TEMTADS2x2, we found that a detection array with coils arranged linearly cross-track (side-by-side) produces a minimum amplitude response for a target directly between two receiver coils. An obvious approach to increasing the detection threshold is to increase the cross-track density of receiver coils. For example, a pyramid geometry with a leading central coil and two trailing coils significantly raises the detection threshold relative to a three coil linear array.

These considerations can similarly be applied to the optimization of multistatic sensors for detection. For example, we might envision a linear array of receiver cubes arranged cross-track to maximize the detection threshold. Alternatively, a rotation of the existing 2x2 geometry can increase the detection threshold by increasing the cross-track receiver density (figure 6). This option

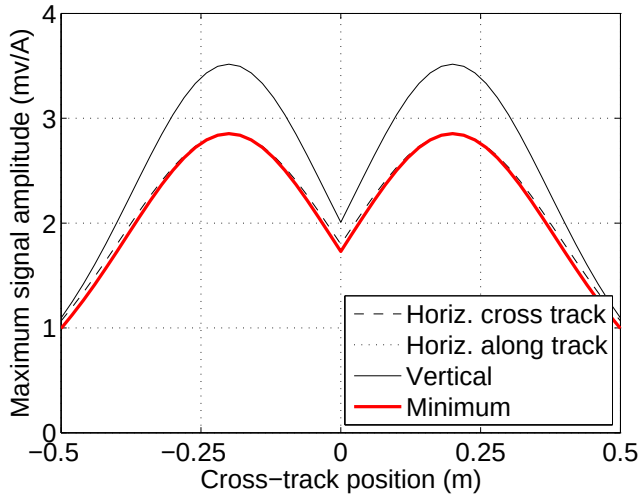


FIGURE 5. TEMTADS2x2 detection profile showing maximum signal amplitude at channel 1 (0.137 ms) over all vertical component receivers, as a function of cross-track target location. Profiles are generated for target in horizontal along-track, horizontal cross-track, and vertical orientations. The minimum over all target orientations defines the detection threshold as a function of cross-track location.

would require a redesign of the 2x2 sensor cart, but still allows for a system geometry be used for both dynamic and static surveys.

Returning to detection with an unrotated sensor (figure 5), we can use the detection profile to determine the detection threshold for a specified line spacing, as shown in figure 7. For a line spacing $\mathcal{L}$, the maximum cross-track location of a detected target is $\mathcal{L}/2$. The minimum amplitude in the detection profile for cross-track locations satisfying $|x| \leq \mathcal{L}/2$ then defines the detection threshold for line spacing $\mathcal{L}$.

As illustrated in figure 7 for the case of the TEMTADS2x2, any $\mathcal{L}$ that is less than the width of the sensor produces a detection threshold determined by the response minimum directly between the transmitters. As the line spacing is increased beyond the width of the sensor array, detected targets may fall outside of the sensor footprint and the detection threshold must be lowered. A sensible guideline for dynamic surveys with this sensor is to restrict line spacing to less than the width of the sensor (0.8 m). Accounting for the inevitable jitter in sensor track during dynamic surveys, a line spacing of 0.7 m or less is advisable.

With a specified detection threshold, we can also calculate the maximum clearance depth for other TOI that may be encountered at a site. We intuitively expect that larger (smaller) TOI will be cleared to deeper (shallower) depths. This is confirmed in figure 8, which shows a clearance depth analysis for selected TOI using the threshold defined for 37 mm projectiles. In fact, it is the amplitude of the secondary polarizabilities that determines the clearance depth for a TOI, since the minimum amplitude response is determined by excitation of the transverse target response. For example, the small ISO in figure 8 has almost identical secondary polarizability amplitudes to a 37

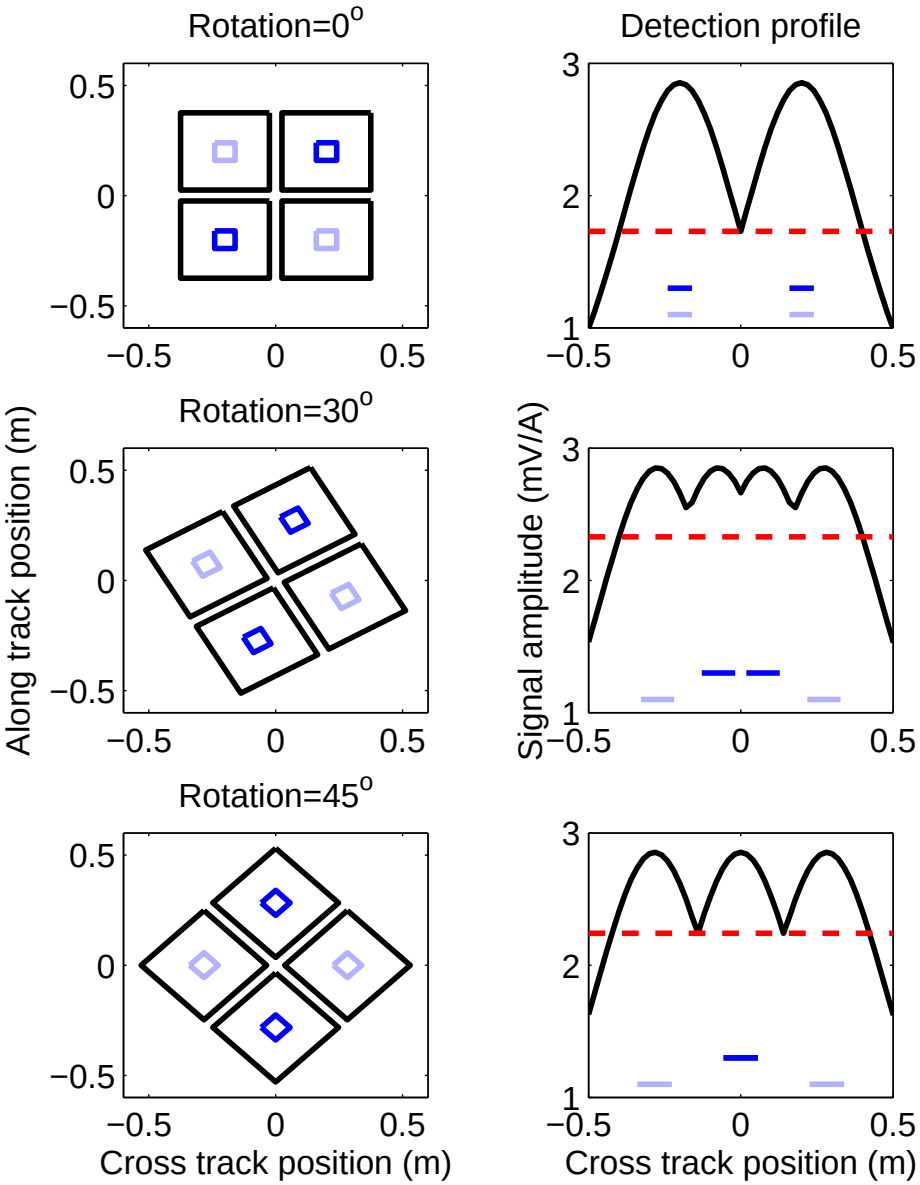


FIGURE 6. Effect of TEMTADS2x2 orientation on detection profile. Left column shows orientation of 2x2 array relative to cross-track direction. Right column shows the resulting detection profile for a 37 mm projectile at 30 cm depth. Blue lines indicate the cross-track positions of receiver cubes. Dashed red lines show the minimum signal amplitude within the footprint of the array.

mm projectile at early times. Consequently, the predicted clearance depth for this target is very close to the specified 30 cm depth for a 37 mm projectile.

Because the detection threshold within the 2x2 sensor footprint is constant (see figure 7), the predicted clearance depths are also constant for line spacings less than the sensor width. Again, we recommend that line spacing be restricted to less than 0.8 m - in which case the clearance depth for each item is a constant.

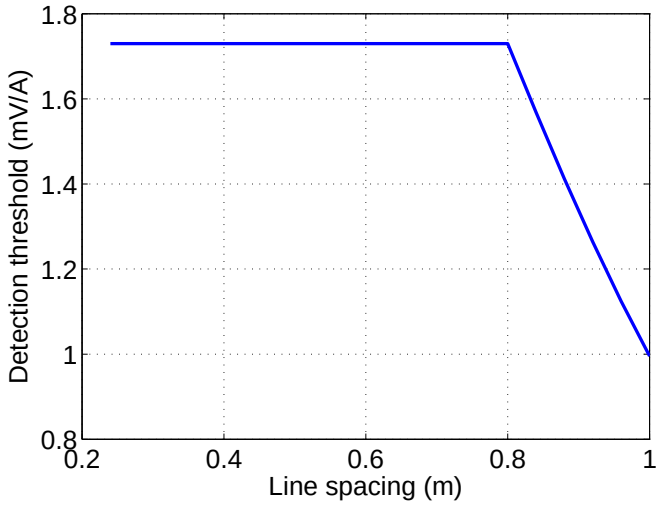


FIGURE 7. TEMTADS2x2 channel 1 (0.137 ms) detection threshold for a 37 mm projectile at 30 cm depth, as a function of line-spacing.

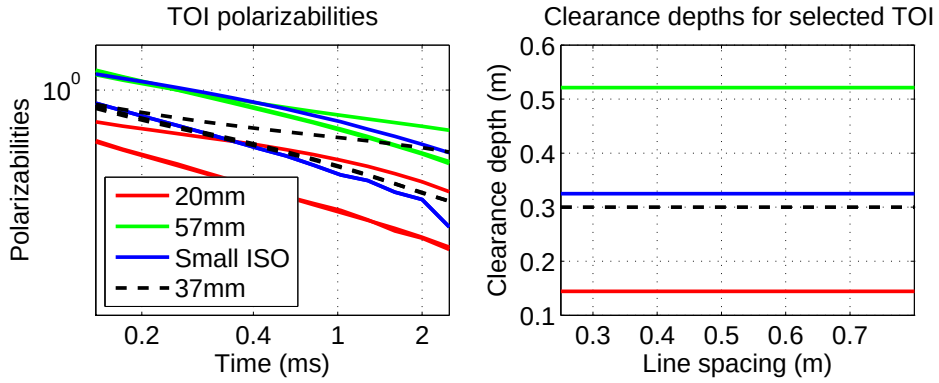


FIGURE 8. Predicted TEMTADS2x2 clearance depths for selected TOI. Left plot show selected TOI polarizabilities for TEMTADS2x2 sensor. Predicted clearance depths (right plot) depend upon amplitude of polarizabilities relative to the TOI used to set the detection threshold. For this example, the threshold is defined for 37 mm projectile (dashed line).

3.2. **MetalMapper detection analysis: effect of target azimuth and polarizability ratio.** Threshold analysis for the TEMTADS2x2 is relatively simple: for line spacings less than the width of the sensor the minimum response is defined by a horizontal target positioned under the middle of the array.

In contrast, the geometry of the MetalMapper complicates detection analysis for this sensor. The offset of receivers from the center of the transmitter produces minima that do not correspond to canonical (cross-track, along-track, or vertical) orientations of the target. This effect is illustrated in Figure 9 which shows the dependence of the along-track data measured in a single cube on the azimuth of a horizontal target. In the case where we pass directly over the target (9a), the maximum of the along track profile does not depend on the azimuth of the target. This is because the maximum occurs when the target is directly under the center cube. At this location, the primary field is entirely vertical and so there is no excitation of the primary polarizability for a target in a horizontal orientation, regardless of target azimuth. However, when the target is offset in the cross-track direction, the maximum of the along track profile exhibits a strong dependence on target azimuth (9b)

In Billings and Beran (2016), we differentiate the dipolar scattered magnetic field with respect to the target azimuth to derive an expression for the target azimuth (θ_{min}) that produces the minimum amplitude vertical field in a receiver loop

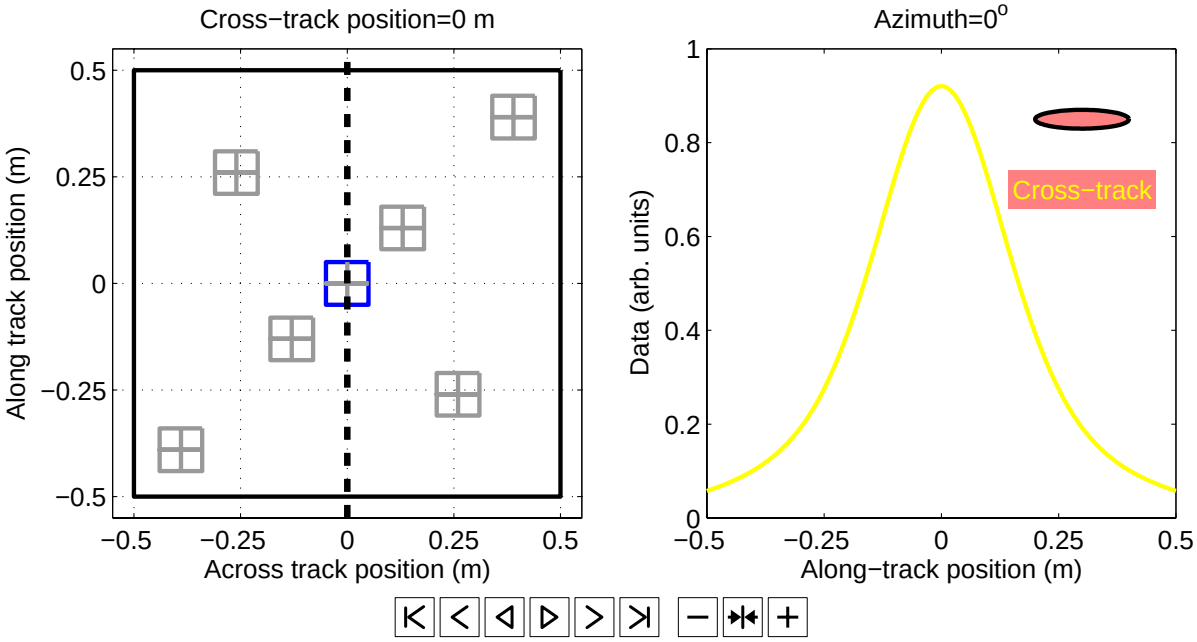
$$\tan(2\theta_{min}) = \frac{B_x\tilde{r}_y + B_y\tilde{r}_x}{B_x\tilde{r}_x - B_y\tilde{r}_y}, \tag{10}$$

with $[\tilde{r}_x, \tilde{r}_y]$ functions of the position vector from target to receiver loop, and $[B_x, B_y]$ the horizontal components of the primary field at the target. In the limiting case of a point measurement of the dipolar secondary field an identical expression is obtained, with $\tilde{\mathbf{r}}$ replaced by the vector between source and measurement location. The angle θ_{min} notably has no dependence on the target polarizabilities. The above expression in fact defines an extremum of the vertical field, taking a second derivative yields the following condition satisfied at a minimum:

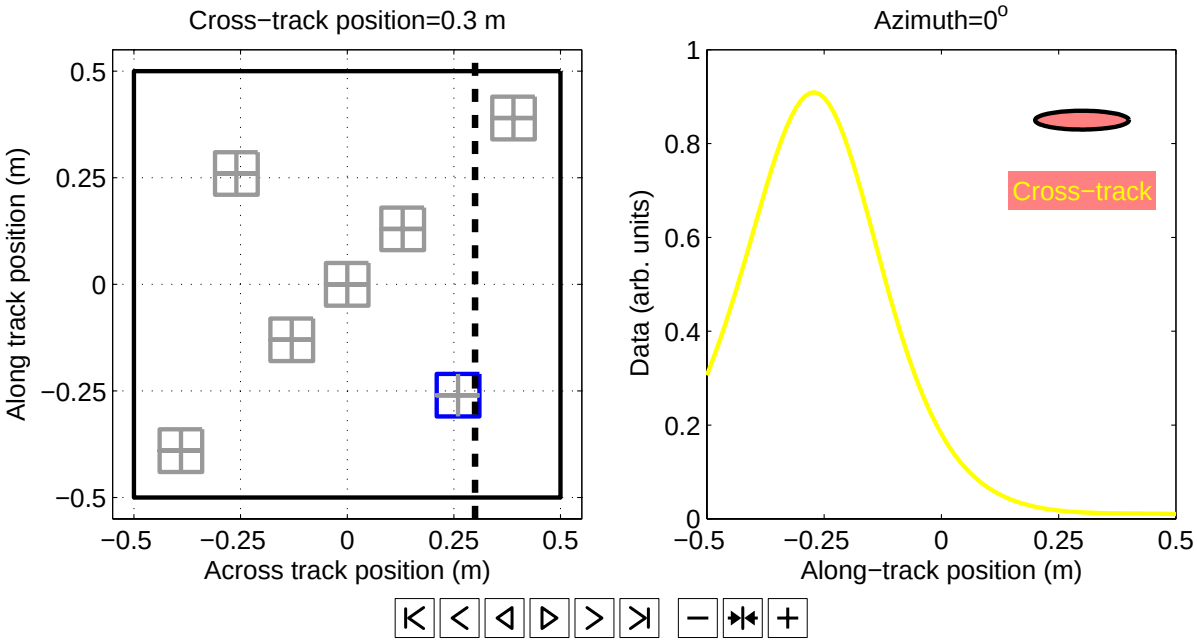
$$(B_x\tilde{r}_x - B_y\tilde{r}_y)\cos(2\theta_{min}) + (B_y\tilde{r}_x + B_x\tilde{r}_y)\sin(2\theta_{min}) < 0. \tag{11}$$

Testing this condition at θ_{min} and $\theta_{min} + \pi/2$ identifies the azimuth producing a minimum of the vertical magnetic field. Figure 10 shows the dependence of θ_{min} on target location relative to the MetalMapper. Discontinuities in this plot arise when the receiver closest to the target location changes. As described further in Billings and Beran (2016), the analytic expressions in equations 10 and 11 can be used for efficient evaluation of the azimuthal dependence of the fields in each receiver.

A second consideration for threshold analysis with the MetalMapper is the dependence of the detection profile on the ratio of primary and secondary polarizabilities. Figure 11 illustrates this effect by considering the detection profile for a single target at two different time channels. For the earlier time channel (11a), the profile is characterized by local maxima if the target is directly



(a) Cross-track location = 0 m



(b) Cross-track location = 0.3 m

FIGURE 9. Dependence of MetalMapper along track profile on target azimuth. Left: MetalMapper geometry, with dashed line indicating cross-track location of the target. Profiles on right show the vertical component data measured in the cube highlighted in blue. Right: Along track profile acquired in highlighted cube, as a function of target azimuth. **Animations must be viewed in Adobe Acrobat Reader**.

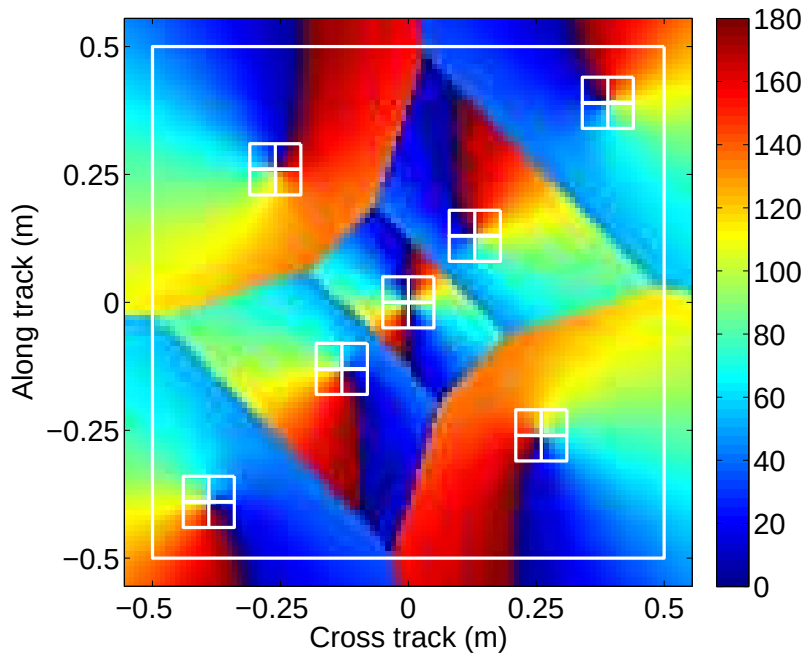
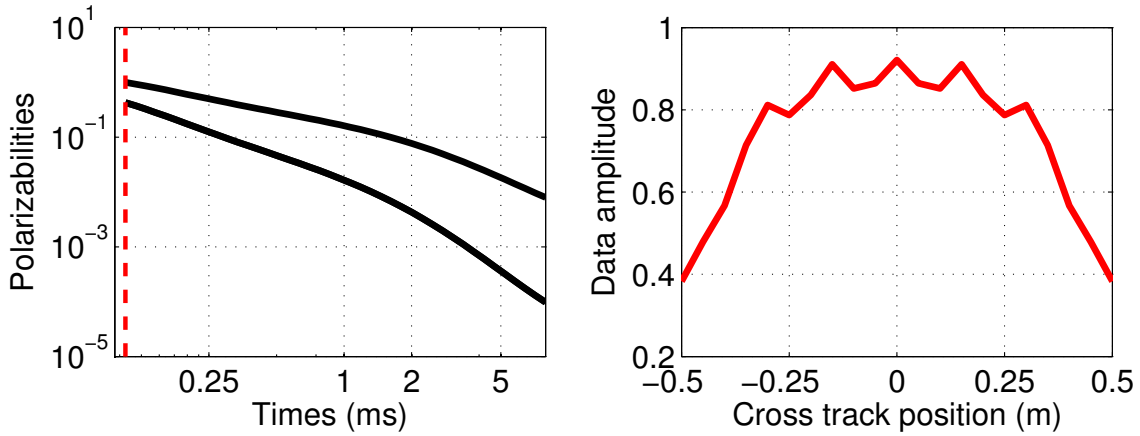


FIGURE 10. Dependence of target azimuth (in degrees) producing a minimal z-component response (θ_{min}) on target location relative to the MetalMapper. Discontinuities occur when the receiver closest to the target location changes. Color scale also introduces discontinuities for azimuths of 0° and 180°, which correspond to the same target orientation.

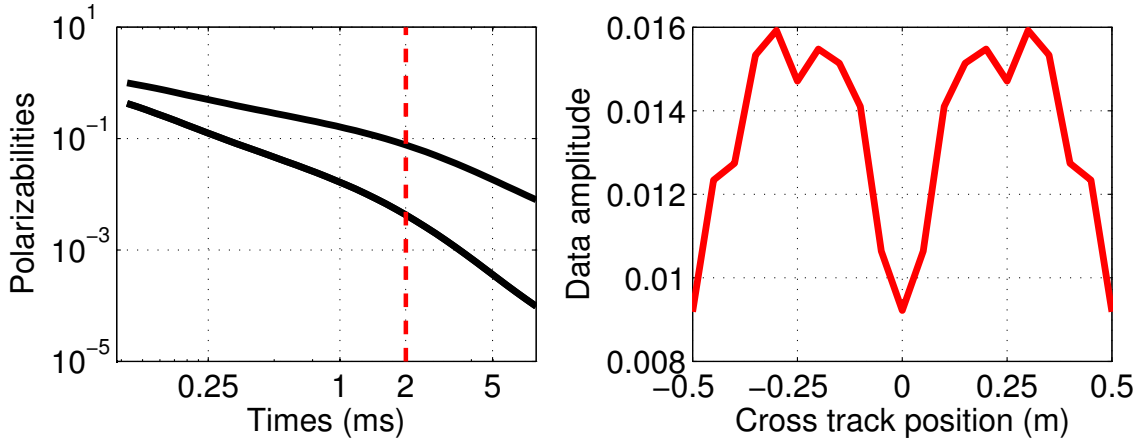
beneath the center three cubes. Outside of these cubes, the detection threshold decreases monotonically with increased cross-track offset. For a line spacing satisfying $\mathcal{L} \gtrsim 0.6$ m, the detection threshold in this case is defined by a target at the maximum cross-track displacement $\mathcal{L}/2$.

In 11b, we consider a later time channel where the polarizability ratio (L_2/L_1) is decreased relative to the earlier channel. This has a strong effect on the detection profile: there is now a minimum when the sensor passes directly over a target (cross-track position = 0 m). This minimum defines the detection threshold for $\mathcal{L} \lesssim 1$ m in this example.

The cross-track location of the minimum MetalMapper response can vary with time channel or, equivalently, with target aspect ratio. This is likely an important consideration when designing a dynamic detection survey with the MetalMapper, and in the next section we develop an automated approach for selecting a detection channel by optimizing the worst-case SNR.



(a) 0.1 ms



(b) 2 ms

FIGURE 11. Dependence of MetalMapper detection profile on polarizability ratio L_2/L_1at two time channels. Left: polarizabilities for a 2.36" rocket. Vertical dashed line indicates the time channel at which the detection profile is calculated. Right: MetalMapper detection profile.

3.3. **Optimizing the detection channel.** Selection of a single time channel for target detection is generally guided by intuition: a later time (e.g > 2 ms) with sufficient SNR is suitable for detection of larger ordnance and can reduce the number of picked targets by attenuating fast-decaying anomalies. However, if small ordnance are present then an early time (< 1 ms) is advisable to ensure that signal is well above the noise floor. Given time-dependent noise standard deviations and detection thresholds, we can formalize this intuition by optimizing the SNR of a linear combination of channels. We adopt a "maximin"strategy (Berger, 2013): we seek to maximize SNR in the worst case scenario defined by the detection threshold. This approach is also presented in Beran and Billings (2016).

We define a detection channel as a weighted sum of the received data for a given field component measured at M time channels

$$d_{detect} = \sum_{i=1}^{M} w_i d_i \tag{12}$$

with d_i the detection threshold at the ith time channel, respectively. The weights w_i are estimated by maximizing the SNR of the detection channel

$$\mathrm{SNR} = \frac{d_{detect}^2}{\sigma_{detect}^2}. \tag{13}$$

Assuming independent (white) noise at each channel, the variance of the noise for the detection channel is

$$(\sigma_{detect})^2 = \sum_{i=1}^{M} w_i^2 \sigma_i^2, \tag{14}$$

and σ_i the noise standard deviation at the ith channel. In the more general case with correlated noise between time channels, the data covariance S propagates to the detection channel as

$$(\sigma^{detect})^2 = \mathrm{w}^T \mathrm{S} \mathrm{w}. \tag{15}$$

Maximizing equation 13 with respect to the weights, we obtain

$$w_i \propto \frac{d_i}{\sigma_i^2}. \tag{16}$$

Normalizing such that $\sum_{i=1}^{M} w_i = 1$ gives us the optimal weights that maximize the SNR. Figure 12 illustrates this procedure for selected MetalMapper polarizabilities and a synthetic noise model. As expected, slower decaying polarizabilities will result in more weight at later times.

At sites where there is ubiquitous clutter with consistent polarizabilities, we can also consider a linear combination of time channels that maximizes the signal to clutter ratio (SCR)

$$\mathrm{SCR} = \left(\frac{d_{detect}(\mathrm{TOI})}{d_{detect}(\mathrm{non-TOI})} \right)^2 \tag{17}$$

with $d_{detect}(TOI)$ and $d_{detect}(non-TOI)$ the weighted detection channels (equation 12) for TOI and non-TOI (clutter), respectively. Unlike equation 13, we cannot assume statistical independence of

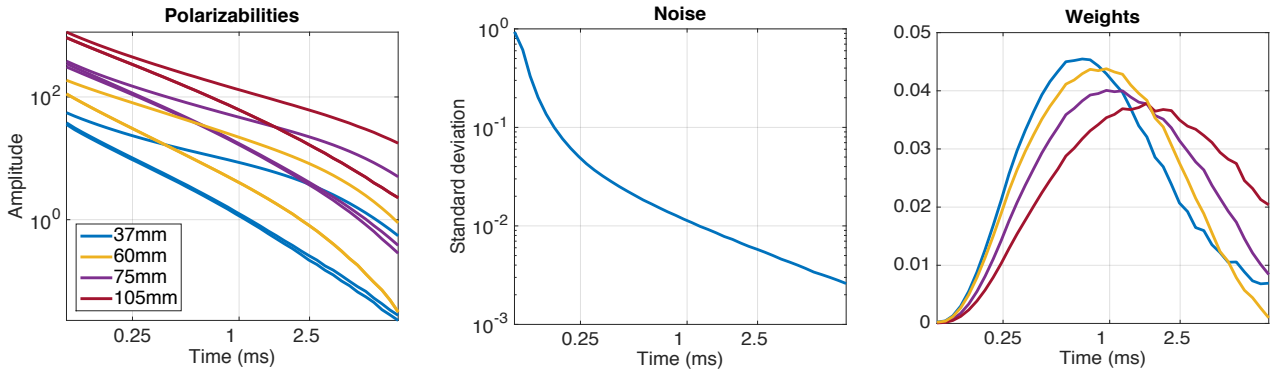


FIGURE 12. Optimal channel weights for target detection with the MetalMapper. Left: Polarizabilities for example TOI. Middle: Noise standard deviations for dynamic MetalMapper data. Right: Optimized channel weights.

the terms in the denominator. This produces a different condition for the optimal weights, with all weight given to the channel that maximizes the ratio of the TOI and non-TOI detection thresholds

$$w_i = \begin{cases} 1 & \text{if } i = \underset{i}{\operatorname{argmax}} \left(\frac{d_i(\text{TOI})}{d_i(\text{non-TOI})} \right) \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

An intuitive approach to jointly maximizing SNR and SCR is to then select the single channel which is given the most weight when maximizing the SNR (e.g. the peaks of the weight distributions in figure 12). This can reduce detections of fast-decaying clutter while still ensuring high SNR for detected TOI.

3.4. **Detection modeller software.** We have transitioned the threshold analysis developed in the previous section to a Windows compatible software application (figure 13). This tool can be used for threshold analysis with MetalMapper and TEMTADS2x2 sensors operating in either dynamic or cued modes. The analysis predicts the minimum amplitude response for a specified target of interest at a maximum clearance depth. Functionality for the reverse analysis is also available: given a detection threshold we can predict the maximum clearance depth for selected TOI in their worst case orientations. Additionally, noise estimates can be imported into the software in order optimize the detection channel, as described in the previous section. Polarizabilities can be imported directly from hdf5 format files in the DoD standardized polarizability library. The Detection Modeller software will be delivered to SERDP together with this final project report. The user manual is included in the software installation and is also attached in Appendix D.

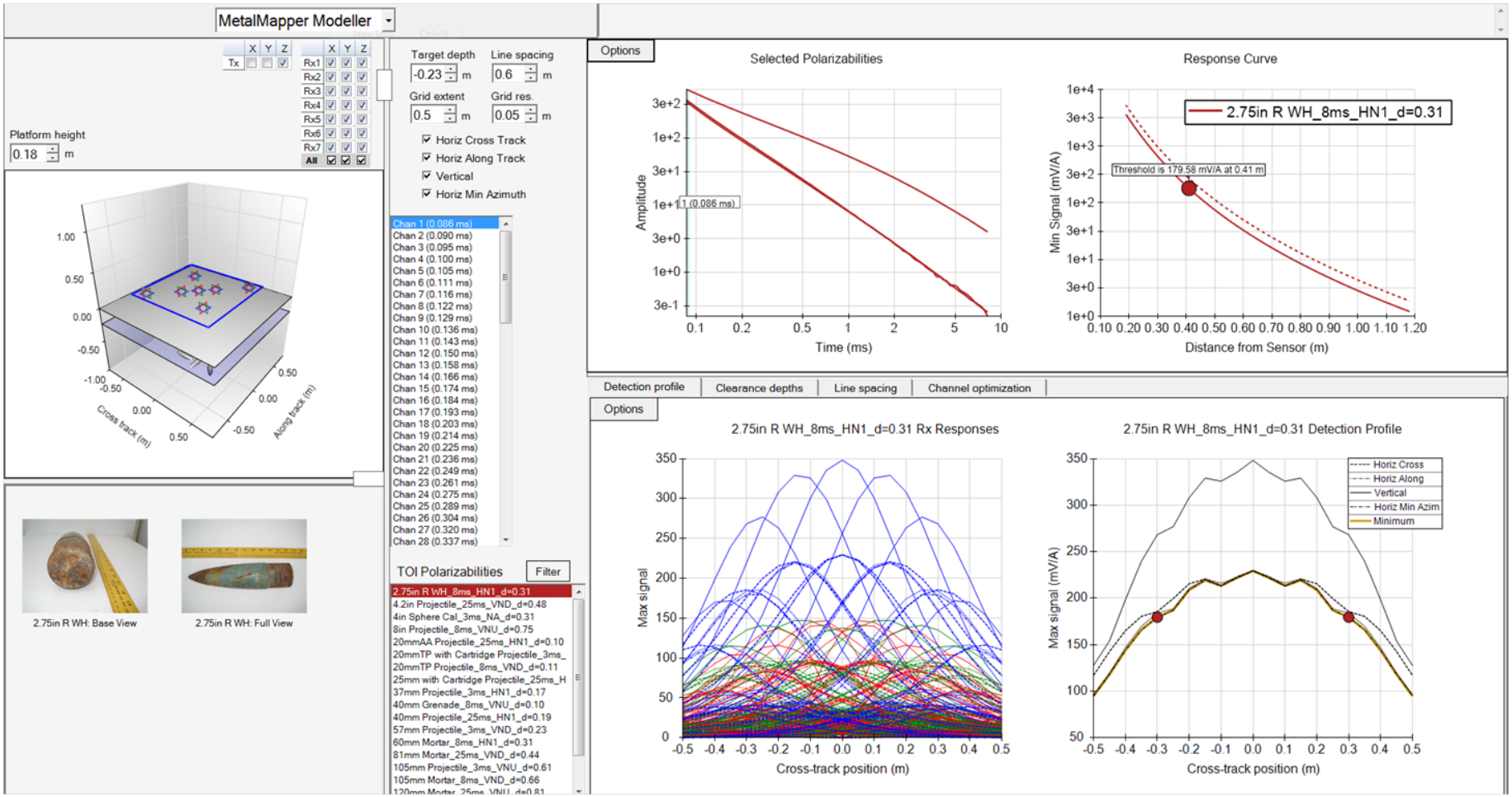


FIGURE 13. Detection modeller interface. Left panel shows sensor geometry and selected target orientations (cross-track, along-track, and vertical). Top right panel shows selected target polarizabilities and a plot of minimum signal (over all receivers and components) versus distance of target from sensor. Bottom right panel shows cross-track profile of all received data (left) and maximum signal as a function of cross track position in each of the three target orientations.

3.5. **Delineating "mag and dig" regions in dynamic detection data.** Picking targets in dynamic detection data is, in practice, often more complicated than simply identifying isolated peaks above a threshold. In particular, a high density of metallic clutter at bombing or artillery targets can produce a high amplitude region without isolated anomalies. Figure 4 shows an example of a high density region in Camp Ellis EM-61 and MetalMapper data sets. As discussed above, the small area of the MetalMapper receiver cubes produces more isolated anomalies in the detection map, whereas the larger EM-61 receiver coil smears the anomalies.

Figure 14 shows a second example of a high density region encountered in TEMTADS2x2 data collected at the 2015 Camp Hale, CO demonstration. While there is a moderate anomaly density throughout most of the site, classification may not be feasible in an extended area of high amplitude response in the center of the survey. For this project, a polygon was manually drawn to define a "mag and dig" region; inside this area no classification was applied and clearance used analog detection only.

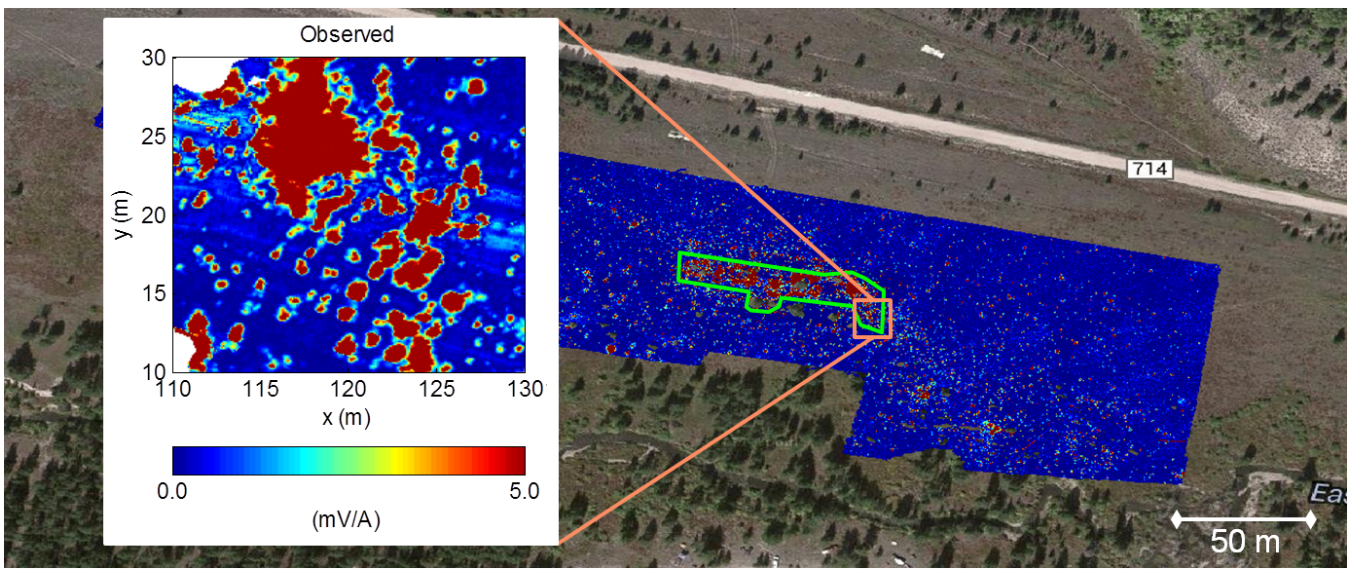


FIGURE 14. Monostatic z-component data acquired with the TEMTADS2x2 at Camp Hale, CO, overlain on satellite imagery. Green polygon shows a manually-delineated mag and dig region used for this demonstration. Inset area shows a close-up of part of the high anomaly density region.

In this section we develop and compare objective methods for identifying high density mag and dig regions where classification cannot be applied. An intuitive approach to this problem is to simply mask out contiguous, spatially-extended areas with elevated amplitudes in the observed dynamic data. As illustrated in figure 15a, we divide the survey area into pixels with widths corresponding to a multiple (usually 1 or 2) of the sensor width. We then define a pixel mask with pixels set equal to 1 if any datum within the pixel exceeds a predefined threshold (15b). Any component of the received data can be used - monostatic, vertical component data is the most straightforward choice for the TEMTADS2x2. Given a mask of pixels with elevated response, we then identify regions of contiguous pixels and trace out the perimeters of the individual regions (15c). Small regions can be

eliminated by requiring a minimum number of pixels in regions, or can be manually merged by the addition of connecting pixels. The resulting pixellated boundary in 15d is likely too complicated for dig teams to accurately follow. We therefore simplify the boundary by tracing out the convex hull (15e) or a non-convex enclosing polygon (15f). The latter can slightly reduce the total mag and dig area relative to the convex hull, but may sometimes produce an undesirably complicated boundary.

The above processing approach is not restricted to a threshold on the observed data, and can be similarly applied to arbitrary transformations of the data. In the next sections, we explore different criteria for delineating mag and dig regions.

3.5.1. *Target density estimation.* Instead of thresholding on data amplitudes, we can calculate a target density from a map of dynamic target picks and mask out contiguous regions where the estimated density exceeds a threshold. The efficacy of this approach will depend on the details of the target picking algorithm. Simple bump picking on a gridded image of monostatic data cannot handle ambiguities that arise with closely spaced anomalies. For example, two adjacent peaks in monostatic data can result from two separate vertical targets, or from a single horizontal target. Therefore a more sophisticated analysis may be required to estimate target density.

Given high resolution multistatic detection data, multi-object inversion can often resolve individual target locations and polarizability tensors when there is a moderate target density. However, when multiple, small fragments are in close proximity (e.g. in a "frag pit"), the observed data may still be reproduced by one or two dipoles. This makes reliable estimation of the number of sources present in the data quite difficult. Model selection criteria - e.g. Minimum Description Length (MDL) or Akaike information criterion (AIC) - can be used to estimate the model order (number of sources) in EMI data (Song et al., 2009). These criteria balance the fit to the data with a penalty on the complexity of the model. Application of these criteria to synthetic and real multistatic data sets showed that, except in the most obvious cases, model selection criteria cannot reliably predict the number of sources. In classification processing we now entirely circumvent estimation of the number of sources by generating an ensemble of possible models with progressively increasing numbers of sources. All of these models are used in the subsequent classification analysis.

Inversion-based target picking with a dipole filter approach goes a long way to overcoming ambiguities in the observed data. By fitting a physical model to the data we can eliminate non-dipolar anomalies from the target list. However, even with inversion it is often difficult to reliably determine the number of sources contributing to the data. This is illustrated in figure 16, which shows the dipole filter results for synthetic multi-object scenarios. For closely-spaced dipolar sources, a gridded image of the correlation coefficient produces a rescaled image of the monostatic data. While the number of sources in these examples is clear for source separations $\gtrsim 0.2$ m, at smaller separations the dipole filter output cannot discern the individual targets.

3.5.2. *Singular value decomposition of dynamic data.* Here we investigate criteria that are sensitive to the presence of multiple targets within the sensor footprint, but do not require an *a priori*

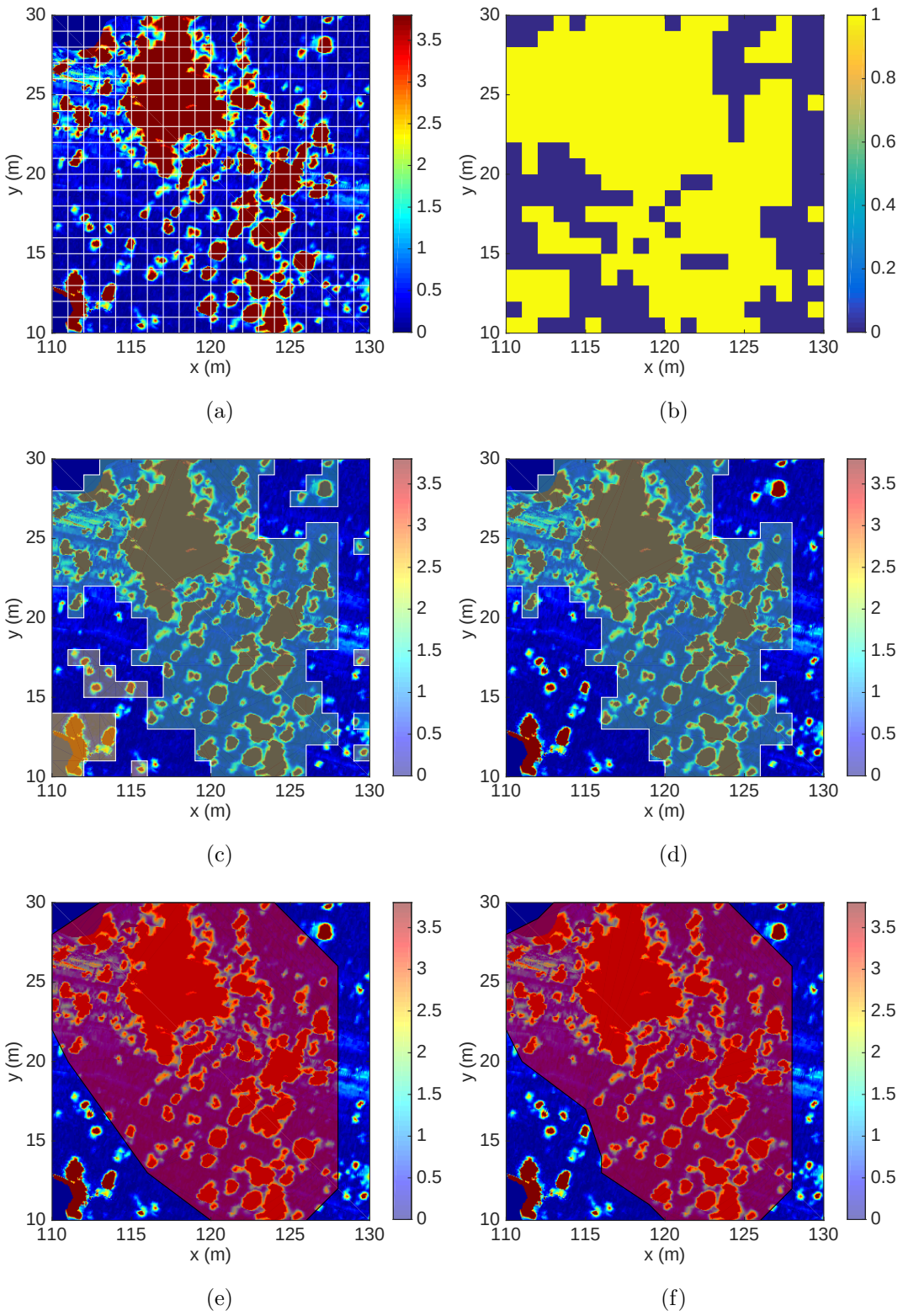
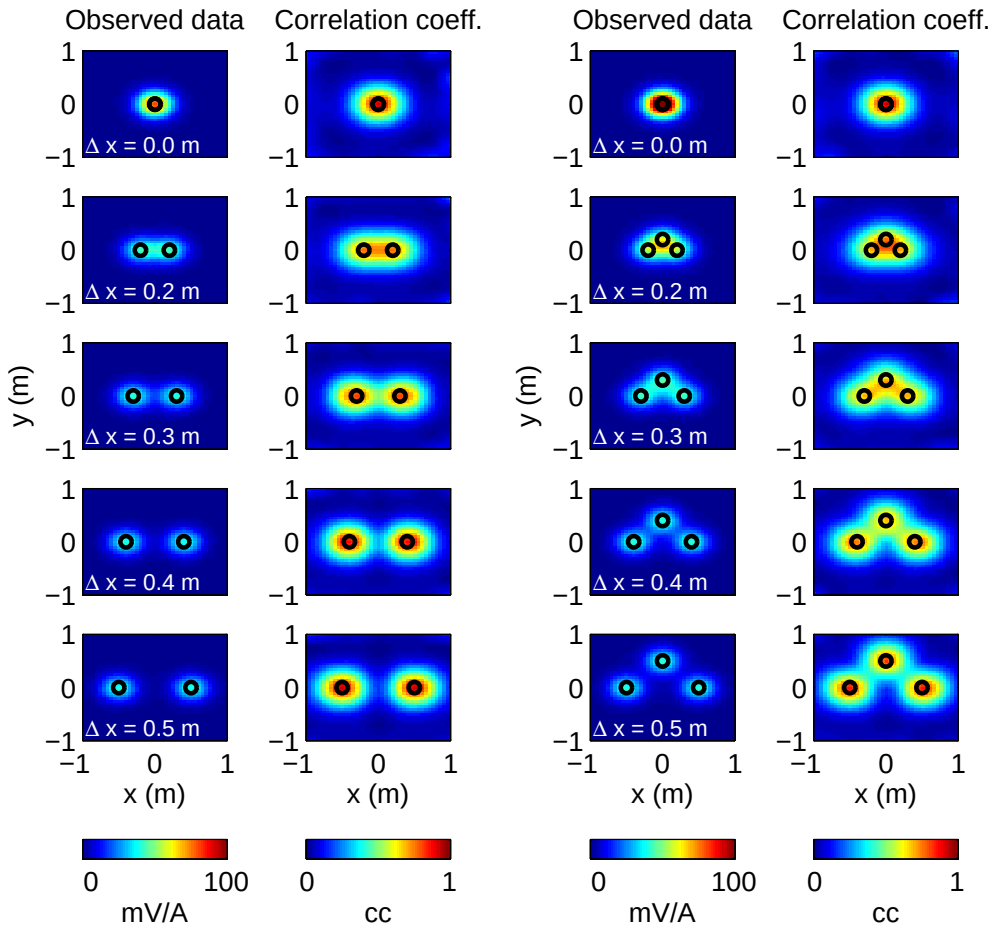


FIGURE 15. Application of pixel-based region delineation algorithm to a subset of monostatic z-component TEMTADS2x2 data from Camp Hale. (a) Division of area into a regular pixel grid. (b) Binary masking: if data within a pixel exceed a threshold, pixel is assigned a value of one. (c) Tracing contiguous regions from the binary mask. (d) Elimination of smaller regions with fewer than a minimum number of pixels. (e) Convex hull of region. (f) Non-convex enclosing polygon.



(a) Two object scenarios (b) Three object scenarios

FIGURE 16. Dipole filter applied to synthetic multi-object scenarios in dynamic TEMTADS2x2 data. In all examples sources are 40 mm projectiles at 30 cm depth below the sensor. Markers indicate source locations. "cc" is correlation coefficient between observed data and data predicted by a dipole filter source. Δx is the displacement of each target away from the origin.

assumption of the number of sources. Our approach employs a singular value decomposition (SVD) analysis of the dynamic data.

As described in Song et al. (2012); Shubitidze et al. (2012), observed multistatic data at a single time channel can be organized into a "multistatic Response Matrix" (MRM) with elements $MRM_{i,j}$ denoting the ith measurement acquired for the jth transmitter firing. For example, a cued TEMTADS2x2 sounding has an MRM of dimension 12 rows (4 receiver cubes $\times$ 3 components) by 4 columns (transmitters).

Assuming that the observed data arise from dipolar sources, the singular values of the MRM correspond to the principal polarizabilities of the sources, multiplied by a scaling factor. Song et al. (2012) therefore term the singular values "apparent principal polarizabilities" (APPs). Joint diagonalization can be applied to the MRM across multiple time channels to recover estimates of

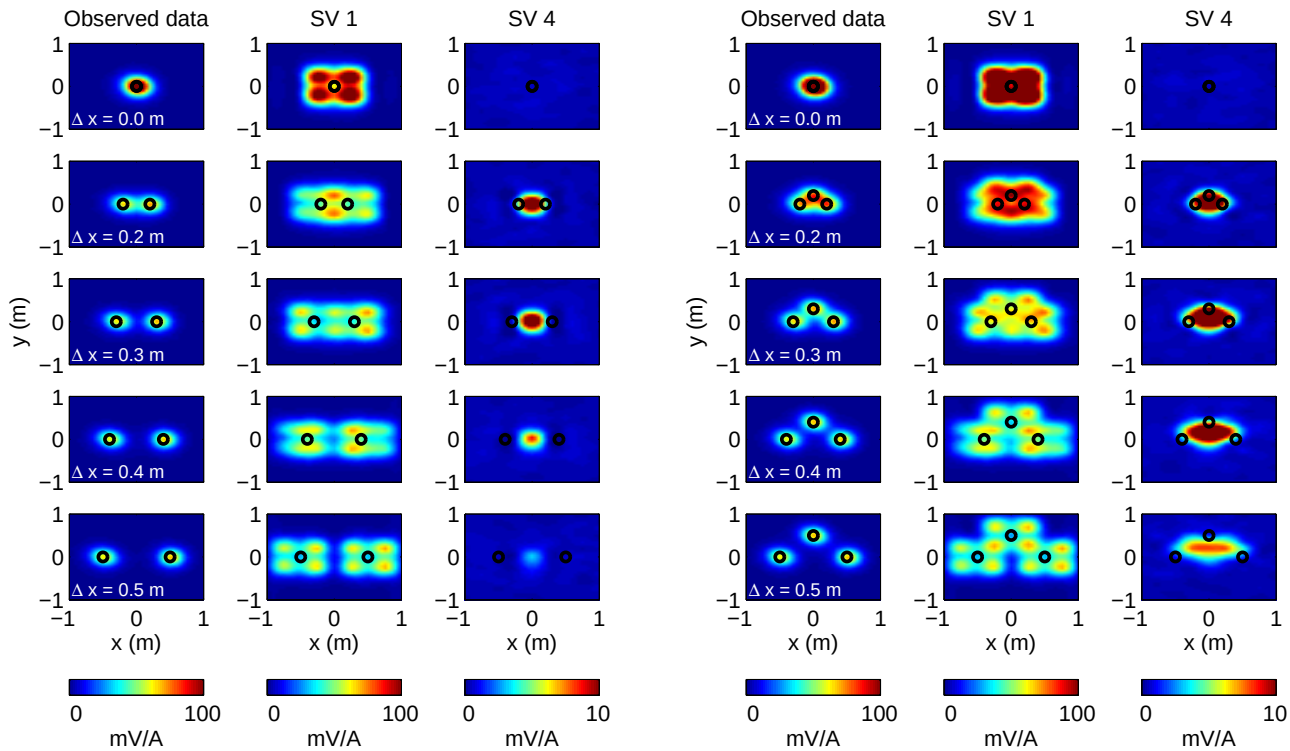
the APPs with time-independent eigenvalues. For diagonalization of the MRM at a single time channel, the eigenvalues are the squared singular values.

Singular value decomposition of the MRM has been used as an in-field or pre-inversion processing step to identify possible multi-object scenarios in cued measurements. Likely TOI can also sometimes be flagged on the basis of apparent principal polarizabilities. Detailed simulations of APP analysis of cued MetalMapper data presented in Pasion (2012) indicate that the number of significant singular values does correlate with the number of sources. However, for scenarios involving four or more targets the number of significant singular values effectively saturates and the number of sources cannot be inferred. In the context of density estimation using dynamic data, this limitation on SVD analysis is acceptable - we only require a diagnostic that is sensitive to multi-object scenarios within the sensor footprint.

For application of SVD analysis to dynamic data, we form multiple MRMs, each comprised of a sequence of measurements equivalent to a cued sounding. We then compute the SVD of the MRM at each equivalent sounding location. Joint diagonalization of each MRM over multiple time channels is prohibitively slow, and so we limit the SVD analysis to a single time channel. The penalty for this increased speed is a reduced number of singular values: the MRM at an individual time channel has rank equal to the number of transmitters. Hence for SVD of dynamic TEMTADS2x2 data we are limited to four singular values. In contrast, joint diagonalization can produce more non-zero singular values by requiring a common set of singular vectors over all channels.

Figure 17 shows synthetic examples of single channel SVD analysis applied to TEMTADS2x2 data. We consider the same two and three object scenarios presented in figure 16. Gaussian noise with standard deviations estimated from Camp Hale dynamic data has been added to the synthetic data. When there is one source in the data (top row in 17a and b), the first singular value (SV) has the character of monostatic z-component measurements for the sensor positioned directly over the target. In a single object scenario, the fourth SV is effectively zero. This is because three apparent polarizabilities (singular values) are sufficient to reconstruct the data. As target separation is increased in multi-object scenarios, the fourth SV (denoted SV4) is maximized at sounding locations where multiple sources contribute to the data. If the sources are moved far apart (greater than one sensor width), the anomaly in the SV4 image disappears, indicating that only single sources are present at any given sounding location.

The analysis in figure 17 is for an idealized case where there is no movement of the sensor during measurements comprising an equivalent cued sounding. This is consistent with the assumption that each row of the MRM corresponds to measurements taken with a single stationary receiver. In a dynamic survey, however, the sensor is in constant motion so that the rows of the MRM comprise measurements taken with a single receiver at multiple locations. Figure 18 shows the effect of sensor motion on the SVD analysis. The most notable change is an increase in the amplitude of SV4 for the single object case (top row). This arises because the motion of the sensor over the target in a dynamic survey is equivalent to a multi-object scenario for the static sensor case. Hence SV4 of



(a) Two object scenarios

(b) Three object scenarios

FIGURE 17. SVD applied to synthetic multi-object scenarios in TEMTADS2x2 data. In all examples sources are 40 mm projectiles at 30 cm depth below the sensor. Markers indicate source locations. Δx is the displacement of each target away from the origin. Second and third columns show first and fourth singular values (SVs), respectively. The fourth SV is insensitive to single targets, but shows a clear anomaly at locations where multiple targets contribute to the data.

dynamic data is not entirely immune to the presence of single objects, and some analysis of the minimum expected SV4 amplitude for multi-object scenarios is likely required.

Figure 19 shows a comparison of SV4 and dipole filter analyses applied to Camp Hale dynamic TEMTADS2x2 data. Consistent with the synthetic examples shown above, the SV4 result does reduce the relative amplitude of discrete anomalies that likely arise from single sources. This is highlighted by manually-defined polygons in figure 19 that enclose regions with discrete anomalies. The singular vector image attenuates these anomalies while leaving larger scale anomalies intact. As predicted in figure 18, the movement of the sensor during dynamic data acquisition limits the attenuation of single source anomalies: they are not entirely removed from the image.

In contrast, the dipole filter produces a rescaled image that allows identification of discrete anomalies regardless of their amplitude in the observed data. This rescaling allows us to discern sources that are unresolved in the data or the singular value transformation of the data. Figure 20 shows a comparison of polygons delineating high density regions in each of the images. To define these polygons we use an automated analysis that identifies contiguous pixel regions with gridded image amplitudes exceeding a threshold. As in figure 15, we then trace the boundary of the pixel

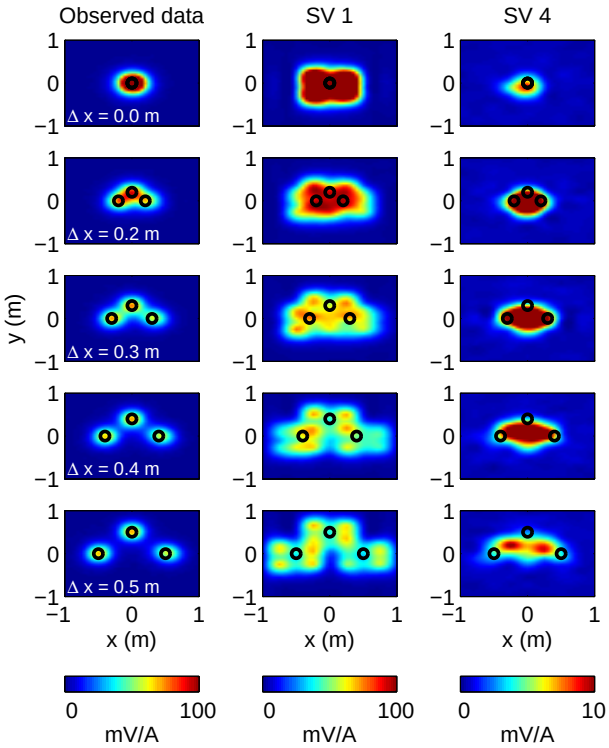


FIGURE 18. Effect of moving sensor on SVD analysis for synthetic three-object scenarios. Measurements are regularly spaced between sounding locations used for simulations in figure 17(b). This produces an elevated response in SV4 for the single object case (top row).

regions to delineate a mag and dig area. The largest area is produced by thresholding on the dipole filter image. This is expected given the large number of anomalies present in this image. The high density region produced by singular value analysis is approximately 30% smaller than the dipole filter region. Filtering out isolated anomalies with singular value analysis can thereby reduce the size of areas designated for mag and dig clearance.

Identification of mag and dig regions by SVD analysis is constrained by the rank of the MRM, which is in turn determined by the number of transmitters. The four TEMTADS2x2 transmitters allow us to identify potential multi-object scenarios via an image of SV4. The dynamic MetalMapper is not so fortuitous in this respect: the single horizontal transmitter coil produces one singular value that is insensitive to the number of sources in the data.

3.5.3. *Multi-object inversion of dynamic data.* We can use multi-object inversion to identify mag and dig areas, regardless of the sensor that is used to collect the dynamic data. We divide the survey area into overlapping cells (typically 1 m x 1 m in size) and carry out a three-object inversion within each cell. The cell overlap ensures a smooth image of the predicted data and residuals when the cells are subsequently stitched together (figure 21). We use three-object inversion based on practical experience: if the observed data in a given cell cannot be predicted by a three-object inversion, then the estimated parameters in that cell likely cannot be used for target classification. Here we assess the quality of the fit to the data by considering the inversion residuals (observed minus predicted

data), together with the detection threshold for the smallest target of interest. If the residuals exceed the detection threshold within a cell, then there is significant signal that cannot be predicted by the modelling and this cell should be flagged for mag and dig clearance. We can then use the same pixel-based region finding algorithm as before to delineate mag and dig regions based on inversion residuals (figure 22). The result is quite similar to the SVD analysis, but is applicable to all sensor types.

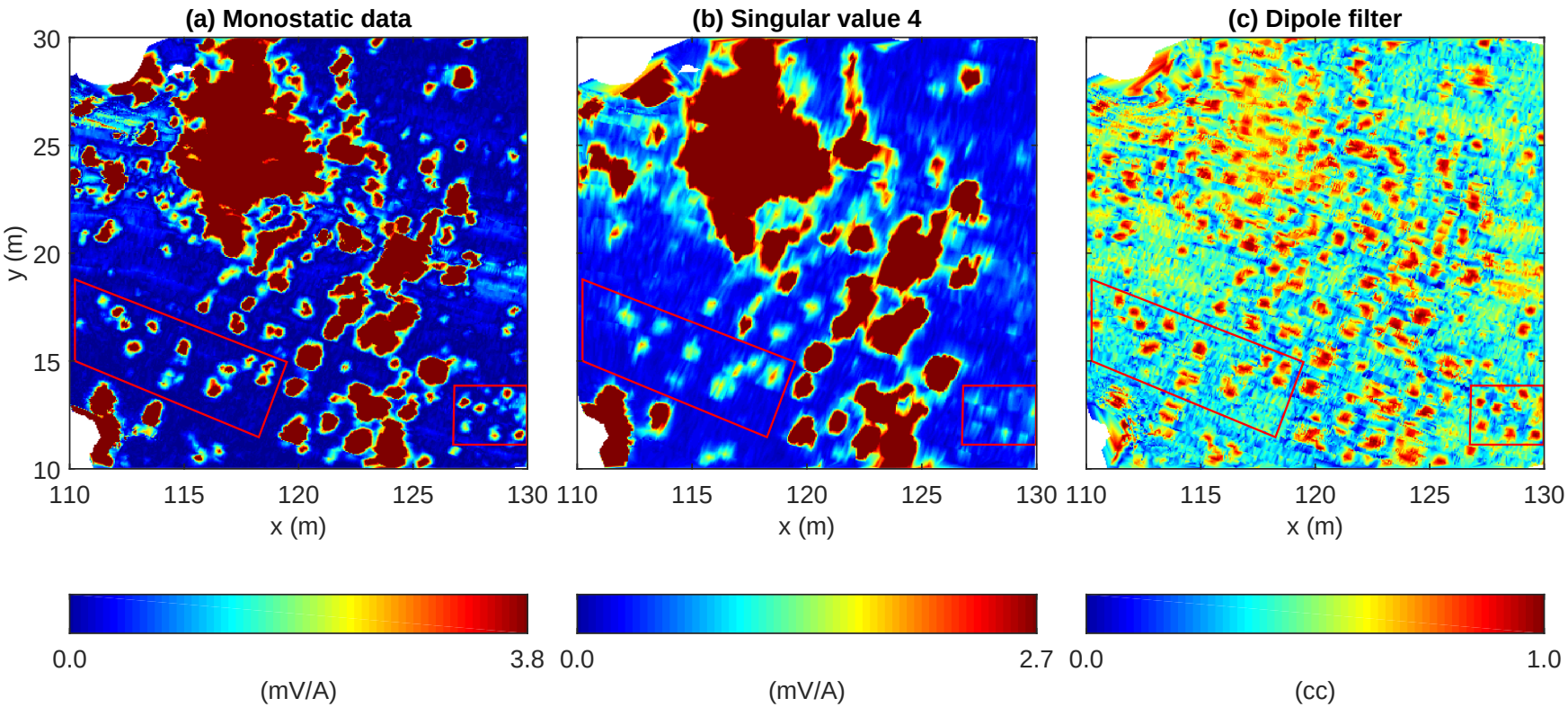


FIGURE 19. Comparison of singular value and dipole filter analysis of Camp Hale dynamic TEMTADS2x2 data. Maxima of monostatic and SV4 colorbars are set to 0.8 of the maximum value in each respective grid. Polygons highlight regions with discrete anomalies.

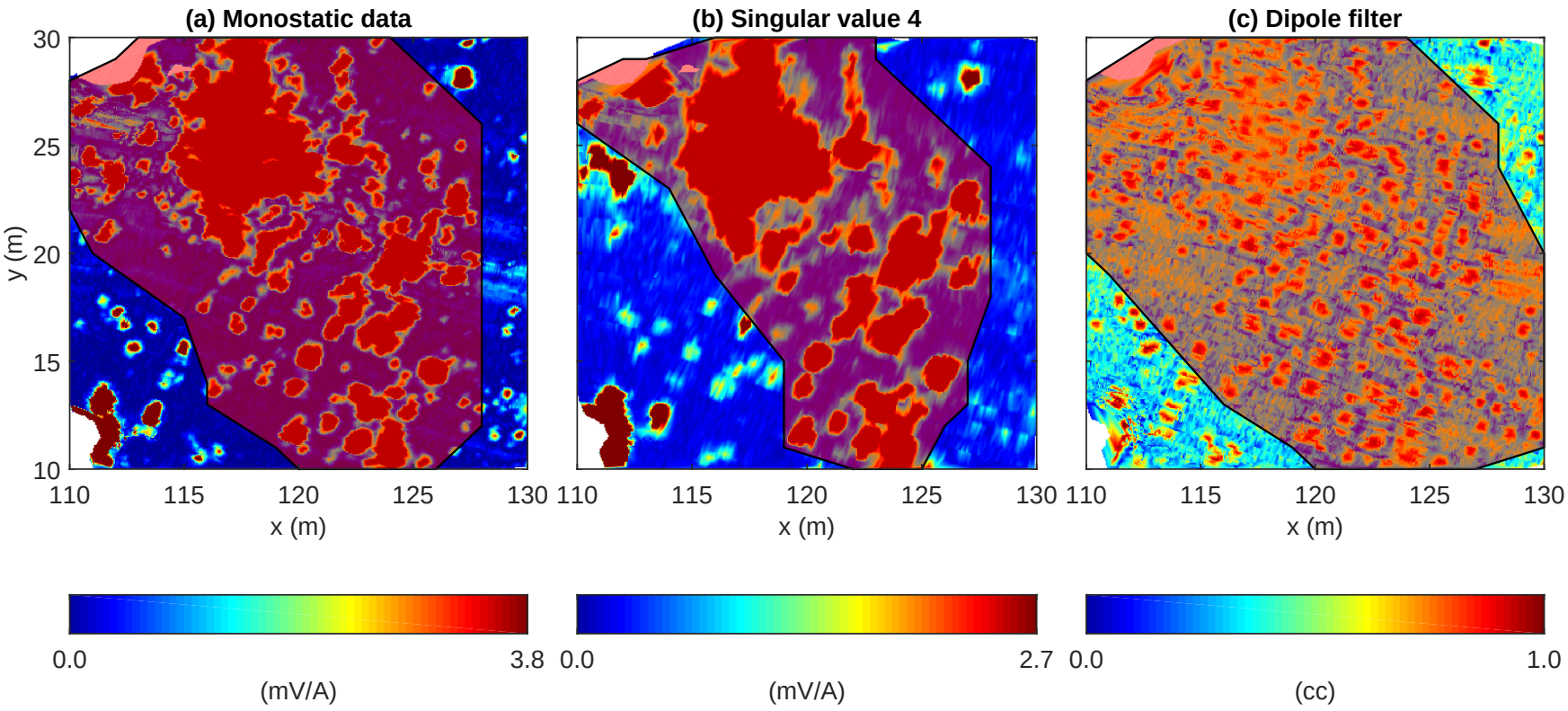


FIGURE 20. High density regions delineated by thresholding on images of monostatic data, SV4 and dipole filter.

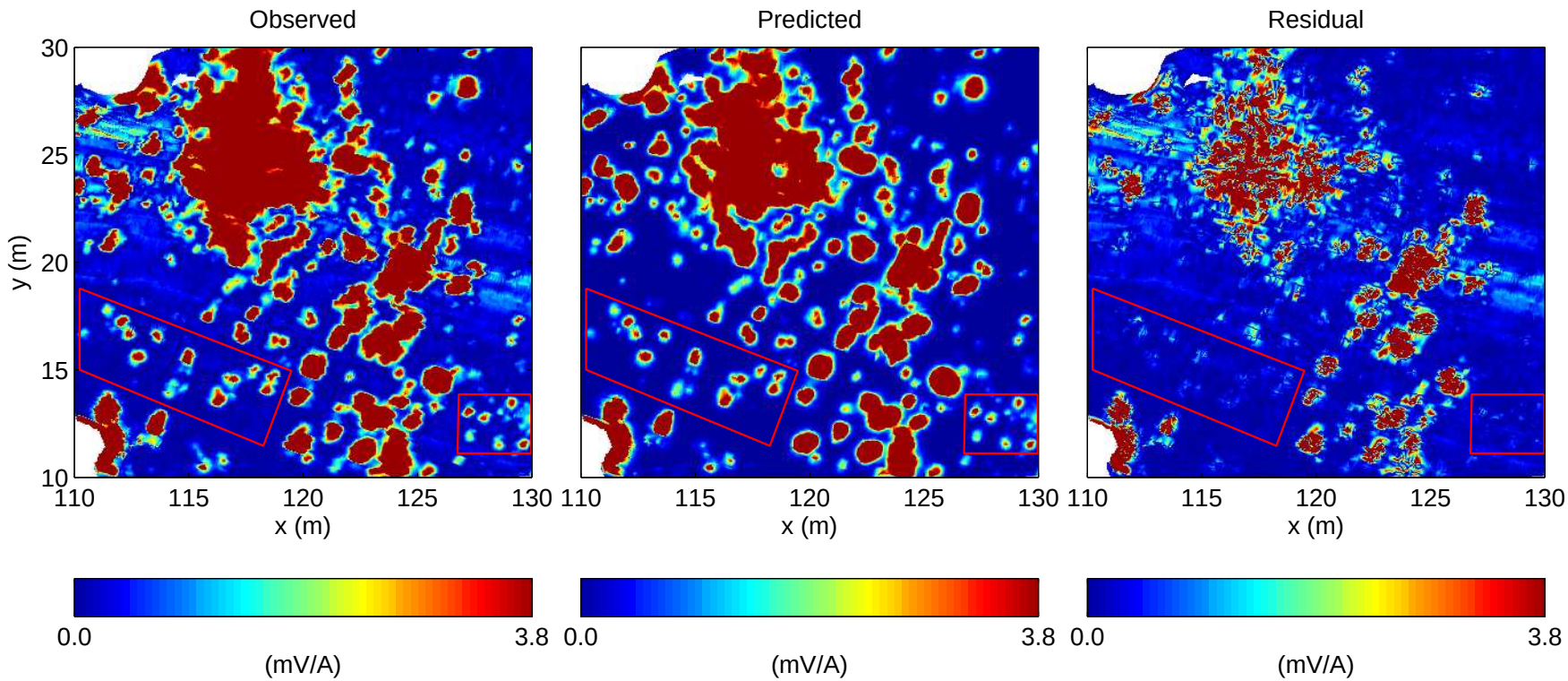


FIGURE 21. Multi-object inversion of multistatic dynamic detection data. Left: observed z-component monostatic data; Middle: data predicted by three-object inversions of observed data. Inversions are carried out in overlapping 1 m x 1 m cells; Right: inversion residuals. Polygons highlight regions with discrete anomalies, as in 19.

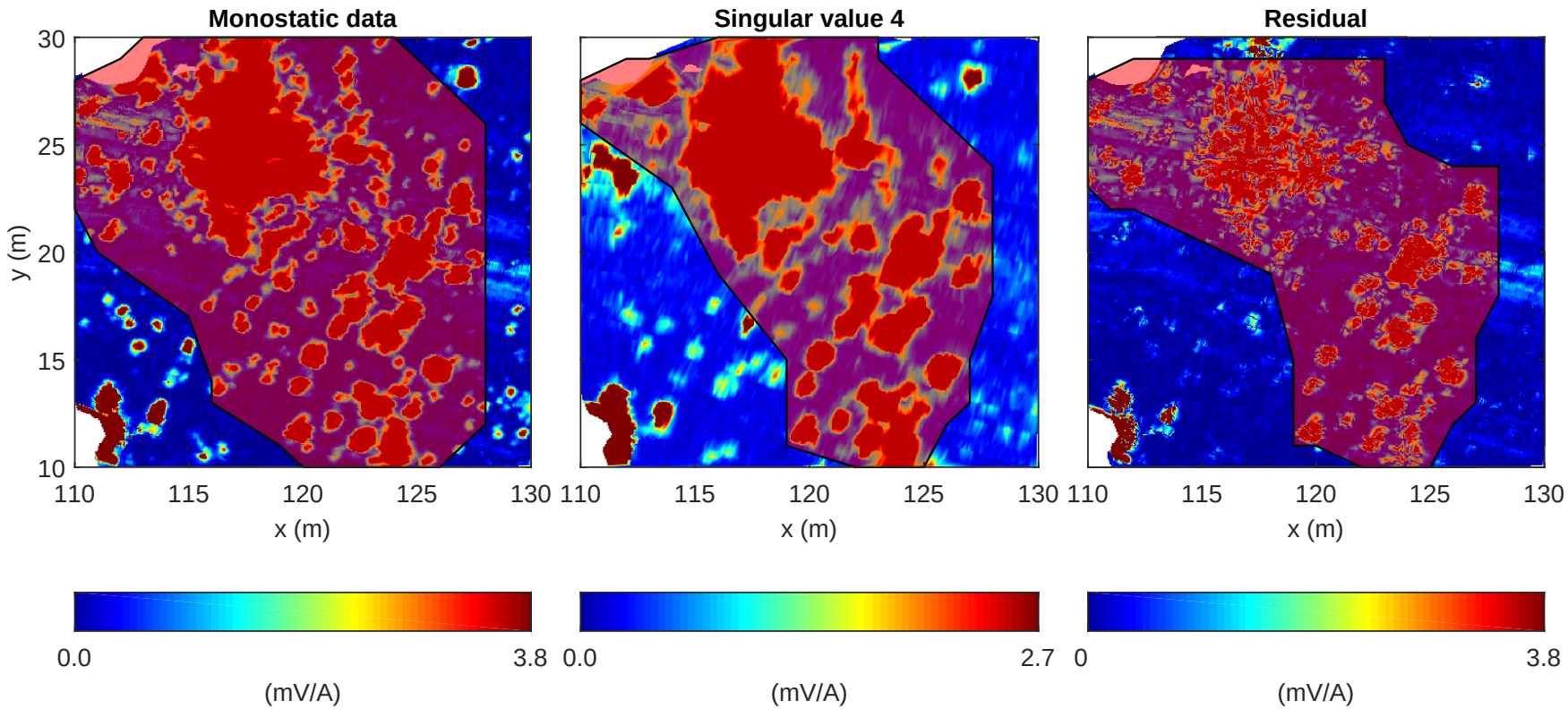


FIGURE 22. Comparison of high density regions delineated by threshold on images of monostatic data, singular value 4 and multi-object inversion residuals.

4. Performance prediction

A major focus of this project has been the development of efficient methods for prediction of classification performance. We initially envisioned a decision support system (DSS) that used site-specific conditions to generate a prediction of the receiver operating characteristic for the problem at hand. In this approach, site-specific conditions include: sensor noise, target depth, the spatial separation between neighboring targets, and the subset of estimated parameters used to classify targets. To this end, we have developed efficient methods for predicting uncertainty in estimated polarizabilities and subsequent classification performance. This work is presented in an unpublished manuscript in appendix B.

Current approaches to this problem use Monte Carlo (MC) simulations that require computationally intensive nonlinear inversion of multiple synthetic data sets. In contrast, our efficient MC algorithm replaces nonlinear inversion with a parametric approximation to the posterior distribution of estimated target locations. We then quantify expected classification performance by considering binary classification of a specified UXO and non-UXO. Using the predicted MC polarizability distributions for both targets, we estimate the probability of correct classification. These methods thereby provide means to assess the feasibility of classification under prescribed, site-specific conditions.

In addition to tools for efficient simulation of performance prediction, we have developed methods to predict classification performance for a specific realization of observed data at a site. To assess whether an observed data set can support advanced classification, we first developed the dataset degree of difficulty (DDD). At the point that all detected anomalies have been inverted, an experienced data analyst can often get a qualitative sense of the site difficulty based on visual QC of inversion results and inspection of the size vs. decay feature space. In Beran et al. (2013b), we codified this subjective judgment and experience into a metric that quantifies the degree of classification difficulty for a dataset. In section 4.1, we assess whether the DDD analysis is predictive of classification performance via retrospective analysis of MetalMapper data sets. We then use regularized linear regression in section 4.2 to develop models for directly predicting the false alarm rate from dataset metrics. Finally, in section 4.3, we extend the regression model to predict performance using a subset of metrics that can be assumed prior to the acquisition of data.

4.1. **Dataset degree of difficulty.** The dataset degree of difficulty combines an ensemble of data and model metrics into a single number. When processing a new data set, we use the DDD to guide our classification strategy. For easy sites (e.g. Pole Mountain), a low DDD (i.e. < 10) indicates that an aggressive classification strategy using all three estimated polarizabilities will likely identify all TOI with a minimal false alarm rate. Conversely, a site with high DDD (> 50, e.g. Vieques) will not support classification and all detected targets should be dug. Table 3 in Appendix C summarizes the DDD metric for a number of cued MetalMapper data sets, note that the values have been rescaled from the initial analysis that was presented in Beran et al. (2013b).

Our recent work with the DDD has examined whether this ad hoc measure is in fact predictive of classification performance, as quantified by the false alarm rate (FAR). In appendix C, we provide

summary statistics for retrospective classification of ESTCP Cued MetalMapper data sets. We also show images of representative TOI at each demonstration site and ROC curves obtained with four classification approaches:

(1) Polarizability matching using all polarizabilities,
(2) Polarizability matching using primary polarizabilities only,
(3) Combined classifier ranking (CCR) using both polarizabilities and size/decay features, see Beran et al. (2013a),
(4) τ metric classifier, using no reference polarizabilities to rank targets (see section 5.6.2).

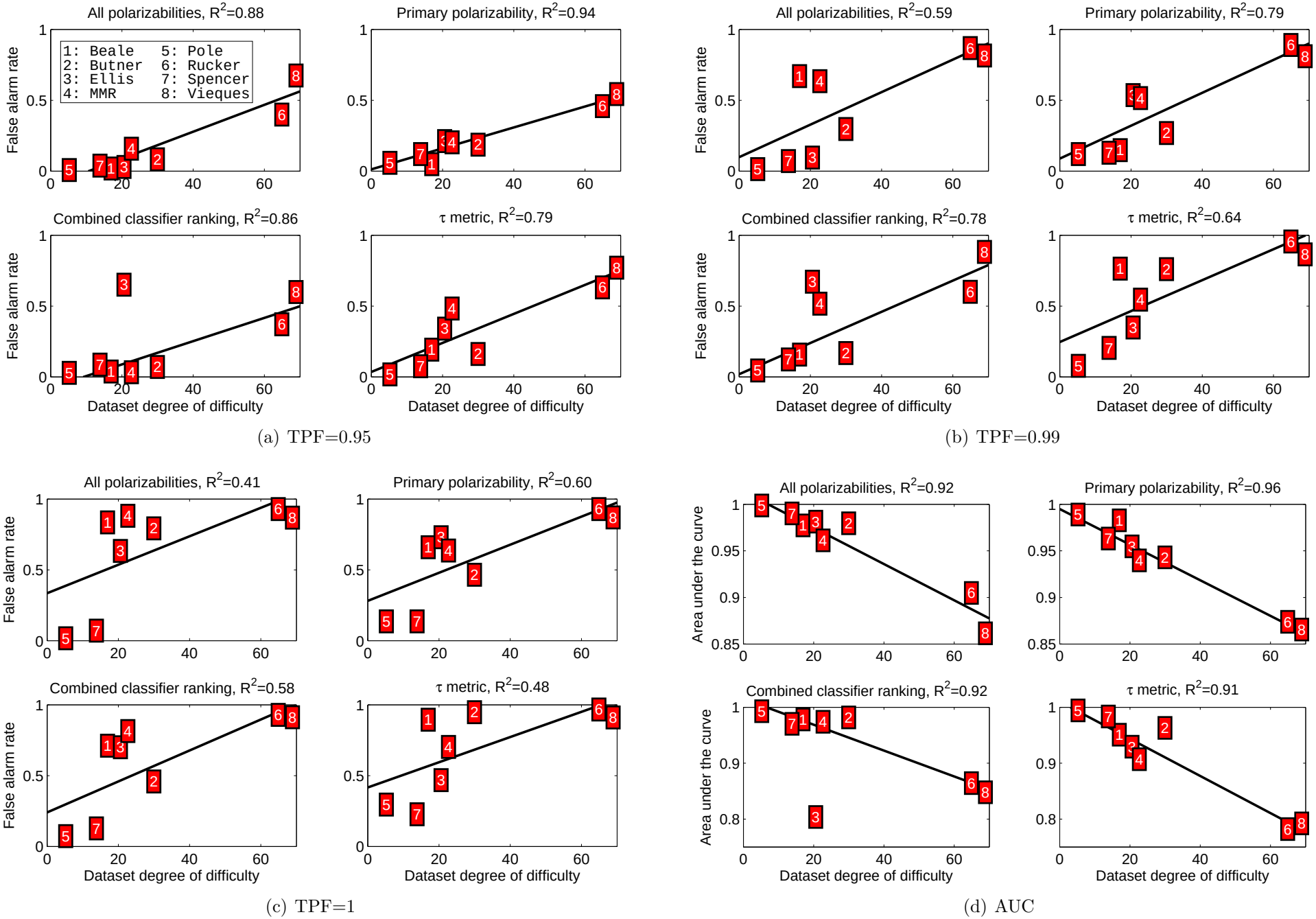
Using these retrospective classification results, we can study how DDD correlates with the observed FAR for each classification approach. In figure 23, we fit a linear trend to FAR versus DDD. We first consider the FAR at operating points identifying 95, 99, and 100% of the TOI in each data set. As quantified by the coefficient of determination (R^2), there is a strong linear relation between DDD and the FAR at which 95% of the TOI are found (figure 23a). Classification using primary polarizabilities produces the strongest correlation with DDD. This classification approach exploits a site-specific polarizability library while guarding against outliers by excluding secondary polarizabilities from the classification decision. In contrast, a classification approach that uses no site-specific information (e.g. the τ metric) results in more variability in classification performance and a lower correlation with DDD.

The Combined Classifier Ranking (CCR) attempts to strike a balance between polarizability matching and prioritization using generic features of TOI model parameters. It is appropriate for scenarios where TOI have large amplitude, slow decaying polarizabilities. At sites with small TOI, however, the CCR can perform quite poorly. This is evident in the Camp Ellis result, which appears as an outlier in the CCR results shown in figure 23a. This site has both large (2.36" rockets) and small (hand grenades) TOI (see Appendix C). These latter items decay relatively quickly and so are ranked low in the CCR diglist. The CCR approach is therefore not appropriate for classification of the Camp Ellis data and this site has been excluded from the regression analysis with CCR.

As we increase the proportion of TOI found to 99 and 100% (figure 23b and c, respectively), DDD becomes less predictive of FAR and R^2 between the two variables decreases for all classification approaches. The DDD is a measure of *average* classification difficulty at a site. Outliers - which dictate the false alarm rate - cannot be reliably predicted using the median sample statistics estimated from all detected targets.

Therefore while the DDD is useful for predicting classification performance for the majority (e.g. 95%) of targets, it cannot reliably predict whether all TOI will be found at a site. This conclusion is supported by the strong correlation between DDD and area under the receiver operating characteristic (AUC) shown in figure 23d, with $R^2 > 0.92$ for all classification approaches. By integrating over the entire ROC, the AUC provides a measure of average classification performance. In the context of UXO classification the AUC is less relevant than the FAR at which all TOI are found. However, in section 5 we show that the AUC can be a useful measure of average classification difficulty when determining the number of ordered digs required for risk assessment.

In the next section, we address the shortcomings of the DDD analysis with a regularized regression model that directly predicts the FAR. By using an enlarged set of sample statistics, the regularized approach has greater sensitivity to factors which give rise to outliers in a diglist.



(a) TPF=0.95

(b) TPF=0.99

(c) TPF=1

(d) AUC

FIGURE 23. False alarm rate (FAR) and Area under the Curve (AUC) vs. Dataset Degree of Difficulty (DDD) for classification of ESTCP MetalMapper datasets. Figures a-c show FAR vs DDD for FAR corresponding to the True Positive Fraction (TPF, i.e. the proportion of TOI found) indicated. R^2 denotes coefficient of determination of linear fit (solid line) for each classification method.

4.2. **Regularized regression for performance prediction.** The dataset degree of difficulty is a heuristic combination of statistics designed to provide a relative measure of classification difficulty. In this section we pursue a more formal regression analysis to learn the mapping between dataset statistics and the false alarm rate. Given performance statistics estimated from cued multistatic data, the regression model developed here can be used to directly predict the false alarm rate at a site. This may be a useful tool for communicating expected performance to regulators and stakeholders prior to starting intrusive operations.

The observed false alarm rate, denoted here as the response variable $\mathbf{y}$, is modelled via the logistic function

$$\mathbf{y} = \frac{\exp\left(\beta^T \mathbf{X}\right)}{1 + \exp\left(\beta^T \mathbf{X}\right)} \tag{19}$$

with $\mathbf{X}$ a matrix of predictors and β the regression coefficients to be estimated. Each row in $\mathbf{X}$ is comprised of performance statistics calculated at each site. The DDD analysis is calculated from median values of metrics related to data and model quality. More description of these metrics can be found in Beran et al. (2013b). In this work we extend the set of predictors to include percentiles of each metric used in the DDD. Specifically, we calculate percentiles $P \in \{10, 25, 50, 75, 90\}$ for each statistic ($P = 50$ denotes the 50th percentile, or median). This additional detail in the predictors allows the regression model to be more sensitive to variations in classification difficulty and, ideally, to predict the presence of outliers in a diglist. Additionally, we introduce a metric related to spatial target density: we compute the mean distance from each detected target to its three nearest-neighbor (figure 24). The inverse of this nearest neighbor distance is then histogrammed over all targets to obtain percentiles that are input into the regression analysis.

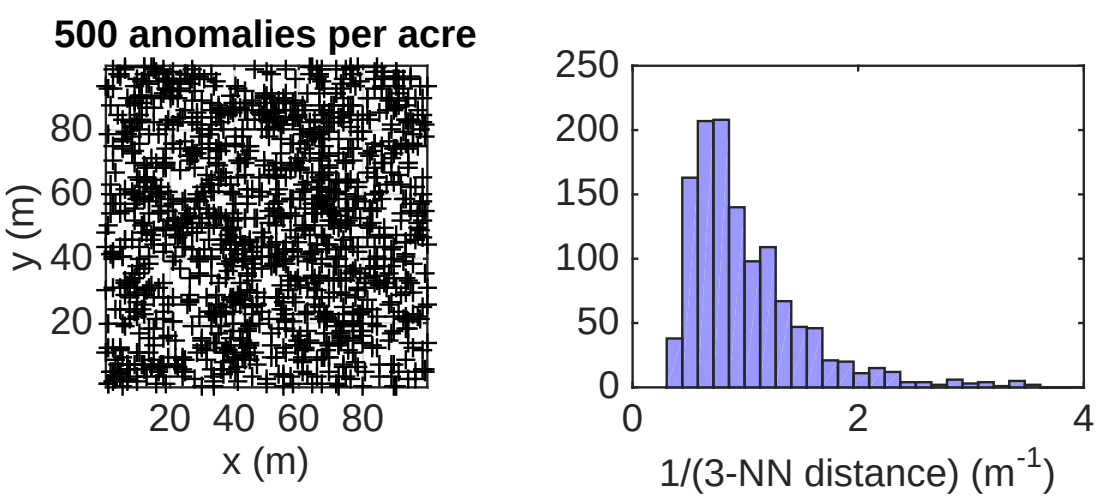


FIGURE 24. Accounting for spatial target density in performance predictions. Left: an example synthetic spatial target distribution corresponding to an average density of 500 targets per acre. Right: distribution of density parameter derived from spatial distribution. The density parameter is calculated as the inverse of the mean distance to each target's 3 nearest neighbors (the 3-NN distance).

We use the "elastic net" method (Zou and Hastie, 2005) to regularize the linear problem

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} \left(\|\mathbf{y} - f(\mathbf{X}, \beta)\|^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2 \right). \tag{20}$$

with $f(\mathbf{X}, \beta)$ denoting the logistic function (equation 19) mapping between predictors and predicted false alarm rate. The parameters (λ_1, λ_2) control the regularization of the regression coefficients. A regularized approach is preferable to ordinary least squares because regularization can produce a simple, parsimonious model that fits the data and is readily interpreted. The elastic net regularization promotes a combination of sparsity via the $L1$ norm and smoothness via the $L2$ norm. The inclusion of the $L2$ norm in the elastic net regularization allows for groups of nonzero coefficients for highly correlated predictors. In contrast, pure $L1$ (lasso) regularization ignores groups and typically selects only one nonzero coefficient for a set of correlated predictors.

A range of models is output by the elastic net algorithm, each corresponding to different values of the regularization parameters. Cross-validation can be applied to select the model that provides an appropriate trade-off between bias and variance (i.e. generalizing to prediction with new data sets vs. fitting the data). For fixed $\alpha = \lambda_1/(\lambda_1 + \lambda_2) = 0.5$, we use leave-one-out cross-validation to estimate the regularization parameter λ_2 providing minimal mean squared error.

Figure 25 shows a cross-validated elastic net regression model for classification of cued MetalMapper data using polarizability matching using primary polarizabilities only. The observed false alarm rates are well explained by the regularized model. Also shown in Figure 25 are the coefficients β for each model. In general, parameters related to noise on the data (e.g. "data badness") and separation between TOI and non-TOI in feature space are most predictive of classification performance. For example, a large feature distance is diagnostic of TOI and non-TOI that can be readily distinguished with advanced classification and so this parameter correlates negatively with the predicted false alarm rate. We also remark that an increase in the highest percentiles of the density parameter (corresponding to an increase in the proportion of multi-object scenarios at a site) correlates positively with the predicted false alarm rate.

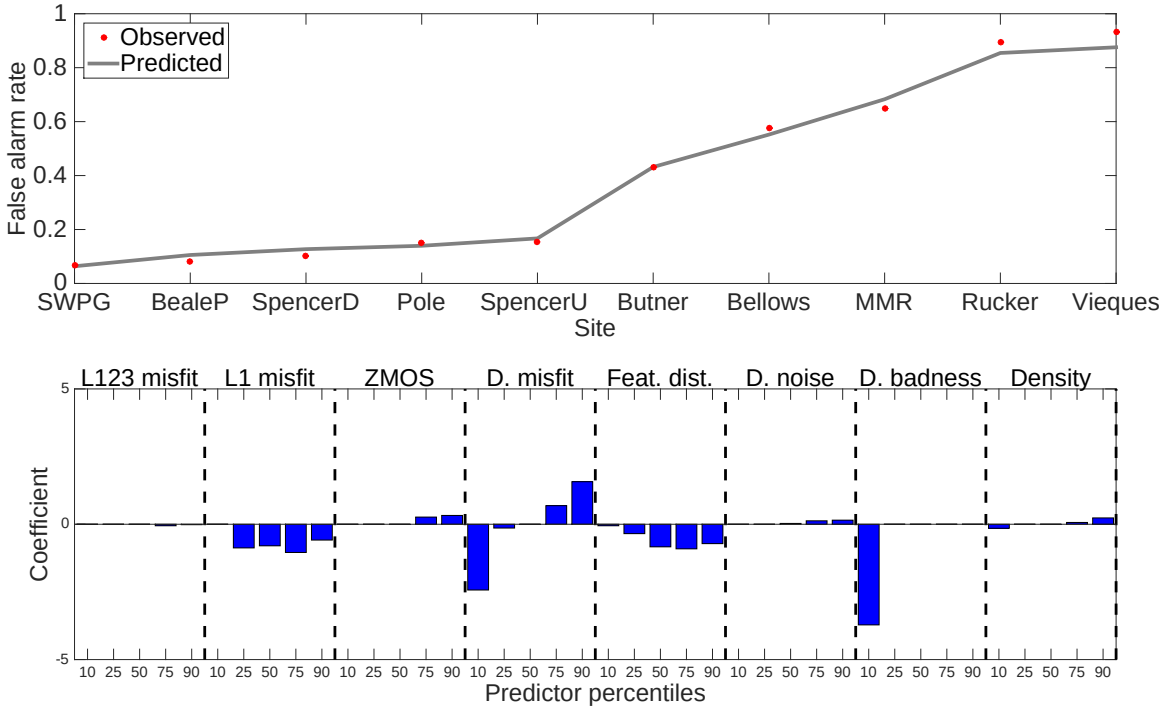


FIGURE 25. Elastic net regression of false alarm rate for classification of MetalMapper data using primary polarizabilities. Note that for Camp Beale (denoted BealeP) we have eliminated an outlying TOI (target 2277, attributed to a groundtruth error in Pasion et al. (2012)). This significantly reduces the FAR at this site relative to analyses presented in the previous section. Top: observed and predicted false alarm rates. Bottom: coefficients of elastic net model.

4.3. ***A priori* performance prediction.** The dataset degree of difficulty and the regression model described in the previous sections are useful tools for predicting classification performance once data have been processed but before intrusive operations begin. Here we consider how performance predictions can be generated prior to acquisition of cued data. We term this *a priori* performance prediction. This is a more challenging problem since we are limited to information in the conceptual site model and possibly some initial dynamic transects acquired at the site. Although site-specific information is limited, predictions at this stage of a munitions response project may be useful for understanding how target density and target type will affect predicted false alarms rates.

We extend the regularized regression model to generate predictions of classification performance by imputing (replacing) missing information with known values from previous sites. Some statistics related to similarity of TOI and non-TOI (i.e. feature distance) or spatial density can be assumed *a priori*, but parameters related to noise cannot be reliably assumed and must be imputed using values from previous sites. We generate a set of predicted false alarm rates $(y|x_i)$ with each prediction conditional on the imputed values (x_i) from a previous site. We then weight the conditional predictions based on the similarity of the previous and current site conditions via the weights

$$w_i \propto \exp\left(\frac{y|x_i - y_i}{\sigma_y}^2\right), \tag{21}$$

with y_i the false alarm rate at site i, and σ_y a scale parameter. Taking the weighted average of the conditional predictions

$$y = \sum_i w_i y|x_i \tag{22}$$

yields the *a priori* prediction of the false alarm rate.

To illustrate this approach to a priori performance prediction, in figure 26 we show an assumed distribution of size/decay features for an ongoing munitions response project. This distribution was generated by analysis of dynamic data transects that indicated only moderate overlap between clutter and TOI (20 mm and 37 mm projectiles) features at the site. Given this assumed feature space, we can generate an expected distribution of feature distance and input this into the regression model, together with imputed parameters. Figure 27 shows how the predicted false alarm rate at this site depends on the smallest TOI and the mean target density. If 37mm projectiles are considered the smallest target of interest at the site, then we expect a low false alarm rate for target densities below approximately 3000 anomalies per acre. At higher densities, we predict that classification will require excavating a significant proportion of clutter in order to find all TOI. This is consistent with physical intuition: densities of approximately 3000-4000 anomalies per acre correspond to 3-4 objects within the sensor footprint and are at the effective resolution limit of inversion algorithms.

When the smallest TOI is a 20 mm projectile, there is an increase in false alarm rate relative to the easier scenario. At densities less than 1000 anomalies per acre, classification of 20 mm projectiles produces an elevated, but acceptable, false alarm rate of approximately 0.2. However, going after 20

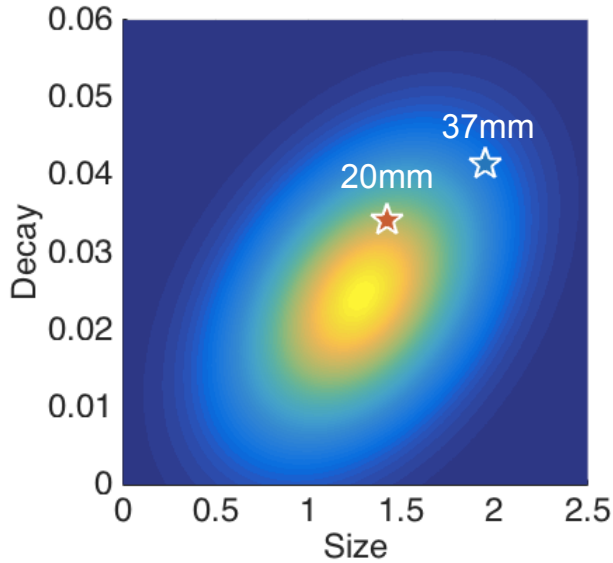


FIGURE 26. Assumed size/decay features for *a priori* performance prediction.

mm projectiles will necessarily raise the detected target density since smaller amplitude anomalies will be included in the cued target list.

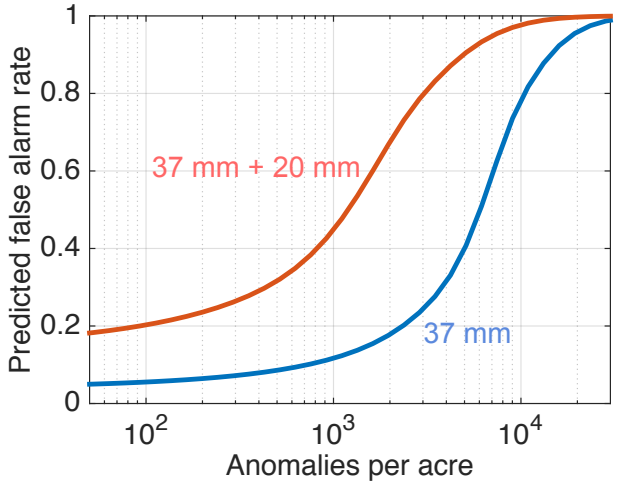


FIGURE 27. *A priori* performance predictions for classification with the assumed size/decay distribution in figure 26. We consider scenarios where the smallest TOI is a 37 mm projectile (blue) or a 20 mm projectile (red).

5. Risk assessment

A second focus of this project has been development and testing of methods for risk assessment. These methods aim to minimize the probability that TOI have been missed following the application of advanced classification. Missed TOI (outliers) in the diglist can be categorized as:

(1) **Near outliers**. Items belonging to a known ordnance class, but which have polarizabilities with an unexpectedly large deviation from reference polarizabilities. These instances can arise in low SNR or multi-object scenarios and will typically have poorly constrained L_2 and L_3 polarizabilities.

(2) **Far outliers**. Items for which acquired data do not support reliable parameter estimation and subsequent classification. These outliers usually occur when there is a failure in the sensor hardware (e.g. a faulty receiver), or a large background response that is unaccounted for in processing. Careful quality control during processing is the best approach to preventing far outliers, and will not be directly addressed here.

(3) **Novel TOI**. Items belonging to an unknown ordnance class. Successful classification relies upon a complete library of TOI polarizabilities, and all TOI classes are ideally identified in the initial training stage. At this stage, the data analyst tries to identify new TOI classes by clustering of estimated polarizabilities. However, small and rare TOI can sometimes be missed by an analyst.

Figure 28 shows a motivating example with near outliers on the ROC curve. This result was derived from processing of MetalMapper data acquired at Camp Beale, CA in 2011 (Pasion et al., 2012). In this case, both missed items correspond to low signal to noise ratio (SNR) scenarios where the recovered polarizabilities are poorly constrained. Representative items and reference polarizabilities from all ordnance classes encountered at this site are shown in figure 29.

We have developed a number of objective approaches to assessing whether all TOI have been found. Most of this work has already been published, see Beran (2014); Beran and Zelt (2014a,b). In this report we extend published results and briefly discuss software implementation of risk assessment algorithms.

We first review the random compliance sampling approach that has been suggested for UXO risk assessment. This approach is then extended to the case where random sampling is replaced with ordered digging. In section 5.3, we discuss methods for identification of near outliers via modelling of the ROC. We then discuss how seeded items emplaced for quality control can be used to increase confidence in the classification process by constraining the ROC model. Finally, in section 5.6 we turn to the problem of identifying novel TOI with prioritized validation digs.

5.1. Compliance sampling.

The approach developed in Hathaway et al. (2009) and implemented in Visual Sample Plan (http://vsp.pnnl.gov/) prescribes random compliance sampling of the targets remaining after classification has been applied. If none of n sampled targets are ordnance, then with confidence level $1-\beta$ we retain the null hypothesis (H_o: no TOI remaining) versus an alternative hypothesis (H_a: at least N_T TOI remain). For a sample size n from a total of N detected

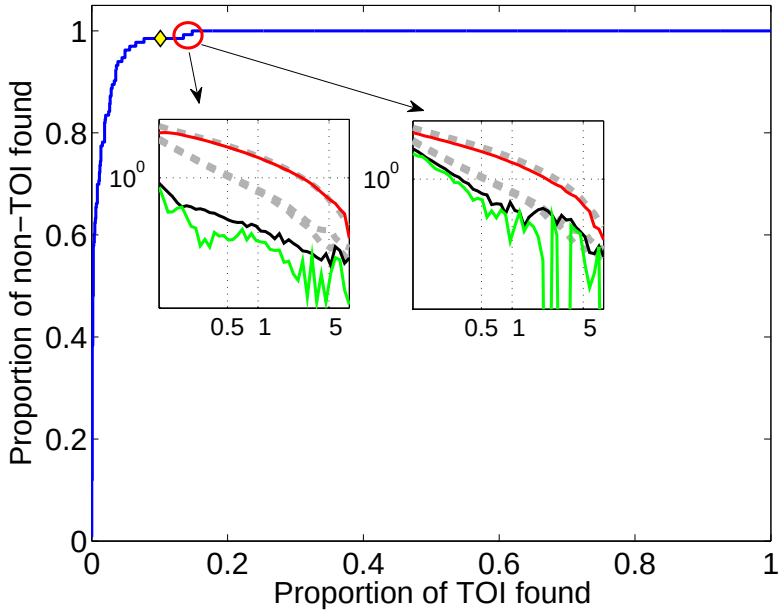


FIGURE 28. Receiver operating characteristic generated by classification of MetalMapper data collected at Camp Beale, CA. Marker indicates the initial stop dig point specified by an analyst. Two outlying TOI occur after the initial stop dig point, highlighted by red ellipse. Inset plots show estimated and reference polarizabilities for the two missed TOI, both targets are ISOs (see figure 29).

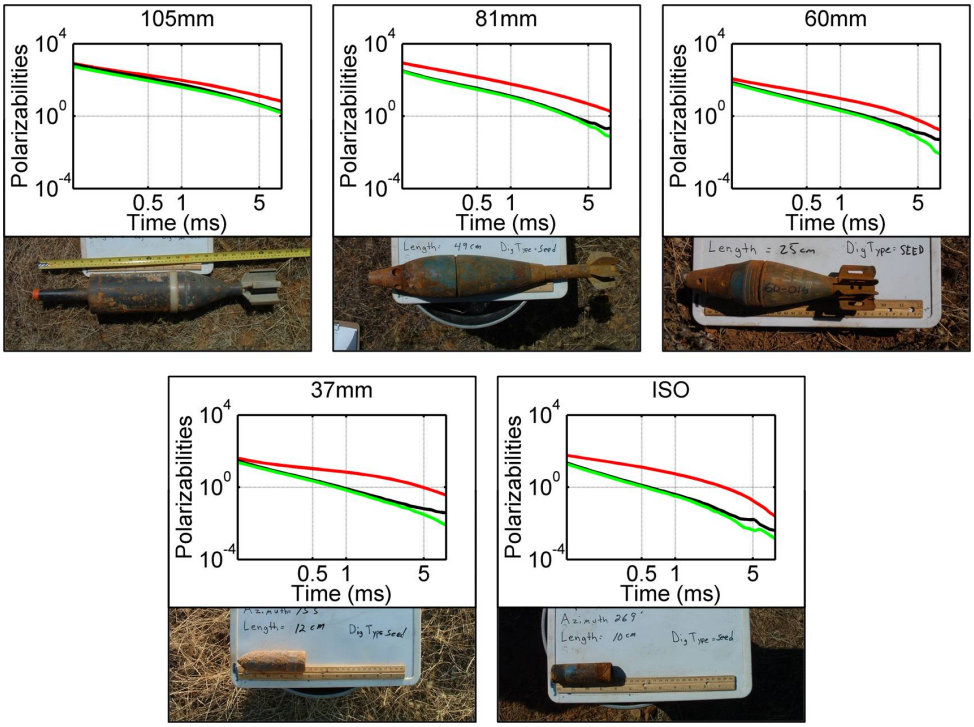


FIGURE 29. Polarizabilities and representative photos of targets of interest at Camp Beale. The industry standard object (ISO) is a 10 cm long section of pipe that is emplaced for quality control of the classification process.

targets, the probability β can be approximated as (Hathaway et al., 2009)

$$\beta \approx \left(1 - \frac{2n}{2N - N_T + 1}\right)^{N_T}. \tag{23}$$

This method makes minimal assumptions about the generating distributions of ordnance and clutter: targets are modelled as discrete samples from a hypergeometric distribution. However, no use is made of the information gained during processing of the geophysical data or during excavation of targets up to the stop dig point. As will be illustrated in the next section (see figure 32), random compliance sampling may require a prohibitive number of samples to achieve an acceptably high confidence.

5.2. **Biased compliance sampling.** The number of compliance samples required to achieve a desired confidence level can be substantially reduced if we replace random sampling with ordered digging of remaining targets, as specified by the classification diglist. In this analysis, we model ordered digging as biased sampling without replacement. The illustration in figure 30 helps to explain the distinction between random and biased compliance sampling.

Whereas random sampling without replacement is associated with the hypergeometric distribution (Hathaway et al., 2009), biased sampling without replacement can be modeled with Wallenius' noncentral hypergeometric distribution (Fog, 2008), which has the form

$$P(\mathbf{x}) = \prod_{i=j}^{2} \binom{\nu_j}{x_j} \int_0^1 \prod_{i=1}^{2} \left(1 - t^{\omega_i/d}\right)^{x_i} dt, \tag{24}$$

with

$$d = \sum_{i=1}^{2} \omega_i(\nu_i - x_i). \tag{25}$$

The vector $\mathbf{x}$ has elements x_i corresponding to the numbers of TOI and non-TOI obtained by random sampling without replacement, from a population initially comprised of ν_i TOI and non-TOI items. The probability of obtaining an item in each class is given by ω_i; unbiased sampling corresponds to $\omega_i = 0.5$, for $i = 1, 2$.

For biased compliance sampling, the odds ω_i can be estimated by applying a ranking method to the subset of targets with available ground truth. The area under the resulting ROC (AUC) - estimated via trapezoidal integration of the ROC curve after normalization of both axes - then corresponds to the probability of correct classification for that ranking method (Hanley and McNeil, 1982).

In practice, we must estimate the bias before generating the ROC and its associated AUC statistic at a given site. We therefore require methods to generate a priori estimates of bias that can be updated as ground truth information becomes available. The dataset degree of difficulty analysis presented in section 4.1 is one avenue for predicting bias for a dataset with minimal ground truth. Figure 31 shows that DDD is predictive of AUC, and so the DDD analysis can be used to generate an initial estimate of bias.

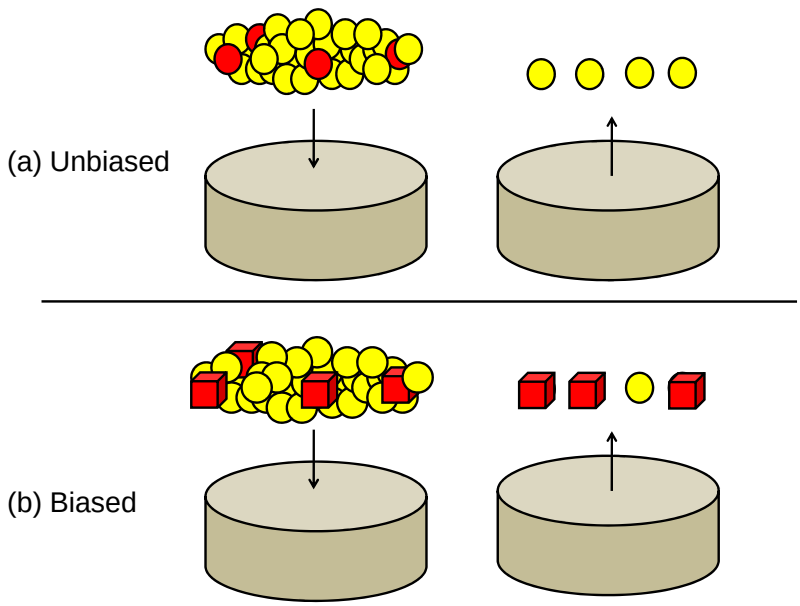


FIGURE 30. Illustration of unbiased and biased compliance sampling. (a) For unbiased sampling we consider a set of balls that can only distinguished by color (red and yellow). The balls are placed in an opaque container. We then randomly sample one ball at a time without replacement, with the result that we are unlikely to choose any red balls. To test whether any red balls are present, we must therefore sample a relatively large proportion of the total set. (b) With biased sampling, we can now distinguish between yellow balls and red cubes in the container on the basis of physical shape. Hence when we sample with the goal of finding cubes, we are very likely to find cubes. Testing whether any red cubes are present in the container will require far fewer samples than the unbiased case.

As in Hathaway et al. (2009), we can then use the noncentral hypergeometric distribution to compute the number of validation samples required to test the null hypothesis (H_o: there are no TOI left in the ground) versus an alternative hypothesis (H_a: there are at least n_T TOI left in the ground), at a specified confidence $1 - \beta$. This analysis yields a very simple and intuitive threshold for setting the stop dig point in a classification dig list: we stop when N_{stop} non-TOI are encountered in sequence in the dig list.

In figure 32 we compare unbiased and biased compliance sampling for the case $n_T < 1$. We consider biased sampling with biases representing the minimum, mean, and maximum AUC derived from classification of all available MetalMapper data sets acquired under the ESTCP demonstration program. In this analysis we employ a relatively conservative strategy that uses only primary (L_1) polarizabilities to classify targets. Biased sampling reduces the proportion of remaining targets that must be sampled by between 50% and 95%, relative to unbiased sampling. By exploiting the information in the geophysical data we can thereby significantly reduce the proportion of targets sampled for validation of the classification process.

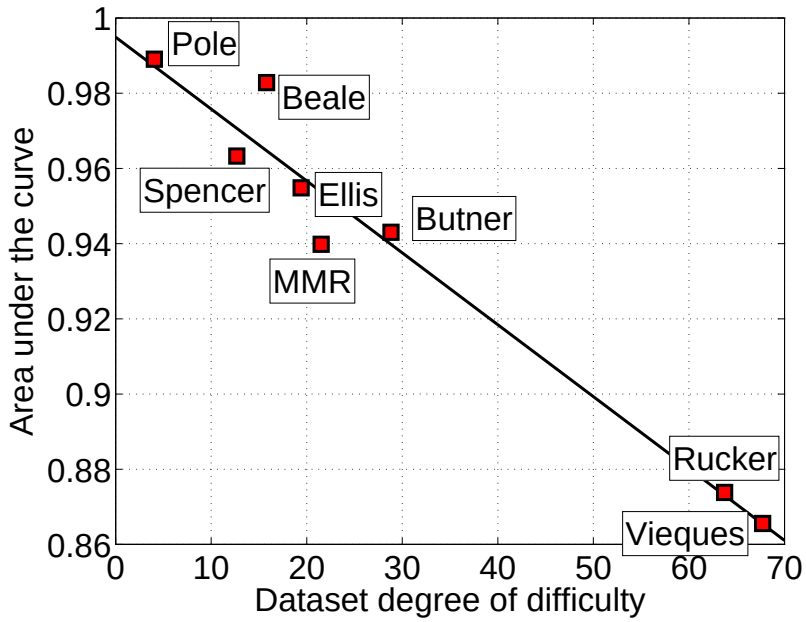


FIGURE 31. Area under the receiver operating characteristic curve versus dataset degree of difficulty for ESTCP MetalMapper data sets. Linear regression (solid line) produces a coefficient of determination of $R^2 = 0.96$. This analysis was carried out using a conservative classification strategy: targets are ranked based on the best match of the estimated primary polarizability (L_1) to primary polarizabilities for reference items.

Figure 33 shows the application of biased compliance sampling to three data sets. We use the DDD for each data set to predict the bias and define a threshold on the number of non-TOI digs encountered in sequence. In the easiest case (Pole Mountain), the stop dig criterion identifies all TOI with only 19 % of non-TOI excavated. This is the ideal outcome for UXO classification, with more than 75 % of the non-TOI left in the ground. For a more difficult site such as MMR, we expect that there will be outliers in the diglist and the final stop dig point identified via biased compliance sampling does not guarantee that all TOI are found. Finally, for a site with a high degree of difficulty such as Vieques, classification is not feasible and all targets must be dug.

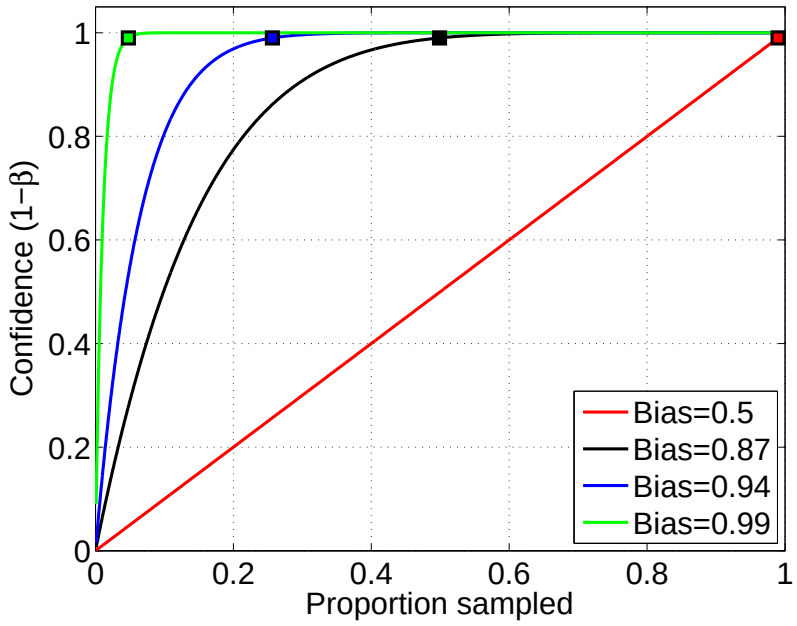


FIGURE 32. Comparison of power $(1-\beta)$ as a function of sample size (expressed as a proportion of the total number of remaining targets) for unbiased ($Bias = 0.5$) and biased ($Bias > 0.5$) compliance sampling. Biased sampling uses bias values representing the minimum (0.87), mean (0.94), and maximum (0.99) bias from retrospective analysis of available MetalMapper data sets. Markers indicate the point on each curve at which 99% confidence that all TOI have been found is achieved.

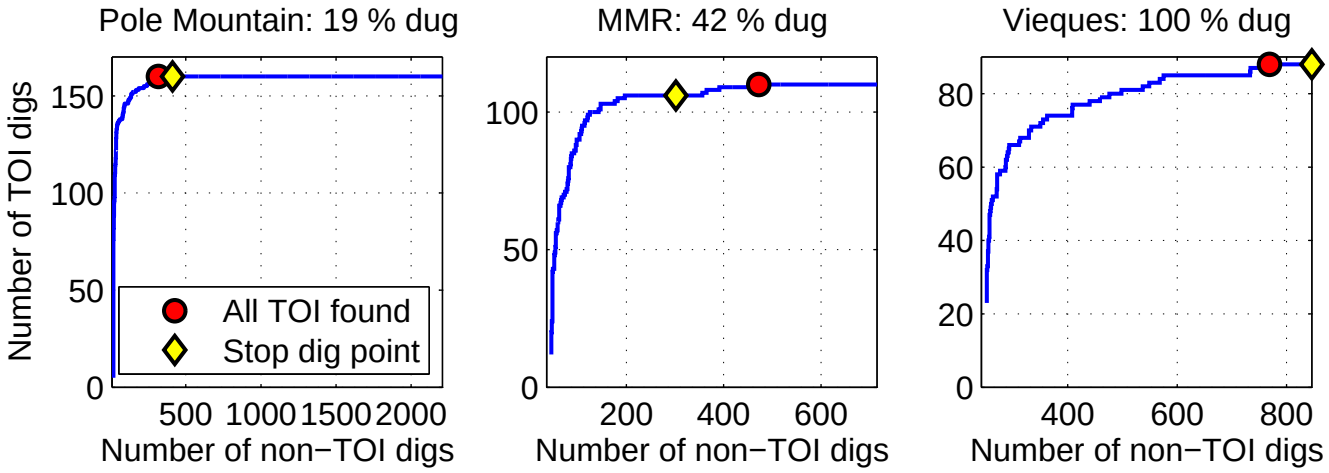


FIGURE 33. Application of biased compliance sampling to three ROC curves from classification of cued MetalMapper data. ROC curves are generated by matching estimated primary polarizabilities with reference polarizabilities. Percentages in each plot title indicate the proportion of non-TOI that are excavated at the final stop dig point.

5.3. **Generative models for risk assessment.** An alternative to compliance sampling for risk assessment is to model the underlying distributions of TOI and non-TOI with respect the classification decision statistic. We term these approaches "generative models", since integration of the class distributions generates the observed ROC.

For example, Walsh et al. (2011) developed a Bayesian approach to risk assessment that accounts for the information obtained during ordered target excavation. They estimate the posterior probability that targets left in the ground are TOI and select a final stop dig point by attaining a desired confidence that no TOI remain. Their confidence predictions are generated by directly fitting a parametric (beta) distribution to the observed distributions of TOI and non-TOI with respect to the decision statistic. Walsh et al. (2011) also formally account for the uncertainty in the proportions of TOI and non-TOI in their confidence calculation .

Similarly, Beran (2014) developed hypothesis tests of the ROC curve that prescribe an objective number of additional excavations required to achieve a specified confidence that all TOI have been found. Both tests fit a "binormal" model to the observed ROC generated from available groundtruth. The binormal model assumes that the distributions of true and false positives (denoted T and F) are normally-distributed. The predicted ROC generated by plotting the cumulative distributions of true and false positives is a function of two parameters (Metz et al., 1988)

$$a = \frac{\mu_T - \mu_F}{\sigma_T}, \; b = \frac{\sigma_F}{\sigma_T}, \tag{26}$$

with μ and σ denoting the mean and standard deviation, respectively.

The binormal assumption is not overly restrictive: when the generating distributions are not normal, the binormal model can often produce a good fit for arbitrary underlying distributions of true and false positives. (Hanley, 1988). This is illustrated in figure 34: the distributions underlying the observed synthetic ROC are decidedly non-normal, but the binormal fit to the ROC is nonetheless quite good.

The estimated binormal model parameters can then be used to calculate the approximate sampling distribution of the point on the ROC at which all TOI are found. Integrating this distribution up to a specified confidence produces a critical point on the ROC that determines how many excavations are required to achieve a confidence threshold. As in Walsh et al., the testing approach is iterative: if additional TOI are found during the validation digs, then the ROC fitting is repeated with the updated ground truth.

Because the ROC is the end product of extensive data processing, these hypothesis tests implicitly use the available information extracted from the geophysical data. This allows us to generate a higher confidence with fewer digs than with compliance sampling. However, the binormal model may not always generate an acceptable fit to the observed ROC. In a retrospective analysis of 32 ROC curves generated from classification of MetalMapper data, the mean correlation coefficient (cc) between observed and predicted ROCs was $\mu(cc) = 0.97$, with standard deviation $\sigma(cc) = 0.03$. Figure 35 shows three examples of high correlations between observed and predicted ROC curves.

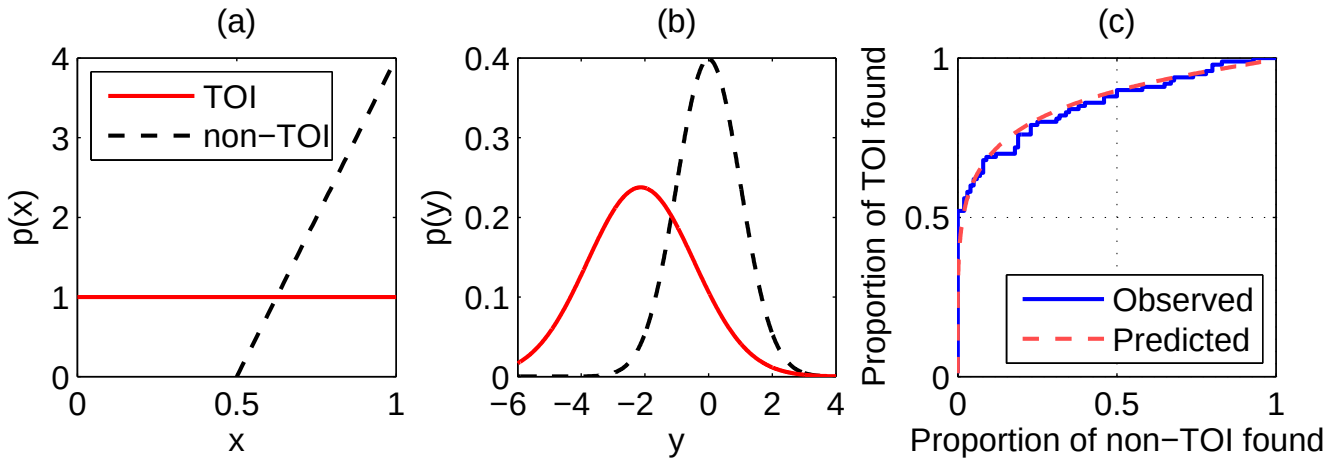


FIGURE 34. Fitting the ROC with a binormal model (after Hanley (1988)). (a) Synthetic distributions of TOI and non-TOI, as a function of decision statistic x. The decision statistic is a single parameter used to rank targets, in this example a small decision statistic indicates that a target is likely a TOI. The synthetic observed ROC in (c) is generated from a sample from these distributions. (b) Normal distributions of TOI and non-TOI estimated by fitting the observed ROC in (c), as a function of transformed decision statistic y. Because we fit the observed ROC directly, rather than the distributions in (a), the monotonic transformation $y = f(x)$ is not required a priori for estimation of the binormal model. (c) Synthetic observed ROC and predicted ROC for estimated binormal model.

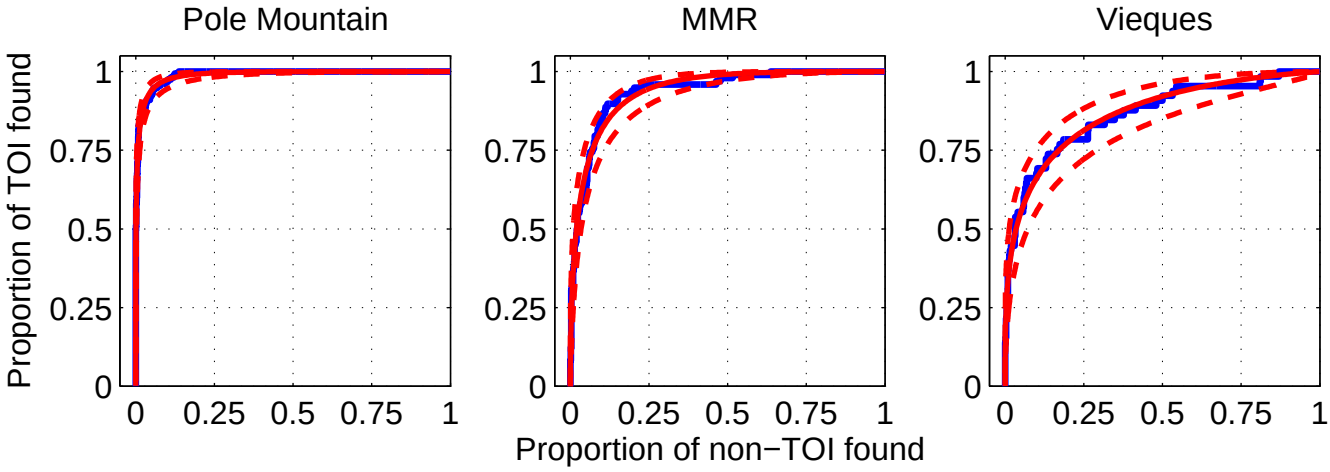


FIGURE 35. Binormal model fits to example ROC curves, all showing high correlation coefficient ($cc = 0.99$ in all cases) between observed (blue) and predicted (red) ROCs. Dashed lines define the 95 % confidence interval on the binormal ROC. All classification examples shown here use matching with the primary polarizability.

In 6 cases we found $cc < 0.95$. These correspond to scenarios where performance is initially quite good, but there are outliers in the last 5% of TOI found that cannot be reliably predicted with the binormal model, e.g. figure 36(a). In addition, there are cases where a change in classification strategy can produce a stepwise (non-smooth) ROC that cannot be fit with the binormal model, as

shown in figure 36(b). In cases where the correlation coefficient (cc) between observed and predicted ROC is low (e.g. $cc < 0.95$), the results should therefore be treated with caution.

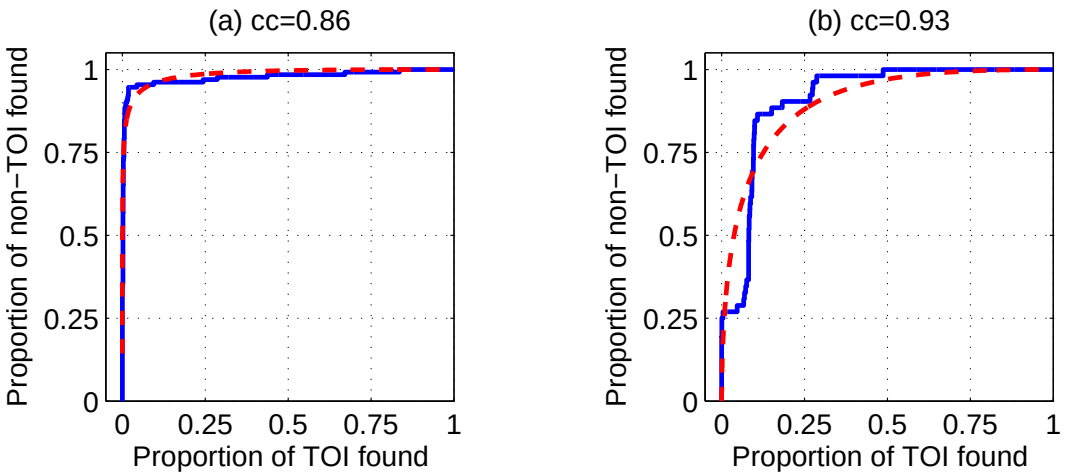


FIGURE 36. Binormal model fits (dashed red lines) to example ROC curves (solid blue lines) showing low correlation coefficients. (a) Classification of Camp Beale MetalMapper data sets with polarizability matching. This dig list uses all three estimated polarizabilities to classify each target, producing a number of outliers in the diglist corresponding to cases with poorly constrained secondary polarizabilities. (b) Classification of Camp Ellis MetalMapper data using a multi-stage approach that uses size/decay features and polarizability matching to rank targets. This approach results in a step-wise ROC curve that cannot be represented with the binormal model.

Because the binormal model does not explicitly depend upon the decision statistic underlying the empirical ROC (see Beran (2014)), it does not require a monotonically-varying decision statistic to determine target ordering. This is useful in cases where a change in classification strategy produces a discontinuous decision statistic that cannot be readily modelled with the approach in Walsh et al. (2011). For example, a common strategy is to use all three polarizabilities to classify high SNR targets, and to then use only the primary polarizabilities for lower SNR targets (Beran et al., 2012).

5.3.1. *Estimating probabilities with the binormal model.* In addition to an ordered ranking of targets, regulators may also require a predicted probability that each target is a TOI as a final output of classification processing. This problem was first addressed in Walsh et al. (2011). They use estimated beta distributions to compute the posterior probability that each target in the diglist is a TOI.

However, in many cases the variable quality of estimated parameters neccessitates a change in classification strategy - and decision statistic - at various points in the diglist (Beran et al., 2012). For example, we may initially classify targets using all three polarizabilities and then revert to using only the primary polarizability for classification of low SNR targets. Consequently the output of classification is not always a single, monotonically-varying decision statistic that corresponds to the ordering of targets in the diglist. Furthermore, some classification algorithms output a discrete, ordinal ranking of targets that cannot be readily modelled with a continuous distribution. This

complicates estimation of the TOI and non-TOI distributions using the approach in Walsh et al. (2011).

In contrast, the binormal model expresses the true positive fraction (TPF: proportion of TOI found) as a function of the false positive fraction (FPF: proportion of non-TOI found):

$$TPF = \Phi[a + b\Phi^{-1}(FPF)] \tag{27}$$

with Φ the standard normal cumulative distribution, and (a, b) the binormal parameters. Because the model has no explicit dependence on the decision statistic, binormal parameter estimates can be obtained for arbitrary ROCs generated from non-monotonic or discrete decision statistics.

Given binormal parameter estimates, we can compute the probability that the ith target in the diglist is a TOI as follows. We can now assume that the normal distributions of TOI and non-TOI underlying the predicted binormal ROC are functions of a continuous, monotonic decision statistic x. The decision statistic x is a pooled sample from TOI and non-TOI classes with distribution

$$p(x) = p(x|TOI)p(TOI) + p(x|non - TOI)p(non - TOI). \tag{28}$$

Here $p(x|TOI)$ denotes the (normal) distribution of TOI, and $p(TOI)$ the prior probability of TOI (similarly for non-TOI). The posterior probability of a TOI at x is

$$p(TOI|x) = \frac{p(x|TOI)p(TOI)}{p(x)}. \tag{29}$$

We assume that the observed ROC is generated by ordering the decision statistic for a sample of size N from $p(x)$. The ith item in this ordered list (the ith order statistic $x(i)$) has the probability distribution Balakrishnan and Cohen (1956)

$$p(x|x(i)) = \frac{N!}{(i-1)!(N-i)!} P(x)^{(i-1)}(1 - P(x))^{(N-i)}p(x), \tag{30}$$

with $p(x)$ and $P(x)$ denoting distribution and cumulative distribution functions, respectively. Marginalizing over x we obtain the posterior probability that the ith target in the ordered list is a TOI

$$P(TOI|x(i)) = \int_{-\infty}^{\infty} p(TOI|x)p(x|x(i))dx. \tag{31}$$

Evaluation of equation 30 is sensitive to numerical errors for larger order statistics ($i > 150$), and so we calculate probabilities in equation 31 via Monte Carlo sampling from the underlying normal distributions. Confidence intervals on probabilities are generated by propagating binormal parameter uncertainties with a linearized uncertainty analysis.

Figure 37 shows predicted probabilities $P(TOI|x(i))$ and associated confidence intervals for targets in the Pole Mountain dig list. At the selected stop dig point, $P(TOI|x(i)) < 0.01$, indicating that targets left in the ground are, with high confidence, non-TOI.

5.4. **Comparison of risk assessment methods.** Figure 38 shows the application of the risk assessment methods discussed here to the Camp Beale classification diglist. We consider a case where an initial stop dig point has missed two near outliers in the diglist. In (a), we show the fit of

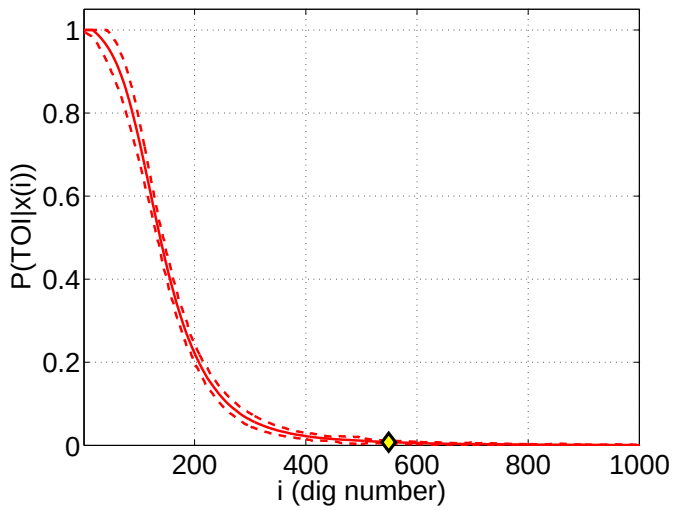


FIGURE 37. Probability that i^{th} dig is a TOI (solid line), for classification of Pole Mountain MetalMapper data. Dashed lines indicate 95% confidence interval on probabilities. Marker indicates stop dig point identified in figure 33.

beta distribution estimates to the empirical cumulative distribution function of the decision statistic for TOI and non-TOI. These estimates are obtained using the decision statistic data up to the initial stop dig point. While the fits in (a) are quite good, the predicted ROC generated by the cumulative beta distributions in (b) is not a good fit to the observed ROC. This is because the parameters of the beta distributions are optimized independently for each class, and the optimizations do not explicitly seek to reproduce the ordering in the classification diglist. In contrast, the binormal model optimizes both distributions simultaneously (via the parameters in equation 26), with a likelihood function designed to reproduce the observed ROC. In this case, the difference in approaches results in the binormal stop dig point identifying the near outliers but requiring an undesirably large number of validation digs. In the next section, we explore how the binormal model can be constrained to increase confidence and reduce the number of digs.

Both the beta model and biased compliance sampling methods find the near outliers in this example. Accounting for model uncertainty is essential with these approaches. As described in Walsh et al., Monte Carlo sampling is used to account for the effect of uncertainties on the the predicted confidence level. Similarly, for compliance sampling we use a nonparametric estimate of the uncertainty in the *AUC* (Hanley and McNeil, 1982) to specify a bias corresponding to the lower bound of a 99% confidence interval.

We recommend that the number of validation digs should first be assessed with a generative model (beta or binormal). If the generative model fails to produce an acceptable fit to the decision statistic (for the beta model) or the ROC (for the binormal model), then biased compliance sampling can be used to objectively determine the required number of digs.

5.5. **Risk assessment with seed items.** An intuitive approach to risk assessment uses seeded items buried across a site for quality control (QC) of the classification process. The identities of QC seeds are initially withheld from the data analysts, and are subsequently revealed once the

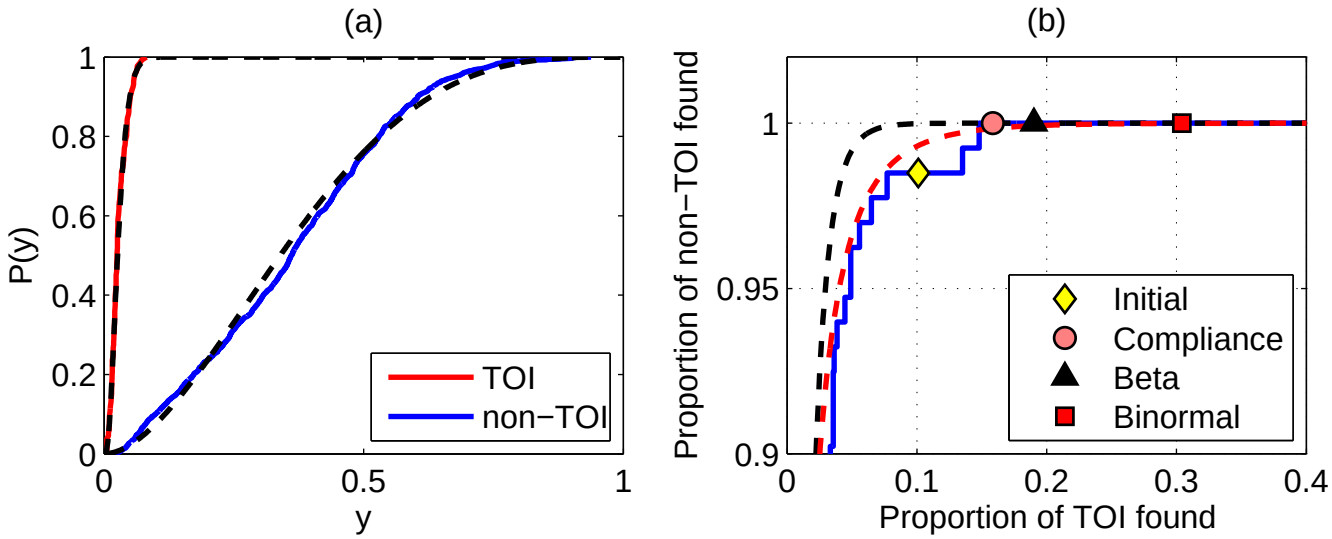


FIGURE 38. Comparison of risk assessment methods applied to Camp Beale classification diglist. (a) Empirical cumulative distributions of the normalized decision statistic y for TOI and non-TOI (solid). Dashed lines are maximum likelihood estimates of cumulative beta distributions. (b) Observed ROC (solid blue) and predicted ROCs for beta (black dashed) and binormal (red dashed) models. Initial stop dig point and stop dig points determined by biased compliance sampling and generative models (beta and binormal) are indicated by markers. All stop dig points achieve a 99% confidence that all TOI are found for the respective method.

classification diglist is finalized. If all QC seeds are identified before the stop dig point, then we have confidence that all actual TOI will also be found via the same processing. Of course, this confidence will depend on the relative difficulty of classifying seeds and TOI.

Seeds are ideally drawn from the same classes of ordnance identified during a remedial investigation (RI). However, all classes may not be identified during the RI, or budget restrictions may prohibit the number and variety of seeds that can be emplaced. This motivates development of objective methods for comparing the relative classification difficulty of QC seeds and TOI, and for quantifying how successful identification of seeds increases confidence in the overall classification process.

At the Camp Beale demonstration all TOI were in fact seeded; no "native" UXO were present in the areas surveyed. For this discussion we therefore treat industry standard objects (ISOs, see figure 29) as QC seeds and all other emplaced targets as TOI. Because ISOs are comparatively easy to obtain, it is likely than many munitions response projects will be primarily seeded with ISOs.

The obvious procedure for comparing classification difficulty between TOI and QC seed items is to compare the receiver operating characteristics generated for each class of item (figure 39). For each target class, the respective ROC is generated by treating only items in that class as true positives. False positives are generated from the same set of non-TOI items in all cases.

At Camp Beale, ISOs are the most difficult target to classify and produce an ROC with maximal false alarm rate. A number of statistical tests can be employed to formalize these observations,

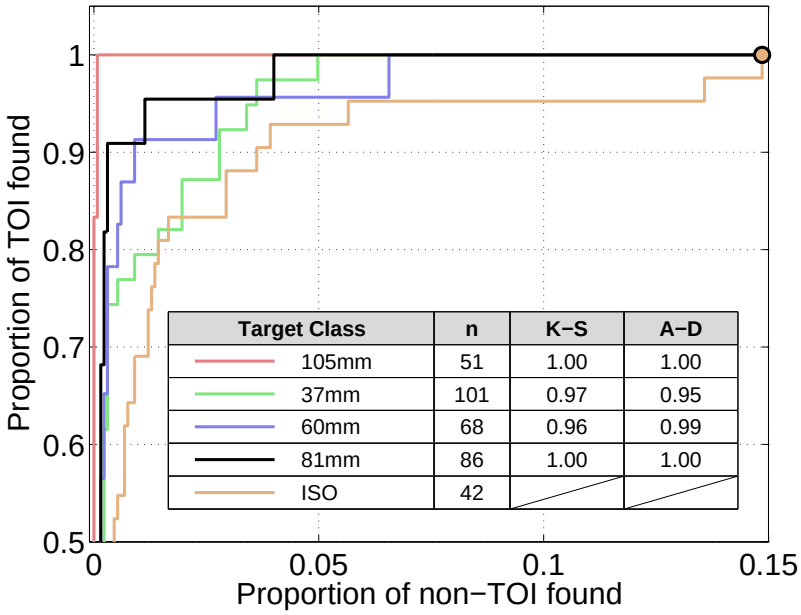


FIGURE 39. ROC curves for ordnance classes at Camp Beale. Dashed line is the combined ROC for all targets. Inset table columns show: number of true positives in each class (n), confidence (1-p) for two sample Kolmogorov-Smirnov (K-S) and Anderson-Darling (A-D) tests. Each statistical test determines whether the distribution of each ordnance class is significantly different from that of seeded ISO.

in figure 39 we give confidence levels for two common non-parametric methods: the Kolmogorv-Smirnov (Press et al., 1992) and Anderson-Darling (Scholz and Stephens, 1987) tests. These tests yield small p-values for the Camp Beale data, indicating that, with high confidence, we can reject the null hypothesis that the underlying distributions of TOI are the same.

The statistical tests confirm our intuition that classification difficulty - and hence the observed ROC curves - will differ between ordnance classes with variable size. However they do not directly address the key question for risk assessment in this context: given that we have found a known proportion of the seeds, what is the probability that we have found all items in each ordnance class?

To answer this question, we develop an extension of binormal ROC estimation that constrains the predicted ROC curve for QC seed items. In figure 40, we fit the ROC for ISO items with unconstrained and constrained models.

The unconstrained model uses the maximum likelihood parameter estimation algorithm described in Metz et al. (1988). Once the labels of QC seeds have been fully revealed, we can constrain the binormal fit by requiring the predicted and observed proportions of true positives to match at the point on the ROC where all seeds are found - as indicated by a marker in figure 40. We do this by progessively increasing the contribution of this point to the likelihood function until the observed and predicted true positive proportions agree within a small tolerance. The resulting constrained binormal ROC is a somewhat poorer overall fit to the observed ROC, but is a much better match at the constraint.

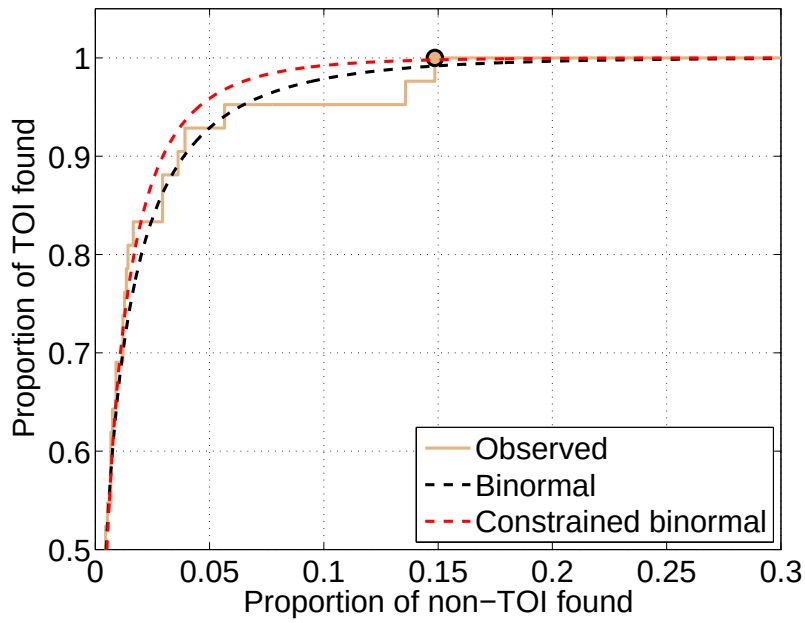


FIGURE 40. Unconstrained and constrained binormal fits for classification of ISO seed items at Camp Beale, CA. Marker indicates constraint point at which all ISOs are found. The constrained binormal fit enforces a close match between observed and predicted ROCs at this point.

We then use the constrained binormal estimates for seed items to model the conditional probability of finding all TOI in class i, given that all seeds have been found, as

$$P(TOI_i|Seed) = hP(TOI_i) + (1-h)P(Seed). \tag{32}$$

Here $P(TOI_i)$ is the probability that all TOI in class i have been found at a specified stop dig point. We use an approximation to this probability described in Beran (2014). If all seeds have been found, then $P(Seed) = 1$, so that the conditional probability in equation 32 satisfies $P(TOI_i|Seed) \geq P(TOI)$.

The weighting h - restricted to the interval $0 \leq h \leq 1$ - quantifies the relative difficulty of classifying TOI and seeds. We define $h(p_{Seed}, p_{TOI_i})$ as the Hellinger distance between the (constrained) normal distribution for seeds (p_{Seed}) and the normal distribution for TOI (p_{TOI}). The Hellinger distance is a measure of the similarity between two probability distributions (e.g. p_A,p_B), and for two normal distributions it is given by (Liese and Miescke, 2008)

$$w(p_A, p_B) = \left[1 - \sqrt{\frac{2\sigma_A\sigma_B}{\sigma_A^2+\sigma_B^2}} \exp\left(-\frac{1}{4}\frac{(\mu_A-\mu_B)^2}{\sigma_A^2+\sigma_B^2}\right)\right]^{1/2}, \tag{33}$$

with μ and σ denoting the mean and standard deviation of each distribution.

Figure 41 illustrates how confidence can be increased by incorporating seed information in the binormal ROC analysis. A final stop dig point can be determined in this analysis by digging until the TOI class with minimal confidence reaches the specified confidence level. We remark that the constrained binormal analysis has the largest effect on confidence for 37mm projectiles. This makes

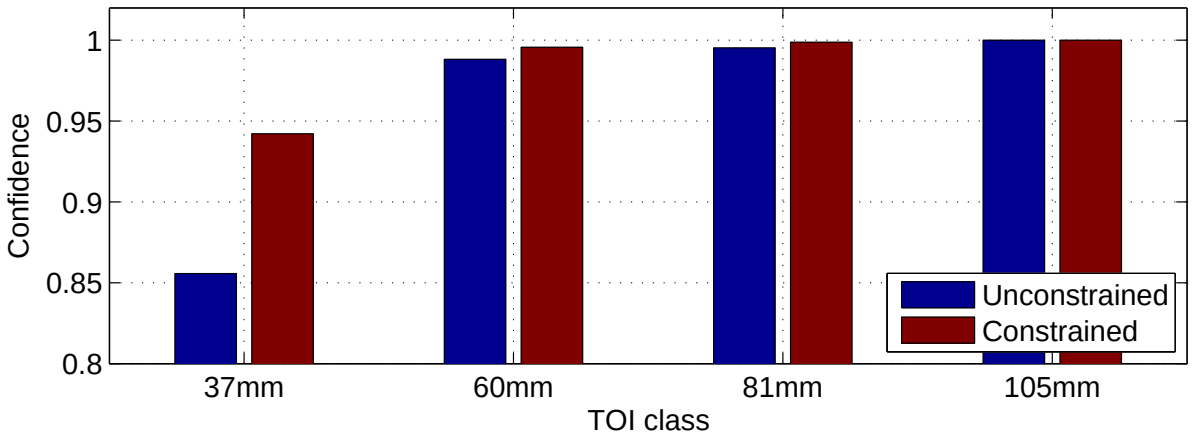


FIGURE 41. Confidence that all TOI in each class have been found at the point in the Camp Beale dig list where all ISO seeds are found. We compare an unconstrained binormal analysis with a constrained analysis that incorporates seed information.

sense: finding all QC seed items has the largest influence on our confidence for TOI that are similar in classification difficulty, as quantified by the Hellinger distance. Conversely, if large and relatively easy items are seeded (e.g. 105 mm projectiles), then finding all of these seeds will have very little effect on the conditional confidence for smaller, more difficult TOI classes.

5.6. **Identifying novel TOI.** The risk assessment methods described above are focused on ensuring that all targets of interest belonging to known classes (e.g. 37 mm projectiles at Camp Beale) have been identified in the classification diglist. However, a risk assessment must also determine whether any TOI classes have been altogether missed in the classification diglist. In this section we discuss two approaches to prioritizing targets for identification of novel TOI.

5.6.1. *Matching with a comprehensive library.* A necessary check to safeguard against missed TOI classes is comparison against an ordnance library compiled from past projects. Figure 42 shows library matching results for the Camp Beale data set.

The library of reference polarizabilities is sensor specific, in this example we use TOI polarizabilities from all previous MetalMapper demonstration projects dating back to 2010. We match polarizabilities using the heuristic misfit function

$$\phi = \sum_{k=1}^{3} w_k \left[\sum_{i=1}^{N} \left(\frac{(L_k^{est}(t_i))^\gamma - (L_k^{ref}(t_i))^\gamma}{1/N \sum_{j=1}^{N} (L_k^{ref}(t_j))^\gamma} \right)^2 \right]^{1/2} \tag{34}$$

with $\gamma \approx 0.1$ and w_k a weighting applied to each of the polarizabilities. An estimated polarizability L^{est} is considered a match to a library polarizability L^{ref} if the misfit is less than a predefined threshold ϕ_{max}. The threshold can be estimated from the observed distribution of ϕ for TOI at the site. In this case we use $\phi_{max} \approx 0.4$ with weights $\mathbf{w} = (1, 1/2, 1/2)$.

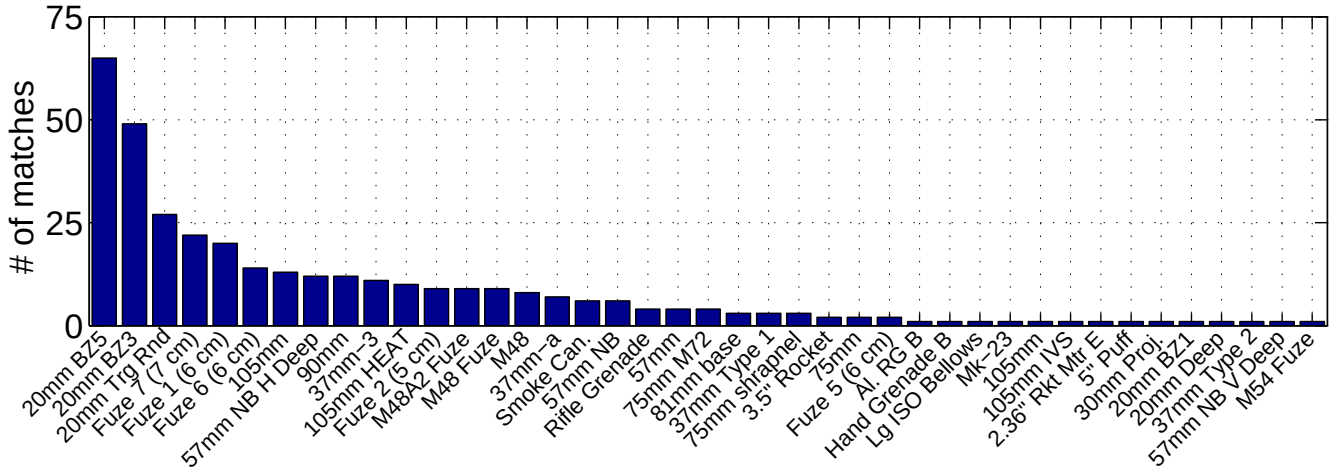


FIGURE 42. Number of matches to each class of item in MetalMapper polarizability library. Only likely non-TOI (i.e targets past the final stop dig point) in the Camp Beale data set are considered in this analysis. A total of 350 targets have an acceptably small misfit (equation 34) with respect to library polarizabilities.

In figure 42, small library items (e.g. 20 mm projectiles) generate the most number of matches at this threshold. Although no 20 mm projectiles were present at Camp Beale, small, fast-decaying non-TOI can produce an acceptably small misfit to this library entry. Naïve comparison against a full library can thereby generate an large number of matches. Where historical records can confidently constrain the ordnance present at a site, the library may be winnowed to consider only plausible ordnance for that site.

5.6.2. *Model metrics.* An alternative to library matching is to use metrics derived from the estimated model to identify potential TOI items. In this analysis we do not match the estimated polarizabilities against a library, but instead use features of the polarizabilities that were diagnostic of TOI at previous sites. For example, TOI are typically characterized by large amplitude, slow decaying polarizabilities relative to non-TOI. In addition, the rotational symmetry of most ordnance is manifested as a small relative deviation between transverse (L_2, L_3) polarizabilities. Table 2 summarizes metrics that exploit these properties - and others that help distinguish TOI from non-TOI - and gives a range of each metric for TOI detected with the MetalMapper.

Metric (m_i)	**Description**	**Formula**	**TOI range** (A_i)
Size	Amplitude of pol.	$\text{size}(t_i) = [L_1^2(t_i) + L_2^2(t_i) + L_3^2(t_i)]^{1/2}$	$1.5 <$ $\log_{10}(\text{size}(t_1)) < 5$
Decay	Decay rate of pol.	$\text{decay}(t_m, t_n) = \text{size}(t_m)/\text{size}(t_n)$	$0.02 <$ $\text{decay}(t_{29}, t_1)$
Model contribution (mc)	Contribution of TOI model i to predicted data for inversion with M objects	$\text{mc} = 100 \times [(d_i^{pred})^2 / \sum_{k=1}^{M} (d_k^{pred})^2]$	$75 <$ $\text{mc} < 100$
Alpha	Difference between log-transformed axial and mean transverse pols.	$\alpha_L(t_i) = 1/3 \lvert \log_{10}\left(\frac{L_1(t_i)}{(L_2(t_i)L_3(t_i))^{1/2}}\right) \rvert$	$0.4 <$ $\sum_{t_i=t_1}^{t_{29}} \frac{\alpha_L(t_i)}{t_{29}-t_1+1}$
Beta	Difference between log-transformed transverse pols.	$\beta_L(t_i) = 1/2 \lvert \log_{10}(L2(t_i)) - \log_{10}(L3(t_i)) \rvert$	$0.1 <$ $\sum_{t_i=t_1}^{t_{29}} \frac{\beta_L(t_i)}{t_{29}-t_1+1} < 0.2$
Skewness	Skewness of pol. tensor	$\text{skewness}(t_m, t_n) = \sum_{t_i=t_m}^{t_n} \frac{2\alpha_L(t_i)(\alpha_L(t_i)^2 - \beta_L(t_i)^2)}{t_m - t_n + 1}$	$0 <$ $\text{skewness}(t_1, t_{29}) < 0.11$
Variance	Variance of pol. tensor	$\text{variance}(t_m, t_n) = \sum_{t_i=t_m}^{t_n} \frac{2/3(3\alpha_L(t_i)^2 + \beta_L(t_i)^2)}{t_m - t_n + 1}$	$0.5 <$ $\text{variance}(t_1, t_{29})$
Asymmetry	Relative difference between transverse pols.	$\text{asymmetry}(t_m, t_n) = \frac{\sum_{t_i=t_m}^{t_n} [L_2(t_i) - L_3(t_i)]}{\sum_{t_i=t_m}^{t_n} L_2(t_i)}$	$0.2 <$ $\text{asymmetry}(t_4, t_{20}) < 0.4$
Shape	Difference between transverse pols. as a function of axial pol.	$\text{shape}(t_m, t_n) = \frac{\sum_{t_i=t_m}^{t_n} [L_2(t_i) - L_3(t_i)]}{\sum_{t_i=t_m}^{t_n} L_1(t_i)}$	$0.15 <$ $\text{shape}(t_1, t_{42}) < 0.2$
Aspect ratio (ar)	Ratio of axial to mean transverse pols.	$\text{ar}(t_m, t_n) = \frac{\prod_{t_i=t_m}^{t_n} L_1(t_i)}{[\prod_{t_i=t_m}^{t_n} L_2(t_i) L_3(t_i)]^{1/2}}$	$50 <$ $\text{ar}(t_1, t_{42}) < 60$
Pol jitter	Total variation of pols. between adjacent channels	$\text{jitter} = \sum_{t_i=t_1}^{t_{N-1}} \sum_{k=1}^{3} \log_{10}\left(\frac{L_k(t_{i+1})}{L_k(t_i)}\right)$	$12 <$ $\text{jitter} < 15$
L_2, L_3 separation	Average separation between transformed transverse pols.	$\text{separation}(\gamma) = \text{mean}(\lvert L_2(t_i)^\gamma - L_3(t_i)^\gamma \rvert) \times \text{mean}(\lvert L_2(t_i)^\gamma \rvert)$	$0.05 <$ $\text{separation}(0.09) < 0.1$
L_2, L_3 amplitude	Mean amplitude of log transformed transverse pols.	$\text{amplitude} = \sum_{t_i=t_1}^{t_N} \frac{\log_{10}[L_2(t_i) \cdot L_3(t_i)]}{2}$	$-50 <$ $\text{amplitude} < 120$
Depth	Estimated target depth		$0.01 <$ $z < 0.9$

TABLE 2. Metrics for identifying novel targets of interest. We abbreviate polarizability as "pol". Axial polarizability - aligned with the axis of symmetry of the target - is L_1 and transverse polarizabilities are L_2 and L_3. The ith time channel of the MetalMapper sensor is indicated as t_i. In our processing we use 42 time channels extending between $t_1 = 0.106$ ms and $t_{42} = 7.912$ ms.

All TOI in MetalMapper data sets acquired at ESTCP demonstration sites between 2010 and 2013 were used to retrospectively determine the representative range of values for each metric. We combine these metrics, each denoted m_i, into a single parameter

$$\tau = \sum_{i=1}^{M} \mathbb{t}(m_i) \tag{35}$$

with the indicator function

$$\mathbb{t}(m_i) = \begin{cases} s(m_i, a_i, c_i) & m_i \in \mathrm{A}_i, \\ 0 & m_i \notin \mathrm{A}_i. \end{cases} \tag{36}$$

The range of expected values for each metric, denoted A_i, is given in table 2. The sigmoidal function is

$$s(x, a, c) = \frac{1}{2}\left(\frac{1}{1 + \exp\left[a(x - c)\right]} + 1\right). \tag{37}$$

The parameters a_i and c_i are selected to smoothly transition $s(m_i, a_i, c_i)$ between 0.5 and 1 as m_i moves from the boundaries of the range A_i to values indicating a high likelihood of TOI. The resulting score τ can then be used to prioritize targets independently of the polarizability misfit to a reference library.

Figure 43 shows the performance of the τ metric ranking for the Camp Beale data, together with example polarizabilities that illustrate the model features prioritized by this metric. Although the classification performance is not as good as ranking based on a polarizability match to a library, the τ metric does a decent job at prioritizing TOI without site-specific information. In this example, a fast-decaying 50 caliber item that is not in the reference library is identified relatively early in the τ metric diglist. This item would not necessarily be found in a diglist based entirely on library matching. We remark however, that some TOI do occur late in the τ metric diglist: polarizabilities for an ISO item in figure 43 are an example of this scenario. This target is one of the near outliers previously encountered in the library matching diglist in figure 28. The τ metric analysis puts a low priority on this target because of its low amplitude, noisy secondary polarizabilities.

Retrospective analysis of available MetalMapper datasets showed that the τ metric finds all TOI with a false alarm rate that is, on average, 10 % larger than is achieved with polarizability matching. We emphasize that library matching and library-independent metrics are complementary approaches for identifying known TOI and novel TOI that may not be in an ordnance library. Ordered sampling from the τ diglist, possibly guided by biased compliance sampling to determine sample size, can provide validation of a library matching diglist.

5.7. **Software for risk assessment.** Following on work presented in this section, we have developed simple, web-accessible tools to support risk assessment. The first tool is a utility for biased compliance sampling analysis. Given an estimate of the classifier bias, the program outputs the number of ordered samples required to achieve a specified confidence that fewer than n_T TOI left in the ground.

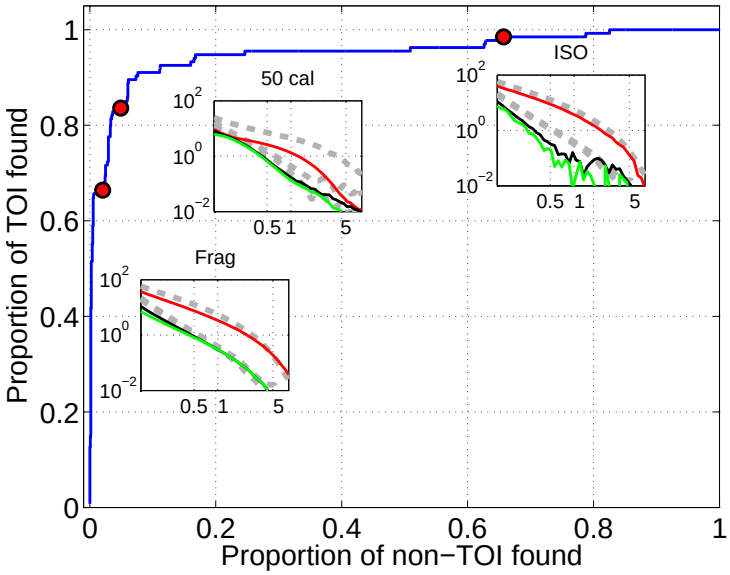


FIGURE 43. ROC for τ metric applied to Camp Beale MetalMapper data. Inset plots show polarizabilities for example targets indicated on the ROC curve, ground truth for each target is indicated by plot title. For comparison, reference polarizabilities for closest-matching library TOI polarizabilities are shown in each inset plot, as dashed gray lines. No library polarizabilities are used in the target ranking.

A second online tool allows upload of ROC data in csv format for non-parametric estimation of the AUC and associated standard error, as well as fitting of the ROC with a binormal model. The posterior probabilities of TOI for all items in the ordered dig list are estimated using order statistics (see section 5.3.1) and can be output to a csv file.

Both of these tools are currently hosted at www.btgeophysics.com/Risk. The underlying computations are carried out using the open-source **R** statistical package, so the underlying source code can also be run locally if necessary.

5.8. **Conclusions: risk assessment.** We have developed and compared a number of approaches for identifying near outliers in the classification diglist. Compliance sampling or generative models can be used to determine the number of excavations required to ensure that all TOI are found at a specified confidence. The latter make stronger assumptions and could be used as a first approach. The biased compliance sampling method developed under this project recasts selection of the stop dig point as a simple criterion on the number of non-TOI digs encountered in sequence. This is appealing in its simplicity and may therefore be useful for justifying this decision point to regulators and stakeholders.

We have extended binormal analysis to the scenario where seeded QC items have been emplaced at the site. When all seeds have been found, an additional constraint can be imposed on the predicted binormal ROC. This has the effect of reducing the number of required excavations, in particular for TOI classes that are similar in size to seed items.

Risk assessment must also consider that novel TOI classes have been altogether missed in the initial training and classification stage. Comparison with a comprehensive TOI library is one strategy to prevent this outcome, but can often generate a large number of false positives to library entries that are not present at the site in question. Furthermore, library matching will likely fail to find items that are not in the library. For this reason, we prefer to prioritize validation digs using a metric designed to identify TOI independently of a match to a library. The τ metric described here is an ad hoc measure that combines a number of metrics (polarizability amplitude, decay, asymmetry, etc.) into a single number that can be used to prioritize validation digs.

Budgetary restrictions may often preclude the additional digs required by the statistical analyses developed herein. Indeed, we expect that the number of validation digs will most often be dictated by the budget, and further work could explore how a limited subset of targets should be selected to provide the best validation of the diglist.

6. Conclusions and Implications for Future Research/Implementation

Extending initial work on modelling detection thresholds for monostatic sensor arrays, we have developed a standalone software tool for the prediction of detection thresholds for MetalMapper and TEMTADS2x2 sensors (section 3.4). The prevailing approaches for target picking with these sensors are model-based (e.g. the dipole filter), but a simple analysis using predicted amplitudes of the observed data is an important QC check on informed source selections. When target picking is carried out on the data, the channel selection procedure presented in section 3 can help to identify a time channel, or linear combination of time channels, producing optimal SNR for detected TOI.

This project initially envisioned a decision support system (DSS) that could be used by site managers and project geophysicists to predict classification performance under assumed conditions. To this end, we developed efficient methods for propagating noise from observed data to estimated polarizabilities, and finally to predict our ability to correctly classify specified items (Appendix B). However, this detailed modelling approach requires significant a priori assumptions regarding noise, target density, and the polarizabilities of TOI and non-TOI present at a site.

A second, more practical approach to performance prediction considered how various data and model metrics can be combined to assess the feasibility of classification given an observed multistatic dataset. Using all available cued MetalMapper data sets acquired under the ESTCP demonstration program, we developed the "Dataset Degree of Difficulty" to objectively assess classification difficulty. This is an ad hoc measure that aimed to encode an expert analyst's judgment. While the DDD can provide a useful indication of relative classification difficulty, retrospective classification of MetalMapper data sets indicated that this approach could not reliably predict the false alarm rate at a site. To address this, we used regularized logistic regression to directly learn the relationship between dataset metrics (e.g. SNR, polarizability misfits, etc.) and classification performance. Initial tests with new datasets suggest that this is a useful approach for predicting classification performance.

The simplest, most intuitive approach to risk assessment for munitions response is to continue ordered digging until no more TOI are encountered. The threshold N_{stop} (the number of non-TOI encountered in sequence) dictates the final stop dig point. This value will, in many cases, be determined by budgetary or regulatory constraints. When necessary, the risk assessment methods developed under this project can provide objective, statistical criteria to justify the specification of N_{stop}. For example, biased compliance sampling can be used to adjust N_{stop} based on the difficulty of the classification problem.

References

N. Balakrishnan and A. C. Cohen. *Order statistics and inference: estimation methods*. Academic Press, 1956.

T. Bell and B. Barrow. Bulk magnetization effects in EMI-based classification and discrimination, SERDP project MR-1711. In *SERDP In-progress Review Meeting*, 2014.

T. Bell, B. Barrow, and J .T. Miller. Subsurface discrimination using electromagnetic induction sensors. *IEEE Trans. Geosci. Remote Sensing*, 39:1286–1293, 2001.

L. Beran. Hypothesis tests to determine if all true positives have been identified on a receiver operating characteristic curve. *Journal of Applied Statistics*, 41(6):1332–1341, 2014.

L. Beran and B. Zelt. Risk assessment for unexploded ordnance remediation. *Stochastic environmental research and risk assessment*, 2014a.

L. Beran and B. Zelt. Setting the stop dig point for unexploded ordnance remediation. 2014b.

L. Beran, B. Zelt, L. Pasion, S. Billings, K. Kingdon, N. Lhomme, L.-P. Song, and D. Oldenburg. Practical strategies for classification of unexploded ordnance. *Geophysics*, 78, 2012.

L. Beran, L. P. Song, S. Billings, B. Zelt, and D. Oldenburg. Addendum to final report: Robust statistics and regularization for feature extraction and uxo discrimination, SERDP project MR-1629. Technical report, SERDP, 2013a.

L. Beran, B. Zelt, and S. Billings. Interim report: Decisions support tools for munitions response performance prediction and risk assessment. SERDP project MR-2226. Technical report, SERDP, 2013b.

L. S. Beran and S. D. Billings. Unexploded ordnance detection with multi-static, multi-component electromagnetic sensors. *Geophysics*, 2016.

J. O. Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.

S. D. Billings and L. S. Beran. Optimizing electromagnetic sensor arrays for unexploded ordnance detection. *Geophysics*, 2016.

J. F. Cardoso. Jacobi angles for simultaneous diagonalization. *SIAM journal on matrix analysis and applications*, 17:161–163, 1996.

A. Fog. Calculation methods for Wallenius' noncentral hypergeometric distribution. *Communications in Statistics: Simulation and Computation*, 37(2):258–273, 2008.

J. A. Hanley. The robustness of the "binormal" assumptions used in fitting ROC curves. *Medical Decision Making*, 8:197–203, 1988.

J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic. *Radiology*, 143:29–36, 1982.

J. Hathaway, R. Gilbert, J. Wilson, and B. Pulsipher. Evaluation of spatially clustered ordnance when using compliance sampling surveys after clean-up at military training sites. *Stoch. Environ. Res. Risk Assess.*, 23:253–261, 2009.

F. Liese and K-J Miescke. *Statistical decision theory: estimation, testing, and selection*. Springer, 2008.

C. E. Metz, B. E. Herman, and J. Shern. Maximum likelihood estimation of reciever operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, 17:1033–1053, 1988.

L. R. Pasion. Selecting optimal models for inverting EMI data. In *SERDP In-progress Review Meeting*, 2012.

L. R. Pasion, S. D. Billings, D. W. Oldenburg, and S. Walker. Application of a library-based method to time domain electromagnetic data for the identification of unexploded ordnance. *Journal of Applied Geophysics*, 61:279–291, 2007.

L. R. Pasion, B. C. Zelt, K. A. Kingdon, and L. S. Beran. Feature Extraction and Classification of EMI Data, Camp Beale, CA, ESTCP Project MR-201004. Technical report, ESTCP, 2012.

W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, 1992.

F. W. Scholz and M. A. Stephens. K-sample Anderson-Darling tests. *Journal of the American Statistical Association*, 82(399):918–924, 1987.

F. Shubitidze, J. P. Fernández, I. Shamatava, B. E. Barrowes, and K. ONeill. Joint diagonalization applied to the detection and discrimination of unexploded ordnance. *Geophysics*, 77(4):WB149–WB160, 2012.

F. Shubitidze, B. E. Barrowes, Y. Wang, I. Shamatava, J. B. Sigman, and K. O"Neill. Advanced EMI models for survey data processing: targets detection and classification. In *Proceedings of SPIE: Detection and Sensing of Mines, Explosive Objects, adn Obscured Targets*, 2016.

L. P. Song, D. W. Oldenburg, L. R. Pasion, and S. D. Billings. Determining equivalent dipole number by information theoretic criteria. In *Proceedings of the UXO Forum*, 2009.

L. P. Song, L. R. Pasion, S. D. Billings, and D. W. Oldenburg. Nonlinear inversion for multiple objects in transient electromagnetic induction sensing of unexploded ordnance: Technique and applications. *IEEE Trans. Geosci. Remote Sensing*, 49(10):4007 –4020, 2011.

L. P. Song, D. W. Oldenburg, and L. R. Pasion. Estimating source locations of unexploded ordnance using the multiple signal classification algorithm. *Geophysics*, 77(4):WB127–WB135, 2012.

J. R. Wait. *Geo-electromagnetism*. Academic Press, 1982.

S. Walsh, K. Anderson, J. Hathaway, and B. Pulsipher. A Bayesian approach to monitoring and assessing unexploded ordnance remediation progress from munitions testing ranges. *Stoch. Environ. Res. Risk Assess.*, 25(6):805–814, 2011. ISSN 1436-3240.

D. Williams, Y. Yu, L. Kennedy, X. Zhu, and L. Carin. A bivariate Gaussian model for unexploded ordnance classification with EMI data. *IEEE Geosci. Remote Sensing Letters*, 4:629–633, 2007.

H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

Appendix A. Project publications and presentations

(1) L. Beran and S. Billings. Interim Report (1/2): Decision support tools for munitions response performance prediction and risk assessment. SERDP project MR-2226, 2012.

(2) L. Beran, S. Billings and B. Zelt. Interim Report (2/2): Decision support tools for munitions response performance prediction and risk assessment. SERDP project MR-2226, 2013.

(3) L. Beran. Hypothesis tests to determine if all true positives have been identified on a receiver operating characteristic curve. *Journal of Applied Statistics*, 41(6): 1332–1341, 2014.

(4) L. Beran and B. Zelt. Risk assessment for unexploded ordnance remediation. *Stochastic environmental research and risk assessment*, 2014.

(5) L. Beran and B. Zelt. Setting the stop dig point for unexploded ordnance remediation. *Proceedings of IGARSS14*, 2014.

(6) L. Beran, L. Pasion, and B. Zelt. Assessing the quality of dynamic and cued EMI data for UXO classification. Presented at *SAGEEP 2015*, Austin, TX.

(7) L. Beran and D. Lutes. Optimizing UXO detection with multistatic sensors. Presented at *SAGEEP 2016*, Denver, CO.

(8) S. Billings and L. Beran. Optimizing electromagnetic sensors for unexploded ordnance detection. submitted to *Geophysics*.

(9) L. Beran and S. Billings. Unexploded ordnance detection with multistatic, multi-component electromagnetic sensors. submitted to *Geophysics*.

Appendix B. Uncertainty analysis and performance prediction for unexploded ordnance classification

Uncertainty analysis and performance prediction for unexploded ordnance classification

Laurens Beran*

Abstract

Data processing for detection and classification of buried unexploded ordnance (UXO) requires careful quantification of data and model uncertainties. The ultimate goal of this data processing is to identify and excavate all buried munitions with a minimum number of false positives. This is accomplished via parametric inversion of geophysical data acquired over each detected target. The inferred model vector is comprised of both extrinsic target parameters (location and orientation) and intrinsic parameters related to physical properties of the target. These estimated intrinsic model parameters - termed polarizabilities - are used to make inferences about the class of each detected target (UXO/non-UXO) and to prioritize excavations.

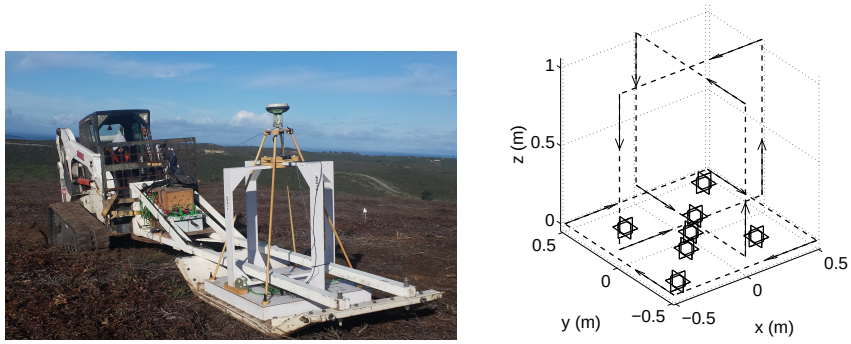
We develop efficient methods for predicting uncertainty in estimated polarizabilities and subsequent classification performance. Current approaches to this problem use Monte Carlo (MC) simulations that require computationally intensive nonlinear inversion of multiple synthetic data sets. In contrast, our efficient MC algorithm replaces nonlinear inversion with a parametric approximation to the posterior distribution of estimated target locations. We then quantify expected classification performance by considering binary classification of a specified UXO and non-UXO. Using the predicted MC polarizability distributions for both targets, we estimate the probability of correct classification. These methods thereby provide means to assess the feasibility of classification under prescribed, site-specific conditions. These conditions include: sensor noise, target depth, the spatial separation between neighboring targets, and the subset of estimated parameters used to classify targets.

1 Introduction

Unexploded ordnance (UXO) contamination is a widespread environmental problem that occurs both at former defence sites within the United States and in former conflict zones around the world. Buried munitions can be readily detected with handlheld metal detectors, but benign metallic clutter often produces a large number of false alarms. This makes detection with conventional metal detectors inefficient and prohibitively expensive (Delaney and Etter, 2003). Electromagnetic sensors that record digital data and GPS locations over

*Black Tusk Geophysics, #401-1755 W Broadway, Vancouver, BC V6J 4S5 (laurens.beran@btgeophysics.com).

each detected target have a substantially improved capability to discriminate between UXO and clutter. Figure 1 shows an example advanced sensor, the Geometrics MetalMapper, designed specifically for UXO detection and classification (Prouty et al., 2011).



(a) (b)

Figure 1: The Metalmapper electromagnetic sensor. (a) Field deployment of the MetalMapper. (b) Sensor geometry. Transmitters and receivers are indicated by dashed and solid lines, respectively. Receivers are seven "cubes", each measuring three orthogonal components of the scattered fields radiated by a buried target.

The MetalMapper is typically deployed in a "cued interrogation" mode: previously-detected targets from a site-wide survey are revisited with a stationary sensor. The instrument radiates electromagnetic fields from three orthogonal transmitters and measures vector components of the time-varying field scattered by buried metallic targets on multiple receivers.

Subsequent processing involves fitting a physical model to the observed data acquired over each target. The model is usually parameterized in terms of target location, orientation, and intrinsic parameters that can be used to make classification decisions. As illustrated in figure 2, the intrinsic, time-dependent parameters, termed *principal polarizabilities*, are matched with those of known UXO to generate a prioritized diglist of targets for excavation. The non-negative polarizabilities are related to physical properties of the target (size, shape, and material composition) and so are useful features for classification of buried targets (Bell et al., 2001; Pasion and Oldenburg, 2001). A dominant, primary polarizability, denoted L_1, corresponds to excitation of a steel target along its long axis. Secondary and tertiary polarizabilities (L_2, L_3) represent the target response along axes perpendicular to the primary axis. As shown in figure 2, the estimated L_2 and L_3 are smaller in magnitude and can be strongly affected by noise in the data at later times (> 1 ms). This variability may impair classification and motivates the present investigation of polarizability uncertainty.

Figure 3 illustrates the dependence of polarizability estimates on site-specific noise. We compare the ensembles of estimated polarizabilities for industry standard objects (ISOs) emplaced at two demonstration sites. These items are short (approx. 10 cm in length) sections of steel pipe that are used for quality control of the classification process. The increased variability of polarizabilities at Camp Beale, relative to Pole Mountain, is primarily due to increased background noise variance at the former site.

In this article, we consider how the uncertainty in estimated polarizabilities and sub-

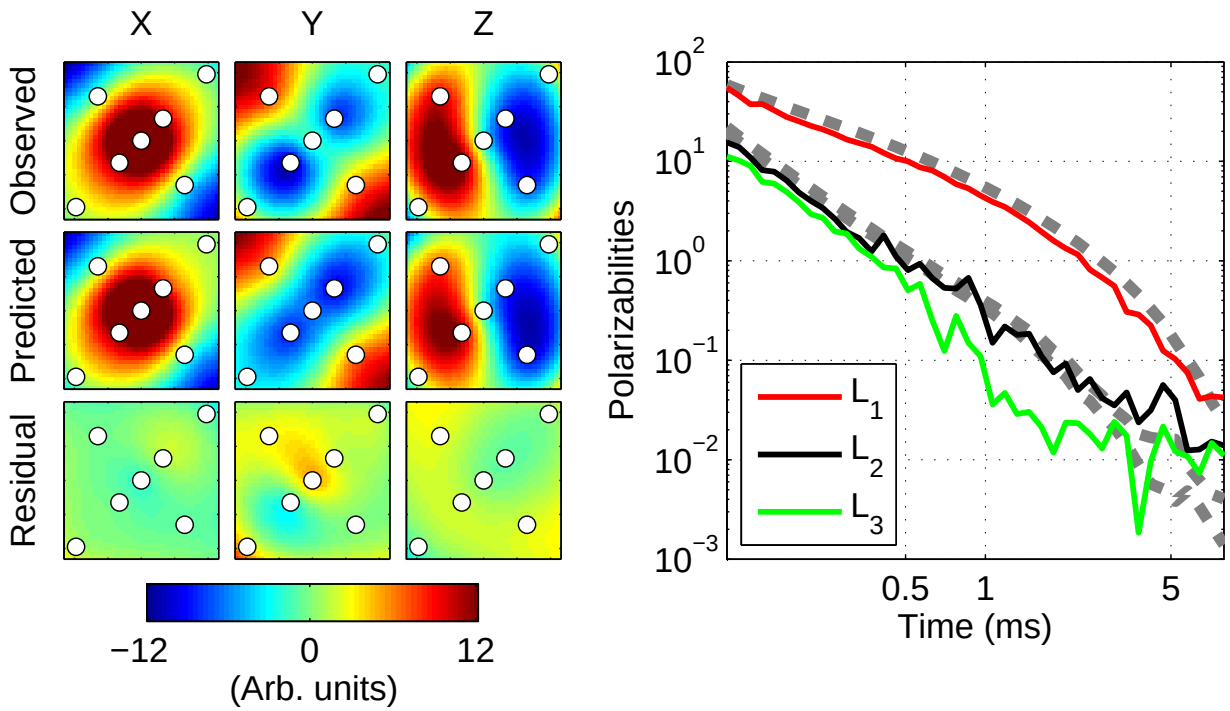


Figure 2: Fitting MetalMapper data. Left: Observed data at the first time channel (top row), and data predicted by fitting a physical model (middle row). Bottom row shows the difference between observed and predicted data. Each column shows the X,Y, and Z components of the measured data, with MetalMapper receiver locations indicated by white circles. Colored images map blue and red to low and high data values, respectively. Right: Estimated polarizabilities at all time channels (solid lines) recovered via fitting, overlain on reference polarizabilities for a small industry standard object (ISO) projectile (dashed lines).

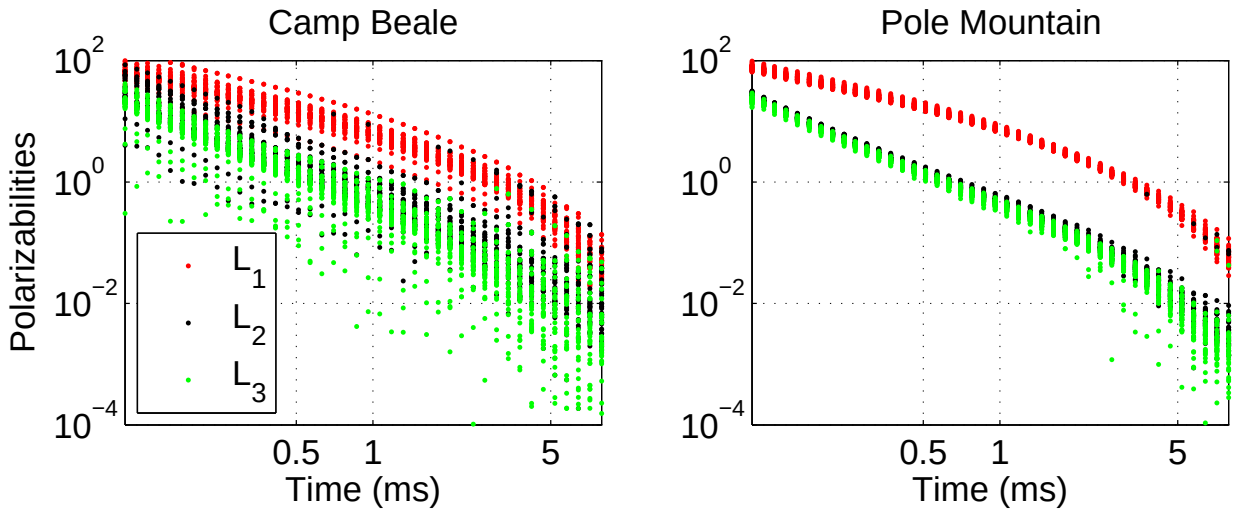


Figure 3: Estimated ISO polarizabilities at Camp Beale and Pole Mountain.

sequent classification are affected by factors including: background noise, target location (relative to the sensor), target orientation, and the presence of multiple targets within the footprint of the sensor. Currently, these effects can be studied with computationally intensive Monte Carlo (MC) simulations involving nonlinear inversions of synthetic data sets (e.g. Bell (2005); Walker et al. (2007); Lhomme et al. (2008)). However, nonlinear inversion of synthetic data sets can take many hours. We reduce this computation time with efficient MC methods that circumvent iterative inversions.

In section 2 we briefly describe the forward modeling, inversion, and uncertainty appraisal required for UXO data processing. In section 3 we develop an efficient Monte Carlo simulation method for estimating polarizability variance. The Monte Carlo sample is then used to predict classification performance in section 4. This analysis provides an objective prediction of classification performance that accounts for the uncertainty in estimated polarizabilities given site-specific noise.

2 Forward modeling, inversion, and uncertainty appraisal

2.1 Forward modeling

Observed electromagnetic data acquired over an electrically conductive object can be fit with a dipole model; details of the underlying physics can be found in Bell et al. (2001); Pasion and Oldenburg (2001). The forward modelling equation can be expressed as

$$\mathbf{d}^{pred} = \mathbf{G}(\mathbf{r})\mathbf{m}. \tag{1}$$

The model is separated into location parameters $\mathbf{r}$, and polarizability parameters $\mathbf{m}$. The vector $\mathbf{m}$ comprises the unique elements of a symmetric tensor $\mathbf{M}$

$$\mathbf{M} = \mathbf{ALA^T} \tag{2}$$

with the nonzero elements of the diagonal eigenvalue matrix L equal to the principal polarizabilities, and $\mathbf{A}$ a rotation matrix parameterized by the target's orientation. The vector of predicted data $\mathbf{d}^{pred}$ in equation 1 is linearly related to the model vector $\mathbf{m}$. The forward matrix $\mathbf{G}(\mathbf{r})$ in turn depends nonlinearly on target location $\mathbf{r}$ (relative to the sensor), with the predicted data decaying approximately as $1/r^6$.

2.2 Inversion

Given a vector of observations $\mathbf{d}^{obs}$, the inverse problem is then to find the set of model parameters - both location ($\mathbf{r}$) and polarizabilities ($\mathbf{m}$) - that best fits the observed data. The model is estimated by minimizing the least squares misfit between observed and predicted data. As described in Song et al. (2011), the inverse problem can be solved with a sequential approach:

1. Solve an overdetermined nonlinear inverse problem for the target location estimate $\hat{\mathbf{r}}$. The model $\mathbf{m}$ can be eliminated from the forward modeling functional by first expressing the least squares estimate for observed data $\mathbf{d}^{obs}$ as

$$\begin{aligned}\hat{\mathbf{m}} &= (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}^{obs} \\ &= \mathbf{G}^\dagger\mathbf{d}^{obs}\end{aligned} \tag{3}$$

with $\mathbf{G}^\dagger = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T$ denoting the pseudo-inverse. The dependence of $\mathbf{G}$ on the estimated location $\hat{\mathbf{r}}$ implied. Substituting equation 3 into 1, we have the nonlinear functional

$$\mathbf{d}^{pred} = F(\hat{\mathbf{r}}) = \mathbf{G}\mathbf{G}^\dagger\mathbf{d}^{obs}. \tag{4}$$

This form of the modeling equations is used for solution of the nonlinear location estimation problem. A single, high SNR time channel is typically used to obtain $\hat{\mathbf{r}}$.

2. Solve an overdetermined linear inverse problem for the polarizability model $\mathbf{m}$ at the fixed location estimate $\hat{\mathbf{r}}$, using equation 3. This problem is solved independently at each received time channel to obtain the time-varying polarizability estimates. The principal polarizabilities used in classification are then recovered from the estimated model via joint diagonalization (Cardoso, 1996). This procedure identifies a rotation matrix $\mathbf{A}$ that approximately diagonalizes the estimated polarization tensors (equation 2) over all time channels.

2.3 Multi-object inversion

At sites with a high spatial density of buried targets, multiple discrete objects can contribute to the measured data at any given location. This neccessitates data processing algorithms that resolve individual target locations and polarizabilities. Shubitidze et al. (2012) use an eigenvalue decomposition of the measured data to estimate the number of discrete sources. Song et al. (2011) generalize the two step inversion process described above to recover parameters for a specified number of targets. Standard processing workflows now invert for one, two and sometimes three objects over each detected anomaly. All estimated models from this processing are subsequently used in classification: the model that most closely matches reference UXO polarizabilities is used to rank each target (Beran et al., 2011).

Song et al. (2011) assume that the response of M targets is additive

$$\mathbf{d}^{pred} = \sum_{j=1}^{M} \mathbf{d}_j^{pred} \tag{5}$$

This can be expressed compactly as

$$\mathbf{d}^{pred} = \mathbf{\Gamma}\mathrm{M}, \tag{6}$$

with $\mathbf{\Gamma}$ a horizontal block matrix comprised of the forward matrices $\mathbf{G}_j$. In a two object scenario, for example, we have

$$\mathbf{\Gamma} = [\mathbf{G}_1\ \mathbf{G}_2]. \tag{7}$$

The column vector M similarly appends vectors $\mathbf{m}_j$ for all targets. Multi-object inversion then proceeds with the sequential approach outlined above.

2.4 Uncertainty appraisal

Uncertainty appraisal seeks to determine the posterior distribution of the model parameters given a particular realization of observed data. Beran et al. (2011) used Gibbs sampling to assess the uncertainty in model parameters estimated from older electromagnetic sensors with a single transmitter and receiver. Relatively small, low SNR data sets acquired with these sensors resulted in multiple local minima in the misfit function, such that the posterior distribution of the model was sometimes multimodal. These difficulties have been minimized with hardware improvements: inversion of field data sets acquired with newer sensors such as the MetalMapper typically produces much better constraints on the estimated polarizability model.

The posterior distribution of the estimated parameters can then be approximated as a normal distribution (Menke, 1989). With a sequential inversion approach, the polarizability parameters in $\hat{\mathbf{m}}$ are conditional on estimated location $\hat{\mathbf{r}}$ and the realization of the observed data, so that the posterior is

$$p(\hat{\mathbf{m}}|\hat{\mathbf{r}}, \mathbf{d}^{obs}) = \mathcal{N}(\hat{\mathbf{m}}, \mathrm{cov}(\hat{\mathbf{m}})), \tag{8}$$

with $\mathcal{N}(\mu, \mathbf{\Sigma})$ denoting a multivariate normal distribution with mean μ and covariance $\mathbf{\Sigma}$. The covariance of the estimated model is

$$\mathrm{cov}(\hat{\mathbf{m}}) = (\mathbf{G}^{\dagger})\mathrm{cov}(\mathbf{d}^{obs})(\mathbf{G}^{\dagger})^T, \tag{9}$$

with $\mathbf{G}$ evaluated at estimated location $\hat{\mathbf{r}}$. For additive, uncorrelated noise with standard deviation σ_i at the ith received time channel, the model covariance simplifies to

$$\mathrm{cov}(\hat{\mathbf{m}}_i) = \sigma_i^2(\mathbf{G}^T\mathbf{G})^{-1}. \tag{10}$$

This linear analysis is appropriate for quantifying the uncertainty in recovered model parameters, given a particular realization of observed data. However, in the context of performance prediction we are interested in a more general case: we seek the distribution of estimated polarizabilities that would be generated over multiple, independent realizations of observed data with prescribed noise statistics. This can be accomplished by marginalizing the joint distribution

$$\begin{aligned} p(\hat{\mathbf{m}}) &= \iint\limits_{\hat{\mathbf{r}},\mathbf{d}^{obs}} p(\hat{\mathbf{m}}, \hat{\mathbf{r}}, \mathbf{d}^{obs})\ d\hat{\mathbf{r}}\ d\mathbf{d}^{obs} \\ &= \iint\limits_{\hat{\mathbf{r}},\mathbf{d}^{obs}} p(\hat{\mathbf{m}}|\hat{\mathbf{r}}, \mathbf{d}^{obs})p(\hat{\mathbf{r}}|\mathbf{d}^{obs})p(\mathbf{d}^{obs})\ d\hat{\mathbf{r}}\ d\mathbf{d}^{obs} \end{aligned} \tag{11}$$

The posterior distribution of the estimated target location can be approximated as

$$p(\hat{\mathbf{r}}|\mathbf{d}^{obs}) = \mathcal{N}(\hat{\mathbf{r}}, \mathrm{cov}(\hat{\mathbf{r}})). \tag{12}$$

The covariance in $\hat{\mathbf{r}}$ obtained via a local sensitivity analysis. We replace the forward matrix in equation 10 with the sensitivity (Jacobian) matrix

$$\mathrm{cov}(\hat{\mathbf{r}}) \approx \sigma_i^2(\mathbf{J}^T\mathbf{J})^{-1}. \tag{13}$$

The elements of the sensitivity matrix, evaluated at $\hat{\mathbf{r}}$, are

$$J_{jk} = \left.\frac{\partial d_j^{pred}}{\partial r_k}\right|_{\hat{\mathbf{r}}}. \tag{14}$$

Evaluation of the multi-dimensional integral in equation 11 is challenging, in particular owing to the nonlinear dependence of the model covariance on estimated location (equation 10). In the next section, we therefore circumvent the marginalization with a novel MC algorithm that directly generates samples from the distribution of $p(\hat{\mathbf{m}})$.

3 Efficient Monte Carlo sampling of the estimated polarizability distribution

To rapidly generate an unbiased sample from the estimated distribution of polarizabilities, we eliminate iterative inversion for location from Monte Carlo sampling with the following approach:

1. Generate synthetic, true data $\mathbf{d}^{true}$ for a target with specified principal polarizabilities at true location $\mathbf{r}^{true}$. Because the predicted data depend on target orientation via equation 2, we model the data for three permutations of target orientation that produce a diagonal tensor $\mathbf{M}$. These correspond to alignment of the target's primary (L_1) polarizability with the sensor's local x, y, and z coordinates. The following steps are then repeated for each assumed target orientation.

2. For the current target orientation, evaluate the sensitivity matrix $\mathbf{J}$ at $\mathbf{r}^{true}$. Draw a realization $\hat{\mathbf{r}}$ from the posterior distribution of estimated target locations given in equation 12, assuming that $E(\hat{\mathbf{r}}) = \mathbf{r}^{true}$.

3. Generate the forward matrix $\mathbf{G}$ at the perturbed location.

4. Estimate the expected polarizability model at the perturbed location

$$E(\hat{\mathbf{m}}) = E(\mathbf{G}^{\dagger}\mathbf{d}^{obs}) = \mathbf{G}^{\dagger}E(\mathbf{d}^{obs}) = \mathbf{G}^{\dagger}\mathbf{d}^{true} \tag{15}$$

   The resulting principal polarizabilities derived from $\hat{\mathbf{m}}$ must be non-negative. This is typically enforced by imposing inequality constraints on the linear inverse problem Song et al. (2011). Here we simply discard any realizations for which the expected principal polarizabilities are negative.

5. Obtain a realization of the estimated polarizabilities with additive model noise

$$\hat{\mathbf{m}} = E(\hat{\mathbf{m}}) + \epsilon_{\mathbf{m}} \tag{16}$$

   with $p(\epsilon_{\hat{\mathbf{m}}}) = \mathcal{N}(0, \mathrm{cov}(\hat{\mathbf{m}}))$ (see equation 13).

Repeating this process over multiple realizations produces a sample of estimated polarizabilities, each conditional on a different orientation, estimated location and realization of noise (propagated via $\epsilon_{\mathbf{m}}$). The sample variance of the ensemble of polarizability models thereby marginalizes over the distributions of estimated locations and data noise. We call this a "*parametric Monte Carlo*" sampling algorithm: iterative inversion for location at each realization is replaced by an assumed parametric (i.e. normal) distribution of estimated locations.

The parametric MC approach decomposes the variance in the estimated polarizabilities about the true model into two terms

$$\begin{aligned} \mathrm{var}(\hat{\mathbf{m}}) &= E[(\mathbf{m}^{true} - \hat{\mathbf{m}})^2] \\ &= E[(\mathbf{m}^{true} - E(\hat{\mathbf{m}}) + E(\hat{\mathbf{m}}) - \hat{\mathbf{m}})^2] \\ &= E[(\mathbf{m}^{true} - E(\hat{\mathbf{m}}))^2] + E[(E(\hat{\mathbf{m}}) - \hat{\mathbf{m}})^2] \\ &= \mathrm{var}(\epsilon_r) + \mathrm{var}(\epsilon_m) \end{aligned} \tag{17}$$

The variance in the estimated model therefore has a contribution from the error in location (ϵ_r) and from the error in polarizabilities at an estimated location ϵ_m.

To validate this procedure, we simulate the distributions of estimated polarizabilities for a 37 mm projectile and small ISO in the following two object scenarios:

1. Both targets directly below the sensor ($x = y = 0$ m), 37 mm at $z = -0.15$ m, ISO at $z = -0.3$ m,

2. Both targets at $z = -0.3$ m and $y = 0$ m. ISO at $x = -0.15$ m and 37 mm at $x = 0.15$ m,

3. Both targets at $z = -0.5$ m and $y = 0$ m. ISO at $x = -0.15$ m and 37 mm at $x = 0.15$ m,

We compare parametric MC sampling with a full MC approach. The latter generates a sample of estimated polarizabilities from multiple, independent realizations of synthetic observed data. Target orientations are generated from a uniform random distribution. We add independent, normally-distributed noise to each realization, with standard deviations that are an average of estimated background noise at Pole Mountain and Camp Beale sites (Figure 4). The full MC sampling then proceeds by inverting each realization of synthetic data with the sequential approach outlined in section 2.2.

Figure 5 shows good agreement between the polarizability distributions obtained with both methods. The full Monte Carlo simulation approach can be accelerated by initializing the nonlinear location search at the true target locations - this considerably speeds up the two-object inversions to approximately 0.5 s per realization. The parametric MC approach averages 10^{-3} s per realization, with the reduction in computation time achieved by altogether eliminating the iterative nonlinear location search.

Figure 6 illustrates the model error decomposition described in equation 17 for the third multi-object scenario depicted in Figure 5. As predicted by equation 10, the variance component $var(\epsilon_{\mathbf{m}})$ is a scaled version of the noise on the data (Figure 4). In contrast, errors in target location result in a percent error on the estimated polarizabilities, so that $var(\epsilon_{\mathbf{r}})$ in

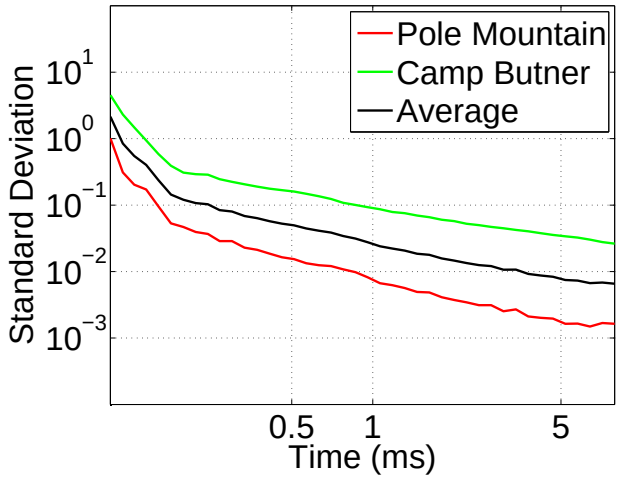


Figure 4: MetalMapper noise standard deviations estimated from background measurements in two field data sets (Pole Mountain and Camp Butner). Synthetic noise with a standard deviation equal to the geometric mean of the field data standard deviations at each channel is added to the full Monte Carlo simulation data.

Figure 5 resembles the primary polarizability for each target. For the 37 mm projectile, the error in the polarizabilities arising from perturbations in estimated location, is a significant contribution to the total model error at time channels ranging between 0.1 and 5 ms. This highlights the importance of accounting for errors in target location when predicting the distribution of estimated polarizabilities.

4 Predicting classification performance

Given an efficient algorithm for generating a sample of estimated polarizabilities, we can now objectively quantify expected classification performance. For classification with MetalMapper data, the decision statistic ϕ is typically a misfit of transformed estimated polarizabilities L^{est} with respect to some reference, or library, polarizability L^{ref} Beran et al. (2012). We often use a heuristic misfit function of the form

$$\phi = \sum_{i=1}^{N} \left(\frac{(L_i^{est})^\gamma - (L_i^{ref})^\gamma}{1/N \sum_{j=1}^{N} (L_j^{ref})^\gamma} \right)^2 \tag{18}$$

with $\gamma \approx 0.1$. For polarizability parameters estimated from high signal to noise ratio (SNR) anomalies we often use all three polarizabilities in the misfit calculation, whereas for noisier cases we might only use the primary polarizability to compute the decision statistic and rank targets.

Classification performance depends not only on the variability of estimated polarizabilities arising from background noise in the data, but on the similarity of the UXO and

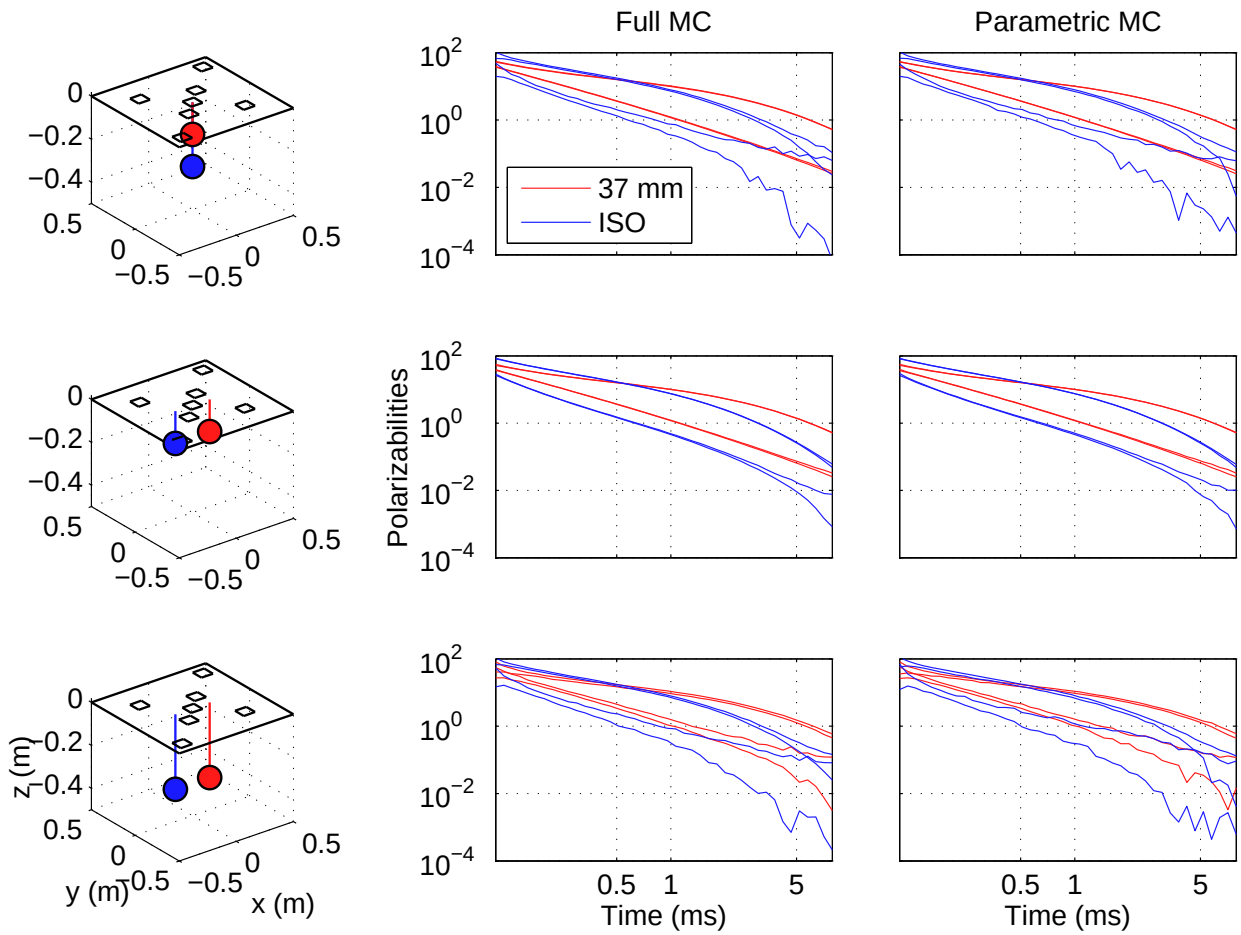


Figure 5: Comparison of polarizability variance for three multi-object scenarios. Left column: location of 37 mm and ISO targets relative to sensor. Middle column: 99% confidence intervals for polarizabilities obtained with full MC sampling. Right column: 99% confidence intervals for polarizabilities obtained with parametric MC sampling.

non-UXO we are trying to classify. A priori, it is impossible to predict the physical size and shape distribution of non-UXO at a given site, and so we must consider how classification performance varies under a range of scenarios for non-UXO. We can parameterize the true polarizabilities for UXO with the function Pasion and Oldenburg (2001)

$$L_i(t) = kt^{-\beta} exp(-t/\gamma). \tag{19}$$

The polarizabilities for hypothetical clutter items are then generated by varying the size (k) and decay parameters (γ) about their true values, as illustrated in Figure 7.

The scenarios presented in Figure 5 consider targets in a fixed location relative to the sensor. In general, performance prediction must consider a range of target locations relative to the sensor. We therefore extend the MC analysis to a grid of points within the sensor footprint. As illustrated in Figure 8, the symmetry of the MetalMapper receiver array allow the the grid to be reduced to the subset of points satisfying $x \geq y$. Additionally, in

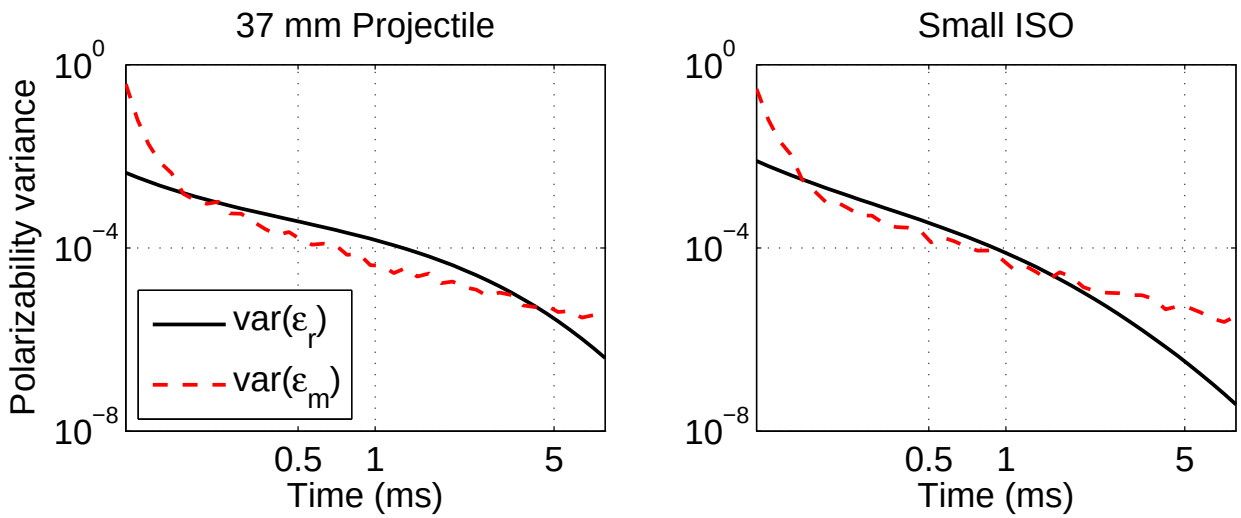


Figure 6: Polarizability variances for primary polarizability (L_1), as defined in equation 17. Variances are estimated from parametric MC samples for the third multi-object scenario depicted in Figure 5 (both targets at $z = -0.5$ m).

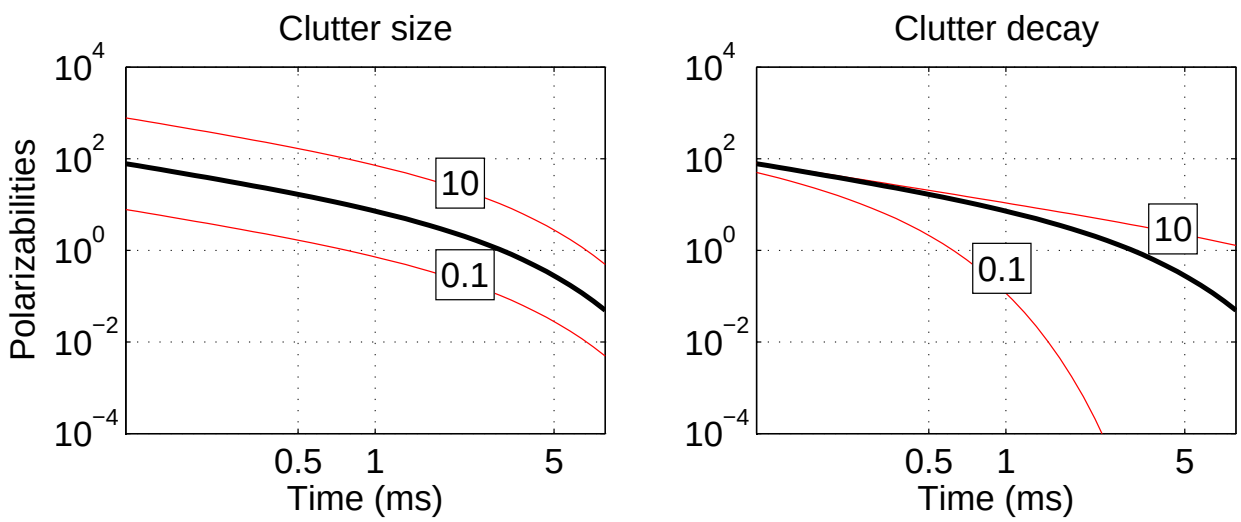


Figure 7: Dependence of primary polarizability on size (k) and decay (γ) parameters. The true polarizability (black line) is compared to polarizabilities generated by independently varying size (k) and decay (γ) parameters in equation 19.

this example we restrict the grid to $|x| \leq 0.3$ m and $|y| \leq 0.3$ m. In-field quality control procedures have improved target reacquisitions such that the target can be expected to be within the assumed bounds.

We simulate the distributions of polarizabilities for single targets at a specified maximum depth and uniformly distributed on the grid of points. This procedure is repeated for a UXO

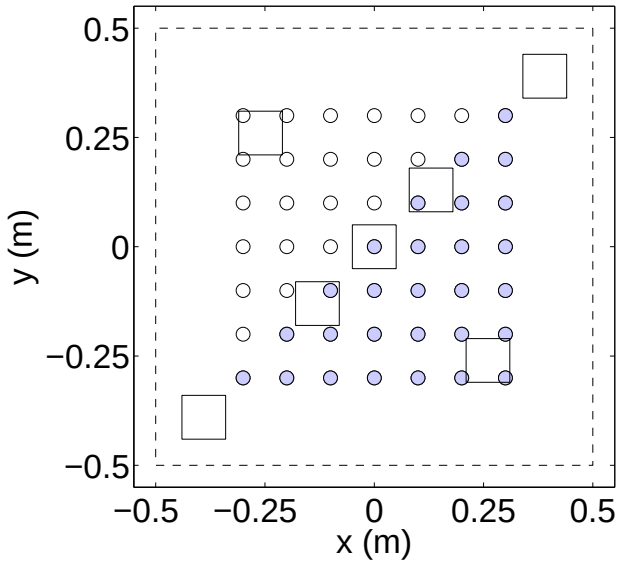


Figure 8: Regular grid of hypothetical target locations (circles). The symmetry of the MetalMapper receiver array (solid lines) allows the grid to be reduced to filled circles only for MC simulations. Dashed line indicates horizontal extents of MetalMapper transmitters.

item and for a range of hypothetical non-UXO with polarizabilities parameterized in terms of size and decay parameters. For each set of non-UXO polarizabilities, we then compute the probability of correct classification by

1. Compute the decision statistic ϕ (equation 18) for UXO and non-UXO polarizabilities generated with MC sampling

2. Prioritize realizations in the pooled (UXO and non-UXO) sample in order of increasing ϕ

3. Generate the receiver operating characteristic for the pooled sample by tracking the cumulative proportion of UXO and non-UXO identified in the prioritized sample.

4. Compute the probability of correctly classifying a UXO vs. a specified non-UXO as the area under the receiver operating characteristic (AUC). Hanley and McNeil (1982) show that the AUC is equivalent to the probability of correctly distinguishing between randomly-sampled true and false instances (UXO and non-UXO).

Figure 9 summarizes this analysis for classification of an ISO target and non-UXO at depths of $z = -0.3$ m and $z = -0.5$ m below the sensor. In these simulations we generate the distributions of UXO and non-UXO for single object scenarios: each target type is modeled separately on the grid of points. The MC samples from the targets are then merged to estimate classification performance as enumerated above.

In all examples in Figure 9 we see a decrease in classification performance down to $P_c \approx 0.5$ when clutter size and decay approach one. This represents the limiting case where the non-UXO has identical polarizabilities to the UXO and so classification can do no better

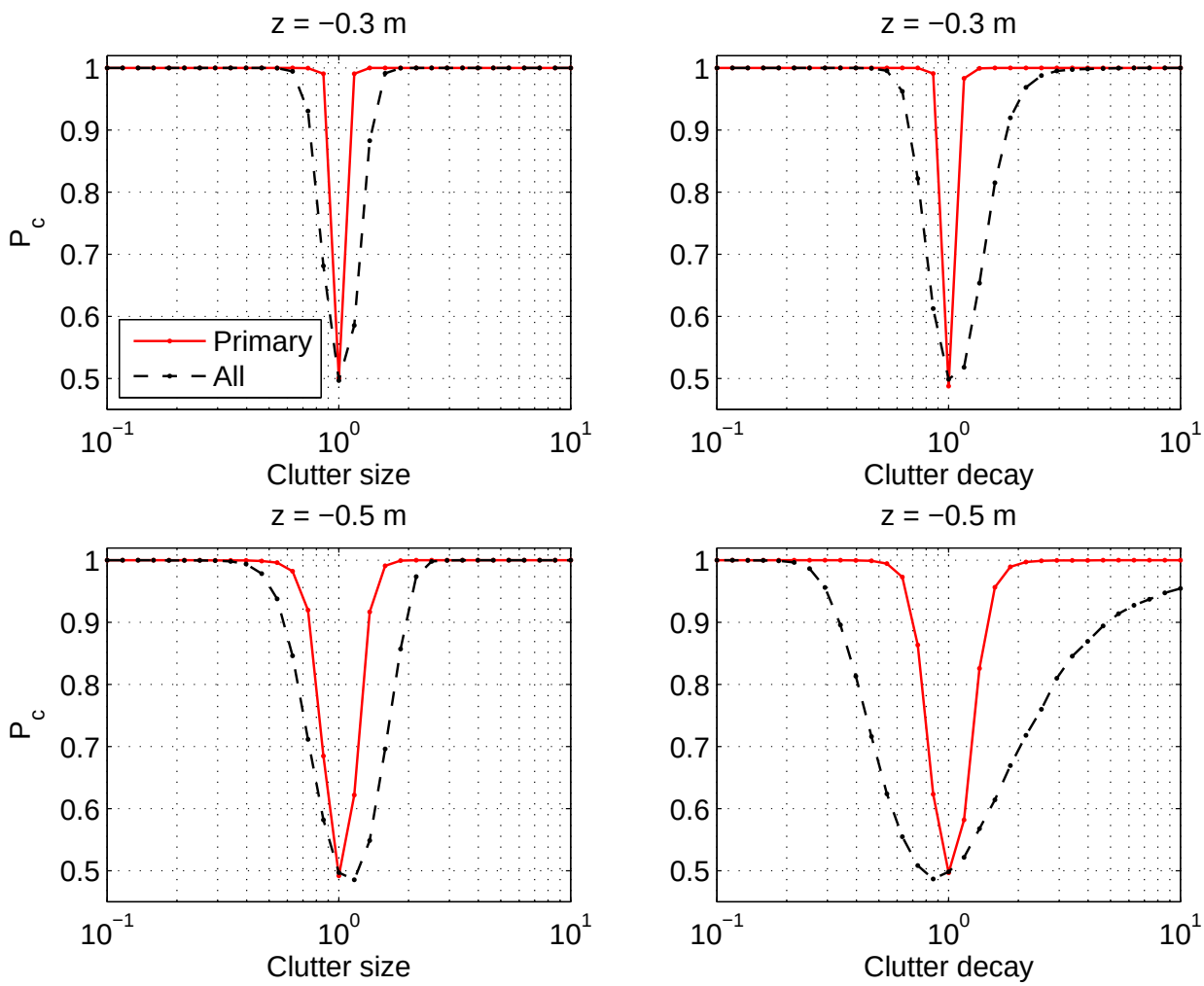


Figure 9: Probability of correct classification (P_c) of an ISO as a function of clutter size and decay. Top row considers classification of targets uniformly distributed at a depth of $z = -0.3$ m. Bottom row: both targets at $z = -0.5$ m. Clutter size is defined as the ratio of the size parameter (see equation 19) for non-UXO and UXO: $k_{non-UXO}/k_{UXO}$. Similarly, clutter decay is $\gamma_{non-UXO}/\gamma_{UXO}$. In all scenarios, classification performance is predicted using only the primary polarizability in the decision statistic calculation (equation 18), and using all polarizabilities.

than random guessing. For relatively shallow items ($z = -0.3$ m), classification in these scenarios is relatively easy, as reflected in the narrow region for which $P_c < 1$. As target depth is increased, classification difficulty increases, and the width of the region with $P_c < 1$ increases. We also compare classification performance using the primary polarizability versus all polarizabilities. The latter approach always produces worse classification performance because there is a larger relative uncertainty in the secondary polarizabilities.

5 Conclusions

We have developed performance prediction analysis techniques that account for the effects of sensor noise, target depth and orientation, and spatial target density on the uncertainty in estimated polarizability parameters. Polarizability variance cannot be fully described with a linear uncertainty appraisal that assumes perfect recovery of target location and we show that errors in estimated location translate to correlated shifts in the recovered polarizabilities.

The methods developed here provide the basic functionality for software tools that can be used to assess the feasibility of classification under site-specific conditions. Ongoing work will aim to couple performance predictions with expected cost

Acknowledgments

This work was supported by Strategic Environmental Research and Development Program project MR-2226, "Decision support tools for munitions response performance prediction and risk assessment"

References

T. Bell. Final report: Geo-location requirements for UXO discrimination, ESTCP Project MM-0413. Technical report, ESTCP, 2005.

T. Bell, B. Barrow, J. Miller, and D. Keiswetter. Time and frequency domain electromagnetic induction signatures of unexploded ordnance. *Subsurface Sensing Technologies and Applications*, 2:153–175, 2001.

L. Beran, B. Zelt, L. Pasion, S. Billings, K. Kingdon, N. Lhomme, L.-P. Song, and D. Oldenburg. Practical strategies for classification of unexploded ordnance. *Geophysics*, 78, 2012.

L. S. Beran, S. D. Billings, and D. W. Oldenburg. Incorporating uncertainty in unexploded ordnance discrimination. *Geoscience and Remote Sensing, IEEE Transactions on*, 49(8): 3071–3080, 2011.

J. F. Cardoso. Jacobi angles for simultaneous diagonalization. *SIAM journal on matrix analysis and applications*, 17:161–163, 1996.

W. P. Delaney and D. Etter. Report of the Defense Science Board on Unexploded Ordnance. Technical report, Office of the Undersecretary of Defense for Acquisition, Technology and Logistics, 2003.

J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic. *Radiology*, 143:29–36, 1982.

N. Lhomme, D. W. Oldenburg, L. R. Pasion, D. Sinex, and S. D. Billings. Assessing the quality of electromagnetic data for the discrimination of UXO using figures of merit. *Journal of Engineering and Environmental Geophysics*, 13:165–176, 2008.

W. Menke. *Geophysical Data Analysis: Discrete Inverse Theory.* Academic Press, 1989.

L. R. Pasion and D. W. Oldenburg. A discrimination algorithm for UXO using time domain electromagnetic induction. *Journal of Environmental and Engineering Geophysics*, 6:91–102, 2001.

M. Prouty, D. C. George, and D. D. Snyder. MetalMapper: A Multi-Sensor TEM System for UXO Detection and Classification, ESTCP Project MR-200603. Technical report, ESTCP, 2011.

F. Shubitidze, J. P. Fernández, I. Shamatava, B. E. Barrowes, and K. ONeill. Joint diagonalization applied to the detection and discrimination of unexploded ordnance. *Geophysics*, 77(4):WB149–WB160, 2012.

L. P. Song, L. R. Pasion, S. D. Billings, and D. W. Oldenburg. Nonlinear inversion for multiple objects in transient electromagnetic induction sensing of unexploded ordnance: Technique and applications. *IEEE Trans. Geosci. Remote Sensing*, 49(10):4007 –4020, 2011.

S. E. Walker, L. R. Pasion, D. W. Oldenburg, and S. D. Billings. Investigating the effect of data quality on time domain electromagnetic discrimination. *Journal of Applied Geophysics*, 61:254278, 2007.

Appendix C. Retrospective analyses of cued MetalMapper data sets

Site	**Year**	**DDD**	**FAR**	**AUC**
Butner	2010	28.8	0.79/0.47/**0.46**/*0.95*	0.98/**0.94**/*0.98*/0.96
Pole Mtn	2011	4.1	**0.02**/0.14/0.07/*0.30*	*1.00*/**0.99**/0.99/0.99
Beale P	2011	15.8	0.84/**0.66**/0.71/*0.90*	0.98/*0.98*/0.98/**0.95**
MMR 1	2012	21.5	*0.88*/**0.64**/0.81/0.70	0.96/0.94/*0.97*/**0.91**
Spencer U	2012	12.7	**0.07**/0.14/0.13/*0.23*	*0.99*/**0.96**/0.97/0.98
Vieques	2012	67.7	**0.87**/0.87/*0.91*/0.91	0.86/*0.87*/0.85/**0.79**
Ellis	2013	19.4	0.63/*0.73*/0.70/**0.47**	*0.98*/0.95/**0.80**/0.93
Rucker PoP	2013	63.7	**0.93**/0.93/0.93/*0.97*	*0.90*/0.87/0.86/**0.78**

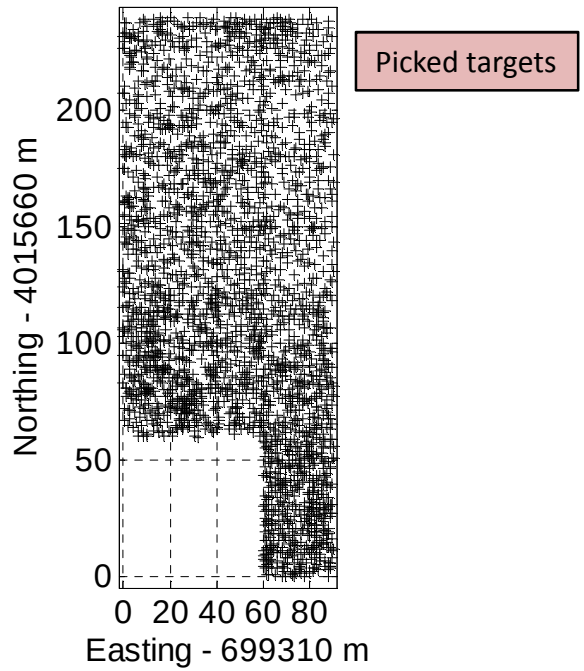
Table 3. Summary statistics for retrospective analyses of cued MetalMapper data sets. DDD denotes dataset degree of difficulty. Values in **blue**/*red* indicate best/worst values for each performance statistic, respectively. For false alarm rate (FAR), smaller values are better. For area under the ROC (AUC), larger values are better. The four values tabulated for each site's FAR and AUC represent the classification results obtained with the following algorithms:

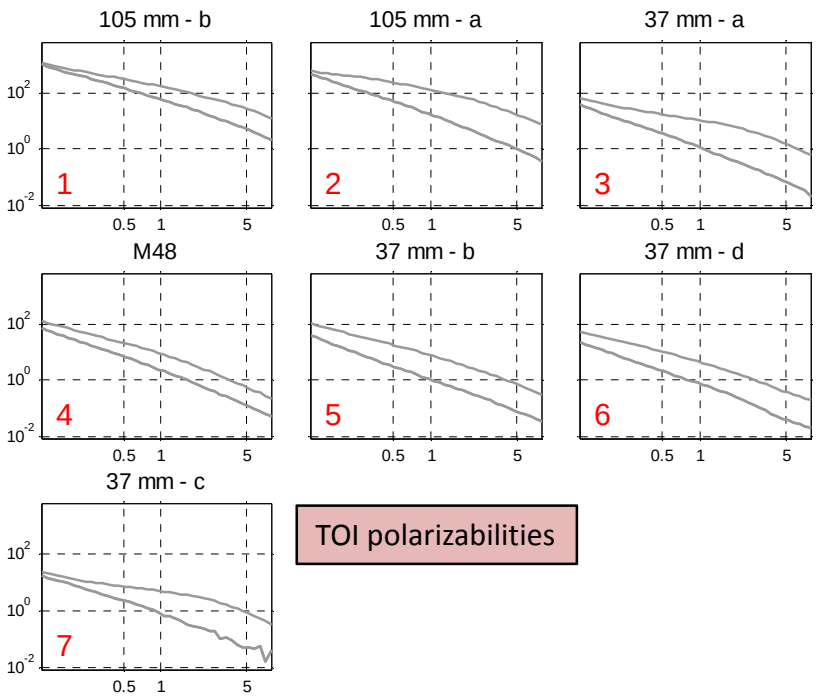
(1) Polarizability matching using all polarizabilities,
(2) Polarizability matching using primary polarizabilities only,
(3) Combined classifier ranking (CCR) using both polarizabilities and size/decay features,
(4) τ metric classifier, using no reference polarizabilities to rank targets (see section 5.6.2).

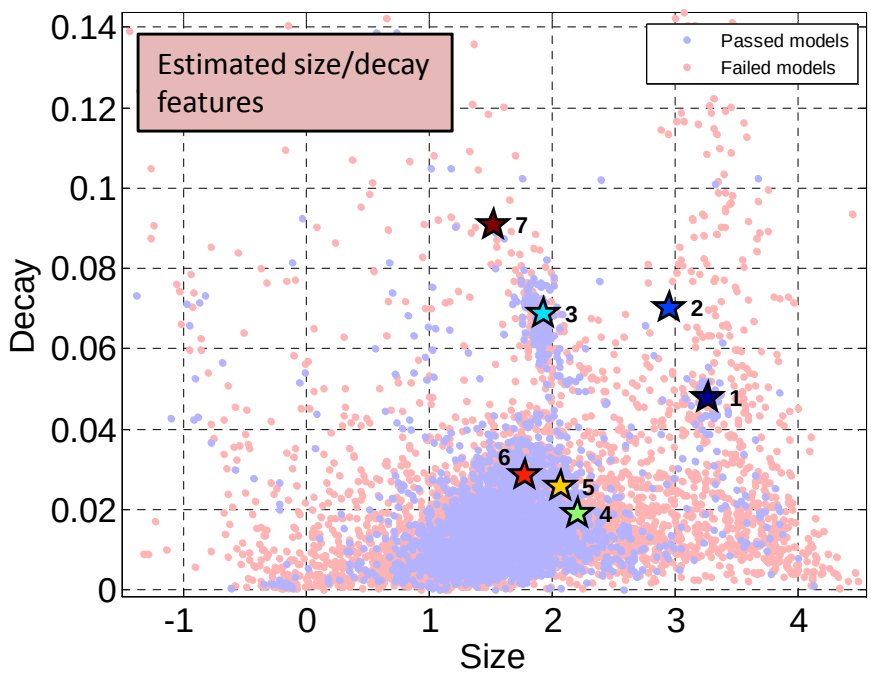
Butner (2010)

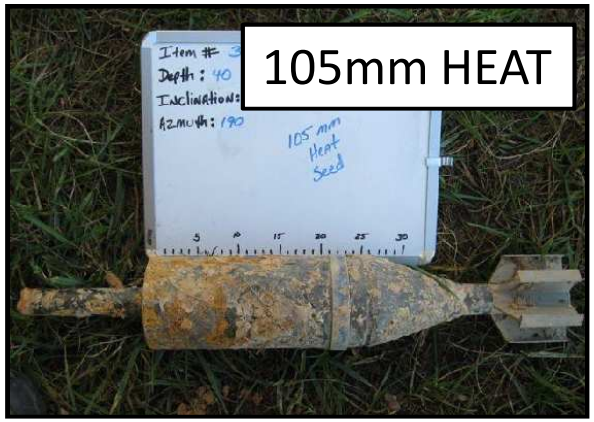
MM Cued

- 2304 targets
- 171 TOI
- Data collected by two groups with two MMs. Data quality from one group noticeably poorer, resulting in some difficult to find TOI.





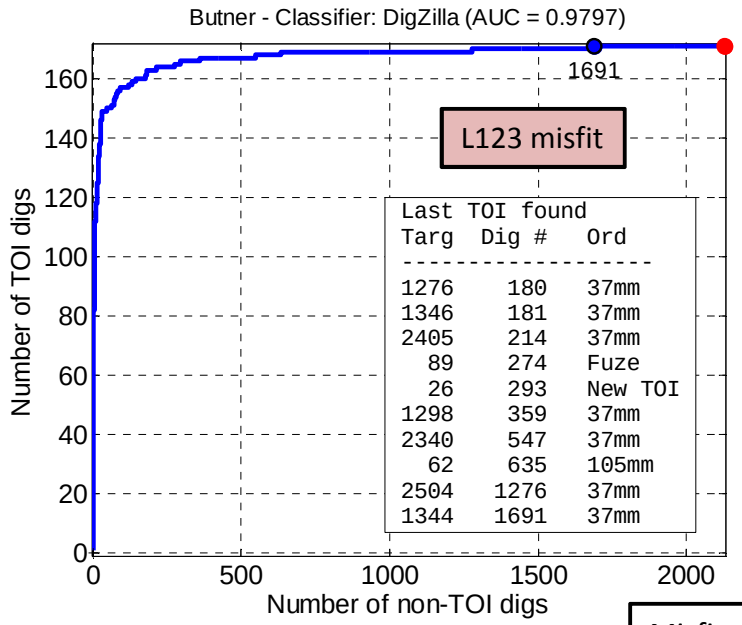


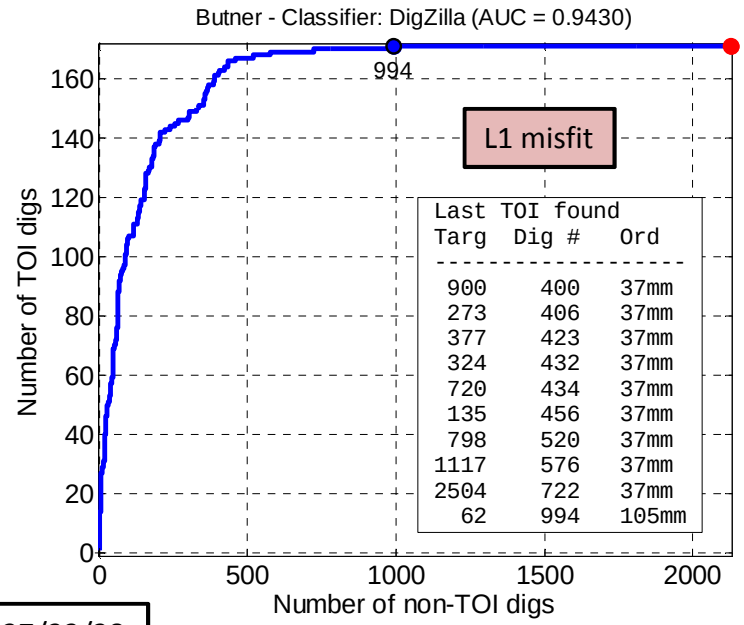


Butner

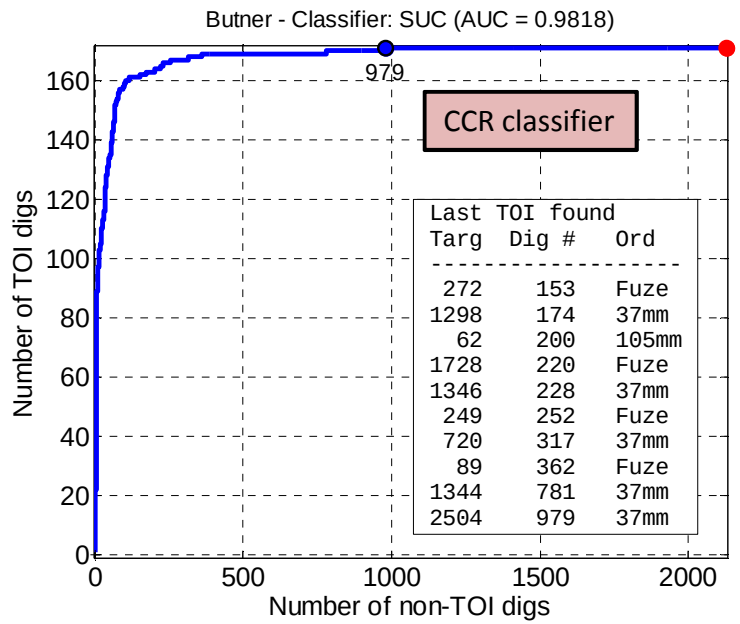
TOI	Number
37mm	121
Fuze	29
105mm HEAT	13
105mm HE	13
Total	176

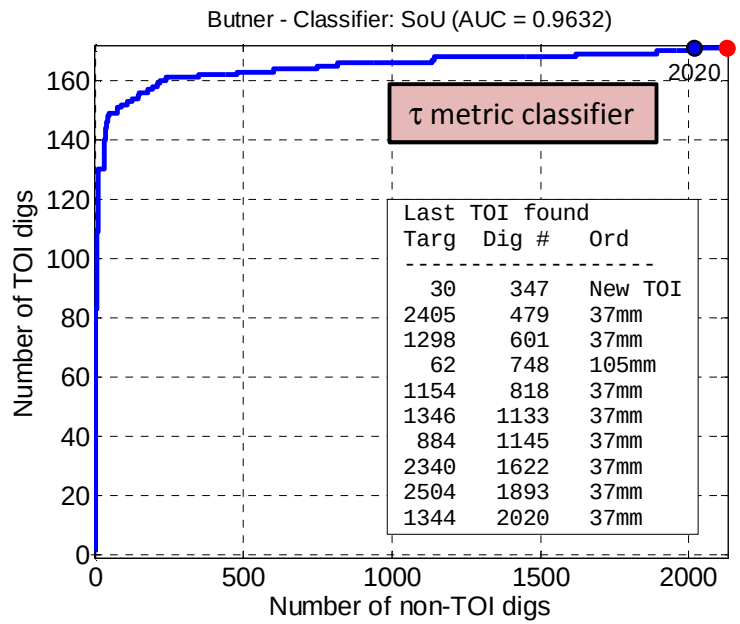
Butner TOI





Misfit channels: 37/33/28

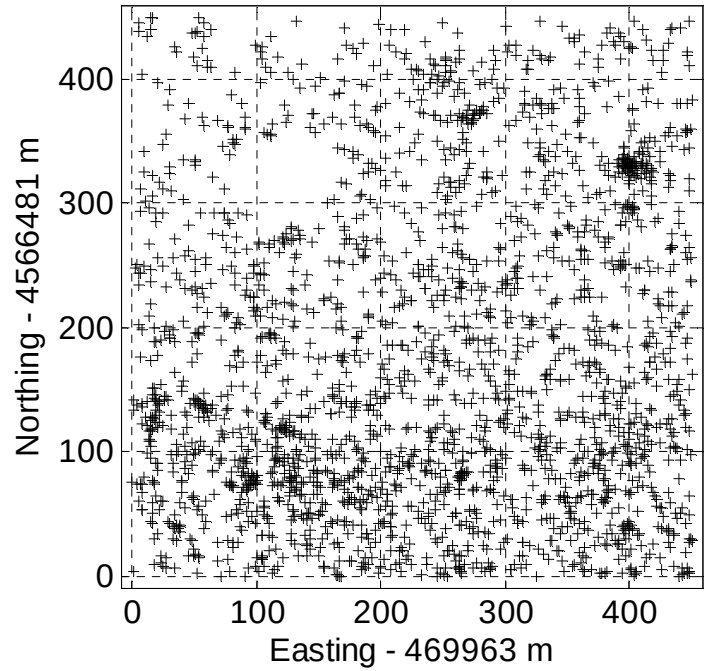


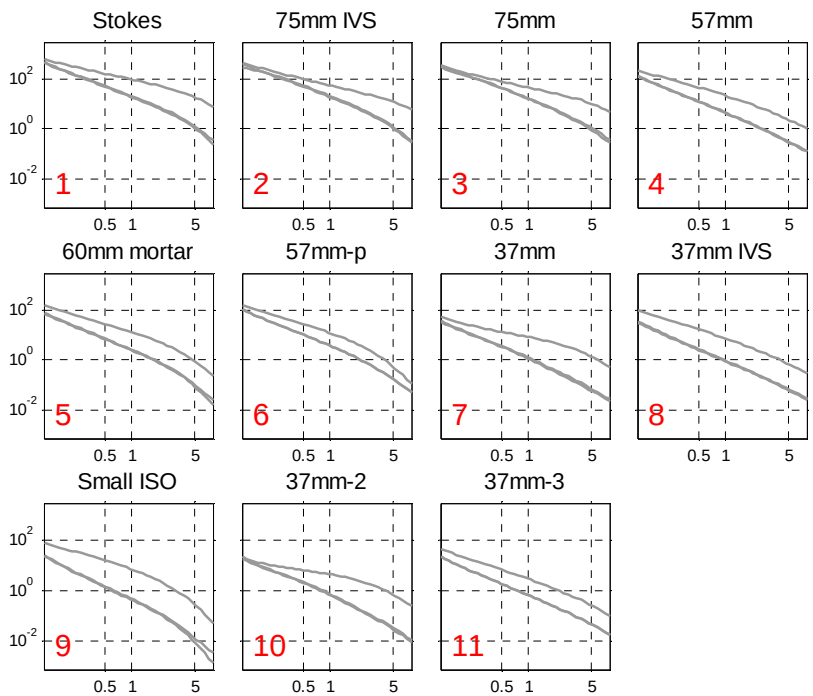


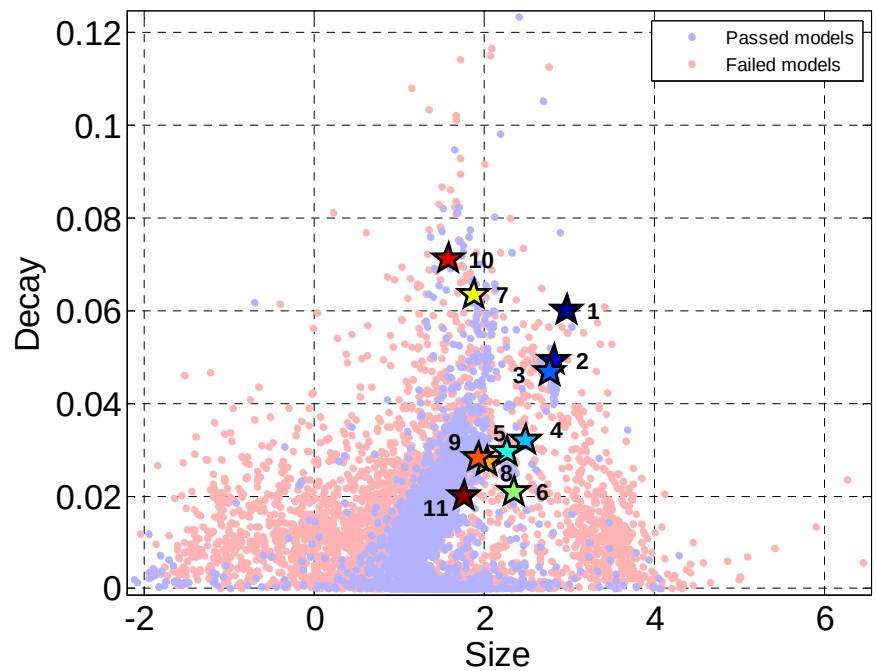
Pole Mountain (2011)

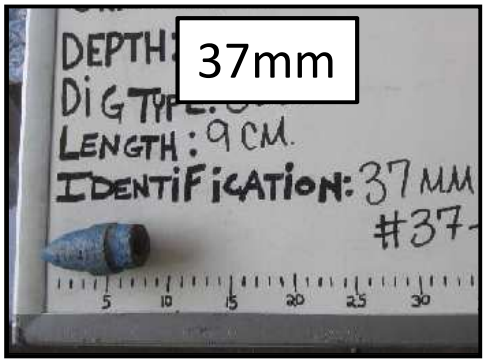
MM Cued

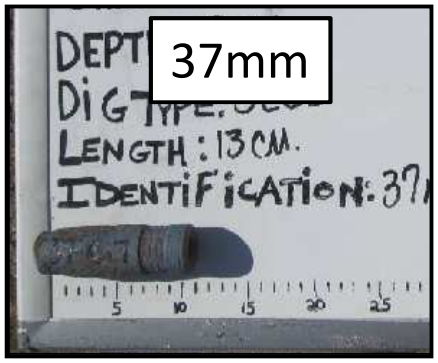
- 2370 targets
- 162 TOI
- Excellent data quality. Widely recognized as a very easy site for classification.

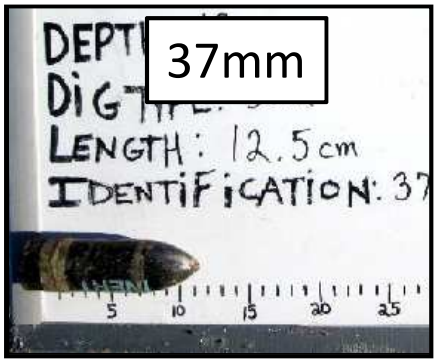


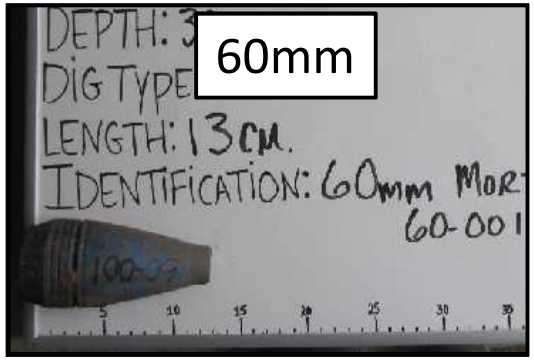


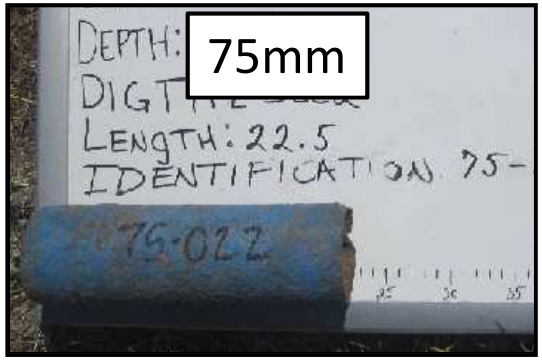


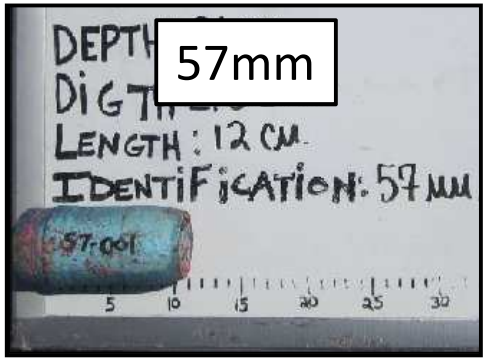








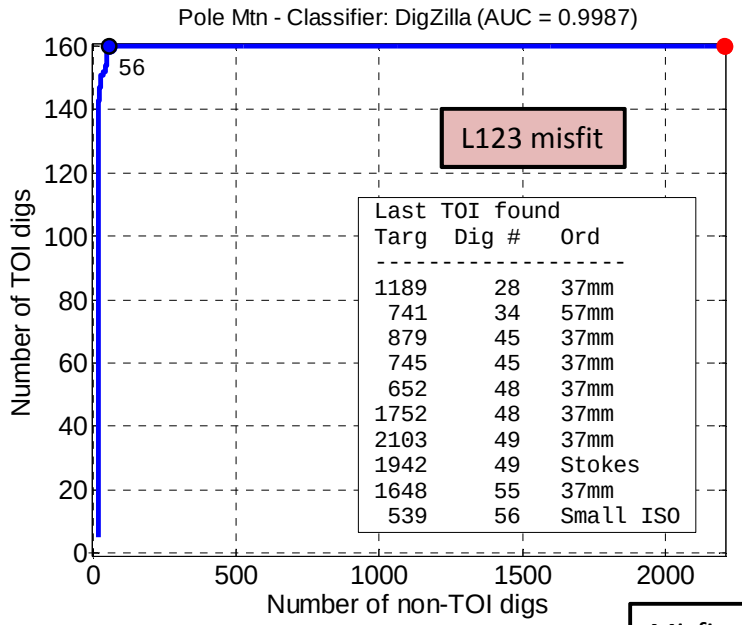


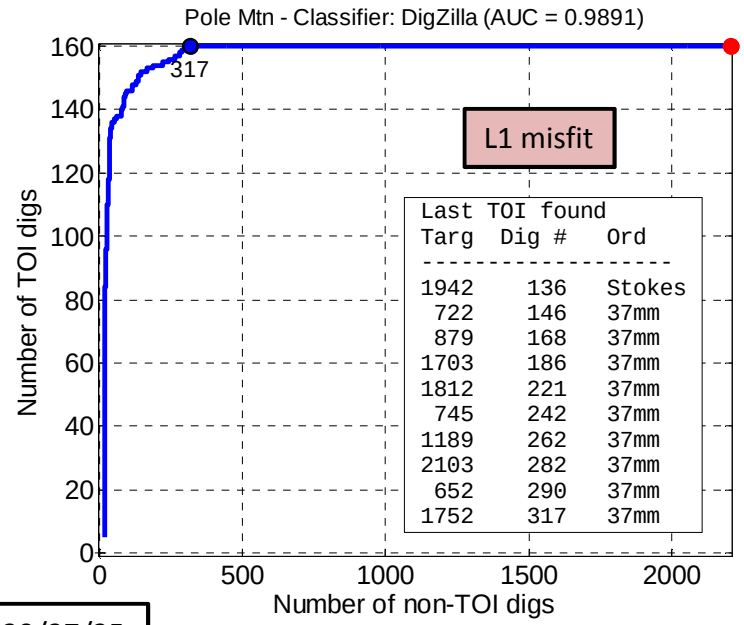


Pole Mountain TOI

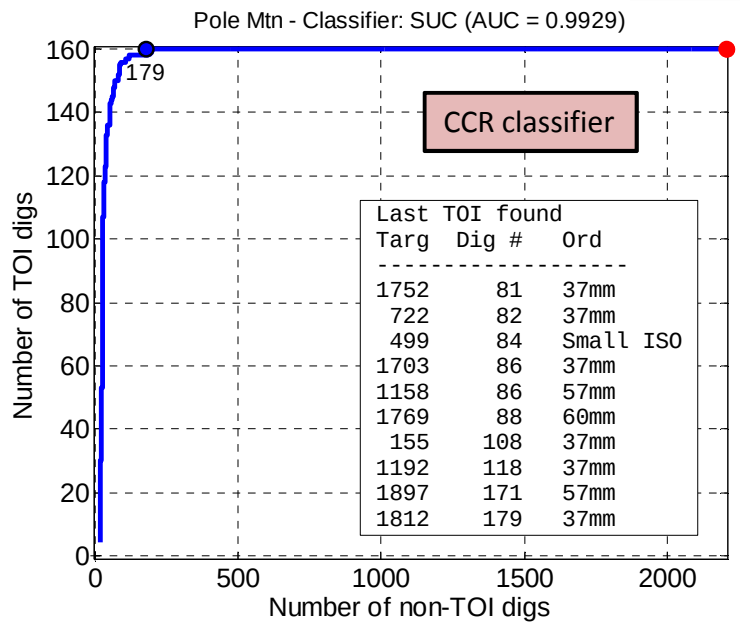
Pole Mountain

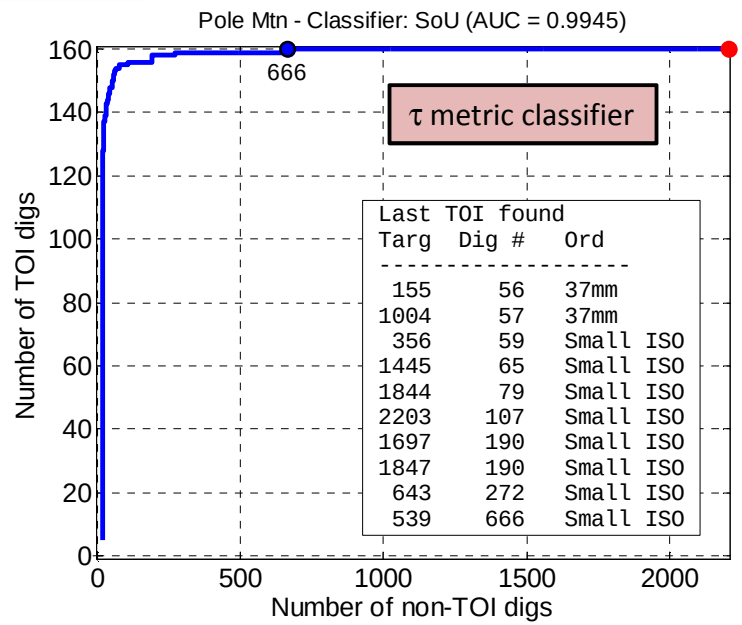
TOI	Number
Small ISO	42
37mm	43
60mm	41
75mm	25
57mm	10
Stokes	1
Total	162





Misfit channels: 39/37/35

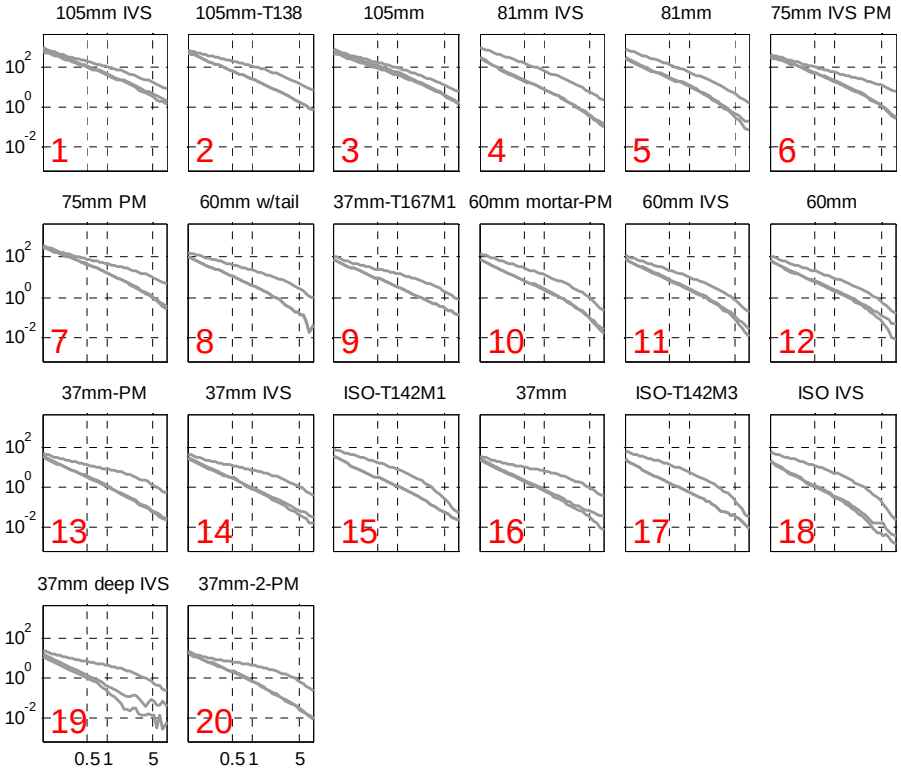


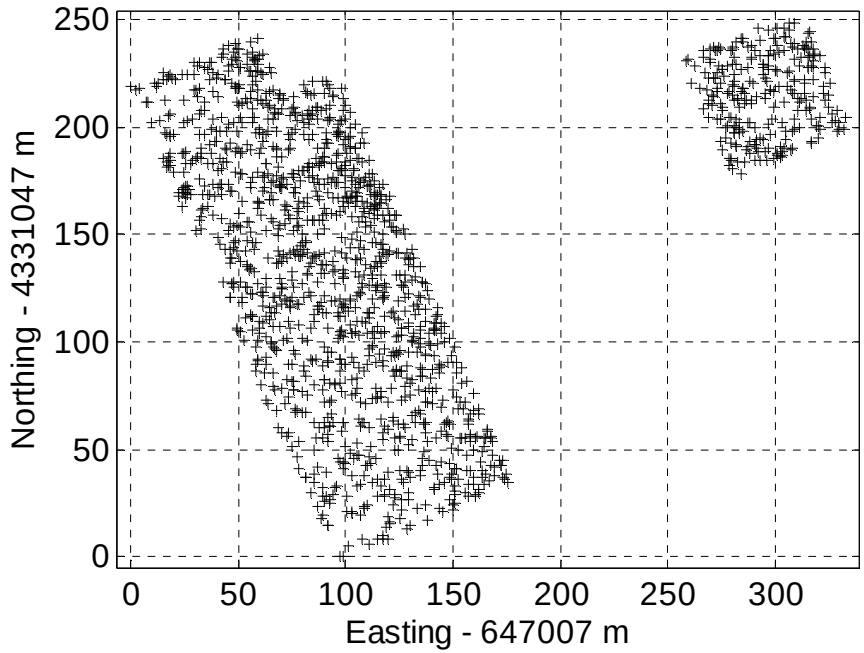


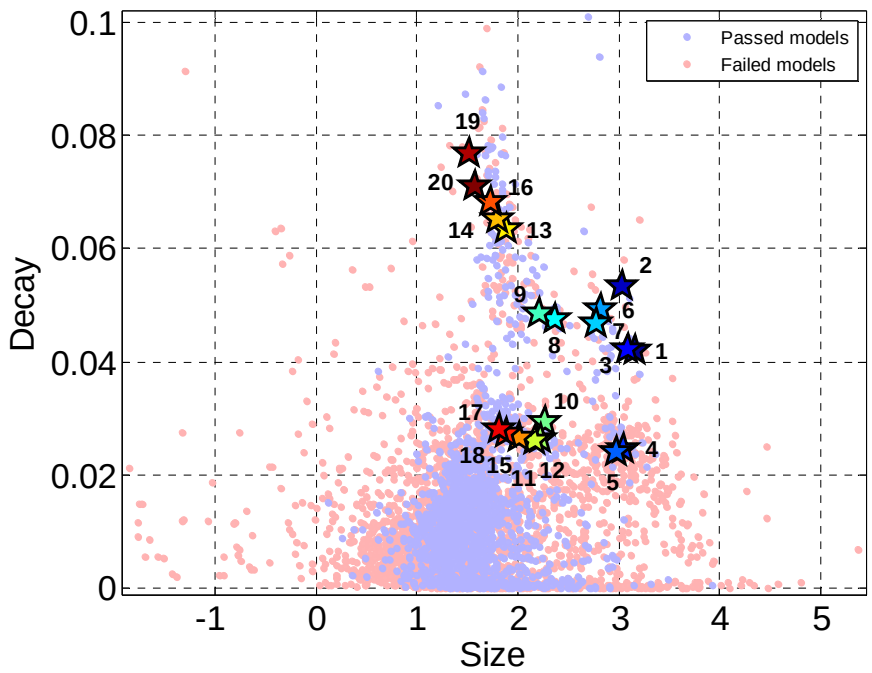
Beale (2011)

MM Cued

- 1490 targets
- 137 TOI
- Originally small fuzes were counted as TOI. Because they were very difficult to find they were eventually treated as non-TOI. First use of ISOs as seed items.





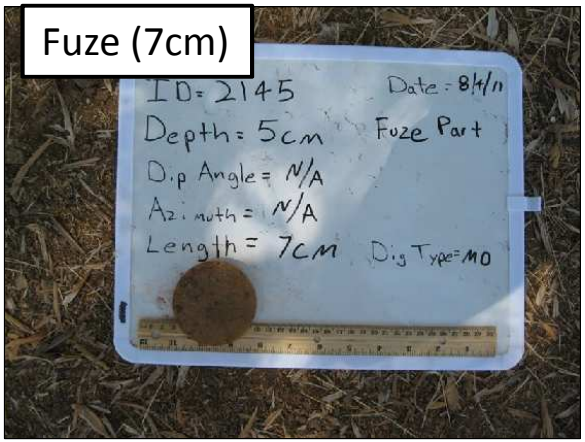


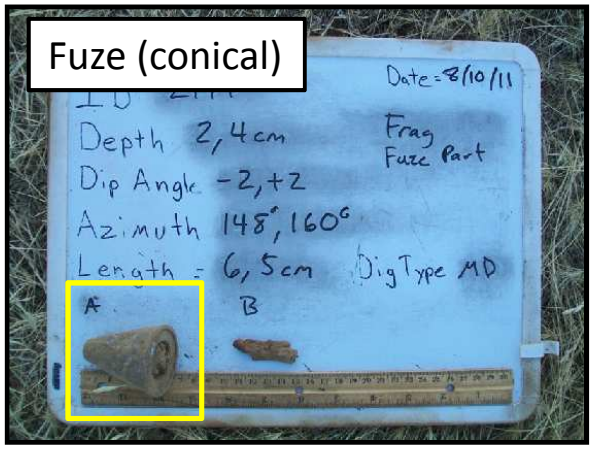
Beale

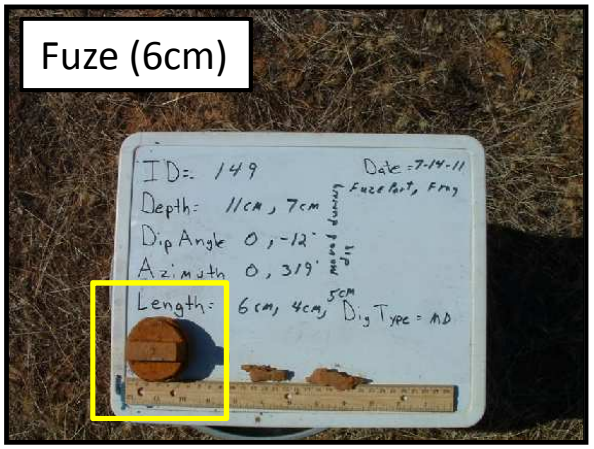
TOI	Number
Small ISO	47
37mm	40
60mm	24
81mm	21
105mm	6
Total	138

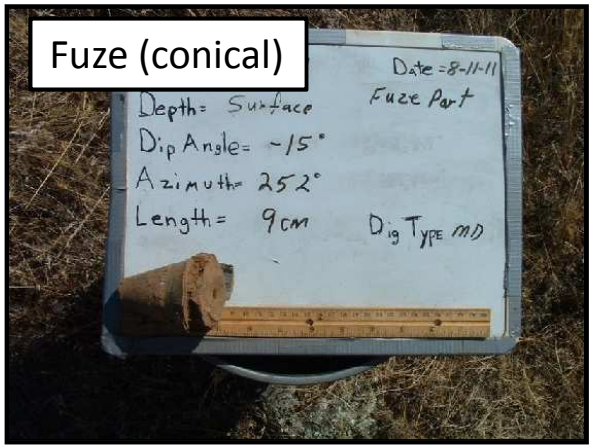
Beale TOI

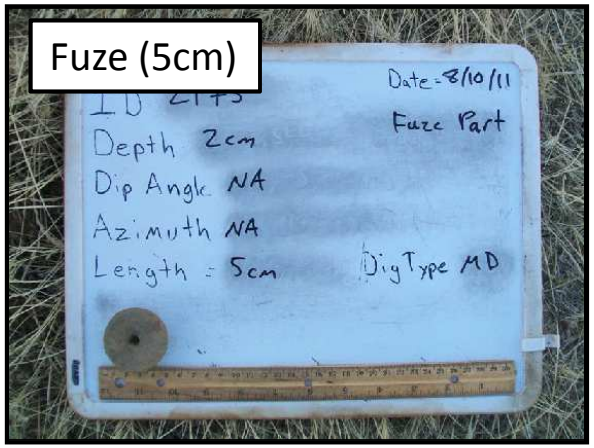
Excluding small fuzes

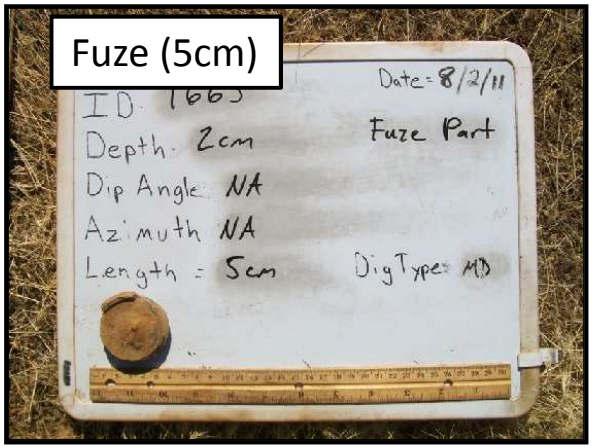


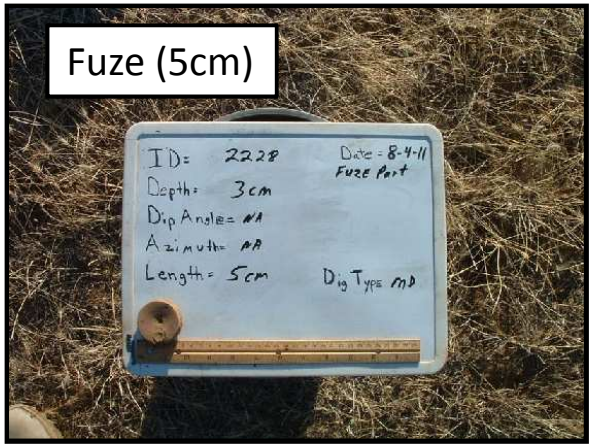


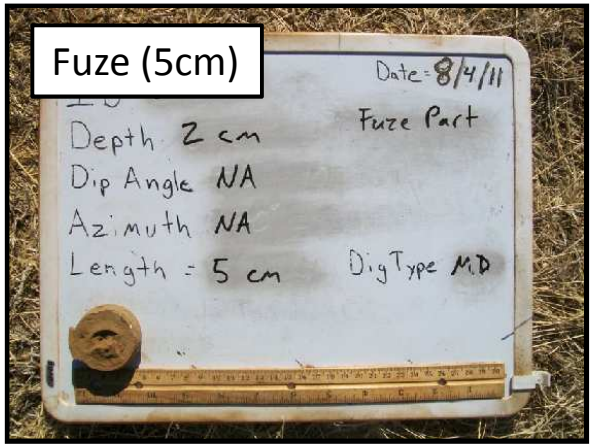






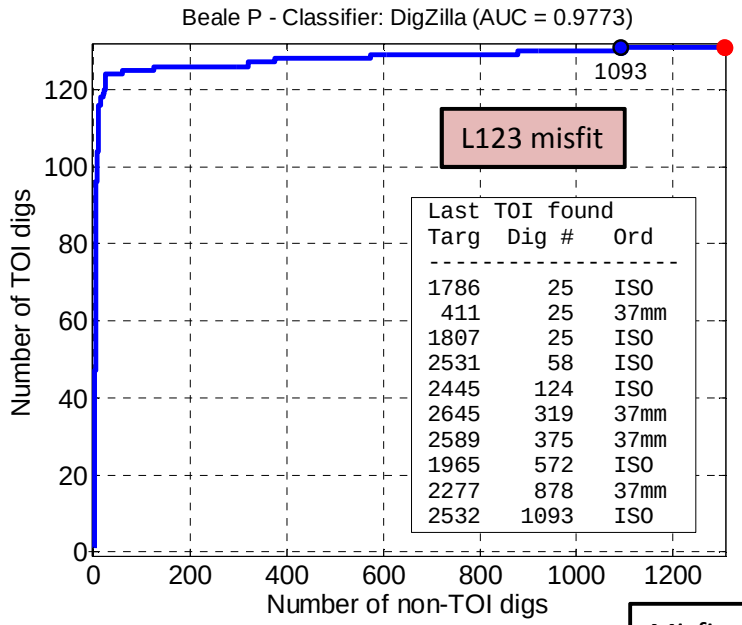


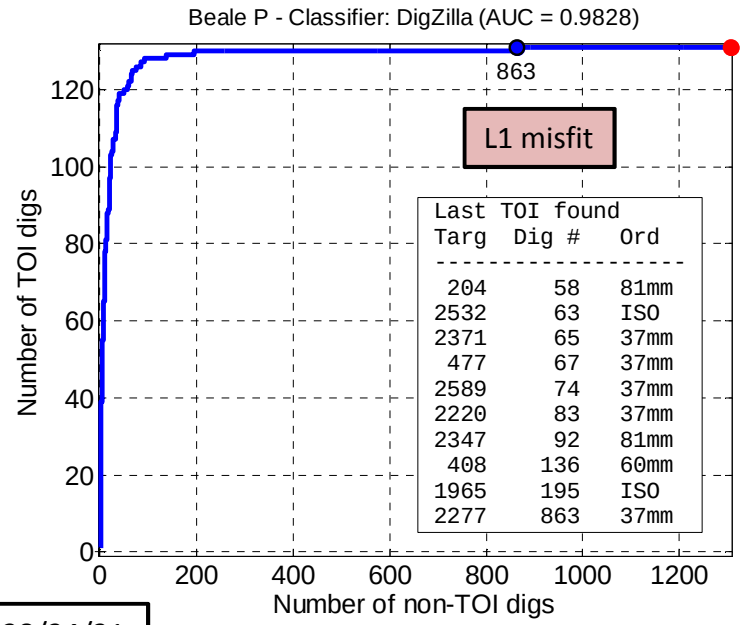




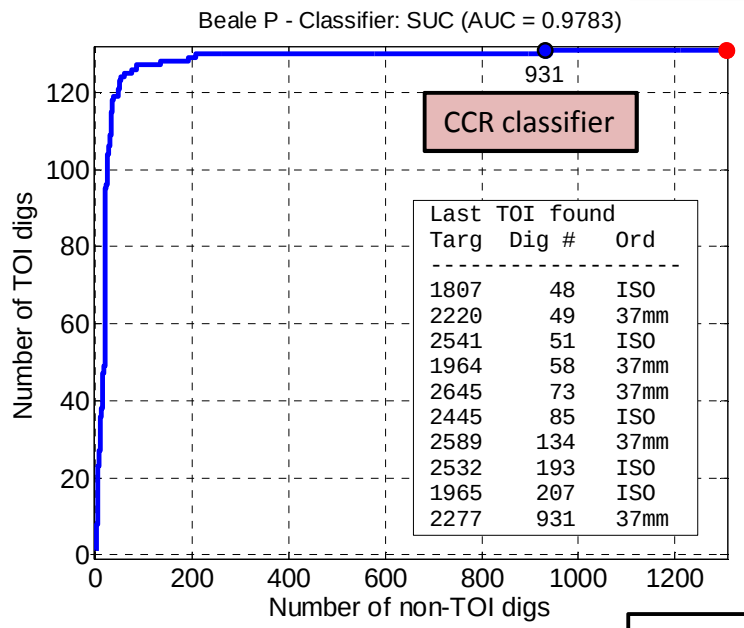
Beale TOI

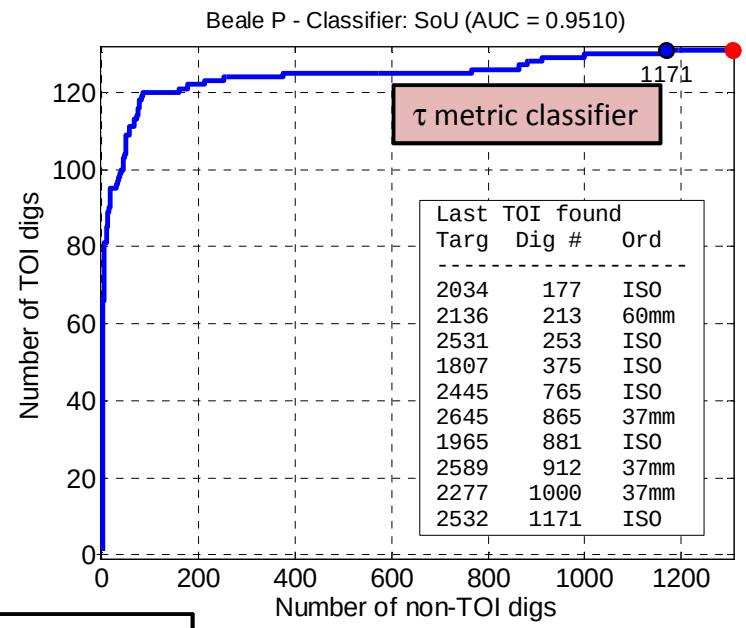
Small fuzes





Misfit channels: 38/34/31



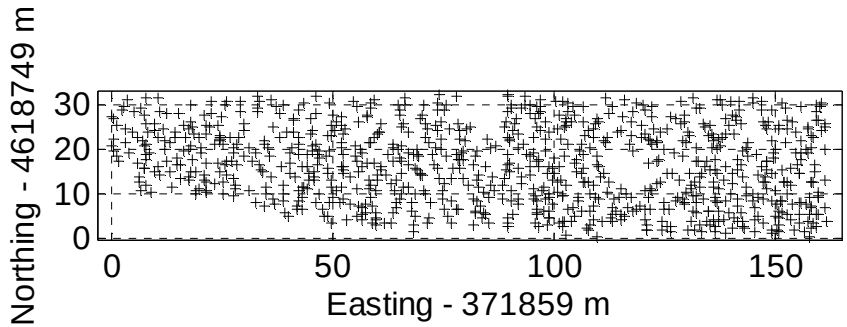


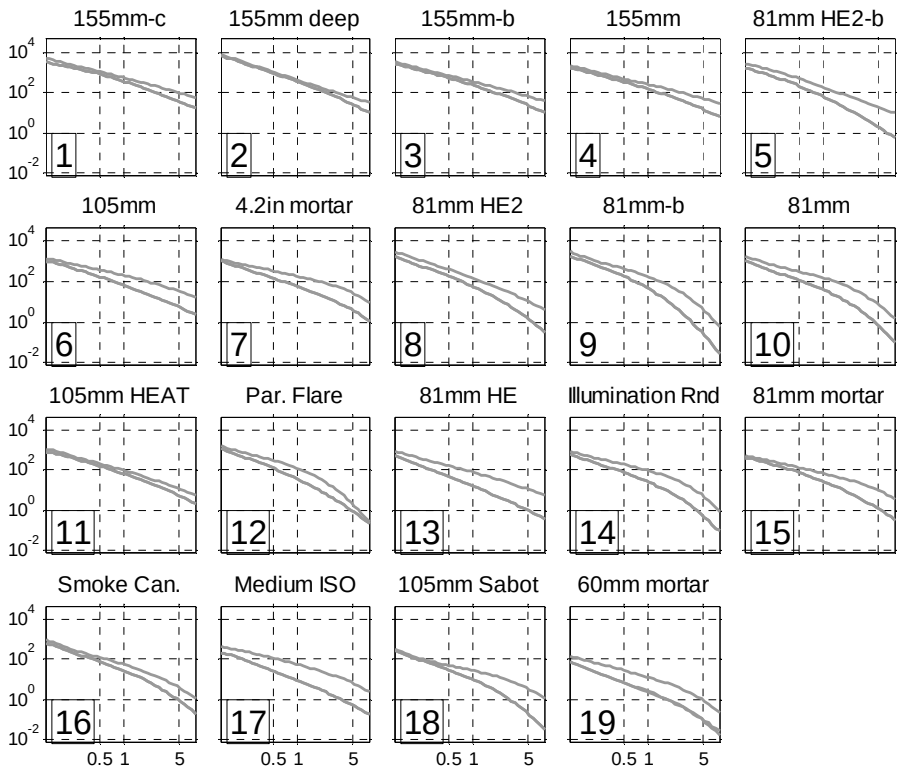
Ground truth: fuzes as clutter

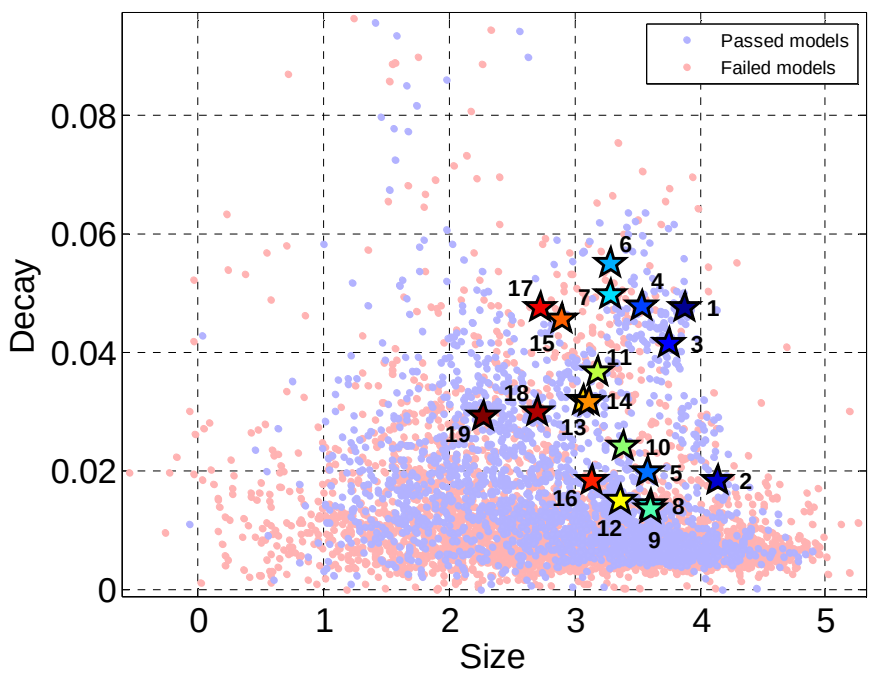
MMR Phase 1 (2012)

MM Cued

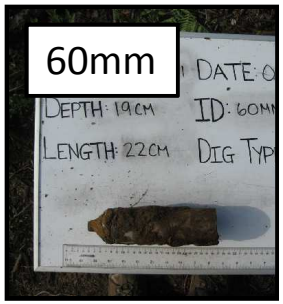
- 881 targets
- 121 TOI
- Mainly large TOI. Large number of 155mm. Lots of scrap that looks like large TOI makes this site challenging.

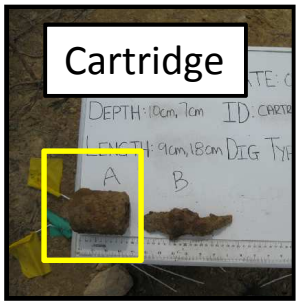






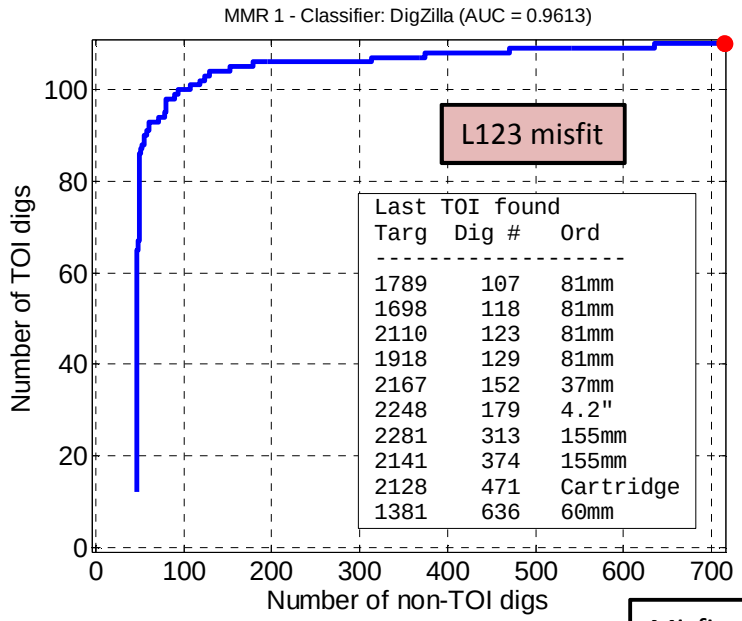
MMR1 TOI

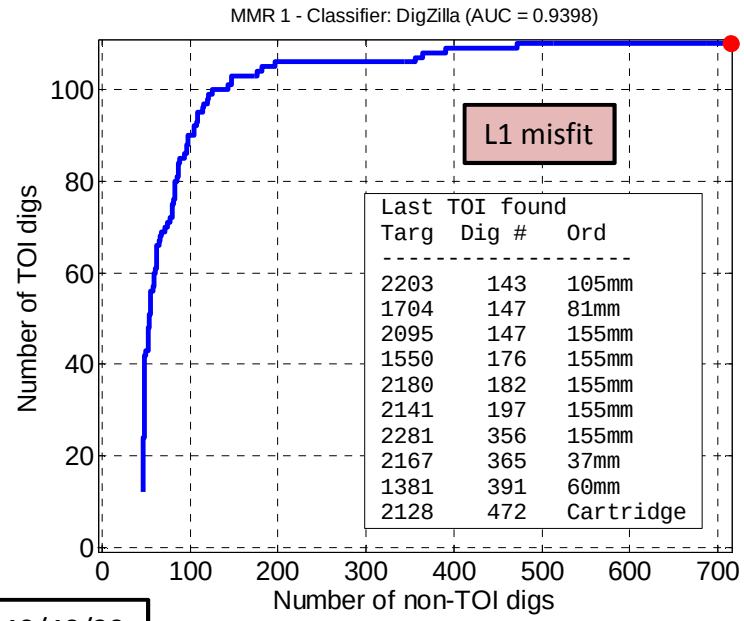




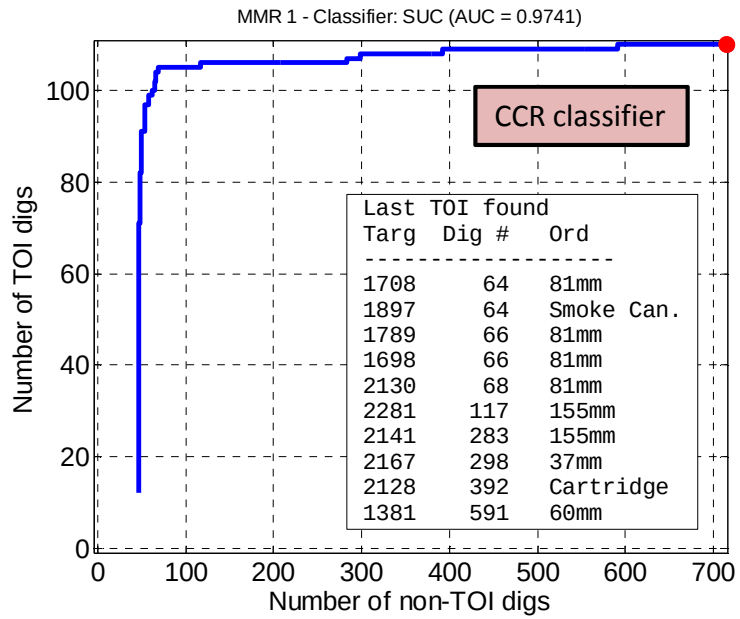
MMR1

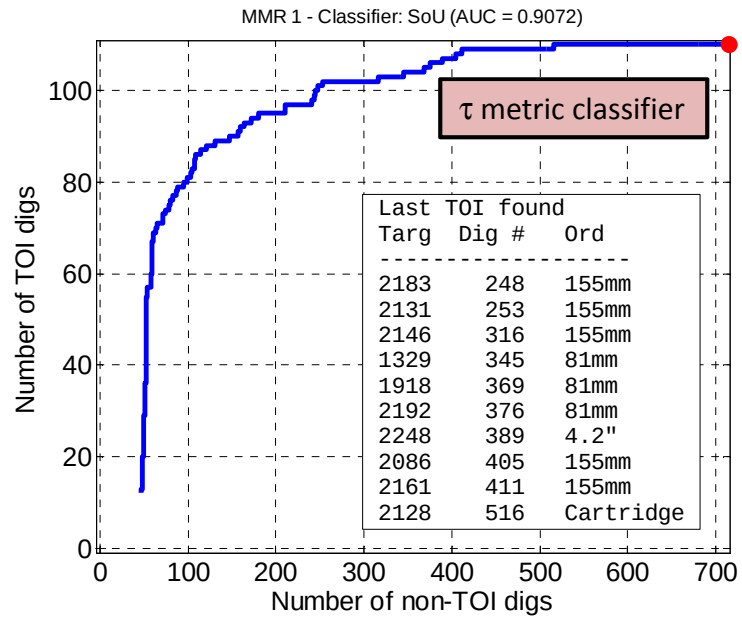
TOI	Number
155mm	71
81mm	14
Illum Rnd	13
105mm	7
4.2"	6
75mm	2
37mm	1
60mm	1
Cartridge	1
Medium ISO	1
Par. Flare	1
Smoke Can.	1
105mm Sabot	1
Total	120





Misfit channels: 40/40/39

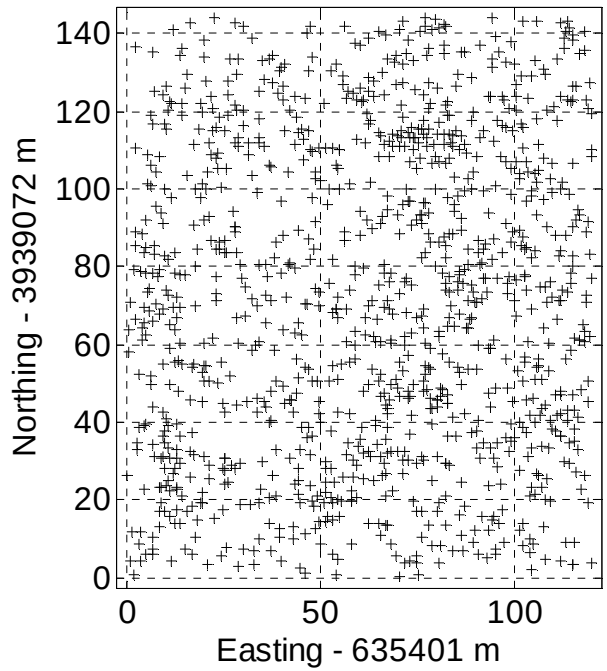


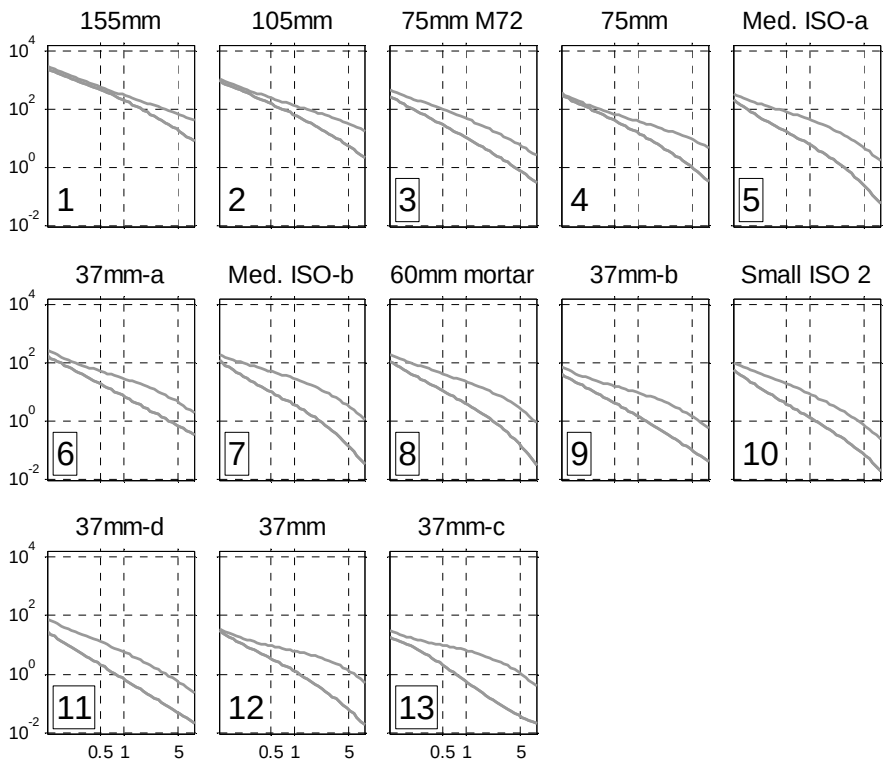


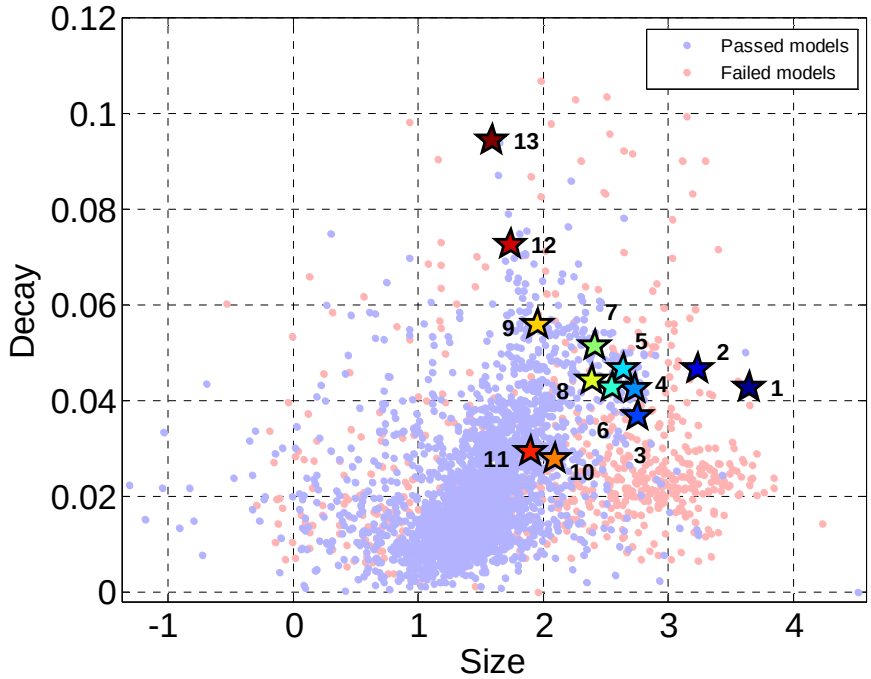
Spencer (2012)

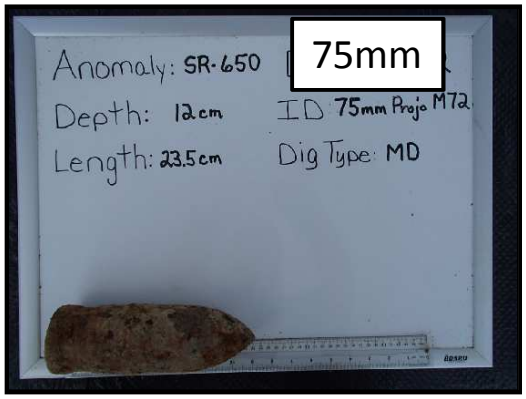
MM Cued

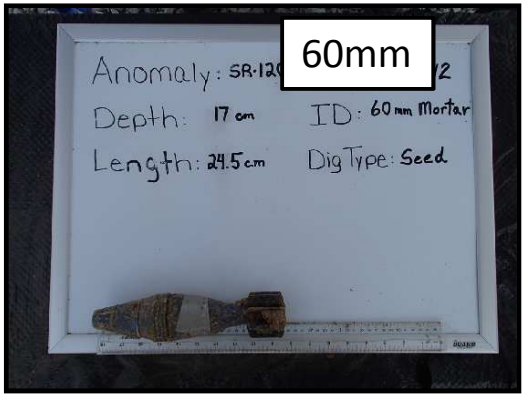
- 1104 targets
- 86 TOI
- Generally regarded as an easy site. Good data quality. Two one-off items were both large (155mm and 105mm).

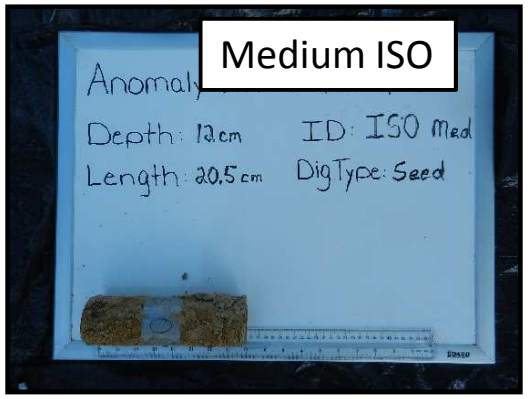


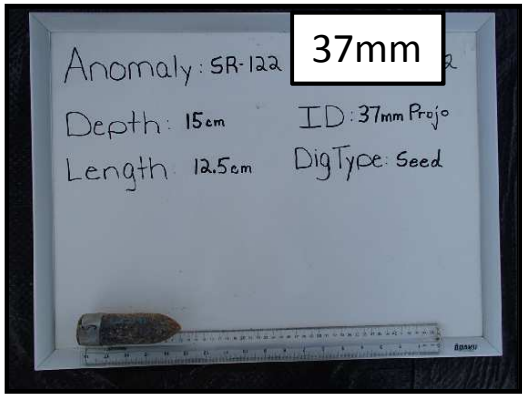


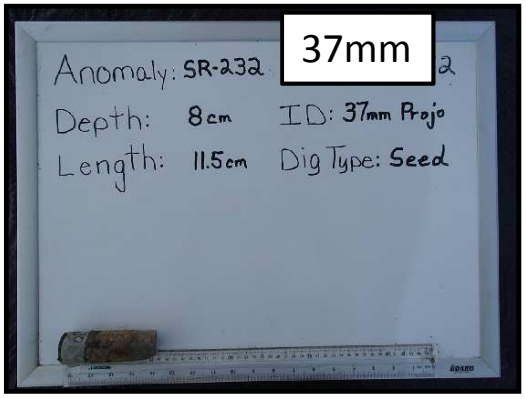


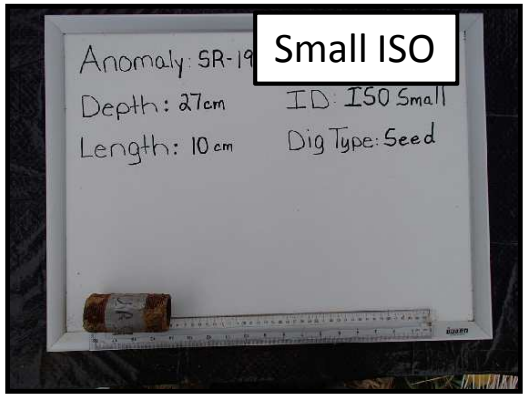








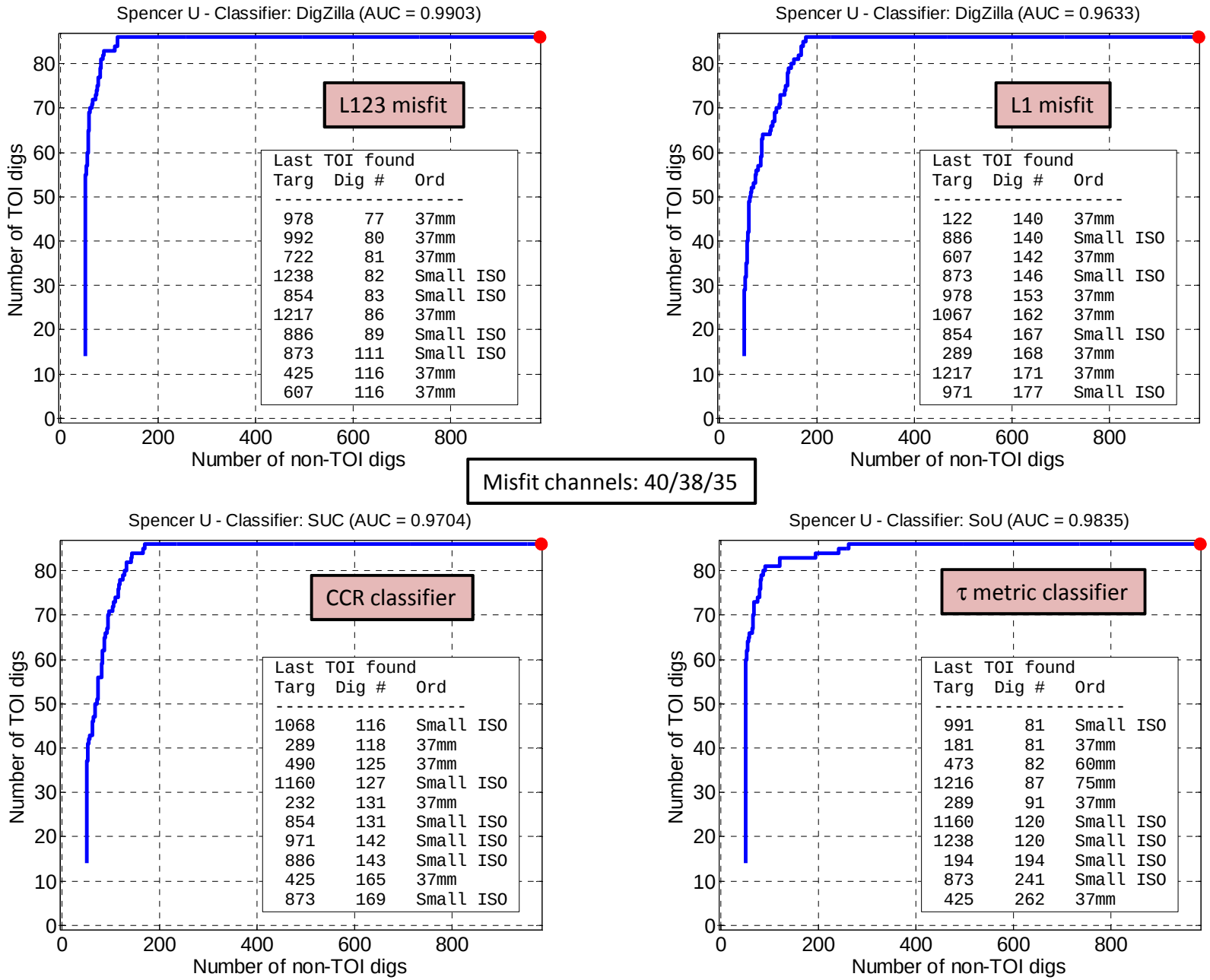




Spencer

TOI	Number
37mm	30
Small ISO	28
75mm	16
60mm	5
Medium ISO	5
105mm	1
155mm	1
Total	86

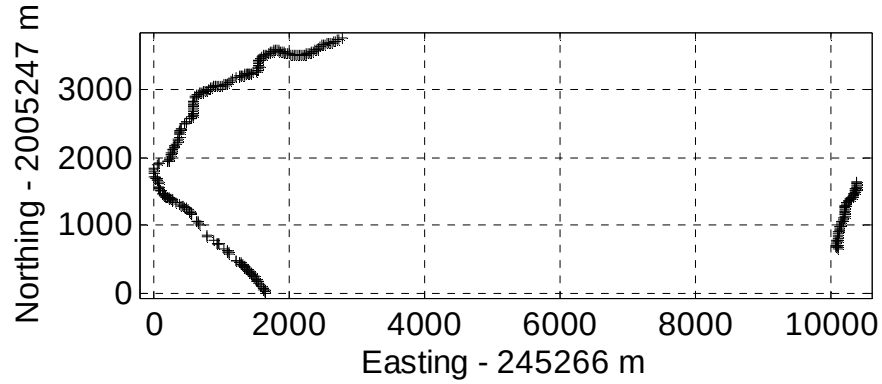
Spencer TOI

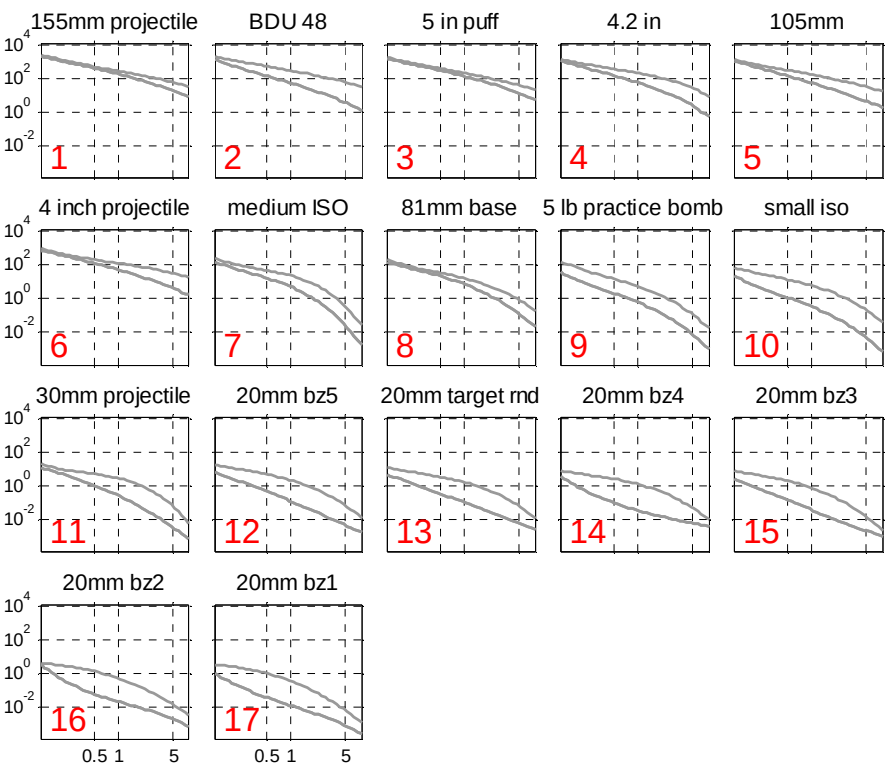

Spencer U - Classifier: DigZilla (AUC = 0.9903)
L123 misfit
Last TOI found
Targ Dig # Ord
978 77 37mm
992 80 37mm
722 81 37mm
1238 82 Small ISO
854 83 Small ISO
1217 86 37mm
886 89 Small ISO
873 111 Small ISO
425 116 37mm
607 116 37mm
Number of TOI digs
Number of non-TOI digs
Spencer U - Classifier: DigZilla (AUC = 0.9633)
L1 misfit
Last TOI found
Targ Dig # Ord
122 140 37mm
886 140 Small ISO
607 142 37mm
873 146 Small ISO
978 153 37mm
1067 162 37mm
854 167 Small ISO
289 168 37mm
1217 171 37mm
971 177 Small ISO
Number of TOI digs
Number of non-TOI digs
Misfit channels: 40/38/35
Spencer U - Classifier: SUC (AUC = 0.9704)
CCR classifier
Last TOI found
Targ Dig # Ord
1068 116 Small ISO
289 118 37mm
490 125 37mm
1160 127 Small ISO
232 131 37mm
854 131 Small ISO
971 142 Small ISO
886 143 Small ISO
425 165 37mm
873 169 Small ISO
Number of TOI digs
Number of non-TOI digs
Spencer U - Classifier: SoU (AUC = 0.9835)
τ metric classifier
Last TOI found
Targ Dig # Ord
991 81 Small ISO
181 81 37mm
473 82 60mm
1216 87 75mm
289 91 37mm
1160 120 Small ISO
1238 120 Small ISO
194 194 Small ISO
873 241 Small ISO
425 262 37mm
Number of TOI digs
Number of non-TOI digs

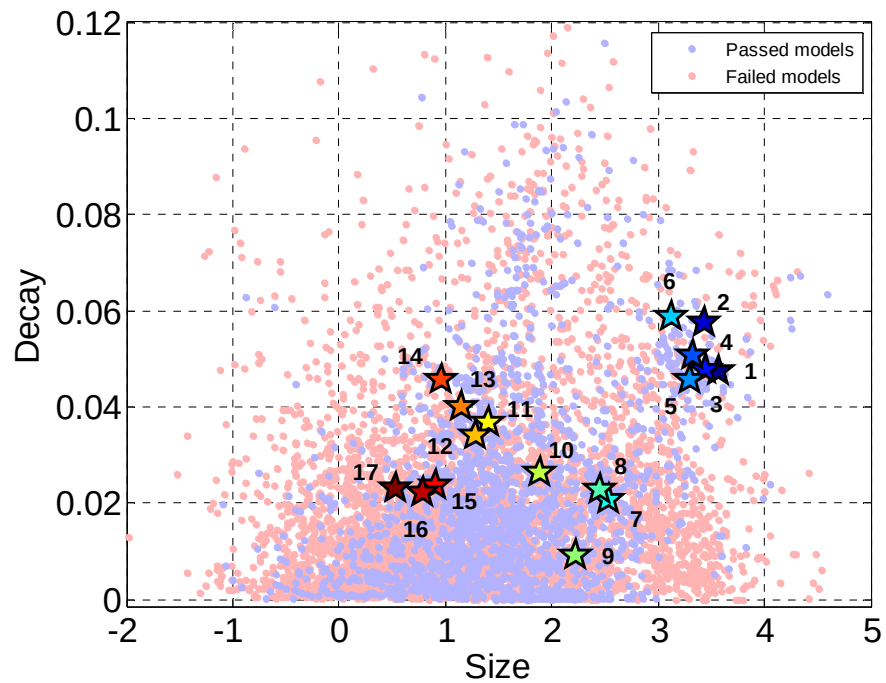
Vieques (2012)

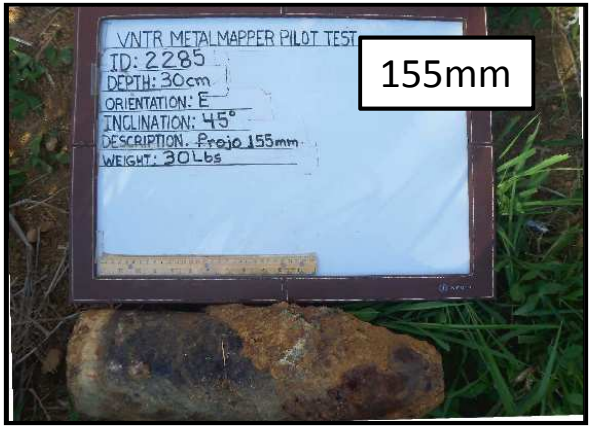
MM Cued

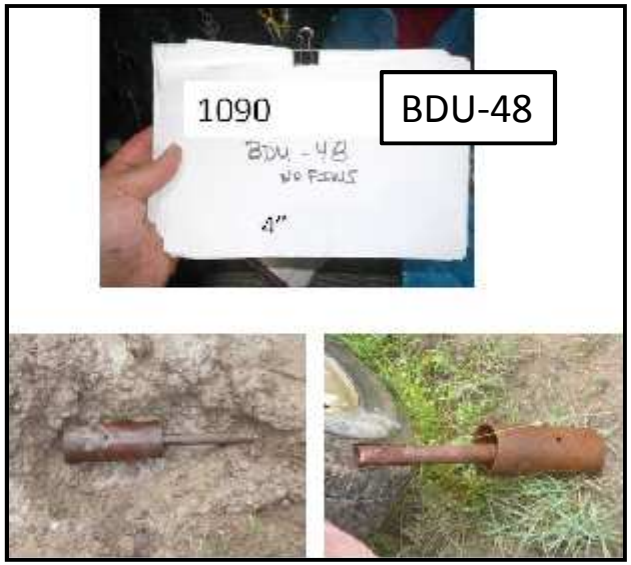
- 1020 targets
- 92 TOI
- Difficult site. Wide range of targets, including several 20mm. Magnetic soil. Poorly centered MM in many cases. Many “cannot analyze” anomalies.

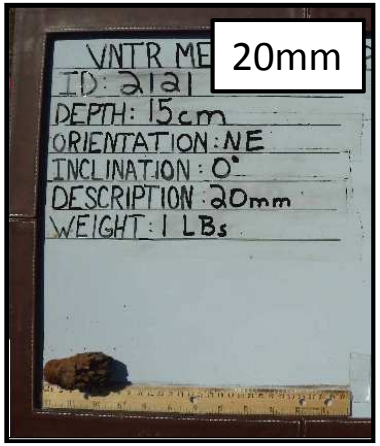










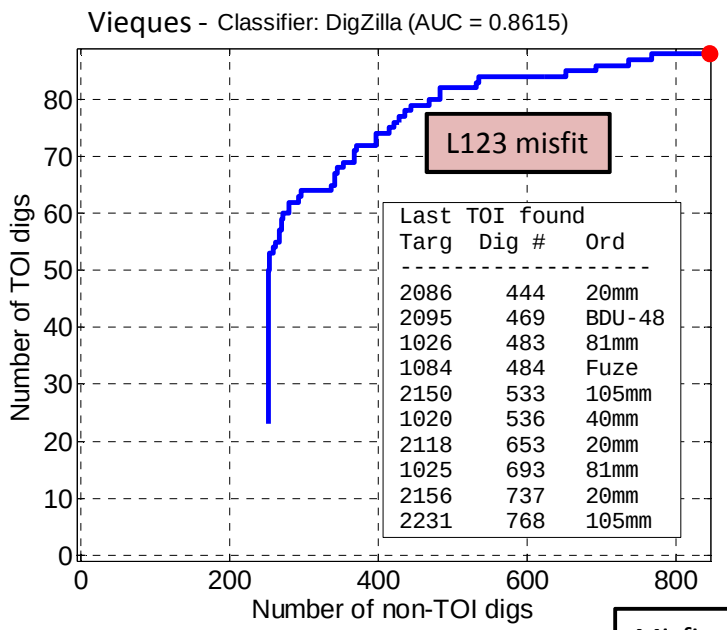


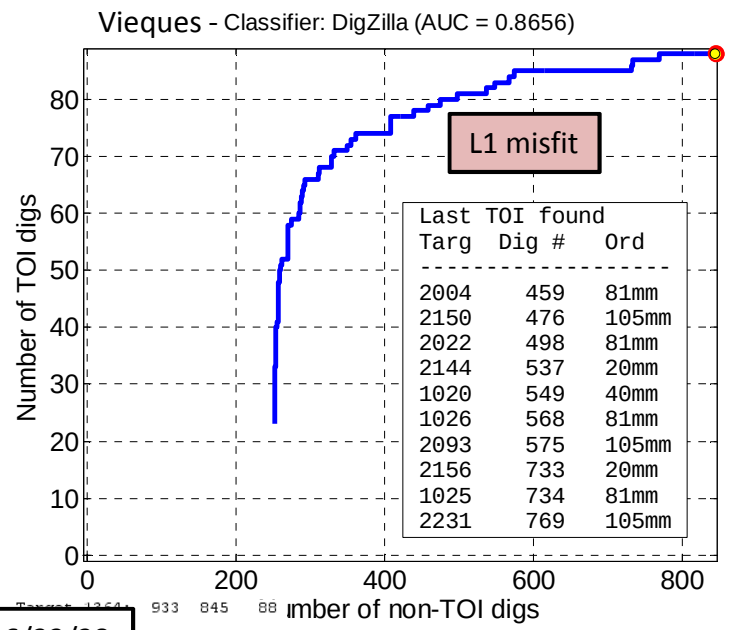
Vieques

TOI	Number
20mm	28
81mm Ilum. Rnd.	11
105mm	10
BDU-48	9
155mm	6
4" Projectile	6
5" Projectile	6
30mm	4
4.2" Mortar	4
8" Proj.	4
Medium ISO	4
40mm	2
5 lb Practice	2
A-MN Fuze	1
Total	97

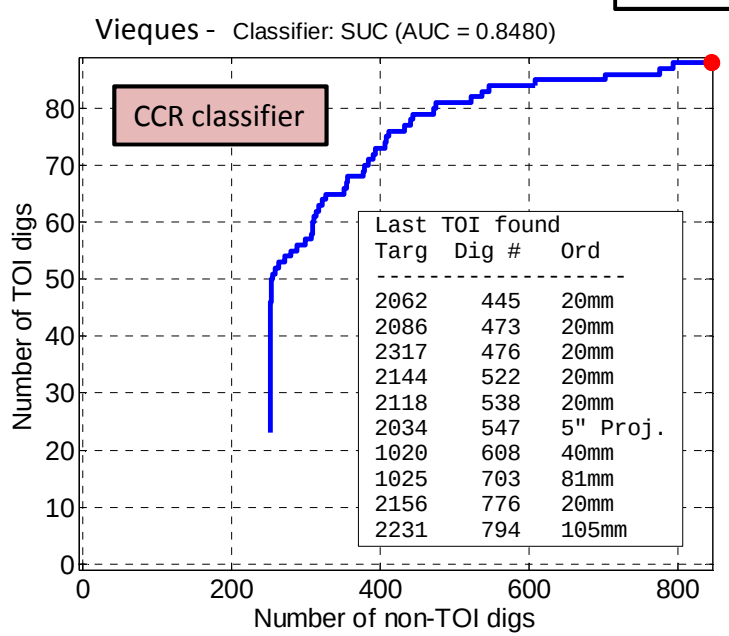
Vieques TOI

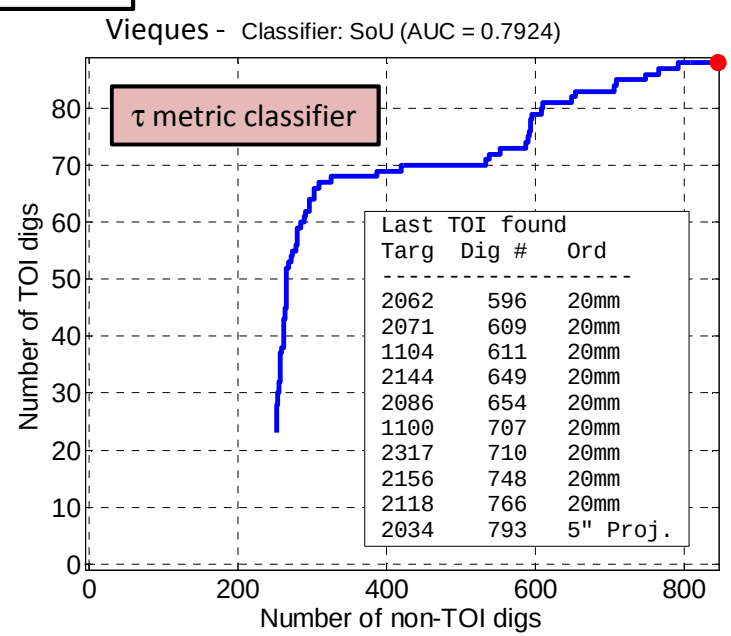
Not all TOI have photos





Misfit channels: 36/33/28

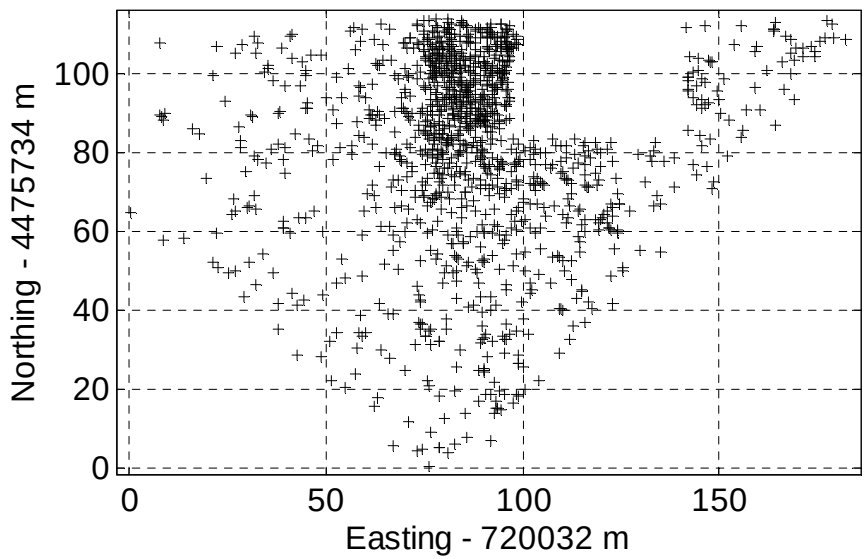


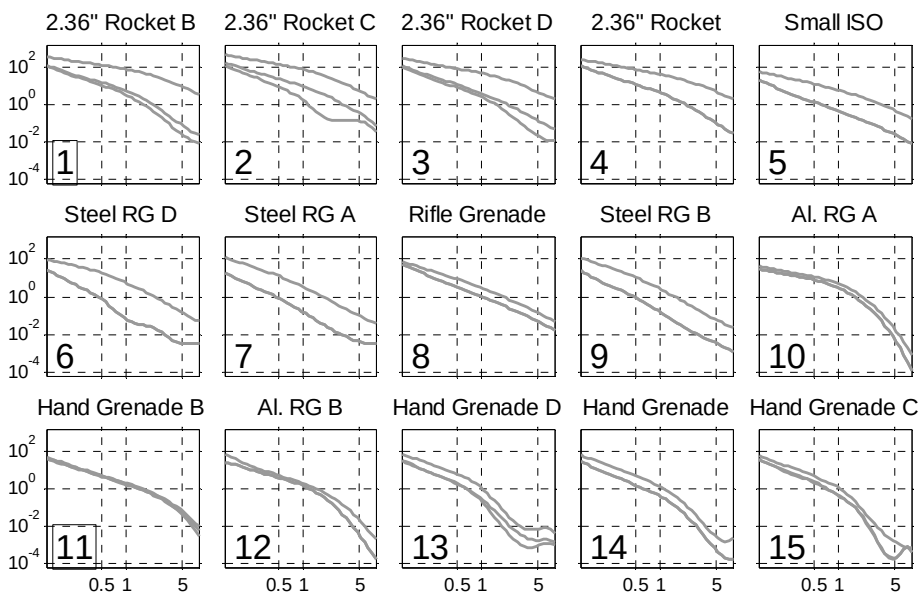


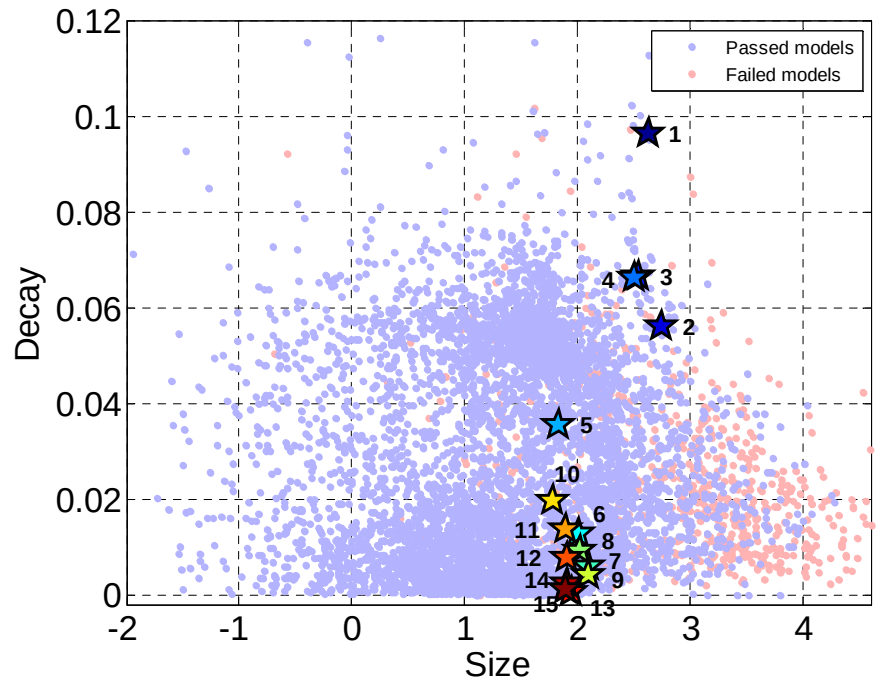
Ellis (2013)

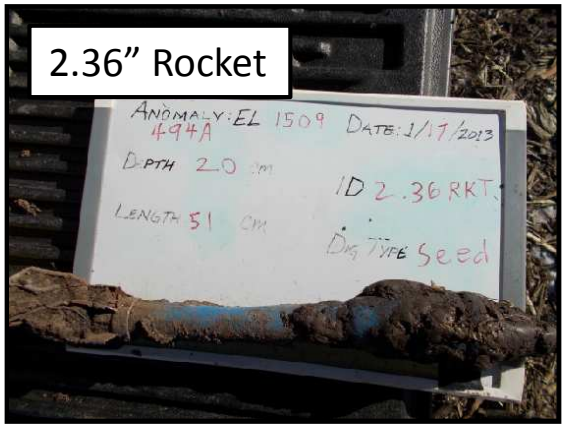
MM Cued

- 1195 targets
- 65 TOI
- Different TOI relative to past ESTCP demos: rockets, rifle and hand grenades. Large number of non-TOI rocket weights (690).





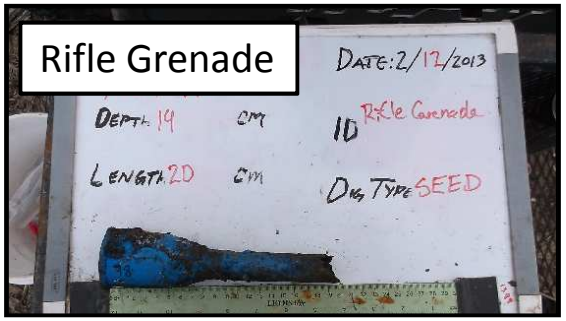


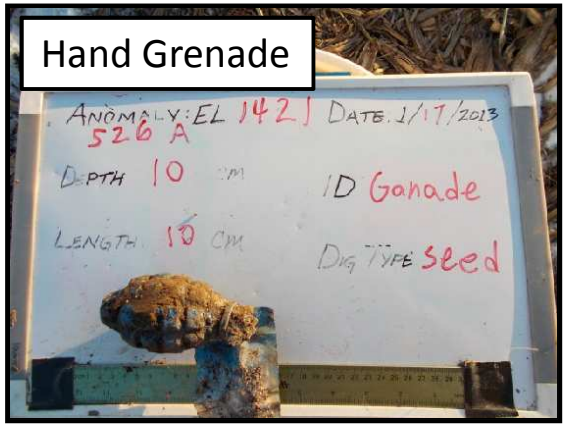


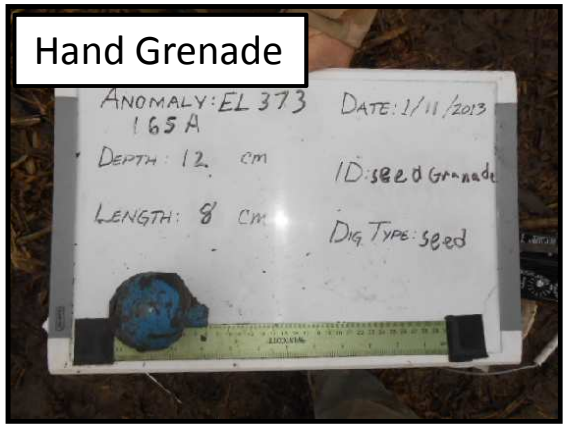
Ellis

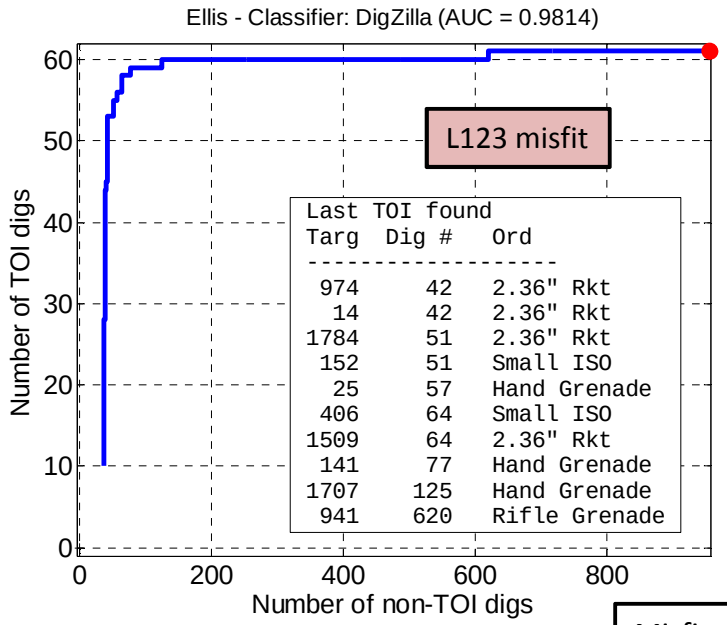
TOI	Number
Small ISO	35
2.36" Rocket	17
Hand Grenade	10
Rifle Grenade	3
Total	65

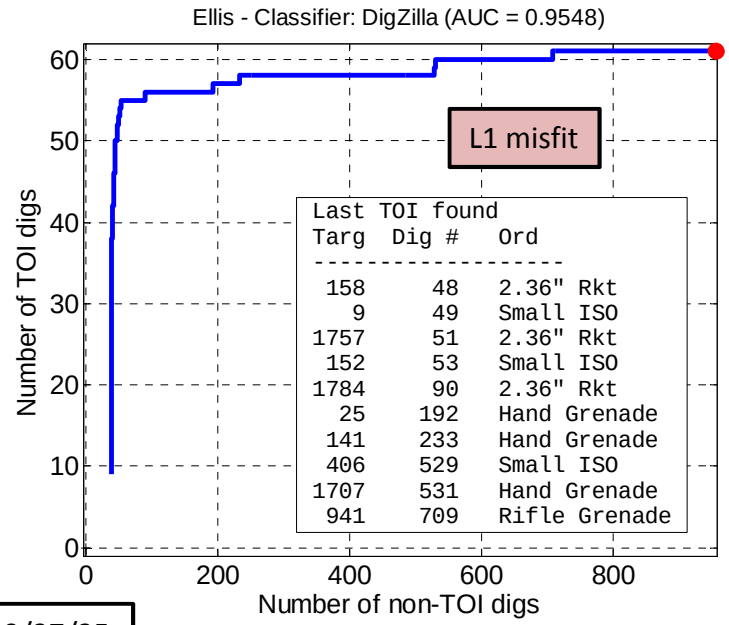
Ellis TOI



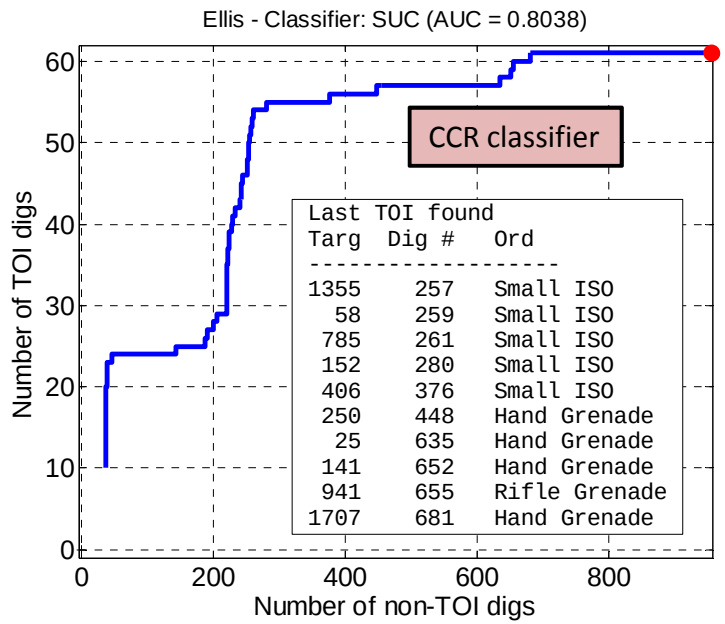


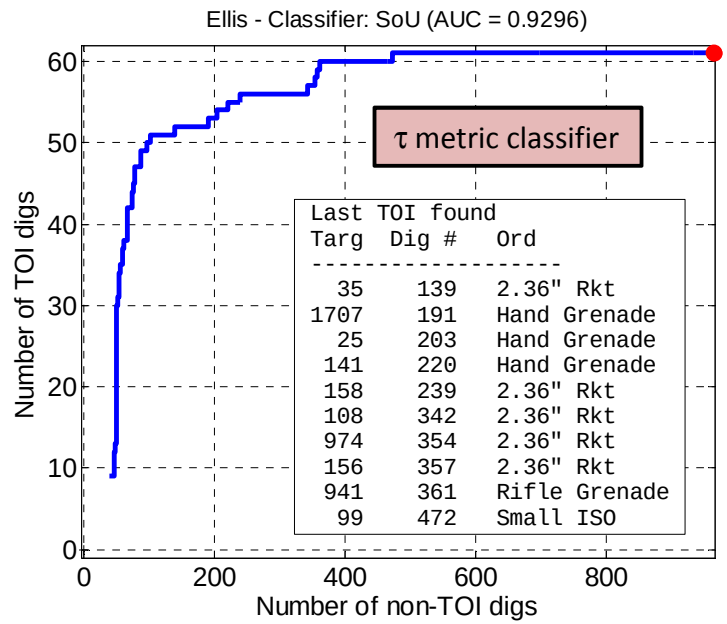






Misfit channels: 39/37/35

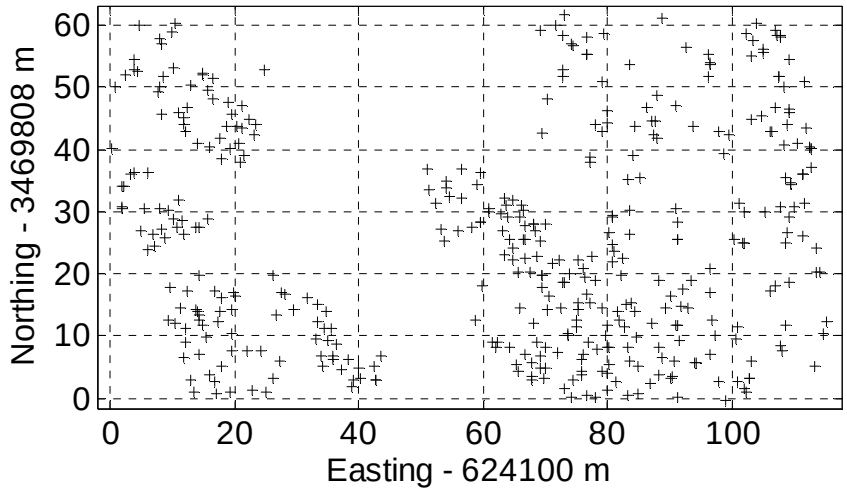


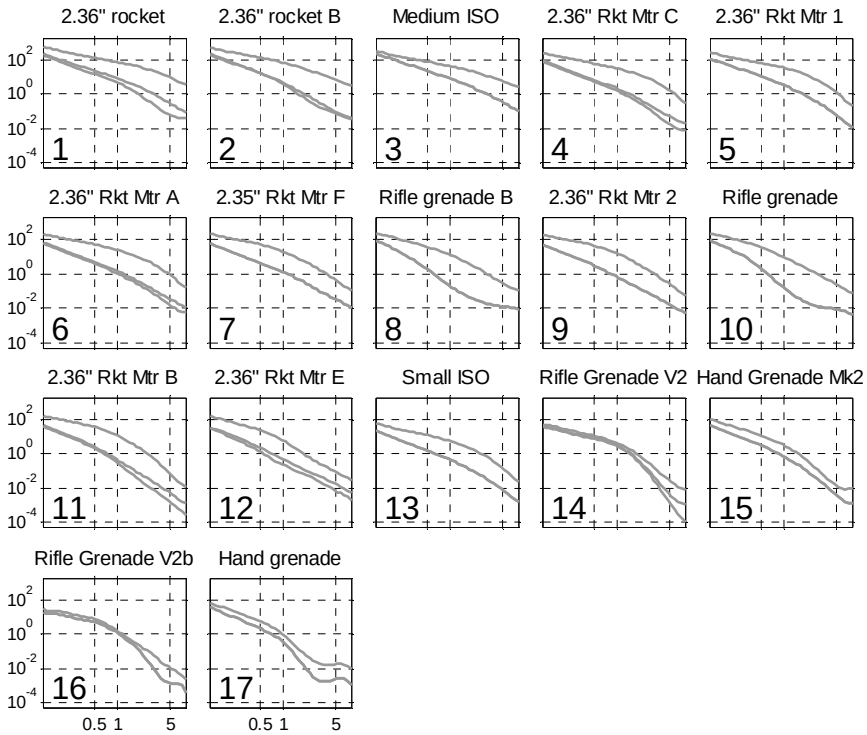


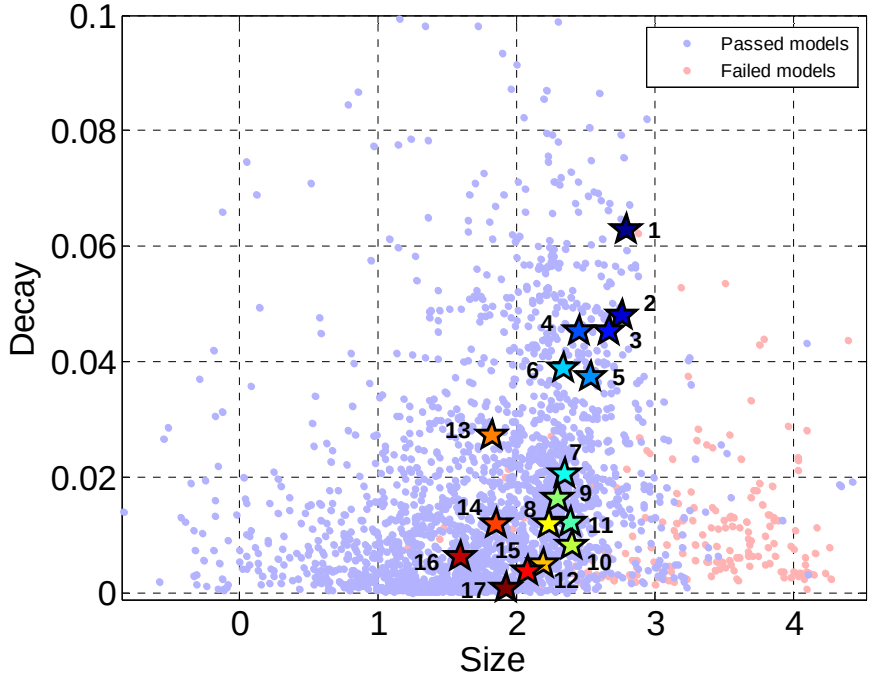
Rucker (2013)

MM Cued

- 407 targets
- 201 TOI
- Rocket motors counted as TOI. Much of the scrap looked like TOI. Very difficult. Generally considered not an appropriate site for classification.







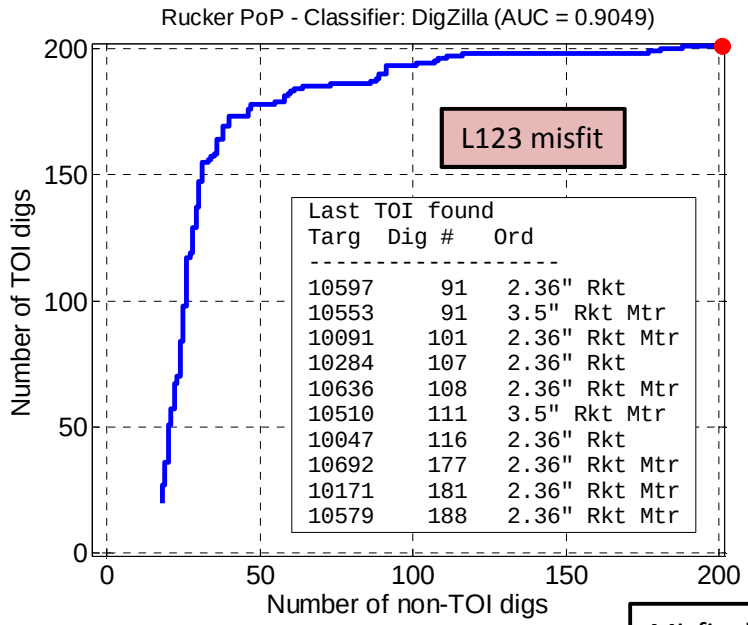
Rucker PoP

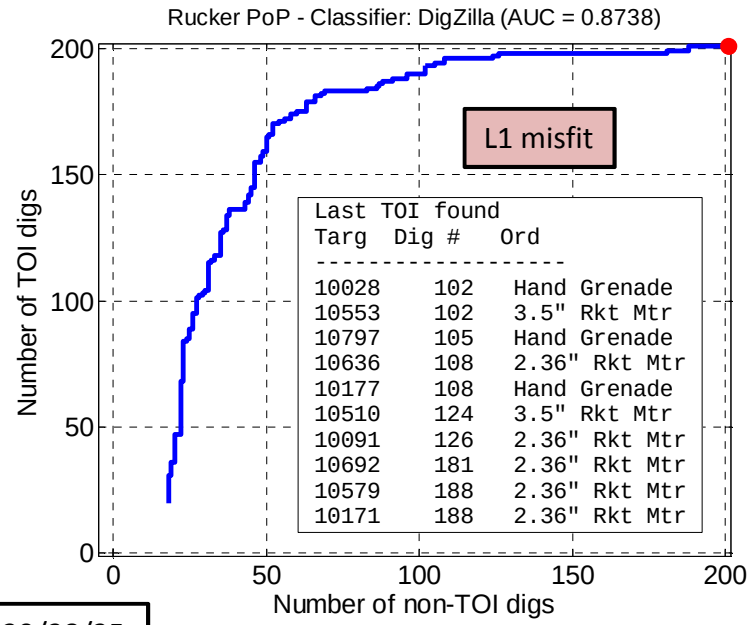
TOI	Number
2.36" Rocket Motor	155
2.36" Rocket	23
3.5" Rocket Motor	11
Rifle Grenade	6
Hand Grenade	5
ISO	1
Total	201

Rucker TOI

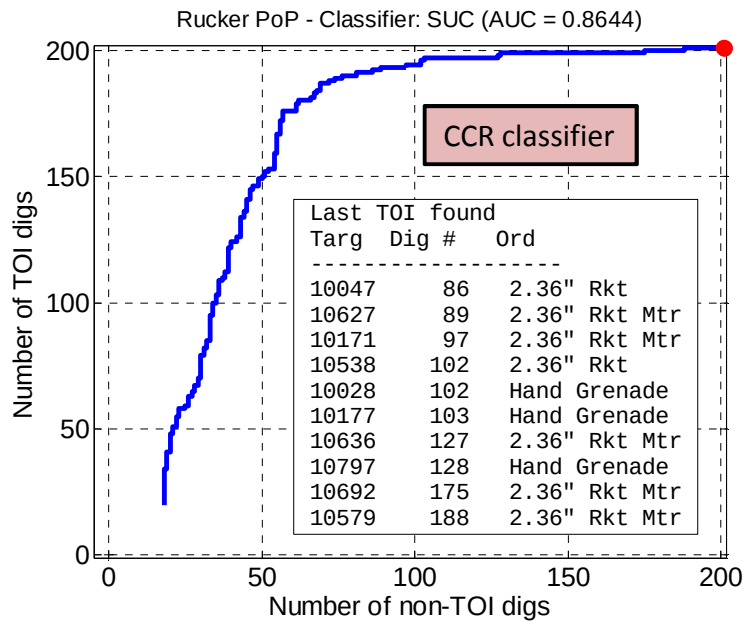
Proof of Principle Area

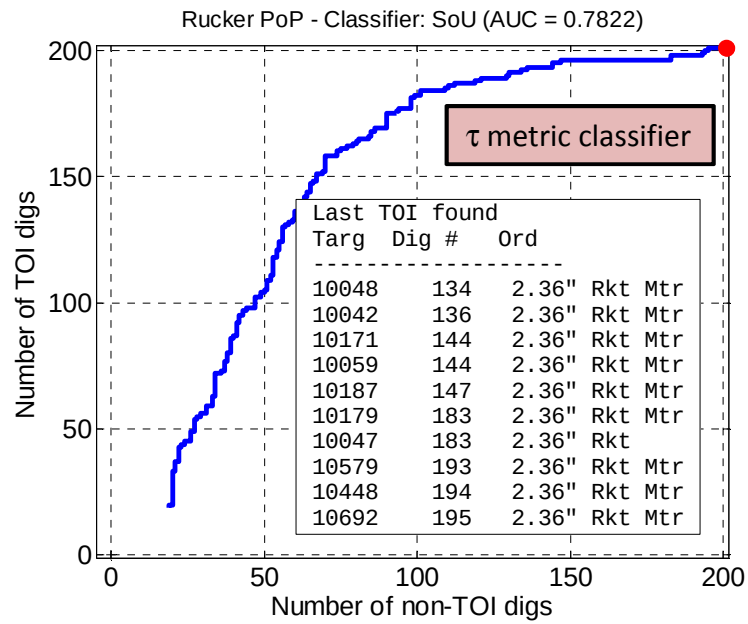
No photos





Misfit channels: 39/38/35





Appendix D. Detecion modeller user manual

Detection modeller quickstart guide

1. Installation & system requirements
2. Detection modeller (DM) interface
3. Transmitter/receiver specification
4. Survey parameters
5. Response curves
6. Clearance depth analysis
7. Importing polarizabilities
8. Querying the polarizability library
9. Optimizing the detection channel
10. Exporting results
11. Support

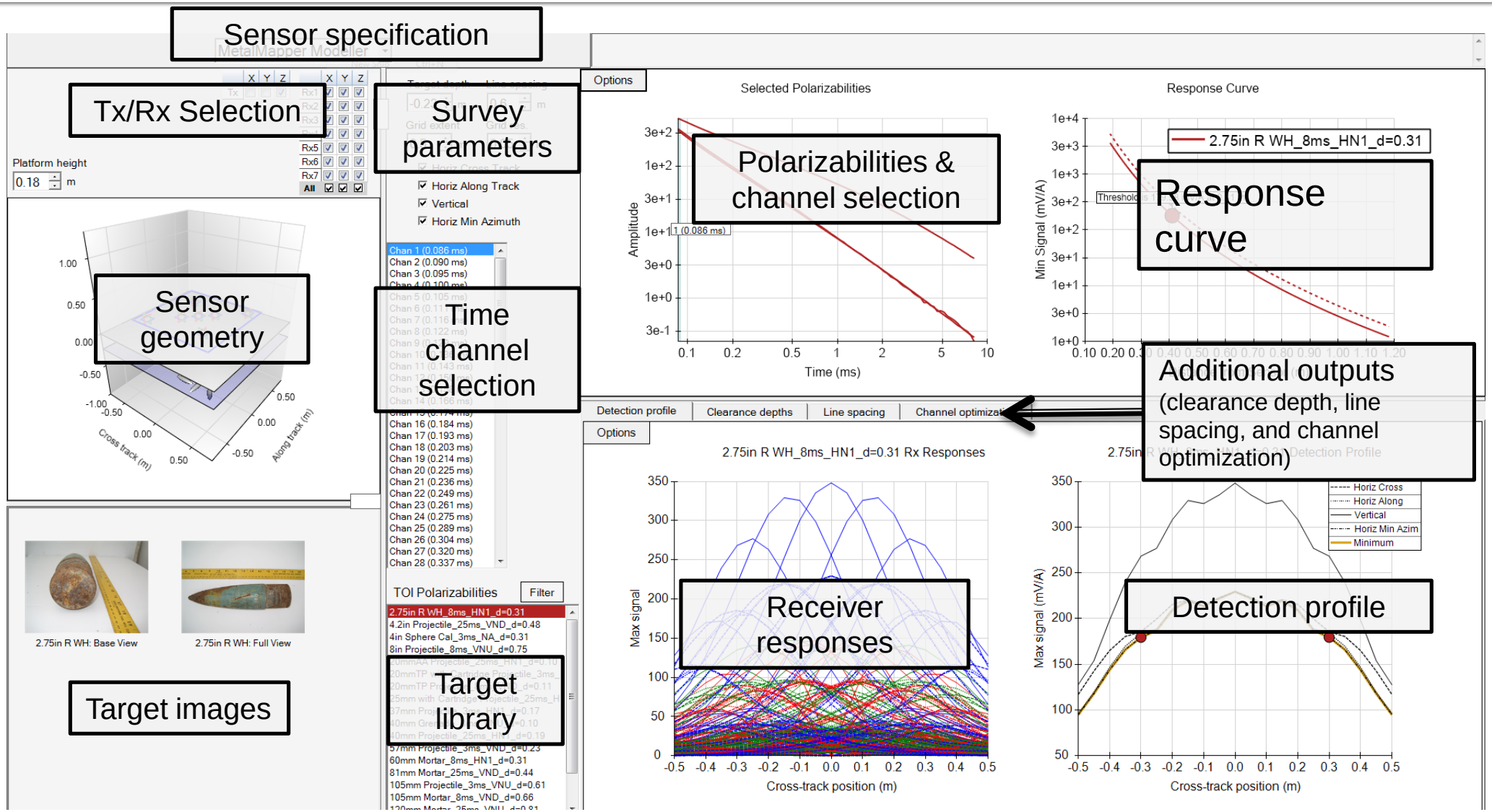
Software development funded under SERDP MR-2226: Decision support tools for munitions response performance prediction and risk assessment

V 2.0 June 2016

Installation and system requirements

- Detection modeler runs on Windows 7, 8.1, and 10 (64 bit versions)
- To install, run the msi installer on your machine
- First installation may take up to 15 minutes to install SQL server.

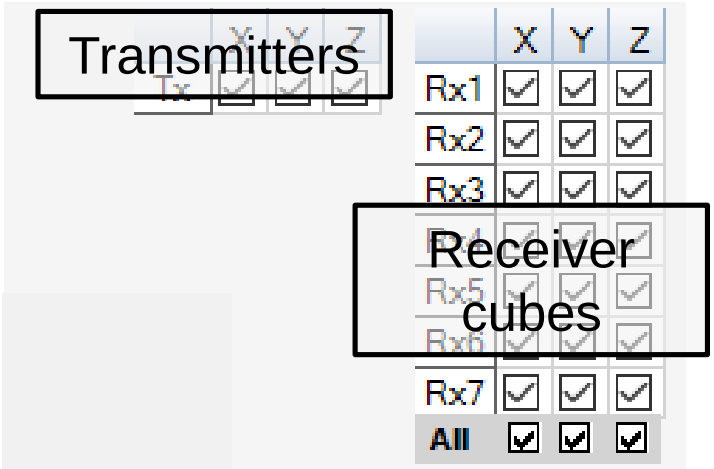
Detection modeller interface



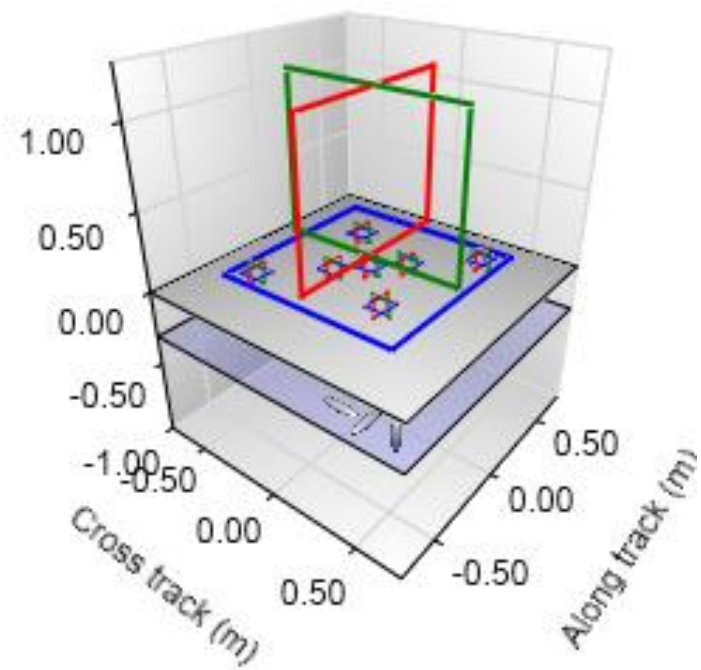
Options for display, file import and export can be accessed by clicking the options button at the top left of each panel, or by right clicking within each panel.

Transmitter/Receiver specification

Threshold can be calculated for a selected subset of transmitters and receivers

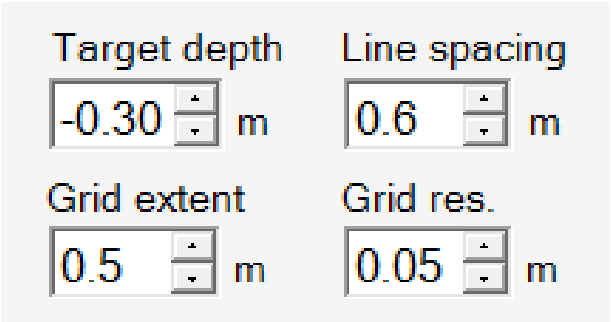


Selected transmitter and receiver combinations are highlighted in sensor geometry plot

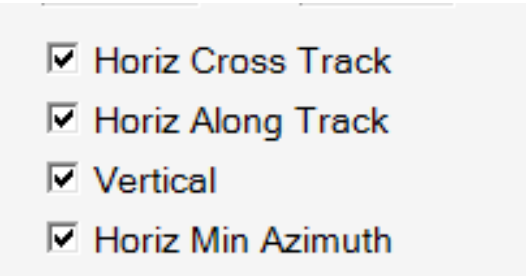


Typically we are interested in detection with Z-component measurements for vertical field (Z) transmitters. These tend to give the maximum response, so you will notice that the threshold does not change when you de-select other X and Y transmitters and receivers

Survey parameters



- **Target depth**: depth below ground surface of target (m)
- **Line spacing**: survey line spacing (m)
- **Grid extent**: extent of target location grid in x and y, relative to sensor location (m).
- **Grid res**: resolution of target grid (m)



Orientation permutations:

The detection threshold is calculated over all vertical and horizontal permutations of target orientation.

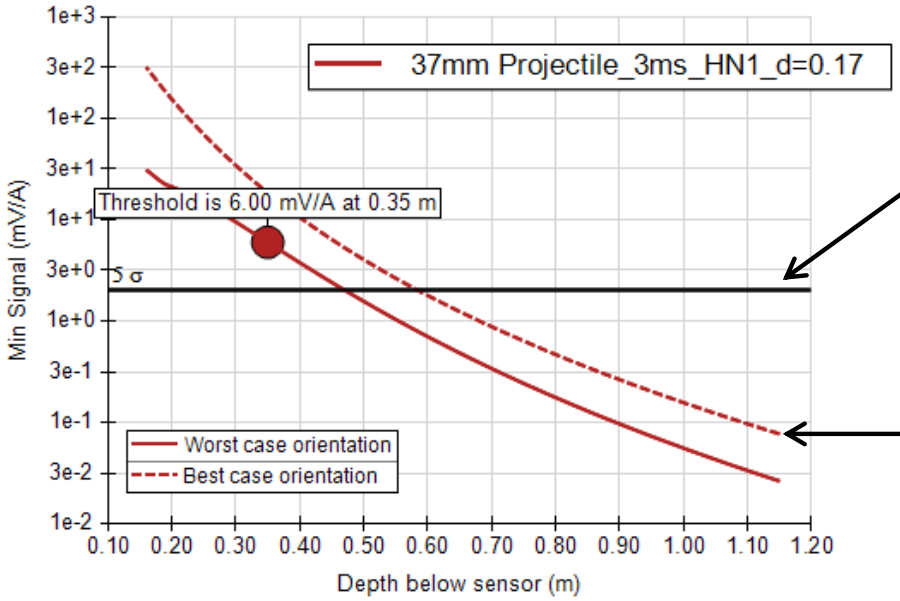
Horizontal minimum azimuth corresponds to a horizontal target with azimuth producing the minimum response for a given cross-track location. For the MetalMapper in particular this azimuth does not necessarily correspond to cross-track or along-track orientations.

All orientation permutations should be selected in order to obtain the worst-case detection threshold

SERDP
DOD • EPA • DOE

Response curves

- The response curve shows the dependence of the detection threshold on the depth of the target below the sensor. This depth is the vertical distance **from the center of the receiver cubes to the target.**



A noise threshold at the selected time channel can be displayed by importing noise:
Options → Import Noise Std. Dev.

The noise threshold can be adjusted by selecting:
Options → Set noise multiplier

Toggle display of best case orientation by selecting
Options → Show thresh. for best orientation

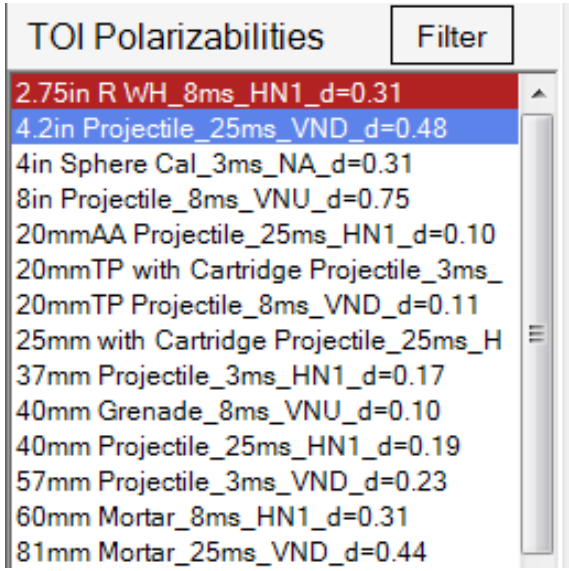
Depth limits on response curve can be adjusted by selecting ***Options → Response Curve Limits***

SERDP
DOD • EPA • DOE

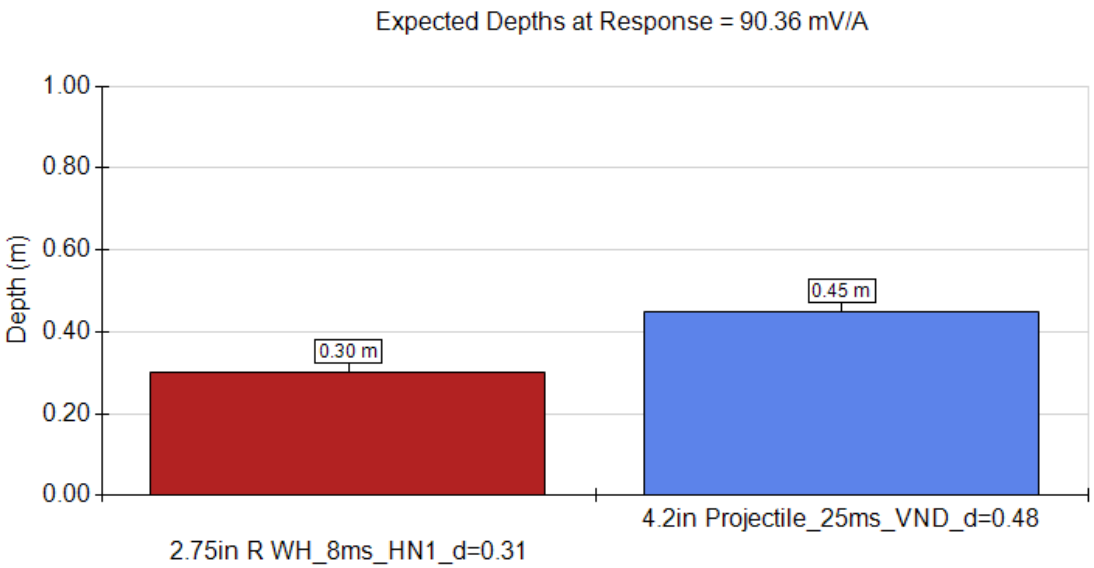
Clearance depth analysis

- Clearance depth analysis predicts the maximum depth at which targets in their worst case orientations will be detected for a defined threshold.

1. In the target list, select multiple targets of interest expected at the site. *The first selected target, at the specified target depth and line spacing, defines the detection threshold*



2. Predicted clearance depths for selected items are shown in the “Clearance depth” tab

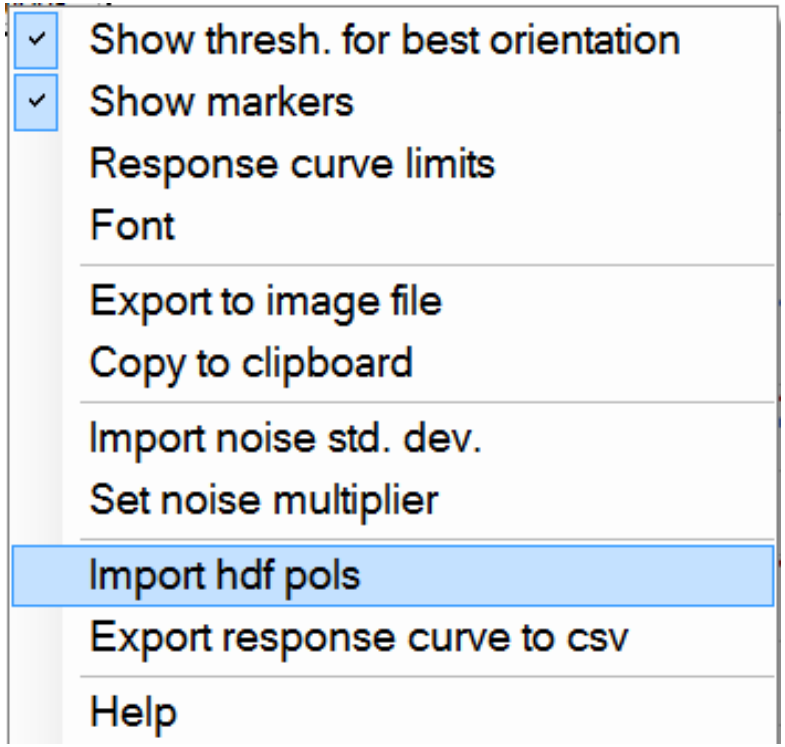


You can adjust the y-axis limits on this plot by selecting ***Options*** → ***Y-axis limits***

SERDP
DOD • EPA • DOE

Importing polarizabilities

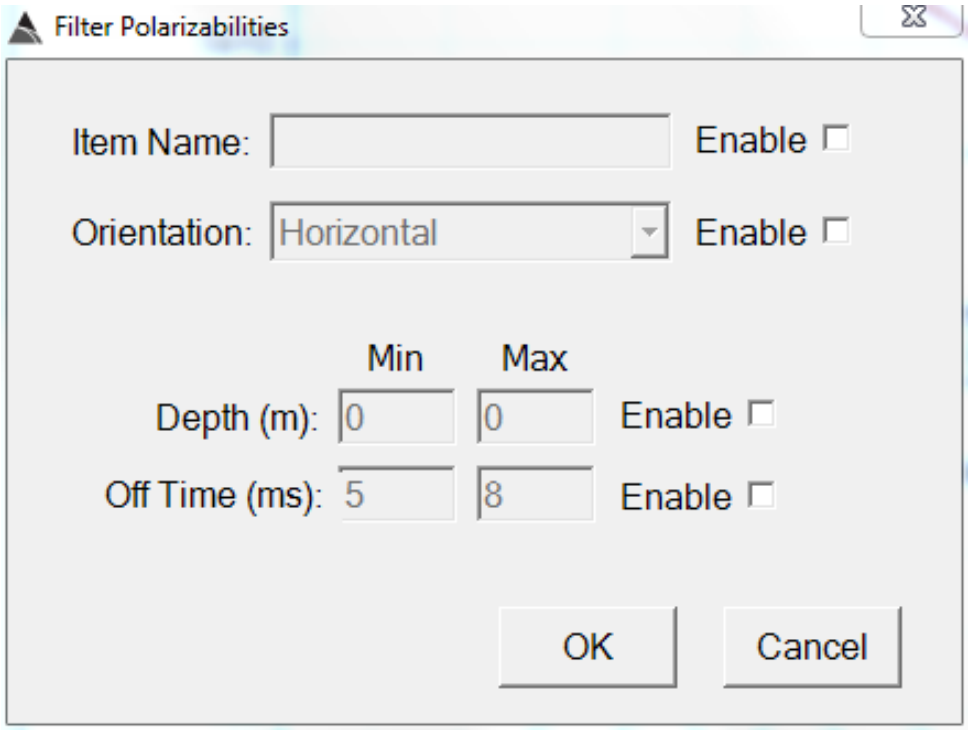
To import new target polarizabilities into the Detection Modeller: select ***Options*** → ***Import hdf pols*** in the upper panel



- *hdf5* files must conform with the file format defined in the DoD standardized polarizability library

Querying the polarizability library

You can refine the list of polarizabilities by clicking on the ***Filter*** button. You search by item name, orientation, depth and off-time. There are defined for each set of target polarizabilities in the DOD library.



Importing noise

- Noise standard deviations can be imported into the Detection Modeller in a csv format file. The file should contain a column of numbers, each corresponding to the noise standard deviation at a time channel, e.g:

```
10.5
5
2.5
1.2
0.5
…
```

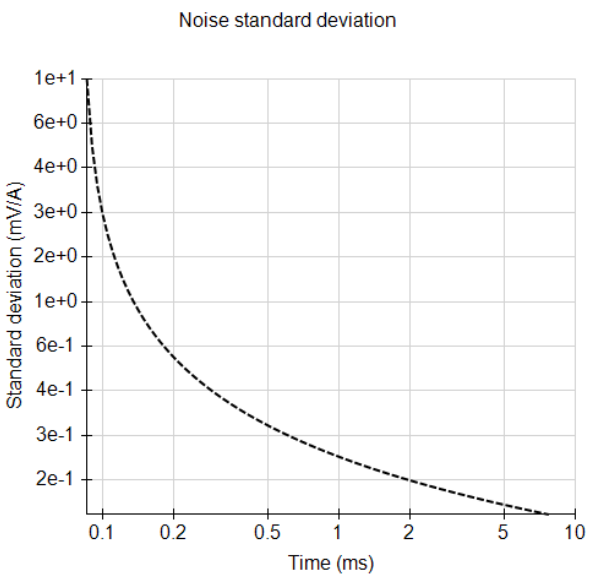
The noise model will only be used if the number of entries matches the number of time gates for the selected polarizabilities

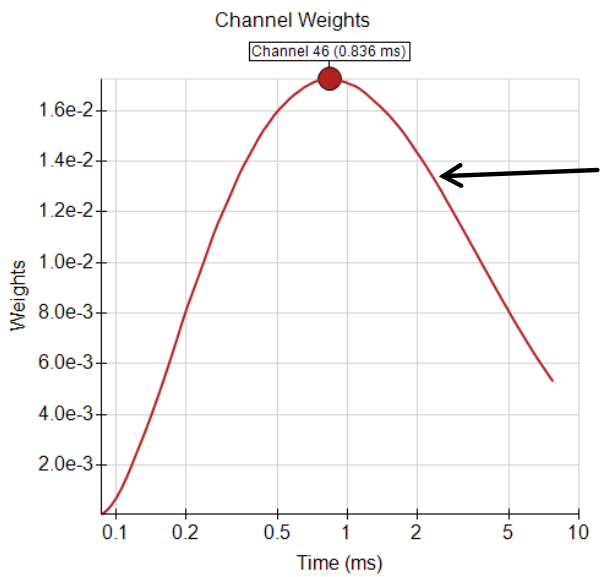
Optimizing the detection channel

- To identify an optimal detection channel for a target of interest (TOI), we maximize SNR for the worst case scenario.

1. Import noise standard deviations from a csv file:
   Options → Import Noise Std. Dev.
   The csv file should contain a list of noise standard deviations. The number of entries in this file must match the number of time channels for the selected TOI's polarizabilities.

2. The algorithm will only consider channels where the detection threshold exceeds the specified noise threshold (usually 5 times the noise standard deviation). You can adjust the noise threshold by selecting ***Options →Set noise multiplier.***



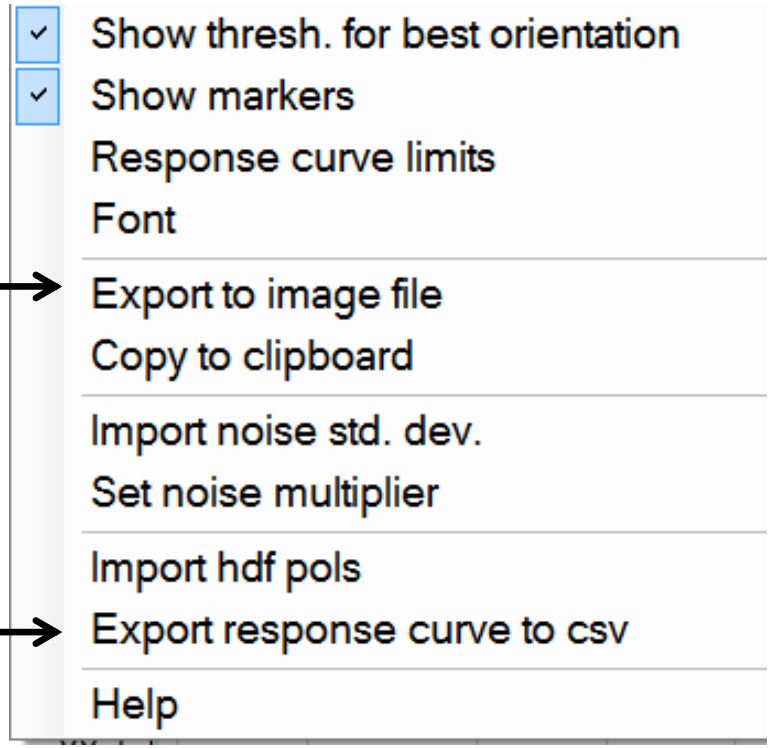


Example weight distribution. Use the channel with the maximum weight for target picking

Exporting results

1. All plots can be exported by selecting "*Options*" (or right-clicking) and selecting "*Export to image file*" (png, jpg, or pdf formats), or "*Copy to clipboard*" from the context menu

2. Response curve data can be exported to a csv file by right clicking in top (response curve) panel and selecting "*Export response curve to csv*"



Support

Please contact

info@btgeophysics.com

to report bugs or for assistance.