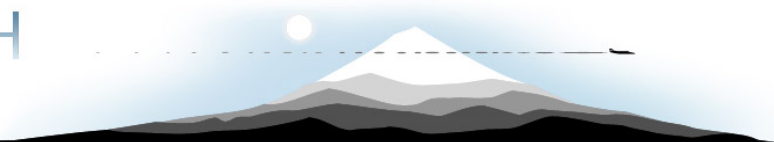


SKY RESEARCH



**Demonstration Report for the
Former Lowry Bombing and Gunnery Range**

**Project 200504: Practical Discrimination Strategies for
Application to Live Sites**

Prepared by

Sky Research, Inc.

445 Dead Indian Memorial Road

Ashland, OR 97520

Prepared for

Environmental Security Technology Certification Program



Submittal Date: March 22, 2007

Final v. 2.0

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
<small>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to the Department of Defense, Executive Services and Communications Directorate (0704-0188). Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</small>					
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.					
1. REPORT DATE (DD-MM-YYYY) 22/03/2007		2. REPORT TYPE Demonstration Report Final		3. DATES COVERED (From - To) 9/2005 - 3/2007	
4. TITLE AND SUBTITLE Demonstration Report for the Former Lowry Bombing and Gunnery Range (FLBGR), Project 200504 Practical Discrimination Strategies for Application to Live Sites				5a. CONTRACT NUMBER W912HQ-05-C-0018	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Dr. Stephen Billings, Leonard Pasion, Jon Jacobson, Stacey Kingsbury - Sky Research, Inc. Dr Douglas Oldenburg, Laurens Beran, Dr. Lin Ping Song, David Sinex, Dr. Nicolas Lhomme - UBC GIF				5d. PROJECT NUMBER ESTCP 200504	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Sky Research, Inc. 445 Dead Indian Memorial Road Ashland, OR 97520				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) ESTCP Program Office Attn: Dr. Anne Andrews 901 North Stuart St. Suite 303 Arlington, VA 22203				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT This project is attempting to demonstrate the application of feature extraction and statistical classification to the problem of UXO discrimination. The demonstration utilized Geonics EM-61 and EM-63 data collected at two sites on the FLBGR. The objectives of the Rocket Range surveys were the discrimination of a mixed range of projectiles with minimum diameter of 37 mm from other items. The 20 mm Range Fan survey objective was to discriminate 37 mm projectiles from 20 mm projectiles and 50 caliber bullets. Discrimination results when using size-based feature vectors were comparable on both-sites. Speed of survey, ease of use and reliability make the EM-61 more suited for this mode of discrimination. The techniques could be immediately transitioned into production field use on the 20 mm Range-Fan. At the Rocket Range additional testing to verify performance against more munitions types would need to be conducted The EM-63 is better suited for a cued interrogation mode where it has the potential to constrain the time-decay properties of the polarization over a wider time-range.					
15. SUBJECT TERMS UXO, discrimination, statistical classification, feature vector analysis, FLBGR					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES 165	19a. NAME OF RESPONSIBLE PERSON Dr. Stephen Billings
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code) 541.552.5185

EXECUTIVE SUMMARY

This demonstration described in this report was conducted under project ESTCP MM-0504 “Practical Discrimination Strategies for Application to Live Sites”. This project is attempting to demonstrate the application of feature extraction and statistical classification to the problem of UXO discrimination. The demonstration utilized Geonics EM-61 MK-II towed array and Geonics EM-63 cart based data collected at two sites on the Former Lowry Bombing and Gunnery Range, Colorado. The demonstrations were conducted with the support of the USACE-Omaha and USACE-ERDC. The objectives of the Rocket Range surveys (8 acres) were the discrimination of a mixed range of projectiles with minimum diameter of 37 mm from shrapnel, junk, 20 mm projectiles and small-arms. The 20 mm Range-Fan survey (2 acres) presented a small-item discrimination scenario where the objective was to discriminate 37 mm projectiles from ubiquitous 20 mm projectiles and 50 caliber bullets. Both EM systems trialed were positioned by a Leica TPS 1206 Robotic Total Station, with orientation information provided by a Crossbow AHRS 400 Inertial Motion Unit. Data processing, feature extraction and statistical classification were all conducted within the University of British Columbia’s UXOLab software package. For the EM-61, 3-dipole instantaneous amplitude models were fit to the available 4 time-channels, while for the EM-63, 3-dipole Pasion-Oldenburg models were recovered from the 26 time-channel data. Parameters of the dipole model were used to guide a statistical classification. Canonical and visual analysis of feature vectors extracted from the test-plot data indicated that discrimination could best proceed using a combination of a size- and a goodness of fit-based feature vector. A Support Vector Machine classifier was then implemented based on those feature vectors and using the available training data.

Two phases of digging and training were conducted at the 20mm RF, and three phases at the Rocket Range. At the Rocket Range, twenty-nine MK-23 practice bombs were recovered, with only one other UXO encountered (a 2.5 inch rocket warhead). At the 20 mm Range-Fan, thirty-eight 37 mm projectiles (most of them emplaced) were recovered, as were a large number of 20 mm projectiles and 50 caliber bullets. For both sites, and for both instruments, the SVM classifier outperformed a ranking based on amplitude alone. In each case, the last detected UXO was ranked quite high by the SVM classifier and digging to that point would have resulted in a 60-90% reduction in the number of false-alarms. This operating point is of course unknown prior to digging. We found that using a stop-digging criteria of $f=0$ (mid-way between UXO and clutter class support planes), was too aggressive and more excavations were typically required for full recovery of detected UXO. Both the amplitude and SVM methods performed quite poorly on two deep (40 cm) emplaced 37 mm projectiles at the 20 mm Range-Fan, exposing a potential weakness of the goodness of fit metric. Retrospective analysis revealed that thresholding on the size of the polarization tensor alone would have yielded good discrimination performance.

At the 20 mm Range-Fan it was found that 50 caliber bullets caused more false-alarms than 20

mm projectiles, even though they are significantly smaller. Retrospective analysis revealed that this was caused by a lower SNR¹ on the 50 caliber bullets. There was insufficient SNR to constrain the depth of the item and inversion solutions tended to be pushed deep due to either flat-objective functions or the presence of multiple locally optimal solutions. Consequently, size estimates of 50 caliber bullets obtained from the amplitude of the polarization tensor varied across four-orders of magnitude and tended to be overestimated. For the larger 20 and 37 mm projectiles, size estimates varied by around two-orders of magnitude, but there was less overlap between the two classes. Relatively poor depth performance on shallow, high SNR MK-23 practice bombs at the Rocket Range indicates that positional errors (and potentially unmodeled dipole components) also cause uncertainty in the object depth (and hence in the object size). We conclude that depth and size are poorly constrained when estimated from single component sensor data obtained with currently available positional precision. However, size estimates may still provide useful information to prioritize digging order.

During the demonstration, feature-vectors derived from the time-decay properties of the polarization tensor were not used to aid discrimination performance of either instrument. The noise-floor decays as $1/t^{0.5}$ while signal falls off more rapidly. This means that the accuracy of time-decay parameters extracted from low SNR anomalies is generally limited. However, retrospective analysis revealed that time-decay properties of the principal polarization tensor could have been used to distinguish MK-23 practice bombs from other items on the Rocket Range (for both instruments). On the 20 mm Range-Fan, the time range of the EM-63 is long enough that the slower decay rate of the 37 mm could have been distinguished from 20 mm projectiles. In contrast, the EM-61 did not sample late enough in time to aid discrimination.

EM-61 and EM-63 discrimination results when using size-based feature vectors were comparable on both sites. Speed of survey, ease of use and reliability make the EM-61 more suited for this mode of discrimination. The techniques could be immediately transitioned into production field use on the 20 mm Range-Fan. At the Rocket Range, additional testing to verify performance against more munitions types would need to be conducted. The EM-63 is better suited for a cued-interrogation mode where it has the potential to constrain the time-decay properties of the polarization over a wider time-range.

¹ Positioning error and sparse data coverage also likely contributed to the inability to constrain size.

TABLE OF CONTENTS

REPORT DOCUMENTATION	i
EXECUTIVE SUMMARY	ii
TABLE OF CONTENTS	iv
LIST OF TABLES	vi
LIST OF FIGURES	vii
ACRONYMS	xiii
ACKNOWLEDGEMENTS	xv
1 INTRODUCTION.....	1
1.1 Background.....	1
1.2 Objectives of the Demonstration	1
1.3 Regulatory Drivers.....	4
1.4 Stakeholder/End-User Issues	4
2 TECHNOLOGY DESCRIPTION.....	6
2.1 Technology Development and Application	6
2.1.1 Creation of a Map of Geophysical Sensor Data.....	10
2.1.2 Anomaly Selection and Feature Extraction	12
2.1.3 Classification of Anomalies.....	15
2.1.4 UXOLab Software	18
2.2 Previous Testing of the Technology	18
3 SITE CHARACTERISTICS AND DEMONSTRATION DESIGN	20
3.1 Performance Objectives.....	20
3.2 Test Site History/Characteristics.....	20
3.3 Pre-Demonstration Testing and Analysis	21
4 DEMONSTRATION SURVEYS AND PROCESSING.....	28
4.1 Demonstration Set-Up and Start-Up.....	28
4.2 Period of Operation.....	28
4.2.1 Sky-Research Five-Element EM-61 Array.....	30
4.2.2 Quad-sensor Magnetometer Survey.....	30
4.2.3 Geonics EM-63 Survey.....	31
4.2.4 Initial data processing	31
4.2.5 Feature Extraction.....	31
4.2.6 Statistical Classification.....	41
4.2.7 Ground-truth obtained.....	42
4.3 Discrimination results	43
4.3.1 Feature selection and choice of classifier	43
4.3.2 EM-61 results on the 20 mm Range-Fan	47
4.3.3 EM-61 results on the Rocket Range	49
4.3.4 EM-63 results on the 20 mm Range-Fan	52
4.3.5 EM-63 results on the Rocket Range	55
4.4 Summary and discussion of classification results.....	57

5	PERFORMANCE ASSESSMENT.....	59
5.1	EM-61 Performance Assessment.....	61
5.1.1	Terrain/Vegetation Restrictions.....	61
5.1.2	Ease of Use (Hardware).....	61
5.1.3	Probability of Discrimination (Pdisc) with a 50% Reduction in False Alarms	61
5.1.4	False-Alarm Rate with PDisc = 1	63
5.1.5	Location Accuracy of Interpreted Item.....	63
5.1.6	Percent Site Coverage.....	64
5.1.7	Processing Time (Initial Processing)	65
5.1.8	Processing Time (Interpretation)	65
5.2	EM-63 Performance Assessment.....	65
5.2.1	Terrain/Vegetation Restrictions.....	65
5.2.2	Ease of Use (Hardware).....	66
5.2.3	Probability of Discrimination (Pdisc) with a 50% Reduction in False-Alarms:...	66
5.2.4	False-Alarm Rate with PDisc = 1	67
5.2.5	Location Accuracy of interpreted item:	67
5.2.6	Percent Site Coverage:.....	67
5.2.7	Processing Time (Initial Processing):	68
5.2.8	Processing Time (Interpretation):	69
5.2.9	Probability of Detection (Pd) of EM-63 Sensor	69
5.2.10	Accuracy of Inversion Parameters.....	69
5.2.11	Survey Rate for EM-63 system.....	69
6	RETROSPECTIVE AND ADDITIONAL ANALYSIS OF PERFORMANCE	70
6.1	Retrospective analysis of the EM-63 results at the 20 mm Range-Fan	70
6.2	Retrospective analysis of the EM-61 data at the 20 mm Range-Fan.....	79
6.3	Additional analysis of the EM-61 results on the Rocket Range	81
6.4	Additional analysis of the EM-63 results on the Rocket Range	85
6.5	Investigation of data covariance matrix.....	91
7	COST ASSESSMENT	97
8	IMPLEMENTATION ISSUES.....	101
8.1	Environmental Checklist.....	101
8.2	Other Regulatory Issues.....	101
8.3	End-User Issues	101
8.4	Advantages and Limitations of the Technology	101
9	SUMMARY AND RECOMMENDATIONS FOR THE NEXT DEPLOYMENT	103
9.1	List of Significant Conclusions.....	103
9.1.1	Discrimination Performance at 20 mm Range-Fan	103
9.1.2	Discrimination Performance at the Rocket Range.....	103
9.1.3	Training Data Requirements	104
9.1.4	Disadvantages of the <i>Misfit/Amplitude</i> ² Feature Vector.....	104
9.1.5	Depth and Size Ambiguity Caused by Low SNR (and Positioning Inaccuracies)	104
9.1.6	The Role of Cooperative Inversion.....	105
9.1.7	Shape and Asymmetry as Discriminators.....	105

9.1.8	Time-Decay Metrics for Discrimination.....	106
9.1.9	EM-61 to EM-63 Comparison	106
9.2	Comments and Recommendations Concerning Feature Extraction	106
10	POINTS OF CONTACT	108
11	REFERENCES.....	109
APPENDIX A	A-1	
APPENDIX B	B-1	
APPENDIX C	C-1	
APPENDIX D	D-1	

LIST OF TABLES

Table 1.	Performance metrics for the demonstration.	3
Table 2.	Sensor and Equipment Summary	8
Table 3.	Previous Inversion/Classification Testing.....	19
Table 4.	Summary of ground-truth items recovered during excavations at three Rocket Range and one 20 mm Range-Fan grid in 2005.....	22
Table 5.	Number of anomalies selected with amplitudes above 10 mV in the 3 rd -time channel of the Sky towed array EM-61 data.	23
Table 6.	Key Project Activities	28
Table 7.	Number of anomalies identified as MK-23, 2.25” rocket, 37 mm projectile, 20 mm projectile, small arms, shrapnel and junk in each of the 10 grids used for this demonstration. For each grid, we identify the phase where ground-truth was released.	42
Table 8.	Different phases of classification on the 20 mm Range-Fan and the Rocket Range and the corresponding training data and feature vectors selected for classification for both the EM-61 and EM-63.	48
Table 9.	Number of targets flagged for digging by discrimination method, rocket range grids, Rounds 2 and 3.....	50
Table 10.	Simple example for computation of Spearman rank correlation between diglists ($\rho = -1$ in this example).	51
Table 11.	Performance assessment of the EM-61 and EM-63 surveys on the Rocket Range and 20 mm Range-Fan.....	59
Table 12.	Performance criteria and the metrics used for evaluation.....	60
Table 13.	EM-61 discrimination performance results at the 20 mm Range-Fan and Rocket Range.	62
Table 14.	EM-63 discrimination performance results at the 20 mm Range-Fan and Rocket Range.	67

Table 15. List of the number of field days required to collect EM-63 data at the Rocket Range and 20 mm Range-Fan.	69
Table 16. Polarization parameters recovered from test-stand data collected over 20 and 37 mm projectiles.	76
Table 17. Assumptions used to compare the different survey methods. The excavation reduction numbers were estimated from the performance at the Rocket Range.	98
Table 18. Comparison of the costs for the different modes of survey using the assumptions in Table 17, and in the bullet points immediately before that table.	99
Table 19. Comparison of the cost of survey for the different methods with percentage of holes to dig and different numbers of anomalies per acre. The smallest reduction in holes to dig that produces a cost less than the standard production method is marked in green.	100
Table 20. Comparison of the cost of survey for the different methods with different % of holes to dig and different amounts of time required for interpretation of each anomaly. The smallest reduction in holes to dig that produces a cost less than the standard produ	100
Table 21. Points of Contact.	108

LIST OF FIGURES

Figure 1. Sky Research's EM 61-MK2 towed array is constructed of composite materials and houses EM sensors, RTS laser positioning (or GPS) sensors, and the Crossbow IMU.	9
Figure 2. SKY man-portable MAG array integrates 4 Geometrics G-823 with the Leica 360° prism at FLBGR. This lightweight and durable system was deployed with a data tether.	9
Figure 3. Modified EM-63 cart collecting discrimination mode data at the Ashland test-site. ...	10
Figure 4. Sky Research utilizes the Leica RTS TPS1206 laser positioning system. This device is setup in over a known point and tracks a prism attached to the geophysical survey equipment. RTS technology out-performs GPS in terms of accuracy, sampling rate and operational ease-of-use.	10
Figure 5. A framework for statistical pattern recognition.	16
Figure 6. Nonparametric density estimate using Gaussian kernels. Kernel centers are shown as crosses. A large kernel width produces a smooth distribution (left) compared to a small kernel width (right).	17
Figure 7. Support vector machine formulation for constructing a decision boundary. The decision boundary bisects support planes bounding the classes.	17
Figure 8. Locations of the Rocket Range and 20 mm Range-Fan sites at FLBGR.	24
Figure 9. Map of time-channel 3 of the EM-61 survey of the FLBGR test-plot with emplaced items overlain.	25
Figure 10. Map of Rocket Range with areas surveyed for this demonstration outlined in red.	26
Figure 11. Map of the 20 mm Range-Fan, with two grids surveyed for this demonstration outlined in red.	27

Figure 12. Image of time channel 5 data on the 21-14 Grid.	33
Figure 13. Comparison of noise in two areas of 21-14 grid. Masks indicating less noisy and more noisy areas of the grid are shown as red rectangles in (a) and (b), respectively. Although these masked areas include targets, it is clear that (b) has a larger the noise-floor.	34
Figure 14. Filtered EM-61 Sky Array data acquired over Grid K-15. The approximately 20 m window was unable to remove the geologic anomaly.	35
Figure 15. Examples of masking data from Grid L-14.	35
Figure 16. Sounding and model fit over the anomaly maximum for an EM-63 anomaly on the Rocket Range. At left, we show the fit using a large mask where the estimated SNR is low and time-channels 11 onwards are not used to constrain the inversion. On right, is the fit when the mask is smaller which leads to a more appropriate SNR estimate and the use of time-channels 1 to 18 to constrain the inversion. The plan views are of time-channel 15.	36
Figure 17. Determining if an anomaly should be inverted as two separate targets. Anomalies 51 and 257 were initially inverted as two separate targets. (a) shows the result when inverting 257. (b) shows the result when the two anomalies are inverted as a single target.	39
Figure 18. Example of Two- and Three-polarization inversions of the same anomaly. The three-polarization model would be used in this instance due to a significantly reduced misfit.	40
Figure 19. Features extracted from FLBGR EM-63 testplot data.	43
Figure 20. Contributions of polarization parameters to first canonical eigenvector for canonical analysis of EM-63 test plot data. The eigenvector provides the maximum separation between 37mm items and small ordnance and explosives (OE) scrap (20mm and fuzes).	44
Figure 21. Comparison of feature spaces spanned by (top-left) misfit, (bottom-left) misfit normalized by anomaly amplitude, (bottom-right) misfit normalized by anomaly amplitude squared. Feature vectors are from test plot data, EM-63.	45
Figure 22. Feature space for classification of EM-63 data from Grid 19-14. “Train UXO” are emplaced 37 mm in 19-14, and “Test UXO” are 37 mm from Grid 21-14 (held out from classifier training). Contours show trained SVM output with $f=0$ corresponding to the decision boundary.	47
Figure 23. EM-61 ROC curves for 20mm Range-Fan grids. Circles show the operating points for each diglist: (on left) Results for 19-14 where the retrospective classifier replaces $\text{misfit}/\text{amplitude}$ by $\text{misfit}/\text{amplitude}^2$; (on right) Results for 21-14.	48
Figure 24. EM-61 ROC curves for the Rocket Range grids. Circles show the operating points for each diglist: (on left) Merged ROCs for Phase I Grids (I-12 and J-13), where the retrospective classifier replaces $\text{misfit}/\text{amplitude}$ by $\text{misfit}/\text{amplitude}^2$; (on right) Merged ROCs for Phase III grids (I-12 and J-13). Retrospective classifier replaces lsum with first canonical eigenvector.	49
Figure 25. Comparison of feature spaces for Grid J-13. Lsum denotes the sum of the instantaneous polarization amplitudes at the first time channel: (on left) Feature space with $\text{misfit}/\text{amplitude}$; (on right) Feature space with $\text{misfit}/\text{amplitude}^2$	51
Figure 26. Feature space used for EM-63 classification at the 20 mm Range Fan. Red targets are below the classifier operating point). Green targets are recommended for excavation, while yellow (21-14 only) are also recommended for excavation but with a lower priority: (on left)	

Feature space for 19-14 (using <i>misfit/amplitude</i> ² in place of <i>misfit/amplitude</i>); (on right) Feature space for 21-14.....	52
Figure 27. EM-63 ROC curves for 20 mm Range-Fan grids. Circles show the operating points for each diglist: (on left) Results for 19-14; (on right) Results for 21-14.....	53
Figure 28. EM-63 ROC curves for 21-14 with ground-truth error corrected for one 37 mm target. Circles show the operating points for each diglist: (on left) “Retrospective” classifier includes time domain information β_1 parameter as an additional feature. “Retrospective merged” combines classifiers for low and high SNR amplitude anomalies; (on right) Comparison of retrospective Pasion-Oldenburg and Lti classifiers.	53
Figure 29. Analysis of features from 19-14: (on left) Contributions to first canonical eigenvector separating 20 mm and 37 mm; (on right).....	54
Figure 30. (on left) Feature data for Grid I12, EM-63. (on right) Observed data for Target 3 identified in the left panel.	56
Figure 31. (on left) Merged ROC for Grids I-13 and J-12, EM-63. Retrospective classifier includes time decay information (β_1). (on right) Observed and predicted data for final UXO found in classifier diglists, Grid I-13.	56
Figure 32. Difference in the ground-truth and inverted EM-61 locations. On the top row, we show the results for the 20 mm Range Fan, split into different object types and include a cumulative distribution curve. Positions were bias adjusted by -4 cm East and -1 cm North.....	64
Figure 33. Difference in the ground-truth and inverted EM-63 locations. On the top row, we show the results for the 20 mm Range-Fan, split into different object types and include a cumulative distribution curve. No bias adjustments were applied to the positions.....	68
Figure 34. Sounding at anomaly maximum for all EM-63 anomalies (with ground-truth) on Grid 19-14 on the 20 mm Range-Fan, along with theoretical (t-0.5) and calculated (obtained through analysis of a signal-free part of the dataset) noise floors. Many of the 37 mm are quite shallow and have signal above the noise floor at late time. In contrast, the signal from most of the 50 caliber projectiles decays to the noise floor between 1 and 10 ms past signal turn-off. The 20 mm projectiles amplitudes are typically intermediate between the 50 caliber and 37 mm.	71
Figure 35. <i>Misfit/amplitude</i> ² versus the integral of the polarization for all anomalies (with ground-truth) on the 20 mm Range-Fan. Also shown are two deep (40 cm) 37 mm projectiles that were blind emplaced by ERDC and were below both the amplitude and original classification thresholds. These two deep anomalies have poor fits and occur late in the excavation order. These poor fits expose a weakness of the <i>misfit/amplitude</i> ² feature vector.	71
Figure 36. Left panel: Histogram of SNRs for 50 caliber, 20 mm and 37 mm anomalies on the 20 mm Range-Fan. The SNR is estimated via the data within the mask defined around each anomaly. The 37 mm anomalies typically have the highest SNR which is one reason that causes the <i>misfit/amplitude</i> ² criterion to work well at this site. Right panel: Histogram showing the number of time-channels used in the inversion for the three different ground-truth items.	72
Figure 37. Fitted versus actual depth for 50 caliber, 20 mm and 37 mm projectiles at the 20 mm Range-Fan (on left). Error in depth estimation versus the estimated object size (on right).	73

Figure 38. Data misfit as a function of depths for three different anomalies in Grid 19-14. Each anomaly has families of solution with different attributes that have optimal solutions at different depths. The top-left panel is for a 37 mm projectile that has a well defined minimum (at approximately the true depth). The top-right panel is also from a 37 mm projectile but this anomaly has a very poorly defined depth. The bottom-left panel is for a 50 caliber bullet with multiple minima. The item was found at 10 cm depth, close to the shallowest minima. However, there are two deeper minima including the global minimum at 88 cm depth..... 74

Figure 39. Polarization tensors recovered by inversion for all anomalies on the 20 mm Range-Fan. Top left panel: Shows the axial and transverse polarization tensors fit to test-stand data over 20 and 37 mm projectiles. Top right panel: Dominant polarizations for 37 mm projectiles. Bottom left panel: Dominant polarization for 20 mm projectiles; Bottom right panel: Dominant polarizations for 50 caliber projectiles..... 75

Figure 40. Normalized polarization tensors recovered by inversion for all anomalies on the 20 mm Range-Fan. Top left panel: Shows the axial and transverse polarization tensors fit to test-stand data over 20 and 37 mm projectiles. Top right panel: Dominant polarizations for 37 mm projectiles. Bottom left panel: Dominant polarization for 20 mm projectiles; Bottom right panel: Dominant polarizations for 50 caliber projectiles..... 76

Figure 41. Recovered Pasion-Oldenburg parameters of the dominant polarization tensor for grids 19-14 and 21-14 in the 20 mm Range-Fan. Left panel: Gamma parameter version the integral of the polarization (indicative of size). Right panel: Gamma versus beta parameters..... 77

Figure 42. ROC curve for Grid 21-14 using a classifier trained on beta, gamma and polarization integral feature-vectors obtained from Grid 19-14. Classifier operating point at $f=0$ of the SVM classifier are shown as circles. 78

Figure 43. Normalized decay curves for the sounding with maximum amplitude on Grid 19-14. Top left panel: Shows the axial and transverse polarization tensors fit to test-stand data over 20 and 37 mm projectiles. Top right panel: Sounding at anomaly maximum for 37 mm projectiles. Bottom left panel: Sounding at anomaly maximum for 20 mm projectiles; Bottom right panel: Sounding at anomaly maximum for 50 caliber projectiles. 79

Figure 44. Comparison of the relative decay rates obtained from the EM-63 (top-row) and the EM-61 (bottom row) on the 20 mm Range-Fan. The relative decay rates are calculated as the ratio of the principal polarization at time-channels N and 1. (a) EM-63 equivalent to EM-61 with $N=12$ (1.1 ms) over $N=1$ (180 μ s); (b) EM-63 ratio ratio at $N=19$ (5.6 ms) over $N=1$; and (c) EM-61 with $N=4$ and a ratio of polarizations at 1.2 ms and 216 μ s. There is clearly better separation between 37 mm and the other projectiles for the EM-63 at later times. 80

Figure 45. Receiver Operating Characteristic for EM-61 data at the 20 mm Range-Fan (Grids 19-14 and 21-14). The digging order is determined entirely by the estimated size of the object, which is obtained by the sum of the polarizations at time-channel 1. 81

Figure 46. (a) Fitted EM-61 dipole depth versus actual depth for different objects at the Rocket Range. (b) Error in depth estimation versus the estimated object size. Calculations were made assuming the EM-61 coils were 25 cm above the ground-surface. 83

Figure 47. EM-61 ROC curves for amplitude and size-based methods applied to all eight grids at the Rocket Range. The red circles are for the items with $L_1(t_1) = 10.25$ and 3 (holes 132 and 347

respectively) which corresponds to the geometric mean and smallest polarizations recovered for a 37 mm at the 20 mm Range-Fan. 83

Figure 48. Comparison of instantaneous polarization parameters recovered from EM-61 data at all eight rocket range grids. (a) A plot of a *shape feature*, $L_2(t_1)/L_1(t_1)$ against a size feature, $L_1(t_1)$. (b) Plot of a *time-decay feature* $L_1(t_4)/L_1(t_1)$ against the same size feature, $L_1(t_1)$. The MK-23 bombs tend to be large and have decay parameter between 0.15 and 0.2. See text for a discussion of the two MK-23 features joined by the dashed red line. 84

Figure 49. Original inversion fit to the MK-23 item identified in Figure 15 and discussed in the text. Plan view of the data at time-channel-3 and a comparison of the modeled and actual sounding at the anomaly maximum. The model returns a low-estimate of the response at time-channel-1 which causes an overestimate of the ratio $L_1(t_4)/L_1(t_1)$ 84

Figure 50. (a) Fitted EM-63-dipole depth versus actual depth for different objects at the Rocket Range. (b) Error in depth estimation versus the estimated object size. Calculations were made assuming the EM-63 coils were 20 cm above the ground-surface. 86

Figure 51. Comparison of polarization parameters recovered from EM-63 data on the Rocket Range. (a) Plot of a *shape-based* feature, k_2/k_1 , against size of the principal polarization (k_1); and (b) Plot of an *asymmetry* feature, $(k_3-k_2)/k_2$, against size of the principal polarization (k_1). For two-dipole models this asymmetry parameter is zero. 86

Figure 52. Comparison of the relative decay rates obtained from the EM-63 at the Rocket Range. The relative decay rates are calculated as the ratio of the principal polarization at time-channels N and 1. (a) Is the EM-61 equivalent ratio at N=12 (1.1 ms) over N=1 (180 μ s); (b) is the ratio at N=15 (2.17 ms) over N=1; and (c) is the ratio at N=19 (5.6 ms) over N=1. The red-dashed line joins the original and recalculated model for anomaly 12 in Grid I-13. 87

Figure 53. Polarization curves obtained by refitting all anomalies at the Rocket Range. (a) Raw principal polarizations for each object type; (b) Normalized principal polarizations for each object type, $L(n)/L(1)$; and (c) Normalized transverse polarizations for MK-23 items. 89

Figure 54. Feature vectors obtained by refitting the EM-63 data at the Rocket Range. (a) Principal polarization, $L_1(t_{12})/L_1(t_1)$ versus $L_1(t_1)$; (b) Transverse polarization, $L_2(t_{12})/L_2(t_1)$ versus $L_2(t_1)$; (c) Principal polarization γ_I versus β_I 90

Figure 55. ROC curve for re-inverted EM-63 data on Phase II and III grids (I-13, J-12, K-15, L-13, L-14 and L-15) at the Rocket Range. The SVM classifier was obtained by training with feature-vectors $L_1(t_1)$, β_I and γ_I on the Phase I Grid, I12 and J-13. ROCs based on amplitude and size, $L_1(t_1)$, are also shown. 91

Figure 56. Error simulation results. Geonics EM-63 first time channel data from a vertical 81 mm target is simulated. The sensor is positioned at $(x,y,z) = (0,0,0.3)$ m, and the 81 mm mortar is located at $(0,0,-0.3)$ m. When there are no errors in sensor positioning and orientation, the sensor would record a value of 529.86 mV. In (a) – (c), we consider contributions of sensor height, location and orientation errors to the data spread. The asymmetry of (b) and (c) is due to is due to the geometry of the problem. The maximum value measured by horizontal loop over a vertical target is directly over the target. In these simulations, any small changes in the orientation or position will produce a decrease in the signal strength. As a result, the most likely value for the data is the true value, and all other data will be lower in magnitude. This produces a one-sided

distribution; (d) shows the result when including all errors. Directly over a target, the error in height contributes most to the variation in the signal value..... 93

Figure 57. Error simulation results. Geonics EM-63 first time channel data from a vertical 81 mm target is simulated. The sensor is positioned at $(x,y,z) = (0.5,0.5,0.3)$ m, and the 81 mm mortar is located at $(0,0,-0.3)$ m. When there are no errors in sensor positioning and orientation, the sensor would measure 30.22 mV. In (a) – (c), we consider contributions of sensor height, location and orientation errors to the data spread. (d) shows the result when including all errors. There errors are nearly Normal..... 94

Figure 58. Percent error over a vertical 81 mm mortar. Sensor height and location uncertainties are Gaussian with $\sigma = 2$ cm. Sensor orientation has Gaussian errors of 2 degrees. The percent error due to sensor positioning errors varies from approximately 5 to 18 mV..... 95

Figure 59. Percent error over a horizontal 81 mm mortar. The same sensor positioning uncertainties of Figure 58 are used. The percent error due to sensor positioning errors varies from approximately 10 to 25 mV. 95

Figure 60. Percent error over an 81 mm mortar at a 45 degree dip. The same sensor positioning uncertainties of Figure 58 are used. The percent error due to sensor positioning errors varies from approximately 5 to 18 mV. 96

ACRONYMS

20mmRF	20 mm Range-Fan
3-D	Three Dimensional
AGL	Above Ground Level
APG	Aberdeen Proving Ground
BOR	Body of Revolution
cm	centimeters
COTS	Commercial Off-the-Shelf
DAS	Data Acquisition System
DoD	Department of Defense
EM	Electromagnetic
EMI	Electromagnetic Induction
EOD	Explosive Ordnance Disposal
ERDC	Engineer Research & Development Center
ESTCP	Environmental Security Technology Certification Program
FAR	False Alarm Rate
FLBGR	Former Lowry Bombing and Gunnery Range
HE	High Explosive
Hz	Hertz
IMU	Inertial Measurement Unit
km	kilometer
lb	pound
MEC	Munitions and Explosives of Concern
m	meter
mm	millimeters
MM	Military Munitions
ms	Millisecond
μs	Microsecond
MTADS	Multi-Towed Array Detection System
mV	Millivolt
OE	Ordnance and Explosives
Pd	Probability of Detection
PDisc	Probability of Discrimination
Pfa	Probability of False Alarm
PNN	Probabilistic Neural Network
P-O	Pasion-Oldenburg
RF	Range-Fan
QA/QC	Quality Assurance/Quality Control
RMS	Root-Mean-Square
ROC	Receiver Operating Characteristic
RR	Rocket Range
RTS	Robotic Total Station
SCAR	Sub-Calibar Aerial Rocket
SKY	Sky Research, Inc.

SNR	Signal-to-noise Ratio
SOP	Standard Operating Procedure
SVM	Support Vector Machine
TEM	Time-domain Electromagnetic
UBC-GIF	University of British Columbia – Geophysical Inversion Facility
USACE	United States Army Corps of Engineers
UXO	Unexploded Ordnance
YPG	Yuma Proving Ground

ACKNOWLEDGEMENTS

Sky Research, Inc. (SKY) performed this work for the Environmental Security Technology Certification Program (ESTCP) under Project MM-200504.

Key SKY personnel were:

Dr. Stephen Billings – Principal Investigator

Leonard Pasion – Data analyst

Jon Jacobson – Data collection

Stacey Kingsbury – Program Manager

Key UBC-GIF personnel were:

Dr. Douglas Oldenburg – UBC-Project oversight

Laurens Beran – Data analyst

Dr. Lin-ping Song – Data processing

David Sinex – Data processing

Dr. Nicolas Lhomme – Data processing

SKY would like to acknowledge the cooperation of the USACE-Omaha district and the USACE-ERDC and in particular, the assistance of Jerry Hodgson (Omaha) and Dr John Cullinane (ERDC).

The work presented herein was accomplished between September 2005 and March 2007.

1 INTRODUCTION

1.1 Background

This report presents the results of a 2006 demonstration of advanced discrimination techniques at the Former Lowry Bombing and Gunnery Range (FLBGR) in Colorado. The demonstration was conducted as part of ESTCP MM-0504 project “Practical Discrimination Strategies for Application to Live Sites.” The objective of the MM-0504 project is to address the need to reduce false-alarm rates at unexploded ordnance (UXO) sites by improving access to advanced modeling and discrimination capabilities. Recent research has focused on the discrimination problem whereby features from physics-based model-fits to anomalies are used to determine the likelihood that the buried item is an UXO. Statistical and rule-based classification techniques for UXO discrimination, when calibrated with good training data, have been shown to be effective at numerous test sites. However, guidelines and standard operating procedures for their application to live sites have yet to be established.

The software and protocols required to apply advanced discrimination techniques to live sites are being developed under this project. Implementation of the techniques within user-friendly software and clear guidance on effective application of the techniques will make advanced discrimination accessible to the non-research community. Furthermore, the enhanced software will provide a reliable framework for effective discrimination at live sites for regulatory and stakeholder acceptance.

1.2 Objectives of the Demonstration

The first demonstration of the methodology defined in this research project was conducted at FLBGR. The focus of the FLBGR demonstration was to verify the underlying methodology for the single and cooperative inversion processes that are used to extract physics-based parameters from magnetic and electromagnetic (EM) induction (EMI) anomalies, as well as the statistical classification algorithms used to make discrimination decisions from those parameters. Thus the demonstration primarily addressed the need to identify the sensor type and survey mode most suited for discrimination under the particular geological setting and military history of the site.

The specific objectives of the demonstration were to validate single and cooperative inversion approaches to UXO discrimination as a function of the following variables:

- Commercial off-the-shelf (COTS) sensors
 - Single sensor, single data-type for inversion
 - Geonics EM-61 as an industry standard COTS time-domain EM (TEM) sensor which provides four time-gates at each sounding
 - Geonics EM-63 as a higher quality COTS TEM sensor that provides twenty-six time-gates spread over a larger range than the EM-61
 - Dual sensor, dual data-type for cooperative inversion

- Geonics EM-61 and magnetometer
 - Geonics EM-63 and magnetometer
- Type of munitions and clutter present
 - Eight grids in the Rocket Range (RR) at FLBGR contain a range of air-delivered munitions from 20 millimeter (mm) projectiles to large bombs;
 - Two grids in the 20 mm Range-Fan (20mmRF) are primarily comprised of 50 caliber bullets and 20 mm and 37 mm projectiles. The ability to distinguish 50 caliber bullets and 20 mm projectiles (considered non-UXO) from 37 mm caliber projectiles (considered UXO) would be a significant advance at the site.
- Target density
 - Four grids in the RR and one in the 20mmRF grid have high target density (> 150 targets per acre);
 - Four grids in the RR and one in the 20mmRF have medium target density (50 to 150 targets per acre).
- Geological conditions
 - Two high density grids in the RR have soils that cause a measurable response in the EM-61 data.

Each of the sensor systems used in the demonstration (EM-61, EM-63 and total-field magnetometer) were positioned by a Leica Robotic Total Station (RTS) TPS 1206, with sensor orientation provided by a Crossbow AHRS400 Inertial Measurement Unit (IMU).

The selection of grids for this demonstration was made by reference to previously collected EM-61 towed array data. Magnetic data and Geonics EM-63 data were collected over the 10 selected grids comprising a total area of around 3.7 hectares (9.2 acres). Within the eight RR grids there were almost 1,200 anomalies selected by reference to the towed-array data, while in the 20mmRF there were 407 anomalies.

As discussed in more detail in Appendix A, there were some problems with the positioning accuracy of the magnetometer data. This meant it was not possible to use the estimated dipole locations from the magnetic data as constraints for the EM inversions. Thus, for this demonstration we were unable to test the cooperative inversion approach.

To evaluate and refine the results of the modeling and discrimination approach, the validation data were withheld from the data analysts and only released after they had submitted interpretations. The verification process was conducted by providing a ranked diglist for two RR and one 20mmRF grids (Phase 1), and then evaluating the diglist for accuracy by comparison with the ground-truth information. That information was then used to modify the statistical classification methodology for two additional RR grids and the remaining grid in the 20mmRF (Phase 2). After release of Phase 2 ground-truth, a final (Phase 3) analysis and interpretation was made for the remaining four grids in the RR. Given the analyst's submissions of anomalies to excavate, the following performance metrics were compared to those for 100% excavation:

- (i) the total number of holes that need to be excavated;
- (ii) the number of false-positives (items incorrectly classified as UXO); and

- (iii) the number of false-negatives (items incorrectly classified as non-UXO).

The analyst's recommended diglist comprised a fixed point in a Receiver Operating Characteristic (ROC) curve. The ranked diglist for all detected items was used to generate ROC curves to capture more detailed performance information on the methodology. The influence of signal-to-noise ratio (SNR) on inverted model parameters was also quantified. This analysis was conducted when there were multiple instances of a given ordnance type, or where previously collected high quality test stand data over an item (e.g., such as the emplaced 37 mm projectiles) were available.

Table 1 lists the specific performance objectives of this demonstration. Specific performance objectives and preliminary results of the demonstration are described in Section 4.3 and criteria and assessment of how well the objectives were attained are described in Section 5.

Table 1. Performance metrics for the demonstration.

Type of Performance Objective	Primary Performance Criteria	Expected Performance (Metric)
Qualitative	Terrain/Vegetation Restrictions	Operator acceptance for use at the site
	Ease of use (Hardware)	Operator and site geophysicist acceptance
Quantitative	Probability of Detection (Pd) of EM-63 sensor	\geq Pd for EM-61 towed array
	Probability of Discrimination (Pdisc) with a 50% reduction in false-alarms	≥ 0.9
	False-alarm rate with PDisc = 1	$> 25\%$ reduction in false-alarms
	Location Accuracy of interpreted items	$< 0.2\text{m}$
	Survey Rate for magnetometer system	1 hectare per day
	Survey Rate for EM-63 system	1/3 hectare per day
	Percent Site Coverage	$> 95\%$
	Processing Time (initial processing)	< 1 day per tile (1 acre)
	Processing Time (interpretation)	< 5 minutes operator time per anomaly
	Accuracy of inversion parameters	Within class variance of cooperative inversion $<$ single inversion

Ongoing geophysical surveys being conducted on behalf of the US Army Corps of Engineers (USACE)-Omaha and a USACE- Engineer Research & Development Center (ERDC) sponsored Congressional Interest project that is evaluating the use of cued-interrogation for discrimination were concurrently conducted with this demonstration. The concurrent surveys allowed cost-sharing between these projects. For example, mobilization and logistical coordination costs were shared with the USACE-ERDC project for data collection, as the EM-63 sensor mobilized for this demonstration also was used for data collection for the ERDC cued-interrogation study. In addition, the costs of collection and processing of the EM-61 towed array data were already covered by the other projects.

The demonstration is being conducted with the support and assistance of the USACE-Omaha district as they recognize the benefits of demonstrating, and ultimately transitioning, advanced technologies for use at the FLBGR site. USACE-Omaha provided UXO escorts for the geophysical survey crews as required. They also covered the costs associated with excavation and removal of all anomalies identified in the survey. This project only covered half of the costs of one UXO technician to collect the ground truth information as the UXO team excavated items (our USACE-ERDC project covered the other half).

1.3 Regulatory Drivers

The Defense Science Board Task Force on UXO noted in its 2003 report that 75% of the total cost of a current clearance is spent on digging scrap. A reduction in the number of scrap items dug per munitions and explosives of concern (MEC) item from 100 to 10 could reduce total clearance costs by as much as two-thirds. Thus, discrimination efforts focus on technologies that can reliably differentiate MEC from items that can be safely left undisturbed.

Discrimination only becomes a realistic option when the cost of identifying items that may be left in the ground is less than the cost of digging them. Because discrimination requires detection as a precursor step, the investment in additional data collection and analysis must result in enough fewer items dug to pay back the investment. Even with perfect detection performance and high SNR values, successfully sorting the detections into MEC and non-hazardous items is a difficult problem but, because of its potential payoff, one that is the focus of significant current research. This demonstration represents an effort to transition a promising discrimination technology into widespread use at MEC contaminated sites across the country.

1.4 Stakeholder/End-User Issues

Applying discrimination techniques results in leaving certain detected items in the ground, thus there is an inherent risk that detected UXO will not be recovered. Therefore, deciding on a discrimination *threshold* involves reaching a compromise between the levels of risk minimization versus the level of cost minimization. This risk-cost tradeoff is a question to be

resolved by stakeholders and the regulatory community. However, they need a framework for estimating the level of risk at a certain discrimination threshold.

This project addresses the critical need for information that can be made available to regulators and to facilitate regulator acceptance for the use of discrimination on military sites. The end results of this project will be a framework for understanding and evaluating confidence levels and risk estimates. In addition, the software components of this demonstration will be used to provide effective quality assurance/quality control (QA/QC) methods that can be used for review and defense of discrimination decisions. The process for facilitating this QA/QC component is a process pathway and reporting tools that would allow a regulator to follow the link from raw data to discrimination decision and ensure that decision was substantiated by the data.

2 TECHNOLOGY DESCRIPTION

2.1 Technology Development and Application

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

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

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

- 1) Creation of a map of the geophysical sensor data: This includes all actions required to form an estimate of the geophysical quantity in question (magnetic field in nanoTesla [nT], amplitude of EMI response at a given time-channel, etc.) at each of the visited locations. The estimated quantity is dependent on the following:
 - a. Hardware, including the sensor type, deployment platform, position and orientation system and the data acquisition system used to record and time-stamp the different sensors;
 - b. Survey parameters such as line spacing, sampling rate, calibration procedures etc.;
 - c. Data processing such as merging of position/orientation information with sensor data, noise and background filtering applied;
 - d. The background environment including geology, vegetation, topography, cultural features, etc.; and
 - e. Depth and distribution of ordnance and clutter.

- 2) Anomaly selection and feature extraction: This includes the detection of anomalous regions and the subsequent extraction of a dipole (magnetics) or polarization tensor (TEM) model for each anomaly. Where magnetic and EMI data have both been collected, the magnetic data can be used as constraints for the EMI model via a cooperative inversion process.
- 3) Classification of anomalies: The final objective of the demonstration is the production of a dig sheet with a ranked list of anomalies. This will be achieved via statistical classification which will require training data to determine the attributes of the UXO and non-UXO classes.

The focus of this demonstration is on the validation of the methodologies for 2) and 3) above that have been developed in UXOLab jointly by Sky Research and the University of British Columbia-Geophysical Inversion Facility (UBC-GIF). The success of the discrimination process will be critically dependent on the attributes of the data used for the feature extraction and subsequent classification (*vis-a-vis*, everything pertaining to the first element described above), in particular, the SNR, location accuracy, sampling density and information content of the data (the more time channels or vector components, the more information that will be available to constrain the fits). Thus, while our intent is to test the algorithms developed in UXOLab, this test cannot be conducted in isolation of the attributes of the geophysical sensor data. Recently, several authors (e.g., Foley *et al.*, 2004; Smith *et al.*, 2004a, b; and Bell, 2005) have shown that very high SNR ratios, centimeter level positional and one degree level orientation accuracy may be required to achieve reliable discrimination performance with current EMI instrumentation.

With the above data quality considerations in mind, during 2005 and early 2006 we made significant advances in our data collection systems, including incorporation of an IMU in our EM-61 towed array, EM-63 cart and magnetometer array; development of an EM-63 cart with suspension; and upgrade to the latest Leica Robotic Total Station (the TPS 1206).

Each of the three key elements of the discrimination process identified above are described in detail in the Demonstration Plan for this project, and are summarized in the following subsections. The sensors and platforms used to collect the data for the demonstration are summarized in Table 2, and the systems are shown in Figures 1 through 4.

Table 2. Sensor and Equipment Summary

Technology/Equipment	Description	Features
EM-61-MK2 Towed Array	5 EM-61 coils Crossbow AHRS400 IMU Leica RTS Mounted on a low-to-the-ground, 8-wheel platform towed by a 6X6 Gator	Platform constructed of composite materials Conforms to different terrain conditions EM coils mounted 14 centimeters (cm) above ground level (AGL)
Quad-Sensor Magnetometer Array	4 Geometrics G-823 magnetic sensors Leica 360 ° prism Man-portable w/ tethered data acquisition system and RTS console	Two operators Configured with sensors 30 cm AGL Has been deployed in terrain with slopes up to 40 degrees
Geonics EM-63 Cart	Multi-channel time-domain EM induction instrument 1 x 1 meter (m) transmitter coil and 3 - 0.5 m ² coaxial receiver loops mounted on a 2 wheel trailer Leica RTS Crossbow AHRS 400 IMU	26 geometrically spaced time gates EM-63 coil mounted 25 cm AGL Air suspension system to moderate rapid terrain changes
Positioning System	Leica RTS TPS1206 High-precision total station system, tracks prism 360 degrees up to 1000 m	3-D position solutions up to 8 Hz Sub-centimeter accuracy Robotic capability tracks system in motion
Orientation Sensors	Crossbow AHRS-400 IMU Crossbow Data Management Units	Measures pitch, roll, and yaw Nine-axis measurement system for complete measurement of system dynamics Successive RTS measurements needed to estimate azimuth for EM systems Onboard digital processing to compensate for deterministic error sources Compute attitude information Zero-velocity update every 10 minutes to recalibrate pitch and roll outputs



Figure 1. Sky Research's EM 61-MK2 towed array is constructed of composite materials and houses EM sensors, RTS laser positioning (or GPS) sensors, and the Crossbow IMU.



Figure 2. SKY man-portable MAG array integrates 4 Geometrics G-823 with the Leica 360° prism at FLGBR. This lightweight and durable system was deployed with a data tether.



Figure 3. Modified EM-63 cart collecting discrimination mode data at the Ashland test-site.



Figure 4. Sky Research utilizes the Leica RTS TPS1206 laser positioning system. This device is setup in over a known point and tracks a prism attached to the geophysical survey equipment. RTS technology out-performs GPS in terms of accuracy, sampling rate and operational ease-of-use.

2.1.1 Creation of a Map of Geophysical Sensor Data

Data collected using each of the sensor systems are mapped to provide estimates of the geophysical quantities measured. The initial steps taken in the data processing flow include:

- *Initial review of collected data:* Validate that data fall within prescribed recording ranges, establish number of points collected, data density, and time-on / time-off.
- *Statistical analysis:* Review XYZ statistics describing survey coordinates and sensor values, etc.
- *Data merging:* The geophysical sensor, positioning and orientation data are merged together using the common time-base established with the SKY-Data Acquisition System (DAS). The procedure involves interpolation or decimation of each sensor stream to that of the geophysical sensor (e.g., 10 hertz [Hz] with the magnetometer system).
- *Sensor positioning and orientation:* The 3 dimensional (3-D) position of the RTS, along with the orientation information from the IMU, are used to accurately calculate the 3-D position and orientation of each geophysical sensor measurement.
- *Base-station correction (magnetics):* The time-varying magnetic field recorded by the base-station magnetometer will be removed from the data.
- *Heading correction (magnetics):* Through the heading correction test conducted each day, heading dependent corrections for each magnetometer will be calculated and then applied to the magnetometer data.
- *Background removal (magnetics):* Initially a constant background will be removed from each survey event (contiguous period of data collection). As much as possible, we will avoid filtering the magnetic data, as any filter will distort the true anomaly shape to some extent. However, if geological trends are present, or if adjacent lines have small offsets (due to limitations in the heading correction), the magnetometer data will be detrended using a recursive median filter with a fixed length of 10 meters. At each iteration, the top and bottom user-specified percentage of background corrected data are excluded when making a new estimate of the background. At the first pass, the background is estimated using all points. Then the background is re-estimated by excluding points that fall outside of the percentage cutoffs in the background corrected data. This new background estimate is then used to determine which points to exclude for the background estimation at the next iteration and so on.
- *Background removal (EM-63):* At the start and end of each survey event, background data are recorded with the EM-63 at least 1.5 meters off the ground. For each recorded time-gate, a background correction will be applied that is a linear interpolation (as a function of time) between the before and after background measurements. If background variations are still present in the data, a recursive median filter with a fixed temporal length (several tens of seconds) will be used to detrend the data.
- *Data gridding:* Filtered data are interpolated using UXOLab onto a 0.125 m grid and reviewed by a geophysicist.
- *Target selection (EM-63; EM-61):* The EM data were automatically picked with a peak detection routine available within UXOLab. Manual picks will be added as required. Time-channel 3 was used for the picking of EM-61 anomalies, and time-channel 1 for the

EM-63. The picking threshold was chosen by reference to the data collected over the FLBGR test-plot.

- *Target selection (magnetics)*: No target picking was conducted using the magnetics, with magnetometer targets inherited from the analysis of the EM data;

2.1.2 Anomaly Selection and Feature Extraction

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

Feature Extraction: Time-domain Sensor

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

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

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

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

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

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

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

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

where we use the convention that $L_1(t_1) \geq L_2(t_1) \geq L_3(t_1)$, so that polarization tensor parameters are organized from largest to smallest. The polarization tensor components are parameterized

such that the target response can be written as a function of a model vector containing components that are a function of target characteristics. Particular parameterizations differ depending on the instrument (number of time channels, time range measured etc) and the group implementing the work. Bell *et al.* (2001) solves for the components of the polarization tensor at each time channel, and this is the procedure we will use for the four-channel Geonics EM-61 MKII. For the EM-63 we will use the *kbg* version of the Pasion-Oldenburg (P-O) formulation (Pasion and Oldenburg, 2001), which does not involve an early time alpha parameter:

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

for $i = \{1, 2, 3\}$ and with the convention that $k_1 \geq k_2 \geq k_3$. For a rod-like body-of-revolution (BOR), $L_2 = L_3$ (Pasion and Oldenburg, 2001) while for a plate-like BOR $L_1 = L_2$.

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

$$p(\mathbf{m} | \mathbf{d}^{\text{obs}}) = \frac{p(\mathbf{m})p(\mathbf{d}^{\text{obs}} | \mathbf{m})}{p(\mathbf{d}^{\text{obs}})} \quad (5)$$

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

$$\mathbf{m}^* = \max_{\mathbf{m}} \{ \log(p(\mathbf{m} | \mathbf{d}^{\text{obs}})) \} \quad (6)$$

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

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

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

During the first phase of this demonstration we made a number of adjustments to the feature extraction routines for the EM sensors, and we describe these in more detail in the results section.

Feature Extraction: Magnetism

For magnetism, the physics-based model most commonly used is a dipole. This comprises a location (horizontal position and depth) and the magnitude and orientation of the dipole moment. More complicated models comprising quadrupoles and octupoles have been developed (Billings *et al.*, 2002a, b). However, in most UXO detection scenarios the sensor is in the far-field of the source body and moments of higher order than the dipole are poorly resolved. The contribution of the dipole to the magnetic field decays as the 3rd power of distance from the object and dominates the far-field,

$$\mathbf{B}(\mathbf{r}) = \frac{\mu_0}{4\pi} \frac{\mathbf{m} \cdot (3\hat{\mathbf{r}}\hat{\mathbf{r}} - \mathbf{I})}{r^3} \quad (8)$$

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

Feature Extraction: Cooperative Inversion of TEM and Magnetic Data

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

$$\text{minimize } \phi(\mathbf{m}) = \sum_j \frac{1}{2\sigma_j^2} (m_j - \bar{m}_j)^2 + \frac{1}{2} \|V_d^{-1/2} (\mathbf{d}^{\text{obs}} - F(\mathbf{m}))\|^2, \text{ subject to } m_i^L \leq m_i \leq m_i^U \quad (9)$$

where j represents the index of parameters whose Gaussian pdfs are assumed to be known. The strategy we will use here for cooperatively inverting magnetism and electromagnetics data, is as follows:

- 1) Invert the magnetism data for the best fit dipole.
- 2) Use the estimated dipole location to define \bar{m}_j (for $j = 1, 2$ and 3 which corresponds to the Easting, Northing and depth of the dipole) and use the standard deviation of the parameter uncertainties to define σ_j . The estimated model parameter standard deviations are obtained from the Gauss-Newton approximation to the Hessian at the optimum model location (e.g., Billings *et al.*, 2002a).
- 3) Invert the EM data using the prior obtained from the magnetism data in Step 2.

At the end of the cooperative inversion process there will be feature vectors from the TEM data (polarization tensor parameters) as well as the magnetism (dipole moment). The critical question

we wish to resolve in this demonstration is whether the improved estimates of polarization tensor parameters returned by cooperative inversion, plus the extra information provided by the dipole moment, outweigh the extra costs of collection and interpretation of two datasets.

Inevitably, there will be anomalies in the TEM that do not have corresponding magnetic fits and vice versa. Where no constraints from magnetometer data were available, the TEM data were inverted using the same procedure as for single inversion. Any magnetic anomalies without corresponding TEM anomalies were placed on the diglist and excavated. Part of the evaluation of the results is to use the ground-truth data that verifies the identity of these magnetic items to attempt to ascertain why the anomaly was not on the EM detection list.

2.1.3 Classification of Anomalies

At this stage in the process, we have feature vectors for each anomaly and now need to decide which items should be excavated as potential UXO. Rule-based classifiers use relationships derived from the underlying physics to partition the feature space. Examples include the ratio of TEM decay parameters (Pasion and Oldenburg, 2001) and magnetic remanence (Billings, 2004). For this demonstration, we will focus on statistical classification techniques which have proven to be very effective at discrimination at various test-sites (e.g., Zhang *et al.*, 2002; Beran *et al.*, 2004).

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

Figure 5 summarizes the supervised classification process within the statistical framework. Given test and training data sets, we extract features from the data, select a relevant subset of these features and optimize the classifier using the available training data. Because the predicted performance of the classifier is dependent upon the feature space, the learning stage can involve further experimentation with feature extraction and selection before adequate performance is achieved.

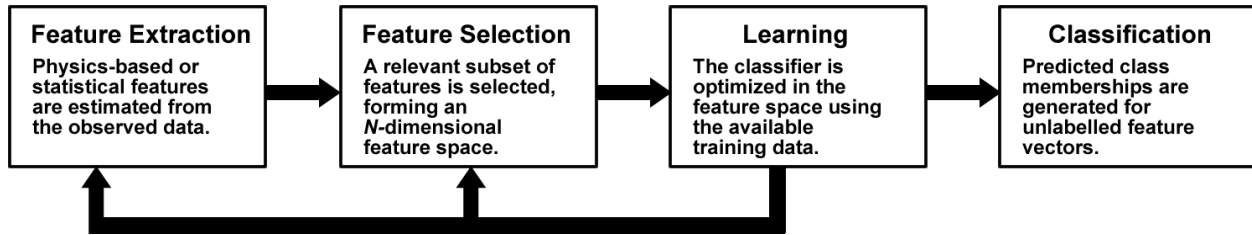


Figure 5. A framework for statistical pattern recognition.

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

$$P(\varpi_i | \mathbf{x}) \propto P(\mathbf{x} | \varpi_i) P(\varpi_i) \quad (10)$$

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

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

Other generative classifiers assume a nonparametric form for the likelihood function. For example, the probabilistic neural network (PNN) models the likelihood for each class as a superposition of kernel functions. The kernels are centered at the training data for each class. In this case the complexity of the likelihood function (and hence the decision boundary) is governed by the width of the kernels (Figure 6).

The *discriminative* approach is not concerned with underlying distributions but rather seeks to identify decision boundaries which provide an optimal separation of classes. For example, a support vector machine (SVM) constructs a decision boundary by maximizing the *margin* between classes. The margin is defined as the perpendicular distance between *support planes* which bound the classes, as shown in Figure 7. The decision boundary then bisects the support planes. This formulation leads to a constrained optimization problem: maximize the margin between classes subject to the constraint that the training data are classified correctly. An advantage of the SVM method over other discriminative classifiers (e.g., neural networks) is that there is a unique solution to the optimization problem.

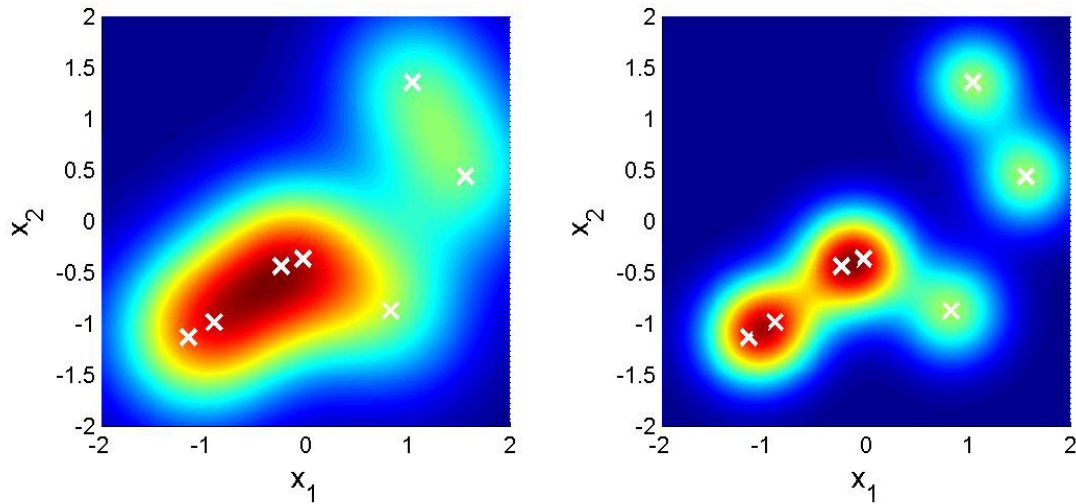


Figure 6. Nonparametric density estimate using Gaussian kernels. Kernel centers are shown as crosses. A large kernel width produces a smooth distribution (left) compared to a small kernel width (right).

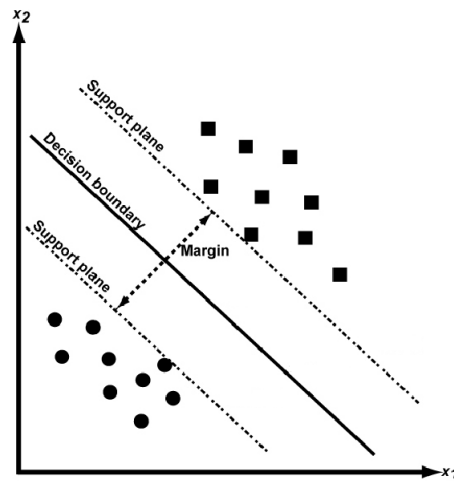


Figure 7. Support vector machine formulation for constructing a decision boundary. The decision boundary bisects support planes bounding the classes.

With all classification algorithms, a balance must be struck between obtaining good performance on the training data and generalizing to a test data set. An algorithm which classifies all training data correctly may produce an overly complex decision boundary which may not perform well on the test data. In the literature this is referred to as “bias-variance trade-off” and is addressed by constraining the complexity of the decision boundary (regularization). In cases such as linear discriminant analysis, the regularization is implicit in specification of the likelihood function.

Alternatively, the complexity of the fit can be explicitly governed by regularization parameters (e.g., the width of kernels in a PNN or Lagrange multipliers in a SVM). These parameters are typically estimated from the training data using *cross-validation*, which sets aside a portion of the training data to assess classifier performance for a given regularization.

2.1.4 UXOLab Software

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

2.2 Previous Testing of the Technology

Several previous tests of the single and cooperative inversion and/or statistical classification abilities of UXOLab algorithms have been conducted. A detailed discussion is provided in the demonstration plan for this project, and Table 3 provides a summary of the previous testing.

Table 3. Previous Inversion/Classification Testing

Inversion/Classification Test and Location	Description	Results
Proof-of-concept of cooperative inversion, Yuma Proving Ground (YPG)	Test of cooperative inversion on EM-63 and magnetometer data collected in 2003. TEM inversions used two decaying orthogonal dipoles, constrained using magnetics data. Three different classifiers (linear and quadratic discriminant analysis, and probabilistic neural network) were applied to the cooperative inversion results.	Classification of cooperatively inverted data is easier than inversion w/o magnetic constraints. Cleaner separation of classes is achieved for k parameters recovered from cooperative inversion; single and cooperative inversion results are similar for β parameters. This test demonstrated the UXOLab capability to perform both cooperative inversion and statistical classification.
Geocenters STOLS EM-61 and magnetometer data at Aberdeen Proving Ground (APG)/YPG	Discrimination ability of the system was marginal due to limitations in positional accuracy (5-10 cm) which is inadequate for advanced discrimination); lack of sensor orientation data; low SNR; no statistical classification algorithms were applied.	Results contributed to system enhancements to SKY sensor systems, including use of RTS for positioning and IMU for sensor orientation. Demonstrated the feasibility of cooperative inversion of large volumes of data with UXOLab.
Single inversion of towed array EM-61 data at FLBGR test grid	Single inversions of a number of items were conducted using a 2-beta polarization tensor parameterization for each time channel. There is considerable redundancy in the feature space, so canonical analysis was used to project the feature vectors into a new space ordered such that the first coordinate maximizes the class separations.	The results show that there is a reasonable separation between different ordnance types. In particular the 20 mm and 37 mm are quite widely separated, which shows that discrimination between the two ordnance types with this system may be feasible.

3 Site Characteristics and Demonstration Design

3.1 Performance Objectives

In terms of performance, there were two sensor configurations that tested under this project:

1. Five-element Geonics EM-61 towed array with single inversion for polarization tensor parameters;
2. Geonics EM-63 cart with single inversion for polarization tensor parameters;

Magnetometer data were collected to support a planned demonstration of the cooperative inversion process for both of these sensors. However, as described in Appendix A, Magnetic Position Problems, there were positional problems with the magnetometer data that precluded their use for the intended purpose.

The ultimate acceptance or rejection of a discrimination methodology at a site depends on the tolerance of risk of the stakeholders. If misclassifying a single detected item is unacceptable, then the performance criteria is Probability of Discrimination (PDisc) = 1. If there is some tolerance of risk, then some detected items may be left in the ground if the cost can be reduced considerably. Any performance criteria will by necessity be subjective. For this demonstration we will use a criterion of:

- $\text{PDisc} \geq 0.9$ with a 50% reduction in excavated false-alarms; and
- at $\text{PDisc} = 1$, there must be a 25% or greater reduction in false alarms.

At both the Rocket Range and the 20 mm Range-Fan, we assumed that the smallest item of concern is a 37 mm projectile. Any excavated item that is a MEC (e.g., fuzes, bursters, spotting charges etc) will be considered to be UXO for the purposes of scoring. This includes half-intact rounds or large pieces of shrapnel that contain sufficient explosives to be considered hazardous. Where multiple items are located in close proximity, the most hazardous item will be used for the assessment of performance (e.g., an UXO item and several pieces of shrapnel would be treated as a UXO).

Table 1 in Section 1.2 lists the performance objectives established for this demonstration.

3.2 Test Site History/Characteristics

FLBGR is located approximately 20 miles southeast of Denver, Colorado, in Arapahoe County. Although the area immediately west of the former bombing range is extensively developed, the site is still primarily grazing land. Evidence of Department of Defense (DoD) use of the bombing range remains at every known range. The gunnery ranges and small arms ranges still contain empty cartridges and projectiles.

FLBGR was originally part of Buckley Field, which consisted of the airfield and bombing and gunnery range and contained 65,547 acres. The status of the various portions of land that comprised Buckley Field changed several times since the land was acquired by the City of Denver beginning in 1937. The airfield and bombing range were used by the Army during World War II. After the war, the airfield became a Naval Air Station and the bombing range came under the custody of Lowry Air Force Base. The bombing range was renamed the Lowry Bombing and Gunnery Range. The bombing range was exscessed beginning in 1960.

In 2005, 45 acres on the Rocket Range, and 6 acres on the 20 mm Range-Fan were surveyed with the Sky EM-61 towed array (Figure 8). These areas were specifically identified by USACE-Omaha as priority areas that are currently being cleared (or will be cleared in the near future). The sites are also representative of the terrain, vegetation and munitions at the site.

The Rocket Range was used for bombing practice with sand-filled practice bombs and high explosive (HE) bombs, rocket practice and gunnery training. Expected MEC in this area include practice bomb debris, HE bomb fragments, 50 caliber ammunition and 20 mm projectiles and practice rockets.

The 20 mm Range-Fan was used for air-to-ground target practice for fixed wing-aircraft firing 50 caliber projectiles, and 20 and 37 mm projectiles.

Within the demonstration areas there is little variation in terrain and vegetation. At both sites the vegetation is a mixture of grasses and Yucca plants. These are dense, low-lying (< 1 meter) plants that caused some survey difficulties to the EM-63 cart in particular.

3.3 Pre-Demonstration Testing and Analysis

In September/October 2005, the FLBGR test grid (Figure 9), 45 acres on the Rocket Range (Figure 10), and 6 acres on the 20 mm Range-Fan (3 acres are shown in Figure 11) were surveyed with the EM-61 towed array with RTS and IMU. These towed array data were used for the EM-61 inversions conducted as part of this demonstration.

Target picking with a 10 millivolt (mV) threshold on time-channel 3 ($V(t_3) > 10 \text{ mV}$), was conducted on towed array data collected on the Rocket Range and the 20 mm Range-Fan. In December 2005, ground-truth data were collected over three of the grids at the Rocket Range (Figure 10) and one grid at the 20 mm Range-Fan (Figure 11). A total of 458 ground-truth items were recovered, with a summary provided in Table 4. The list contains forty MEC with caliber greater than 20 mm, seventy-seven 20 mm projectiles and twenty emplaced 37 mm rounds. These rounds were emplaced by Sky Research personnel to test the detection and discrimination characteristics of the towed array.

This validation exercise was useful from the perspective of this demonstration in several ways. First, it allowed us to test the feasibility of our ground-truth collection methodology. Second, it provided valuable information on the number and distribution of MEC within the survey area. The highest concentration of anomalies and of MEC occurred around Grid K-13.

Table 4. Summary of ground-truth items recovered during excavations at three Rocket Range and one 20 mm Range-Fan grid in 2005.

Anomaly	20-14	H-8	H-9	K-13	Total
Bomb M50 Incendiary (4 lb)			2		2
Bomb M-38 Practice (100 lb)			1		1
Bomb MK-34 Practice				1	1
Bomb MK-23 Practice (3 lb)				34	34
Projectile 57 mm HEAT			1	1	2
Projectile 37 mm (emplaced)	6		7	7	20
Projectile 20 mm	25	6	14	32	77
Shrapnel/fragmentation		7	16	83	106
Small arms	37	1			38
Non-Military Scrap/Junk	1	2	4	129	136
Nothing Found		17	6	18	41
Total	69	33	51	305	458

Counting the number of towed-array anomalies with $V(t_3) > 10$ mV in each grid of the Rocket Range (Figure 10 and Table 5) reveals that Grids L-13 to 15, K-14 to 15 and I-12 to 13 have relatively high anomaly densities. As these are close to Grid K-13 where many MEC were found, we decided to focus our demonstration efforts in that area, specifically we intended to collect data on the eight grids outlined in red in Figure 10 and itemized in Table 5. On arriving at the site in early June 2006, we discovered that Grid K-14 had already been cleared by the incumbent explosive ordnance disposal (EOD) contractor. We therefore substituted Grid L-13 for Grid K-14.

Target picking using $V(t_3) > 10$ mV in each grid of the 20 mm Range-Fan (Figure 11 and Table 5) was also conducted. This target picking reveals that Tile 21-14 has a considerably higher concentration of anomalies than 19-14. Both of these tiles were used for this demonstration.

Table 5. Number of anomalies selected with amplitudes above 10 mV in the 3rd-time channel of the Sky towed array EM-61 data.

Grid	Number Targets	Comments
I-12	93	Medium density
I-13	121	Medium density
J-12	131	Medium density
J-13	118	Medium density
K-14 (not used)	157	High density
K-15	184	High density and geology
L-13 (replaced K-14)	211	High density
L-14	195	High density
L-15	164	High density and geology
Total (RR)	1,217	Does not include K-14
19-14	148	Medium density
21-14	259	High density
Total (20mmRF)	407	
Total (All)	1,624	Does not include K-14

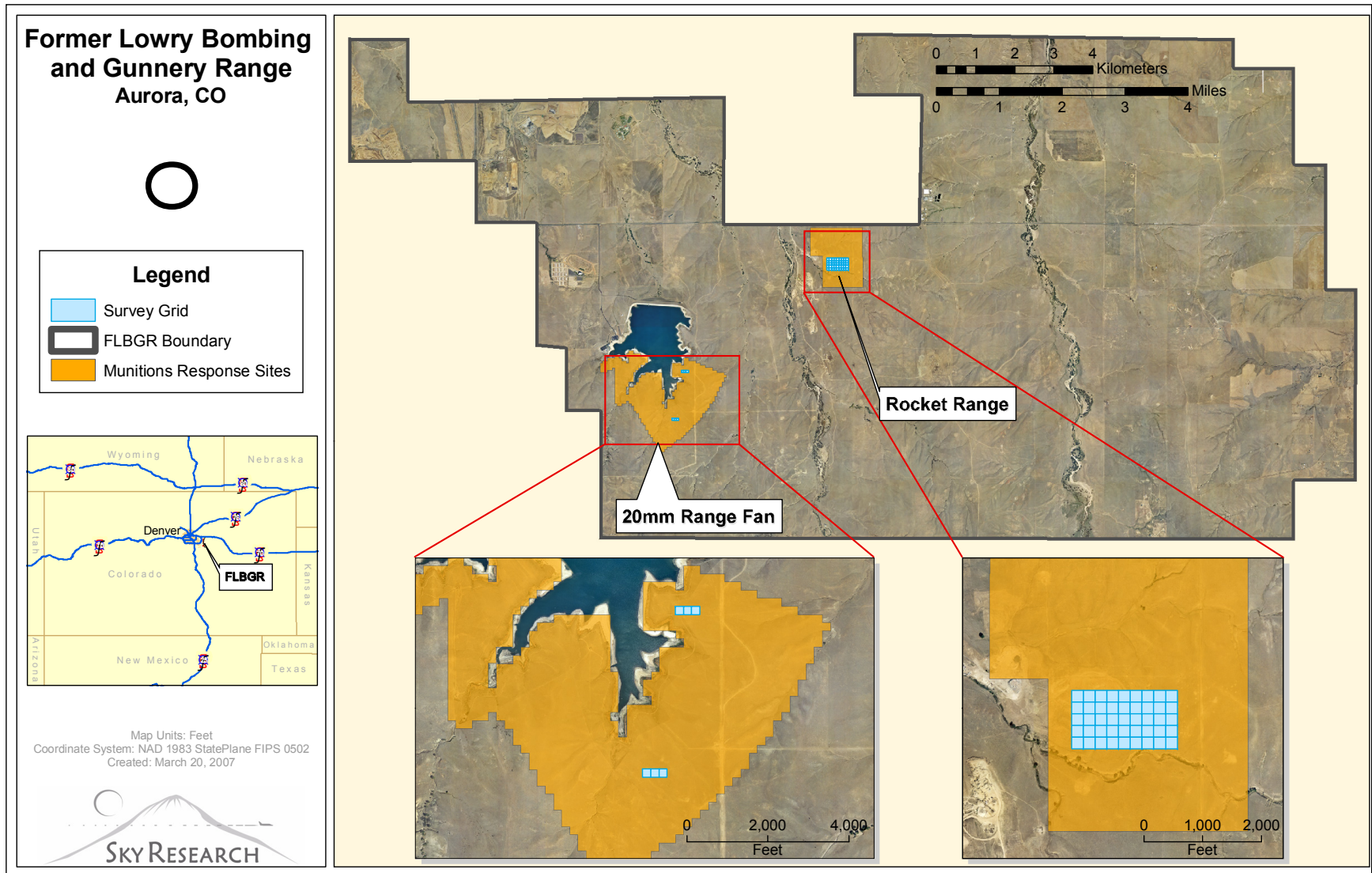


Figure 8. Locations of the Rocket Range and 20 mm Range-Fan sites at FLBGR

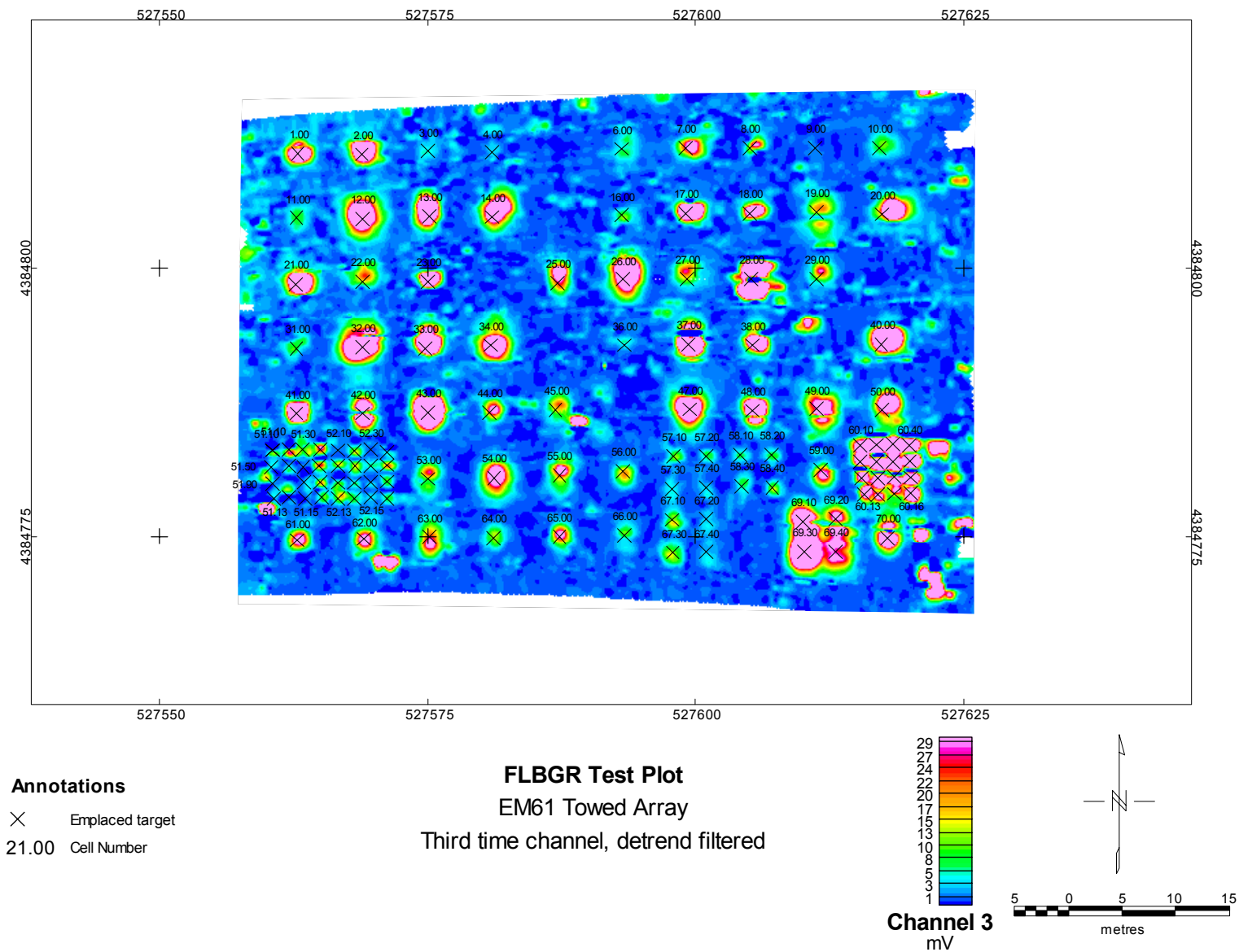


Figure 9. Map of time-channel 3 of the EM-61 survey of the FLBGR test-plot with emplaced items overlain.

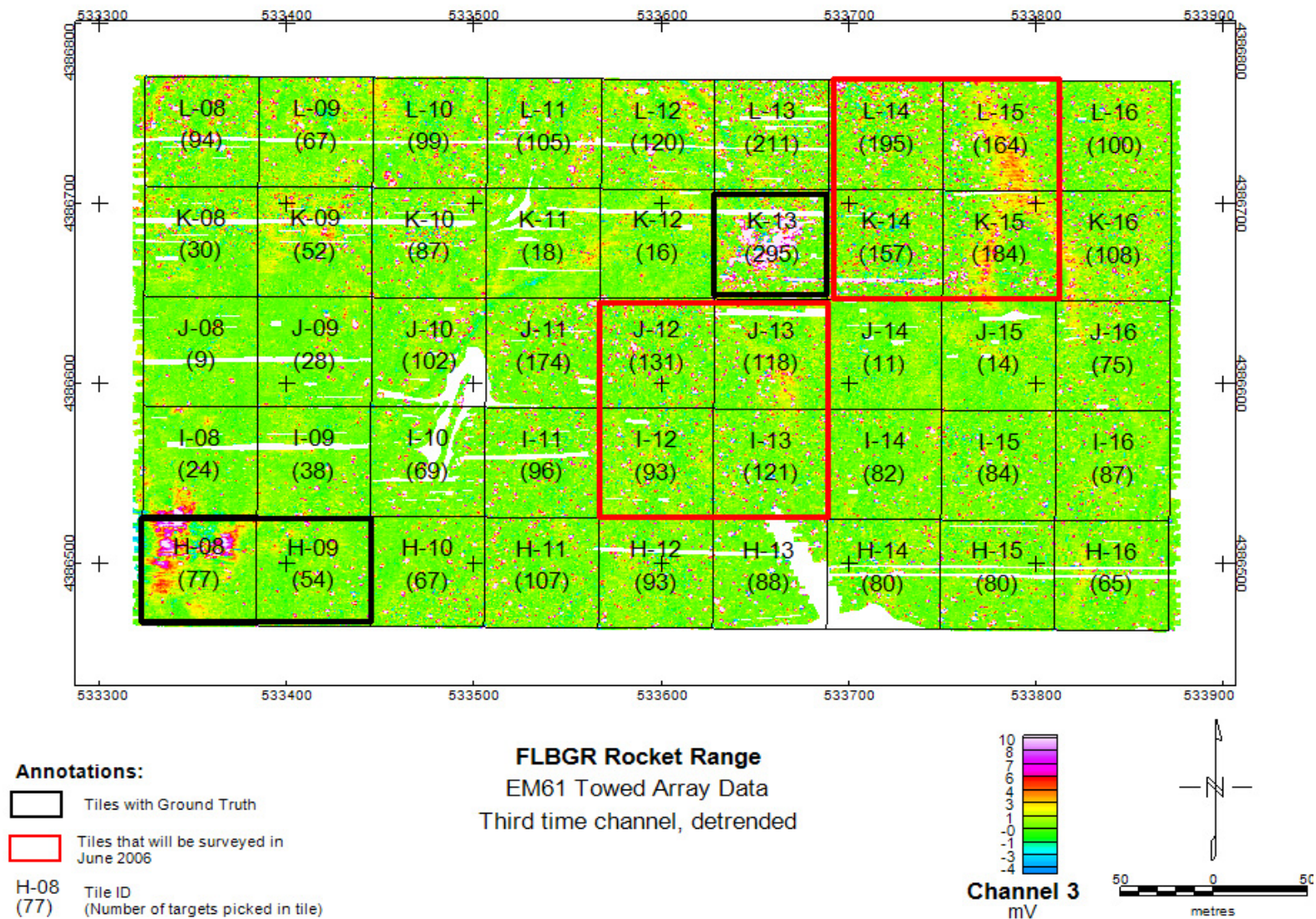


Figure 10. Map of Rocket Range with areas surveyed for this demonstration outlined in red.

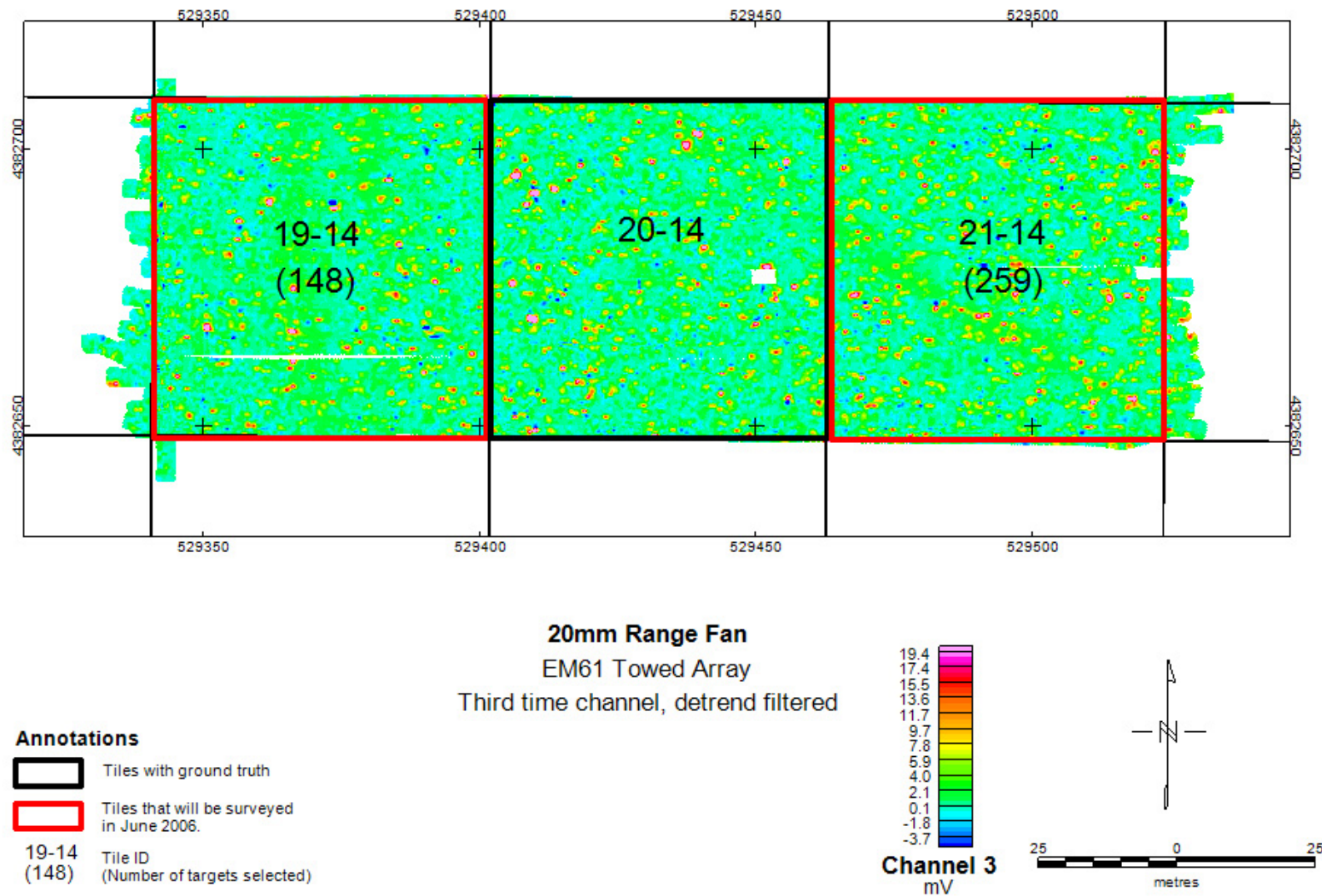


Figure 11. Map of the 20 mm Range-Fan, with two grids surveyed for this demonstration outlined in red.

4 DEMONSTRATION SURVEYS AND PROCESSING

4.1 Demonstration Set-Up and Start-Up

The majority of mobilization activities for this demonstration were completed as part of the concurrent geophysical surveys being conducted for ERDC and USACE-Omaha by Sky Research. Surface clearance was already conducted, and survey control established. Additionally, Sky Research has an on-site trailer and storage compound within a few kilometers of the Rocket Range site. As noted previously, the EM-61 data were collected during a previous ERDC sponsored mobilization in September/October 2005.

Project-specific mobilization comprised the following:

- 1) Mobilization of the EM and magnetometer field crew and associated equipment.
- 2) Mobilization of the Quality Assurance Officer to the site.
- 3) Emplacement of twenty 37 mm projectiles within at the 20 mm Range-Fan. This provided training data for the EM-63 and magnetometer sensors and also served as a test of the detection performance of the sensor systems (against the smallest MEC).
- 4) Standard pre-collection maintenance and calibration procedures were performed for the sensor systems. These included all of the calibrations listed in our Quality Assurance Project Plan in Appendix C of the demonstration plan.

4.2 Period of Operation

In the table below we provide a table of the projects key activities.

Table 6. Key Project Activities

Day	Activity
Pre-survey	
Sep-05	Commence EM-61 data collection over 45 acres on Rocket Range, and six acres on the 20 mm Range-Fan
Oct-05	Complete EM-61 data collection and initial processing
Jan-06	Validation of five-grids (4 on Rocket Range, 1 on 20mmRF)
May-06	Demonstration plan approved
Magnetometer	
5-Jun-06	Arrive at site, unpack equipment, safety brief etc
6-Jun-06	Survey I-12, I-13, J-12 and J-13
7-Jun-06	Survey K-14, K-15, L-14 and L-15 (K-14 was previously cleared)
8-Jun-06	Survey test-plot
12-Jun-06	Survey grid L-13 and demobilize from site

Day	Activity
EM-63	
12-Jun-06	Arrive on site
13-Jun-06	Commence EM-63 survey of J-13
14-Jun-06	Continue with J-13
15-Jun-06	Complete J-13, start J-12
19-Jun-06	Survey test plot with EM-63
20-Jun-06	Troubleshoot problems with CSM EM-63. Down-time while waiting for replacement instrument
23-Jun-06	Receive replacement EM-63 and survey test plot
26-Jun-06	Start surveying J-12 with new EM-63 (repeat areas down with the CSM system)
27-Jun-06	Finish J-12, start and finish I12
28-Jun-06	Recollects on I12, start I-13. Strong winds knock over the RTS base-station and damage it
29-Jun-06	Receive replacement RTS and continue with survey of I-13
30-Jun-06	Demob from site for 1 week (short week due to 4-th July celebration)
9-Jul-06	Redeploy to site
10-Jul-06	Commence EM-63 survey of 19-14
11-Jul-06	Complete survey of 19-14, commence 21-14
12-Jul-06	Complete 21-14, recollects on 19-14, move equipment back to RR
13-Jul-06	Start and complete K-15, commence L-15
17-Jul-06	Complete L-15
18-Jul-06	Start and complete L-14
19-Jul-06	Start and complete L-13 and conclude ESTCP surveys
Processing	
Jun-06	Initial processing magnetics and EM-63
Jul-06	Complete initial processing magnetics and EM-63
Aug-06	Feature extraction commences
Validation	
Aug-06	Delay excavations to wait for a possible MSEM survey
11-Sep-06	Submit Phase I interpretations (I12, J-12, 19-14)
11-Sep-06	Excavation commences of Phase I grids
13-Sep-06	Commence excavation of Phase II grids on 20mmRF (21-14)
14-Sep-06	Phase I ground-truth released
18-Sep-06	Phase II interpretation on 20mmRF (21-14) submitted
19-Sep-06	Phase II ground-truth released for 20mmRF (21-14); Phase II validation of RR commences (L-15)
19-Sep-06	Phase II validation of K-15 commences
21-Sep-06	Phase III validation commences (I-13, J-12)
22-Sep-06	Phase II interpretation on RR (K-15, L-15) submitted
25-Sep-06	Phase III validation continues (L-13, L-14)
5-Oct-06	Phase III interpretations submitted (I-13, J-12, L-13, L-14). Later that day, Phase III ground-truth released

4.2.1 Sky-Research Five-Element EM-61 Array

Sky Research's EM-61MK2 towed array (Figure 1) contains 5 coils, Crossbow AHRS400 IMU, and Leica RTS. Data at the Rocket Range and 20 mm Range-Fan were previously collected by this system. The EM-61 logged data at 10 Hz, the RTS at 4 Hz and the Crossbow IMU at 30 Hz. The data were merged and processed using Sky Research's standard operating procedures for the towed-array. Standard calibration tests were performed and records kept as per the Quality Assurance Project Plan in Appendix C of the demonstration plan.

4.2.2 Quad-sensor Magnetometer Survey

The Sky Research man-portable quad-sensor magnetometer array (Figure 2) with Leica RTS for position and Crossbow IMU for orientation was used for the magnetic survey. The Geometrics G-823 magnetometer sensor separation was 0.375 m, which means the system covered a swath 1.5 meters wide on each transect. This is closer than the usual 0.5 m sensor spacing and 2.0 m swath width used effectively in Montana for projectiles of 76 mm caliber or greater. The array was configured so that the sensors were nominally 30 cm above the ground. The magnetometer collected measurements ten-times a second, which translates to about 11 cm sample spacing at a nominal traverse rate of 4 kilometer (km)/hr. The RTS was operated at around 4 Hz and the Crossbow IMU at 30 Hz. In a nearby, quiet location away from any traffic, a Geometrics G-856 proton precession magnetometer was used to monitor and record temporal changes in the Earth's magnetic field (at a rate of one measurement every 5 seconds). This secular was removed from the magnetic data during processing.

The following calibration/standardization tests were performed

- Background: The magnetic sensors were allowed to warm up to a point where there was <3 nT variation in the sensor readings before data collection began.
- Time calibrations for the sensor and IMU (see description in Appendix C4 of the demonstration plan);
- Measurement of coordinate systems of the IMU and sensor system (see description in Appendix C4 of the demonstration plan); and
- Heading correction (see description in Appendix C4 of the demonstration plan): Apart from the QA benefit of this test (detect if operator is carrying magnetic material) it also provided a mechanism to calculate a heading dependant correction for each sensor.

After demobilization from the site, we realized that there were a number of problems with the positioning of the magnetic data (see Appendix A for more information). The end result was that the magnetic data were not suitable for use in a cooperative inversion process. Therefore, no cooperative inversions were performed as part of this demonstration. We will test the cooperative inversion process in the next demonstration.

4.2.3 Geonics EM-63 Survey

Sky Research's modified EM-63 cart system with Leica RTS for position and Crossbow IMU for sensor orientation was used to conduct the third survey of the area. EM-63 data at twenty-six geometrically spaced time gates (spanning the range 180 μ s to 25.14 milliseconds [ms]) were collected at a 5 Hz rate (the maximum for the EM-63). The RTS was operated at around 4 Hz and the Crossbow IMU at 30 Hz. The EM-63 coil was 25 cm above the ground and was used to collect data along transects spaced 0.5 meters apart. Data were collected while walking slowly at about 2 km/hr so that the along line sample spacing was approximately 10 cm.

The following calibration/standardization tests were performed and are described in Appendix C of the demonstration plan in more detail:

- Background, spike, background test;
- Time calibrations for the sensor and IMU;
- Measurement of coordinate systems of the IMU and EM-63 coil.

The initial phase of EM-63 surveying was conducted with a new instrument supplied by the Colorado School of Mines. After collecting data on the test-plot and the first Rocket Range grid (I12), we noticed that the EM-63 was producing a lot of out-of-sync (i.e., invalid) soundings. Approximately 1 in every 30 soundings was corrupt and could not be used. We temporarily halted data collection while a rental EM-63 was shipped directly to the site from Geonics. This rental EM-63 had a slightly higher noise floor (~ 1 mV in time-channel 1 compared to 0.8 mV for the CSM unit), but a much lower incidence of invalid soundings. Only about 1 in every 150 readings was unusable.

4.2.4 Initial data processing

Initial Data Processing was conducted as described in Section 2.1, *Creation of a Map of Geophysical Sensor Data*. Appendix B describes the test-plot analysis that was conducted to select the target picking thresholds that were used to define the production level dig-sheets. Images of processed data for time-channel 3 for the EM-61 and time-channel 1 of the EM-63 are provided in Appendix D.

4.2.5 Feature Extraction

The inverse problem for estimating dipole parameters can be cast as an optimization problem, as per equations (7) to (9) above. For the data of this study we only consider uniform priors such that our problem is to minimize a data misfit function subject to box constraints:

$$\text{Minimize } \phi(\mathbf{m}) = \frac{1}{2} \left\| V_d^{-1/2} (\mathbf{d}_{\text{obs}} - F(\mathbf{m})) \right\|^2 \text{ subject to } m_i^L \leq m_i \leq m_i^U \quad (11)$$

where i represents model parameters which have upper and lower bounds. Finding a model that minimizes the above equation involves defining a data covariance vector V_d , the data vector

dobs, forward modeling function $F(\mathbf{m})$, the constraints m_i^L and m_i^U , and a numerical optimization procedure. In this section we outline our choices for these components when applying parameter estimation to the FLBGR data sets.

Data Covariance Matrix V_d

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

$$[V_d^{-1/2}]_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ \frac{1}{\delta_i + \varepsilon_i} & \text{if } i = j \end{cases} \quad (12)$$

where δ_i is a percentage of the i^{th} datum

$$\delta_i = \%error \times [d_{obs}]_i \quad (13)$$

and ε_i is a base level error that is present in the i^{th} datum in the absence of a target. In this demonstration we assumed zero %error, so the data covariance matrix is constant in space and determined by the base level errors, which are a function of time. Munkholm and Auker (1996) showed that log-gated and stacked white noise maps onto the TDEM response as errors with a standard deviation exhibiting a $t^{0.5}$ decay. Our strategy for determining the base line error is to mask a section of data within a grid that has no targets, and calculate the data statistics for each time channel within that grid. A single value represents the baseline error for the entire grid. Figure 35 (Section 6.1) shows a calculated noise floor for the EM-63 data in Grid 19-14 along with the theoretical $t^{0.5}$ decay due to Gaussian input noise. The theoretical (when appropriately scaled) and calculated noise floors are comparable.

For this demonstration we used a single data-covariance matrix for each grid. However, after further analysis we found that noise characteristics often changed significantly with survey event (different data collection periods). For example, Figure 12 plots the fifth time channel of data for Grid 21-14. The northerly portion of 21-14, from approximately 88 to 105 m, is less noisy than the southerly portion of the data. A geologic artifact can be seen to run diagonally to the southeast corner to the grid. Figure 13 demonstrates how the noise estimate will change as a function of where the data are sampled. We are currently developing automated methods for estimating a baseline error that varies as a function of position. These techniques will be used in subsequent demonstrations.

Forming the Data Vector \mathbf{d}_{obs}

The inversion procedure assumes that we are dealing with a single target in free space. Sensor drift, background geology, and nearby targets are non-random errors in the data that bias the estimated polarization parameters. By appropriately detrending the data and masking the individual anomalies we can minimize these effects.

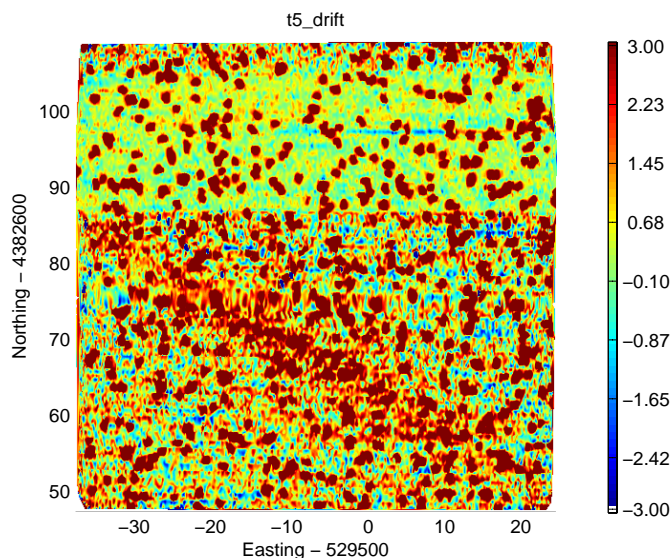


Figure 12. Image of time channel 5 data on the 21-14 Grid.

Detrending Data

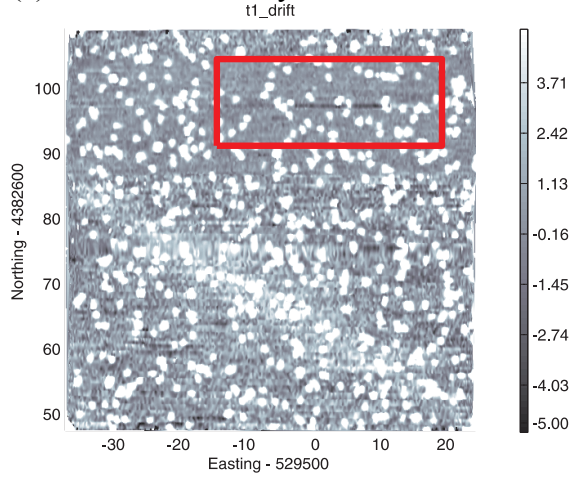
A demedian filter was applied along lines of data to reduce instrument drift, DC offsets, and geologic responses from the data. The default window size for this filter was 201 points. In some instance a 201 point window (approximately 20 m) was too large, and was insufficient to remove the geology (Figure 14). In these cases, the raw data were re-filtered with a smaller window.

Defining the Data to be Inverted 1: Spatial coverage

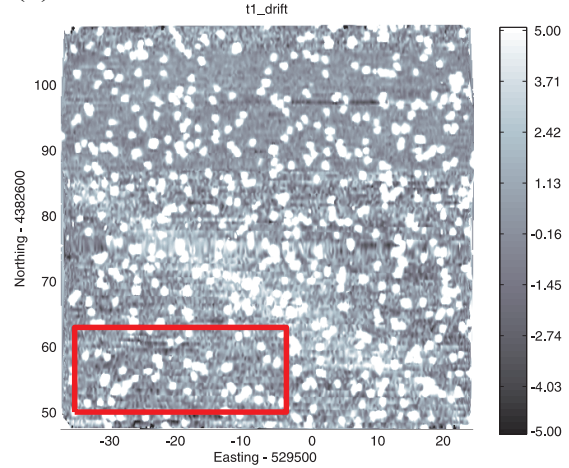
Once data anomalies are identified, a mask is defined that represents the spatial limits of the data to be inverted. Unlike magnetics data, an unconstrained EMI inversion is very sensitive to adjacent anomalies. The masking procedure helps ensure that signal from adjacent anomalies does not affect the inversion results. In addition, from a practical standpoint, inverting the minimum number of observations reduces the computational time.

The default mask is a 2 m radius circle centered on the picked target location (for example, Figure 15(a) anomaly 72). An automated correction to remove overlapping masks is then performed (for example, Figure 15(a) anomaly 159 and 66). Manual re-masking is required in cases where the automatically assigned masked region contains appreciable signal energy from an adjacent target. For example, to the NE of anomaly 257 in Figure 15(b), there is a smaller anomaly that isn't included in the inversion target list. Therefore, the mask overlap removal does not exclude the small anomaly. As a result, the data processor manually redrew that mask to exclude the small anomaly. Re-drawing masks is a tedious and time-consuming process.

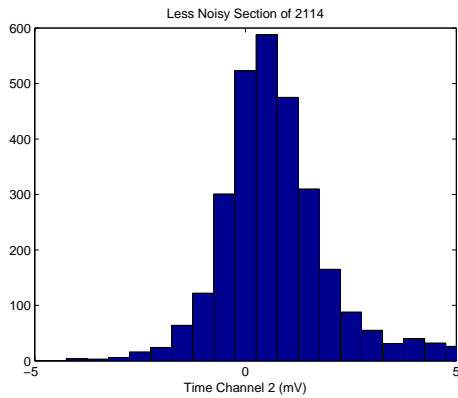
(a) Mask of less noisy area of Grid 21-14



(b) Mask of noisier area of Grid 21-14



(c) Time-channel 2 data histogram for (a)



(d) Time channel 2 data histogram for (b)

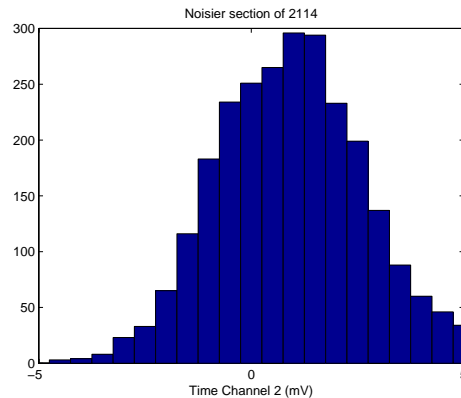


Figure 13. Comparison of noise in two areas of 21-14 grid. Masks indicating less noisy and more noisy areas of the grid are shown as red rectangles in (a) and (b), respectively. Although these masked areas include targets, it is clear that (b) has a larger the noise-floor.

Defining the Data to be Inverted 2: Time Channels

Figure 34 compares the sounding at anomaly maximum from 50 calibers, 20 mm, and 37 mm targets to the base-level noise of the survey. For these targets it is clear that at later times the signal from the target will sit within the base level noise of the instrument. These low-SNR, late-time channels provide little information to constrain the dipole-parameters. For this demonstration, we excluded any channels with a signal-to-noise ratio (SNR) of less than 10 from the inversion. For this procedure to be effective, we found we needed to be very careful with the selection of our mask, and with the magnitude of the errors assigned to each time-channel (Figure 16). On occasions the mask was too large and valid data at later times were not used to

constrain the model. To mitigate the problem, careful visual QC, re-masking and re-inversion were required.

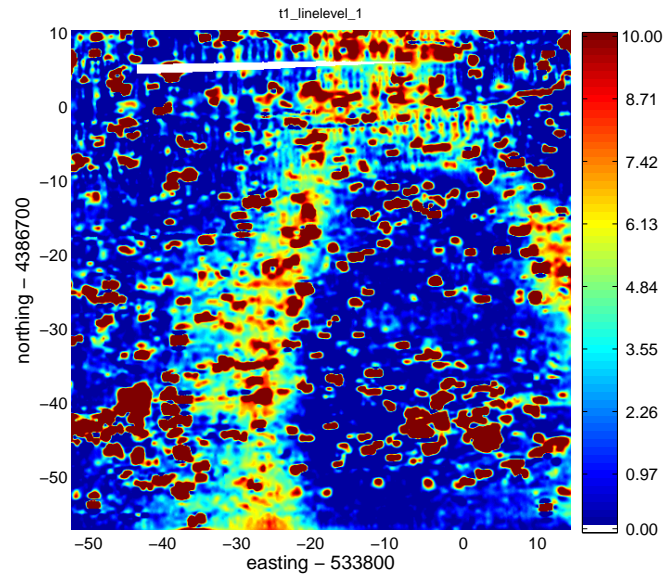
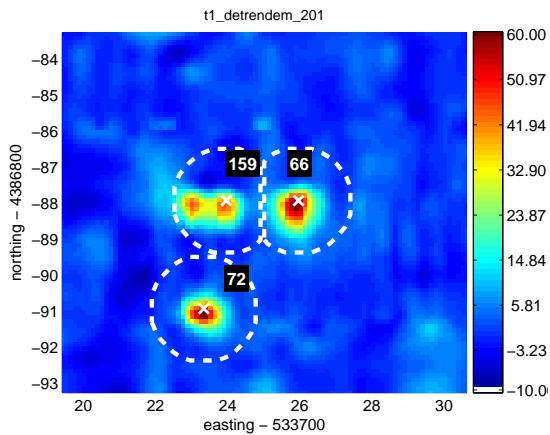


Figure 14. Filtered EM-61 Sky Array data acquired over Grid K-15. The approximately 20 m window was unable to remove the geologic anomaly.

(a)



(b)

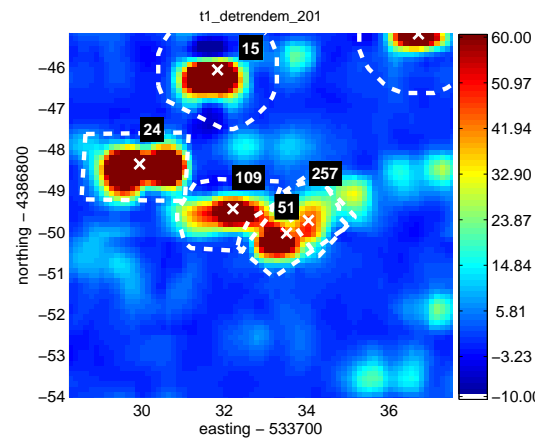


Figure 15. Examples of masking data from Grid L-14.

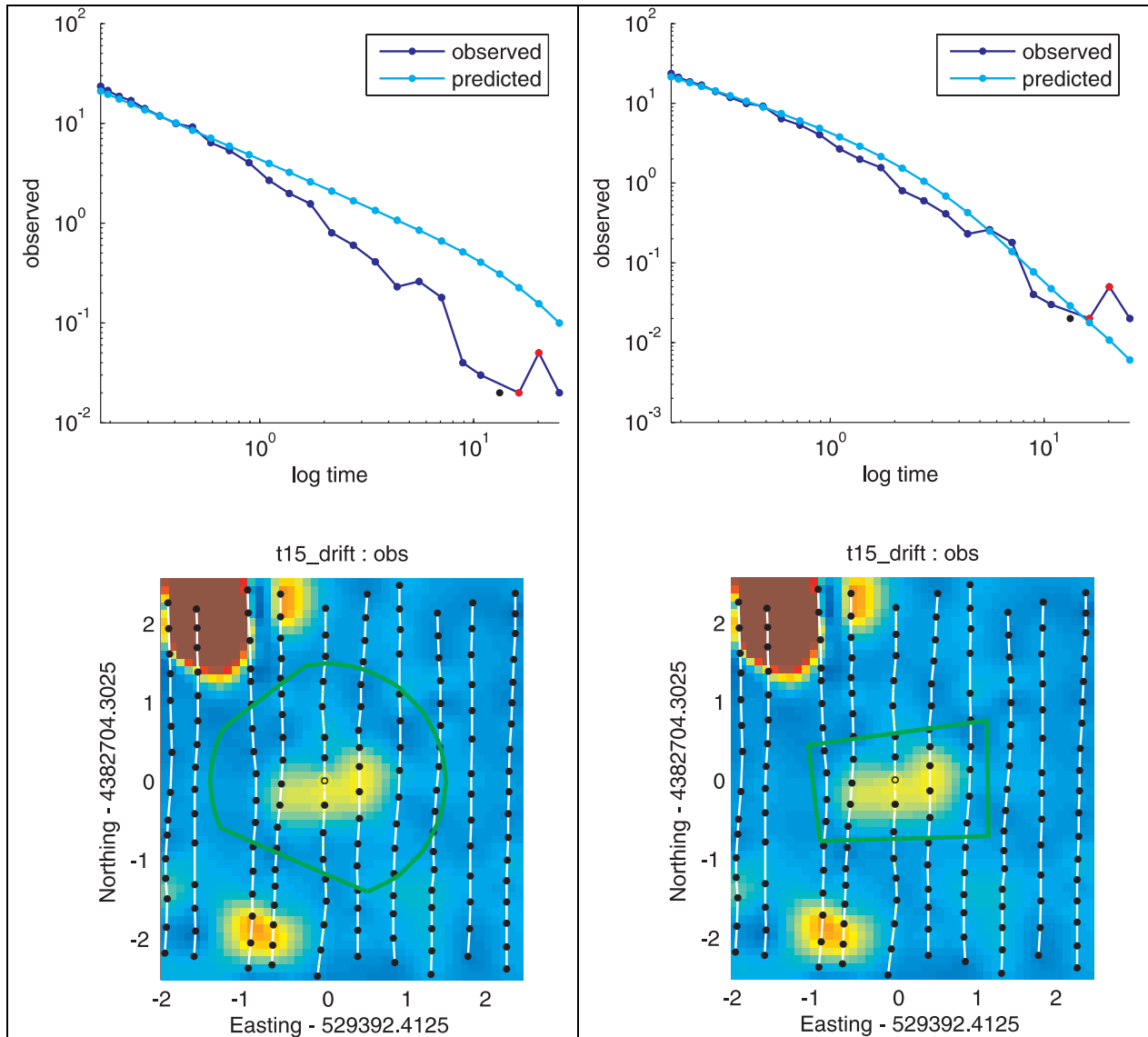


Figure 16. Sounding and model fit over the anomaly maximum for an EM-63 anomaly on the Rocket Range. At left, we show the fit using a large mask where the estimated SNR is low and time-channels 11 onwards are not used to constrain the inversion. On right, is the fit when the mask is smaller which leads to a more appropriate SNR estimate and the use of time-channels 1 to 18 to constrain the inversion. The plan views are of time-channel 15.

Defining the Model $F[m]$

Determining if a double-peaked anomaly should be inverted as a single target or a pair of targets

Visual examination of spatial anomaly pattern can be misleading for determining the number of objects. A horizontal target with a dominant axial polarization (as is the case with rod-like UXO) can lead to an anomaly with two-peaks. The peak separation is a function of the target depth and

transmitter loop size. When processing these anomalies, we need to decide if the anomaly is best fit with a single target, or a pair of targets. For the case of a pair of targets we segmented the anomaly into two separate anomalies, and inverted each masked portion of the anomaly with a single dipole model. A visual comparison of the two results (single object versus two objects) determined which model should be used. Figure 17 demonstrates this process using anomalies 257 and 51 from Figure 15(b). After inverting 51 and 257 individually, the anomaly was inverted as a single target. For this anomaly, it was decided that a single target was the best interpretation. The misfit and correlation coefficient provide a quantitative measure of the quality of the fits of the two different models. However, the final decision on which model to select required a subjective decision by a human interpreter who had previously modeled and viewed many different anomalies.

Determining if a single target should be inverted for 2 or 3 unique polarizations

A 3 x 3 magnetic polarizability tensor characterizes the induced dipole of a metal target. In the principal reference frame, it has three orthogonal polarizations that degenerate to two polarizations for a rationally symmetric or body-of-revolution (BOR). A target that can be characterized by three unique polarization tensor components does not have a symmetry axis, and is not likely intact UXO. For UXO anomalies, when the SNR is low, there may not be enough information in the data to constrain the two transverse components to be equal. In that case, the anomaly may be misclassified as non-UXO if the asymmetry of the polarization components is used as a feature for discrimination.

For this demonstration we used a conservative approach to determine whether to use two or three polarization tensors. Both two and three polarization models were inverted, and the data misfits were compared (Figure 18). If the data misfit of the three polarization inversion was less than 0.85 times the misfit of the two polarization inversion, then we selected the three polarization model. A memo describing this process was submitted to the Program Office in September 2006.

Determining the parameterization for the polarization decay

There are a number of different techniques for parameterization the temporal behavior of the polarization tensor. One common approach is to solve for the polarization value at each time channel (for example the AETC beta model). We apply this approach to EM-61 data. This approach is less efficient for the 26 time-channels recorded by the EM-63. For that sensor we parameterize the polarization decay. The parameterization is inspired by the different decay regimes observed in compact targets. At very early times, the decay of the voltage will follow a $t^{-1/2}$ decay, followed by a steeper power law decay ($t^{-3/2}$ for a sphere). At the late stage the response decays exponentially. For this study we use the following parameterized version of the polarization decay for the Geonics EM-63 data:

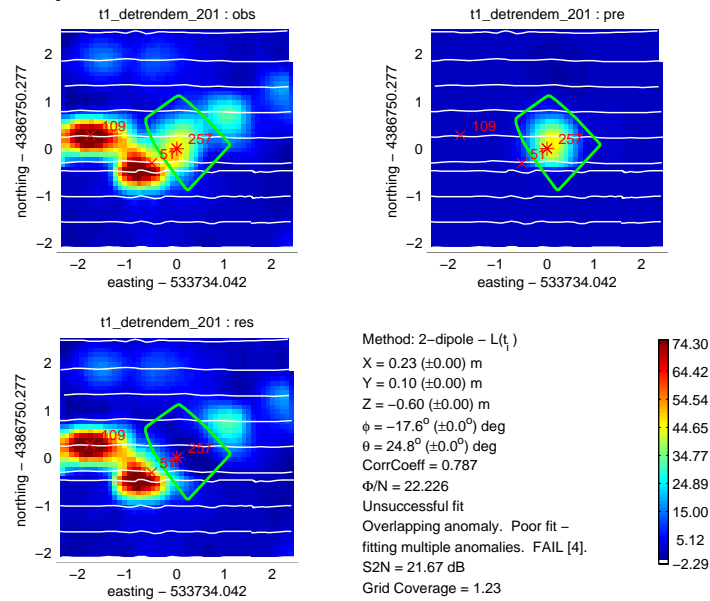
$$L(t) = kt^{-\beta} \exp\left(-\frac{t}{\gamma}\right) \quad (14)$$

Depending on the locations of the TEM sensor time gates and the noise levels of the data, the late time stages may not be seen in sensor data, and therefore model parameters describing these features will be poorly constrained upon inversion. In our initial analysis, we found that for small targets, γ was poorly constrained but had a significant impact on the values of the k and β parameters. The approach we used in this demonstration was to identify situations where γ was poorly constrained, and re-invert those targets with a power law parameterization $L(t) = kt^{-\beta}$. After submission of dig-sheets, we conducted a retrospective analysis and found that many of the problems associated with small targets were due to the lower constraints on the polarization being too large. The local search method produced solutions on the constraints which caused the estimated variances of the remaining parameters to be inaccurate. By reducing the lower constraints of the polarization values, a re-inversion using a power law was unnecessary. A retrospective analysis using these reinverted anomalies is given in Section 6.1.

Optimization: Determining the minimum of $\phi(\mathbf{m})$

The optimization routine we used for inversion is a local Newton-type method that minimizes the least squares objective/misfit. We addressed the problem of local minima by choosing multiple starting models. We first defined a number of depths and orientations. For each combination of depth and orientation, we then solved a linear inverse problem for the polarization tensor at the first time channel by fixing the depth and orientation. The five combinations of orientation and depth with the smallest misfits are then chosen for inversion. For the 2-polarization inversions we bias inversion results towards an axi-symmetric rod-like target by initially inverting for a model where the axial polarization is larger than the transverse polarization. This is achieved by applying a linear constraint to the polarization values. The resulting model is then used as a starting model for to a box-constrained inversion. For the 3 polarization inversion we solve only a box-constrained inversion.

(a) Inversion of Anomaly 257



(b) Inversion of Anomaly 51 when 257 is included

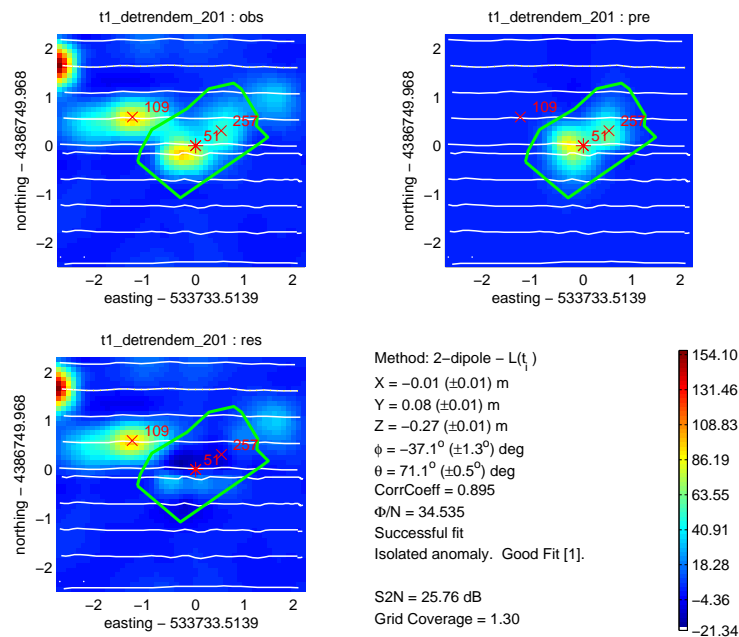
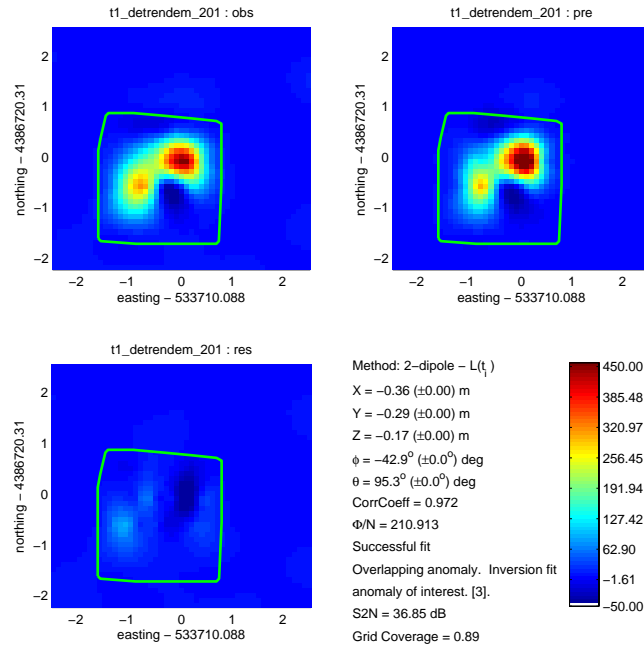


Figure 17. Determining if an anomaly should be inverted as two separate targets. Anomalies 51 and 257 were initially inverted as two separate targets. (a) shows the result when inverting 257. (b) shows the result when the two anomalies are inverted as a single target.

(a) Two-polarization Inversion of Cell 2



(b) Three-polarization Inversion of Cell 2

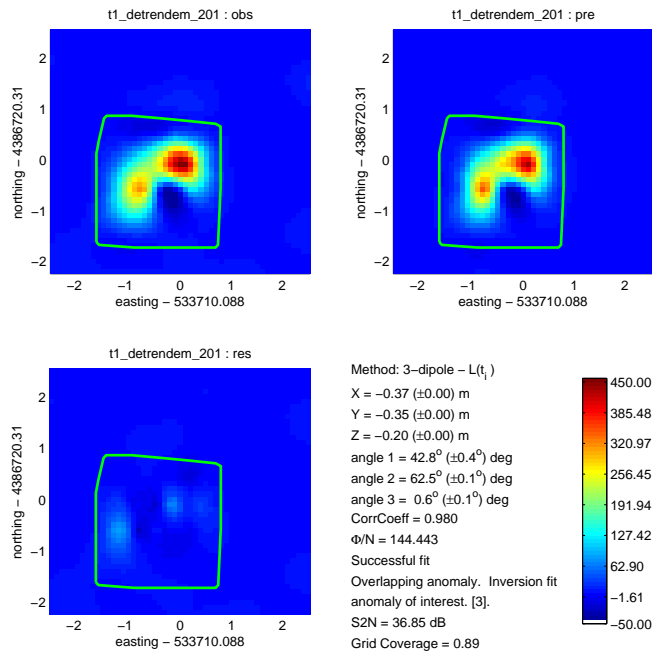


Figure 18. Example of Two- and Three-polarization inversions of the same anomaly. The three-polarization model would be used in this instance due to a significantly reduced misfit.

4.2.6 Statistical Classification

The following general procedures were used for the feature vectors of each sensor combination:

- *Selection of features:* By analysis of the training data, those features that contribute to separation of the different classes (comprising UXO types and clutter) were selected. This selection involved both canonical analysis and visual review to determine the combinations of features that contribute most to class separation.
- *Choice of classification algorithm:* Through analysis of the training data the best performing classifier was selected.
- *Classification:* Anomalies were placed in a prioritized diglist by using the classifier to compute probabilities of class membership for unlabeled feature vectors. The probability of membership of the UXO class was reported on the dig-sheet.
- *Anomalies where feature vectors are unreliable:* Some anomalies inevitably have insufficient SNR or data coverage to constrain the TEM model parameters. This also includes anomalies with overlapping signatures that cannot be isolated and inverted one at a time. All these anomalies were placed in the dig-sheet and given a label of “unknown” and were excavated as suspected UXO. They were not used in the performance evaluation.
- *Note on choice of classifier:* The same Support Vector Machine classifier was used for all three phases. We made slight variations on the feature vectors used after Phase 1 as described in the results section.

The objective in the 20 mm Range-Fan was to discriminate between ubiquitous small UXO (50 caliber bullets) and larger 37 mm projectiles. A number of 37 mm items were emplaced by ERDC in surveyed grids; ground-truth was available for some emplaced items, but not all. In the rocket range, a wider variety of UXO were expected, with items of interest ranging in size from 100 pound (lb) bombs down to 37 mm projectiles. A training data set for both areas was obtained from “geophysical prove-out” data acquired over the FLBGR test-plot that was established several years previously to support production geophysical activities. Emplaced items in the test plot were representative of all types of UXO expected in the two survey areas (Appendix C, Table C-1).

Data from two grids within the Rocket Range (one medium, one high density) and one grid within the 20 mm Range-Fan (medium density) were interpreted using the training data from the test-plot, and ranked diglists were submitted to the Program Office and Don Yule of USACE-ERDC, together with a recommendation for how many of these anomalies to dig as potential UXO. After ground-truth for these grids was collected (for all detected anomalies, not just the ones recommended for excavation) they were released to the analysts then updated the training data and interpret the remaining grids. Complete ground-truth from two-more of the Rocket Range grids (medium and high density) was released and revised dig-sheets submitted for the remaining four grids.

4.2.7 Ground-truth obtained

Ground-truth data were managed by Linda Daehn of Tetra-Tech EMI Inc. of Helena, Montana, and were kept secret from Sky Research. Ground truth was only released after interpretations had been submitted to the Program Office. There were two-phases of ground-truth data release on the 20 mm Range-Fan and three phases on the Rocket Range. In Table 7 we list the total number of items recovered in each grid in each of the following categories MK-23, 2.25" rocket, 37 mm projectile, 20 mm projectile, small arms (50 cal), shrapnel and junk. We consider MK-23s and 37 mms as UXO regardless of whether they are inert or live rounds.

Table 7. Number of anomalies identified as MK-23, 2.25" rocket, 37 mm projectile, 20 mm projectile, small arms, shrapnel and junk in each of the 10 grids used for this demonstration. For each grid, we identify the phase where ground-truth was released.

Grid and (phase)	Bomb MK- 23	Rocket 2.25"	37 mm projectile	20 mm projectile	Small arms	Shrapnel	Junk	Grand Total
19-14 (I)			18	25	91			134
21-14 (II)			20	39	90		1	151
I-12 (I)	5			46	4	28	2	85
I-13 (III)	14			41	8	20	10	93
J-12 (I)	7			44	14	52	17	134
J-13 (III)	3			45	9	34	10	101
K-15 (II)				52	27	44	60	183
L-13 (III)		1		55	23	72	44	195
L-14 (III)				65	14	66	51	196
L-15 (II)				50	18	54	43	165
Grand Total	29	1	38	462	298	370	238	1437

On the 20 mm Range-Fan, the majority of recovered items were either 50 caliber bullets or 20 mm or 37 mm projectiles. There was almost no shrapnel or junk recovered from that site. Of the 37 mm projectiles, the majority were seed items emplaced by Sky Research in known locations (20 rounds) or by ERDC (16 items) in unknown locations.

On the Rocket Range, there are also large numbers of 50 caliber bullets (small-arms) and 20 mm projectiles, but no 37 mm projectiles. All but one UXO were MK-23 practice bombs, with a single rocket warhead recovered. In addition, almost all the MK-23s were found at or near the surface so they had large amplitudes and large signal-to-noise ratio. Unfortunately, for the two Phase II grids (L-15 and K-15) and two of the Phase III grids (L-13 and L-14) there was only one UXO item found (the rocket warhead), and the fit for that item was failed. The number of items on the dig-sheet is the only metric we can use to compare the amplitude and classification methods. The *best method* will be the one that recommends the lowest number of items to be excavated regardless of the validity of the underlying discrimination methodology. Therefore,

results for Phase II Rocket Range grids and L-13 and L-14 in Phase III were not used to compare the performance of the discrimination methods.

4.3 Discrimination results

4.3.1 Feature selection and choice of classifier

No separation between UXO and the available clutter items is evident in a feature space spanned by polarization parameters extracted from Sky Array or EM-63 test plot data (Figure 19). Note that the k parameters have been relabeled so that $k_1 > k_2 > k_3$, so all items appear rod-like in this figure.

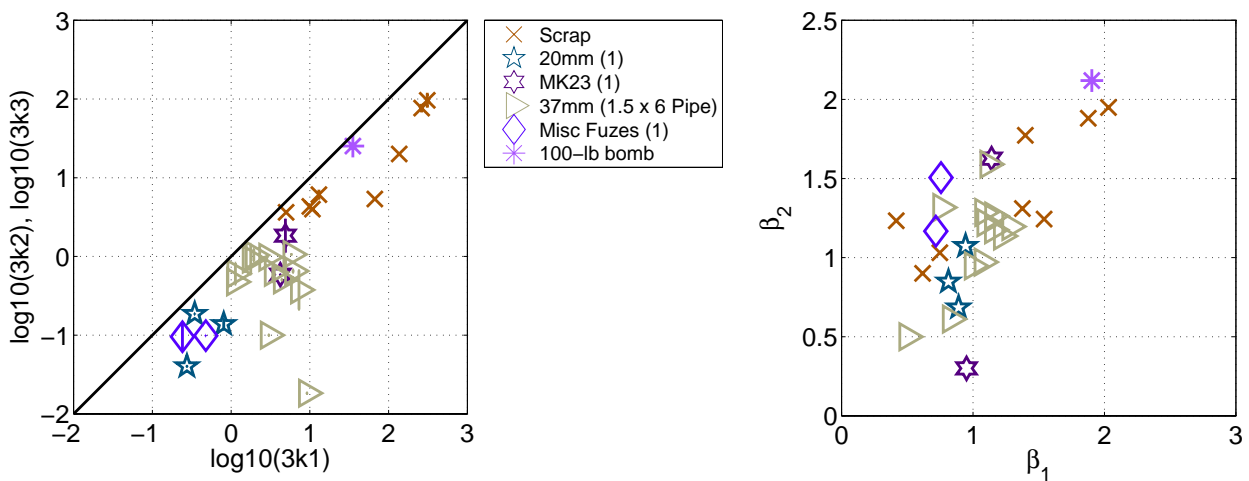


Figure 19. Features extracted from FLBGR EM-63 testplot data.

Many studies have assumed that UXO are rod-like axisymmetric targets that can be discriminated from clutter by considering the difference between the secondary elements of the polarization tensor (i.e., the dipoles aligned perpendicular to the axis of symmetry). By replacing three-dipole models with two-dipole models in cases where the two-dipole misfit is comparable to that of the three dipole model, we enforce an axisymmetric assumption in cases where it is justified by the data. Interestingly, clutter items are symmetric ($k_2 = k_3$) for all targets considered in these data. Furthermore, a number of larger UXO targets can be fit with dipole models with significant asymmetry. This observation is contrary to the hypothesis that UXO targets can be represented as axi-symmetric bodies of revolution. While this hypothesis is almost certainly true in laboratory settings where high quality data can be inverted to obtain parameter estimates, in field situations the data quality (SNR and coverage) may not be sufficient to support a decision scheme which relies upon estimates of target shape.

However, from Figure 19, it is evident that target size can provide discrimination information between 37 mm items and smaller targets such as 20 mm projectiles and fuzes. In particular,

there is (with one exception) good clustering of the simulated 37 mm pipes in a feature space spanned by the k parameters. The classification task for the 20 mm Range-Fan can therefore be formulated as a problem of discriminating between small ordnance or ordnance related-items (e.g., 20 mm and fuzes) and larger items (37 mm and up).

Canonical analysis of the polarization parameters yields few surprises (Figure 20). When trying to discriminate between 37 mm and 20 mm targets, it is primarily the size parameters (k parameters) which provide useful information. Under the assumption of identical within-class scatter (both classes have the same covariance), the classes can be separated with a single eigenvector which is a linear combination of the polarization parameters. The canonical analysis indicates β parameters provide additional classification information and there is some separation between the small ordnance and 37 mm in a feature space spanned by the first two β parameters (Figure 19b). However, there is sufficient overlap between the classes that we discard, at least initially, the β parameters as features for the first round of discriminations.

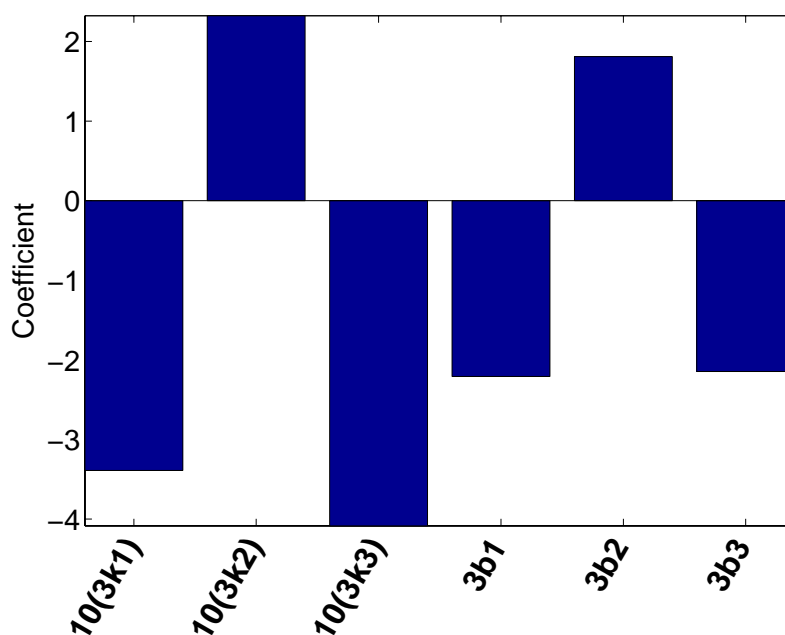
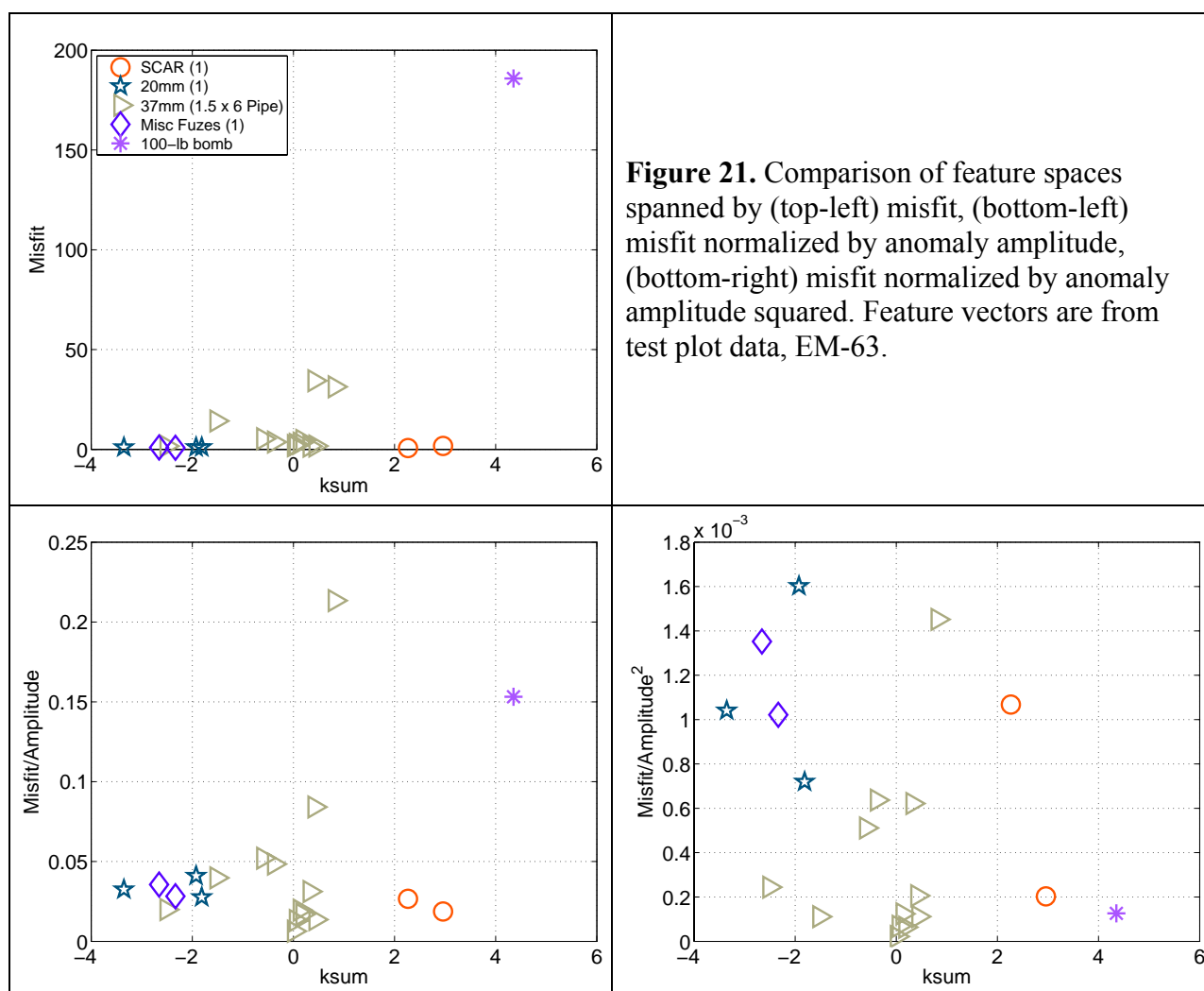


Figure 20. Contributions of polarization parameters to first canonical eigenvector for canonical analysis of EM-63 test plot data. The eigenvector provides the maximum separation between 37mm items and small ordnance and explosives (OE) scrap (20mm and fuzes).

The EM-63 data contained a large number of low amplitude anomalies for which parameter estimates likely had large relative uncertainties. One way to avoid these low SNR fits is to simply use SNR as a feature for discrimination. This is perhaps unwise, since it is possible that UXO of interest may produce a lower SNR than is seen for UXO in the training data. This situation occurs when comparing emplaced 37 mm items in grids 19-14 and 21-14. There are a number of deep (>30 cm) 37 mm in 21-14 with significantly lower SNR than those in 19-14.

A more robust feature for rejecting low amplitude anomalies appears to be the misfit divided by the maximum signal amplitude at the first time channel. Here the misfit is selected by comparing two- and three-dipole models, as outlined previously. The misfit is itself normalized by the number of data points (a “reduced χ^2 ” misfit). This feature has been successfully used in a previous demonstration for discrimination using multi-sensor towed array detection system (MTADS) data (Nelson *et al.*, 2003). In that study, the set of ordnance types was much smaller and contained large items (e.g., 105 mm and 155 mm items) which are relatively easy to identify. Normalizing by the anomaly amplitude scales the misfit so that low SNR anomalies are down-weighted (less-likely to be UXO). Furthermore, for high SNR targets the misfit can be relatively large, even though the fit is quite good. This is best exemplified by the fit for the test plot 100 lb. bomb: the fit is reasonable but the misfit is the largest of all test plot items. Normalizing by the anomaly amplitude places the 100 lb bomb well within the distribution of UXO (Figure 21). Also shown for comparison is the misfit normalized by anomaly amplitude squared. As discussed in the following section, this feature was used for discrimination in rounds two and three.



This approach to comparing fit quality and associated parameter uncertainties, while shown to be viable in this and previous demonstrations, is admittedly somewhat ad hoc. As later outlined in the discussion, careful analysis of spatially and temporally variable data uncertainties is clearly necessary to rigorously incorporate parameter uncertainties into advanced discrimination.

The correspondence between emplaced 37 mm items and 1.5" x 6" test plot pipes, meant to represent 37 mm projectiles, is quite good in a feature space spanned by the k -parameters. However, the relative misfit for one pipe is quite large (well outside the distribution for emplaced 37 mm items) and so the pipes have been discounted from the training data.

For the first round of classifications our ground truth is limited to the test plot but the selection of a classifier can be informed by both statistical analysis and subjective criteria. At this stage, it is important to keep things simple by using only a few relevant features. Following the predicted progress of a proposed classifier through the test data feature space is an important QC step: the digging order should be consistent with the intuitive order which is obtained by "eyeballing" the features.

For Grid 19-14, we use a two-dimensional feature space spanned by *misfit/amplitude* (based on the experience of previous demonstrations) and the sum of the k parameters (based on canonical analysis). The sum of the k parameters is used to provide a feature which has physical meaning (it is related to the volume of a target) and which is supported by statistical analysis (the k parameters are most relevant according to canonical analysis). The training data are emplaced 37 mm (UXO class) vs. 20 mm and OE fuzes (clutter class).

Considerable time was spent experimenting with classifiers to obtain a satisfactory digging order. We found that the statistics for small ordnance (20 mm) taken from the training grid were not particularly representative of unlabelled targets. This is likely because there are many unlabelled targets with low amplitude anomalies and relatively poor fits. We therefore included some additional test plot fuzes with relatively high misfits as part of the clutter class.

Given the small samples available for the first round, a generative classifier, which relies upon having a sufficiently large sample to estimate the probability distribution of the clutter, is probably not a good choice. A support vector machine was therefore used to classify all unlabelled targets. This classifier directly models the decision boundary between classes and is robust for small training sets. The output of this classifier ranges on the interval $(-\infty, \infty)$, here we use the convention that a positive value lies on the "UXO" side of the decision boundary ($f=0$). Figure 22 shows the feature space for EM-63, training and test data and targets selected for digging. Also shown are the contour lines of the trained SVM in this feature space. An excellent correspondence is observed for emplaced 37 mm items from another grid (21-14). All items in this hold-out test set fall on the UXO side of the decision boundary.

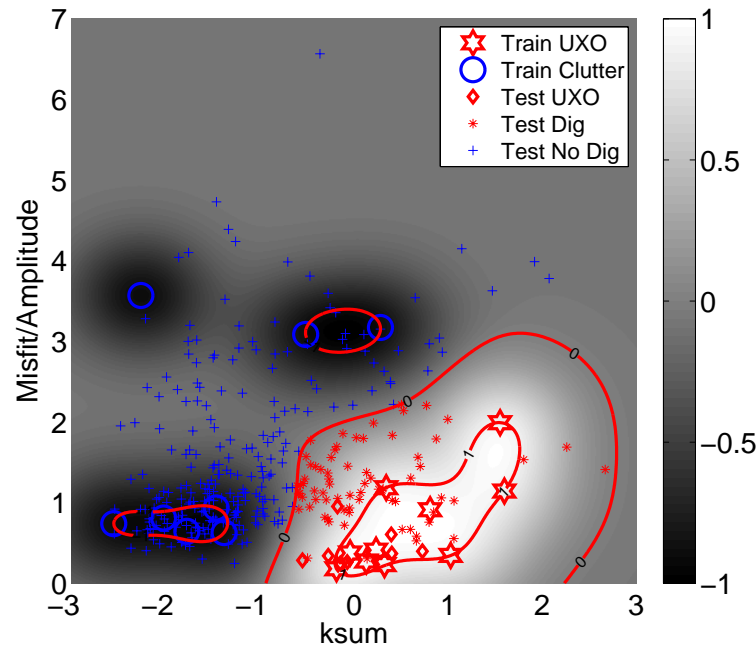


Figure 22. Feature space for classification of EM-63 data from Grid 19-14. “Train UXO” are emplaced 37 mm in 19-14, and “Test UXO” are 37 mm from Grid 21-14 (held out from classifier training). Contours show trained SVM output with $f=0$ corresponding to the decision boundary.

A similar feature space was used for classification of EM-61 data, with the sum of the k parameters replaced by the sum of the instantaneous dipole polarizations at the first time channel. We found that the stopping criterion for the SVM was somewhat conservative for these grids; some very high SNR targets were being missed. Although these targets are outliers to both UXO and clutter classes, it is not sensible to leave anomalies of 800 mV (or more). The SVM threshold was therefore lowered to include these targets in the “Dig” category of the diglist.

A list of the training data available for each phase of classification and the resulting feature vectors are provided in Table 8. As described below, we attempted to maintain the same feature vectors and classifier for all phases, but did need to modify some feature vectors.

4.3.2 EM-61 results on the 20 mm Range-Fan

Figure 23 shows the ROCs obtained for trained classifiers and amplitude thresholding in the 20 mm Range-Fan. All ROCs presented only consider anomalies for which ground truth exists and for which the data fitting was successful. In Grid 21-14 there is one failed fit (a deep round emplaced by ERDC at 40 cm). The SVM significantly outperforms amplitude thresholding in both grids, although we did have a tendency to stop digging too soon. It should be noted that three 37 mm targets (two in 19-14 and one in 21-14) identified by the statistical classifier fall below the 15 mV production threshold for the Sky array and would not be excavated in a

standard setting. The retrospective classifier shown in Figure 23(a) replaces *misfit/amplitude* with *misfit/amplitude*², as discussed later in this section.

Table 8. Different phases of classification on the 20 mm Range-Fan and the Rocket Range and the corresponding training data and feature vectors selected for classification for both the EM-61 and EM-63.

Phase	EM-61		EM-63	
	Training data	Feature vectors	Training data	Feature vectors
20mmRF: I (19-14)	Test plot	<i>misfit/amplitude</i> Lsum	Test-plot, 10 emp. 37 mm in 19-14 Test-data 10 emp. 37 mm in 21-14	Misfit/amplitude ksum
20mmRF: II (21-14)	Test plot and 19-14	<i>Misfit/amplitude</i> ² Lsum	Test-plot, 20 emp. 37 and 19-14	<i>Misfit/amplitude</i> ² ksum
RR: I (I-12, J-13)	Test-plot	<i>Misfit/amplitude</i> Lsum	Test-plot	<i>Misfit/amplitude</i> ksum
RR: II (L-14, K-15)	Test-plot and I-12, J-13	<i>Misfit/amplitude</i> ² Lsum	Test-plot and I-12, J-13	<i>Misfit/amplitude</i> ² log(k ₁)
RR: III (I-13, J-12, L-13, L-14)	Test-plot, I-12, J-13	<i>Misfit/amplitude</i> ² Lsum	Test-plot, I-12, J-13	<i>Misfit/amplitude</i> ² log(k ₁)

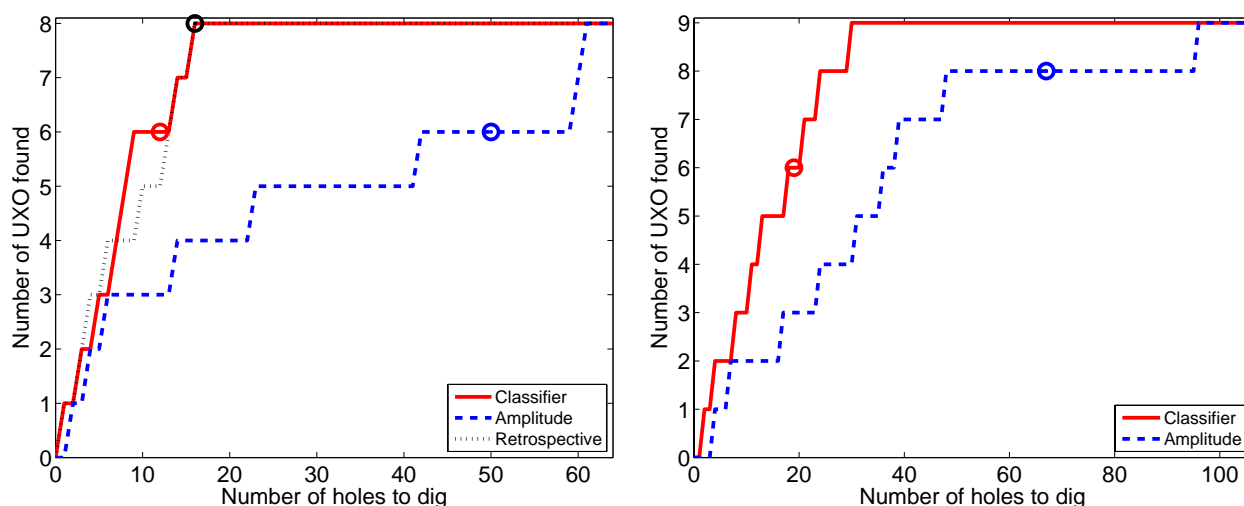


Figure 23. EM-61 ROC curves for 20mm Range-Fan grids. Circles show the operating points for each diglist: (on left) Results for 19-14 where the retrospective classifier replaces *misfit/amplitude* by *misfit/amplitude*²; (on right) Results for 21-14.

4.3.3 EM-61 results on the Rocket Range

Figure 24(a) shows the merged ROC for Grids I-12 and J-13 in the rocket range. The performance of the support vector machine is quite poor relative to the amplitude threshold in both grids. For I-12, the classifier performance is degraded by two poor fits which should not have passed QC. As described in the previous section, these high-misfit, high-amplitude feature vectors motivated us to lower the SVM threshold for digging to ensure that these targets were flagged for excavation in the classifier diglist. These targets are the final UXO found by the submitted and retrospective classifiers in Figure 25.

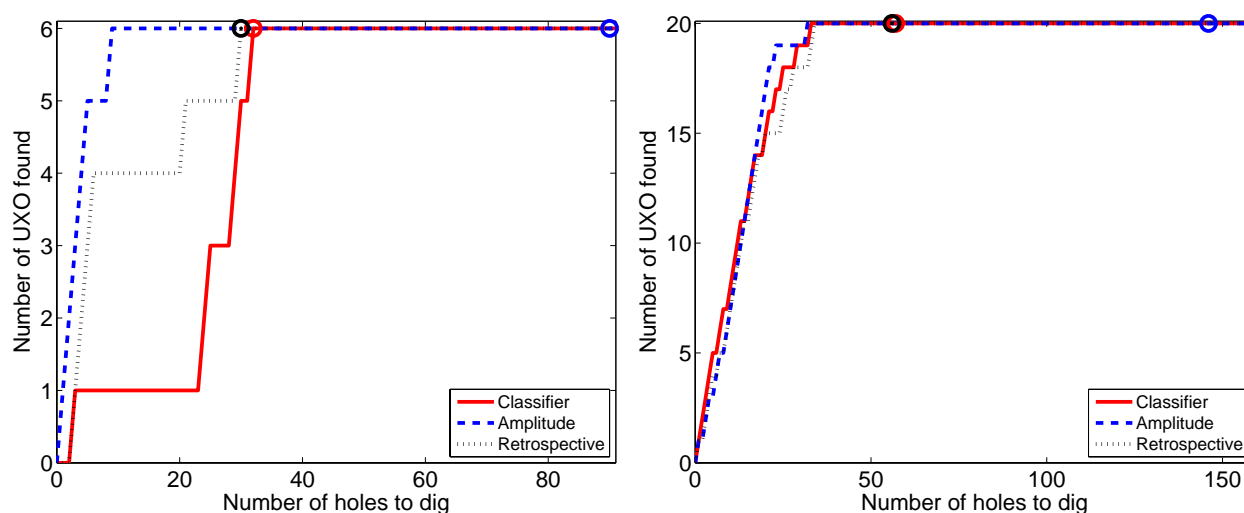


Figure 24. EM-61 ROC curves for the Rocket Range grids. Circles show the operating points for each diglist: (on left) Merged ROCs for Phase I Grids (I-12 and J-13), where the retrospective classifier replaces *misfit/amplitude* by $\text{misfit}/\text{amplitude}^2$; (on right) Merged ROCs for Phase III grids (I-12 and J-13). Retrospective classifier replaces *lsum* with first canonical eigenvector.

Classification results in Grid J-13 are quite poor, requiring more than 20 excavations before the first of 3 MK23 bombs is found. However, retrospective analysis easily improved on this result by replacing the feature *misfit/amplitude* by $\text{misfit}/\text{amplitude}^2$. While the misfit is technically a unit-less quantity after normalizing the data residuals by the estimated standard deviations, it is, in practice, dominated by contributions from the first few time channels. In this demonstration, channel standard deviations were estimated by computing the data standard deviation in a “quiet,” anomaly-free area of the survey grid. However, this estimate of the noise does not capture the dynamic range of the data and so the misfit can be dominated by early time channels, especially when fitting high amplitude anomalies. To compare misfit values between different anomalies, we therefore normalize the misfit by the squared anomaly amplitude at Channel 1. This produces a much tighter clustering of MK23 targets for Grid J-13, as shown in Figure 25.

These results motivated us to revise our feature space for the remaining grids in the rocket range,

and there was a commensurate improvement in classifier performance for these grids. For the Phase III grids I-13 and J-12, ROCs for classification are comparable to amplitude thresholding (Figure 24(b)), which is nearly ideal for these grids. Also shown in these figures are classifier operating points (i.e., stop-digging points) identified in the diglists. The operating points, corresponding to $f=0$ for the SVM, were selected based on the first round rocket range grids (I12, J-13), while a 15 mV cutoff was used for the amplitude diglists. As summarized in Table 9, the operating points chosen for the classifier diglists require far fewer digs than the amplitude threshold. In grids where no UXO were found (K-15 and L-15) this is our only criterion for comparing discrimination methods. Obviously, we could raise the amplitude cutoff and thereby significantly reduce the number of targets to dig. Of course, there is no way to know *a priori* that only shallow, large anomaly amplitude targets (MK23 bombs) are present in these grids. By raising the amplitude threshold, we risk missing deeper or smaller targets which might successfully be identified with advanced discrimination. It is notable that while the classifier and thresholding ROCs for e.g., Grid J-13, are virtually identical, the order in which UXO are found is inversely correlated between the two methods.

Table 9. Number of targets flagged for digging by discrimination method, rocket range grids, Rounds 2 and 3.

Grid	Classifier: Sky Array	Classifier EM-63	Amplitude	Number of UXO
I-13	35	34	116	5
J-12	30	21	138	15
K-15	43	31	188	0
L-13	40	103	202	1
L-14	58	52	203	0
L-15	33	35	175	0
Total	239	276	1022	21

To quantify the correlation between two discrete series of length N with integer elements, we can use a variant of the Spearman rank-order correlation coefficient:

$$p = 1 - \frac{6 \sum D_i^2}{N^3 - N} \quad (15)$$

with D^2 the squared difference of ranks. For example, if we have two diglists

$$a = [1 \ 2 \ 400]$$

$$b = [400 \ 2 \ 1]$$

we replace each element by its rank in the diglists, compute the difference between ranks and sum them, as shown in Table 10.

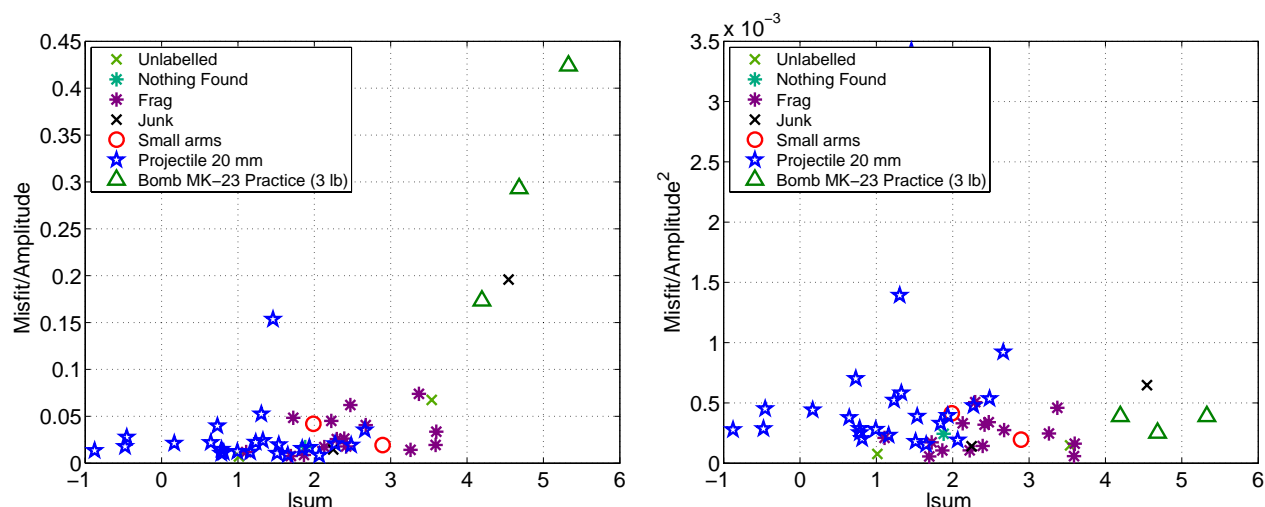


Figure 25. Comparison of feature spaces for Grid J-13. Lsum denotes the sum of the instantaneous polarization amplitudes at the first time channel: (on left) Feature space with *misfit/amplitude*; (on right) Feature space with *misfit/amplitude*².

Table 10. Simple example for computation of Spearman rank correlation between diglists ($\rho = -1$ in this example).

		Element	
	1	2	400
Rank (a)	1	2	3
Rank (b)	3	2	1
D	-2	0	2

For example in Grid J-12, the Spearman correlation between ROCs for amplitude thresholding and classification is $\rho = -0.38$. That is, while amplitude thresholding goes after large amplitude anomalies first, classification tends to identify smaller amplitude anomalies which can be fit well with inversion and which have a size parameter which indicates a target is the right size for a MK23 (given our prior knowledge of MK23).

Figure 24(b) also shows a retrospective ROC for classifiers trained using *misfit/amplitude*² and the first canonical eigenvector obtained from canonical analysis of feature data from the first two rounds. The canonical eigenvector uses information from all time channels and it was hoped that this might provide an improvement in performance compared to previous strategies, which used only the polarizations at the first time channel. However, in both Grids I-13 and J-12 there is no significant difference between ROCs for the two statistical classifiers. This suggests that the first canonical eigenvector is not substantially different from the sum of the polarizations at the first time channel and the contribution of later time channels (or their linear combination) to the separation between UXO and clutter is negligible.

As mentioned earlier Phase II excavations in grids L-13, L-14, L-15 and K-15 yielded only one

UXO (a 2.25 inch rocket). Although the fit to this anomaly was failed, the amplitude was greater than 2000 mV and so the target would have high priority when digging failed items using amplitude thresholding.

4.3.4 EM-63 results on the 20 mm Range-Fan

Figure 26 shows the feature space used for classification for both 20 mm Range-Fan grids. Performance of classifiers in the 20 mm Range-Fan for the EM-63 was comparable to that obtained with the EM-61: statistical classification significantly outperformed amplitude thresholding when searching for 37 mm projectiles (Figure 27). The improvement in performance in Grid 21-14 was especially notable, with several projectiles found very late in the amplitude diglist. A target mislabeling error occurred in this grid: ground-truth for an emplaced 37 mm was mistakenly associated with Target 54, instead of Target 154. The choice of mislabeling was lucky, however, as Target 54 was (unbeknownst to us) in fact a 37 mm projectile. This error was not spotted while training the classifier, since Target 54 was within the region of the feature space expected for UXO targets. Retrospective analysis of classifier performance with Target 54 treated as a test (unlabeled) vector and Target 154 as a training (labeled) vector is shown in Figure 28.

There were actually eight 37 mm with valid feature vectors in 21-14, not six as seemingly indicated by the ROC curves. Two of these were very deep (40 cm rounds) emplaced by ERDC and with amplitudes significantly below the production threshold (20.5 and 22 mV in time-channel 1, compared to the 35 mV threshold). These two deep anomalies were ranked at positions 74 & 115 by classification and at positions 208 & 304 by amplitude.

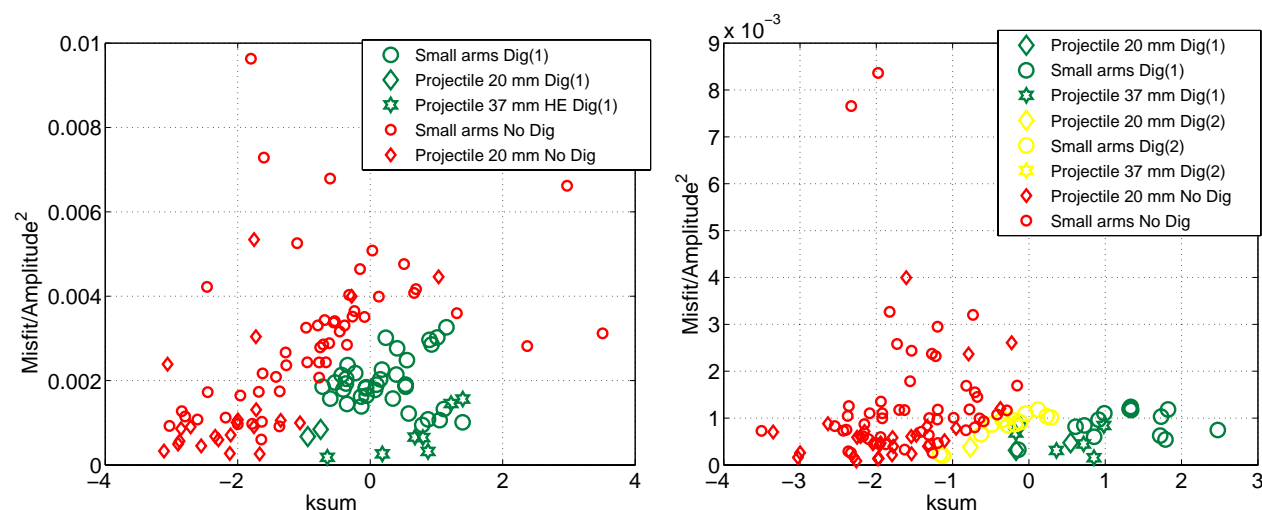


Figure 26. Feature space used for EM-63 classification at the 20 mm Range Fan. Red targets are below the classifier operating point). Green targets are recommended for excavation, while yellow (21-14 only) are also recommended for excavation but with a lower priority: (on left) Feature space for 19-14 (using $\text{misfit}/\text{amplitude}^2$ in place of $\text{misfit}/\text{amplitude}$); (on right) Feature space for 21-14.

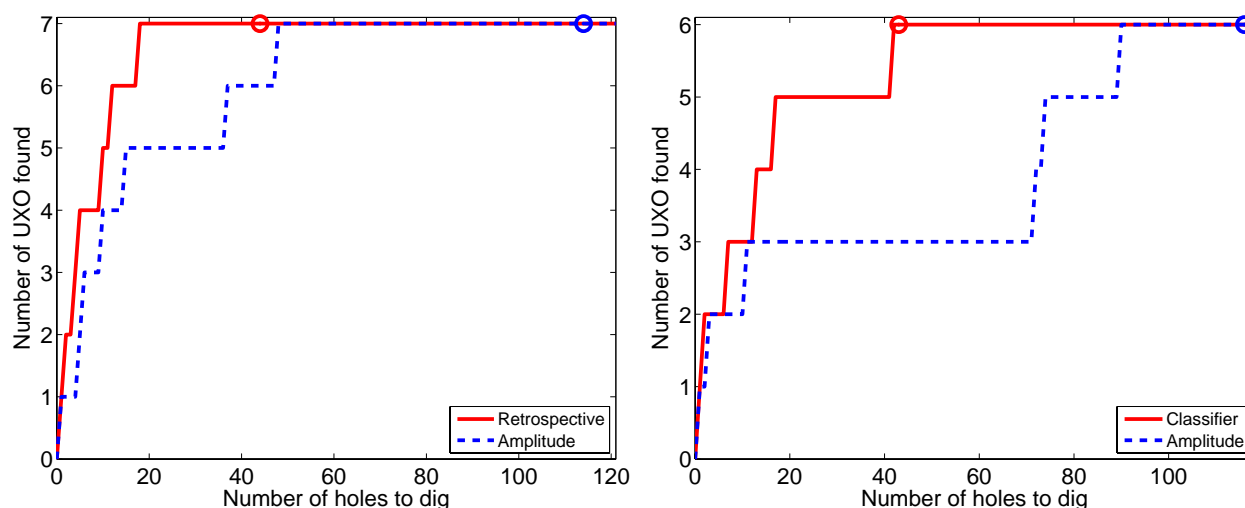


Figure 27. EM-63 ROC curves for 20 mm Range-Fan grids. Circles show the operating points for each diglist: (on left) Results for 19-14; (on right) Results for 21-14.

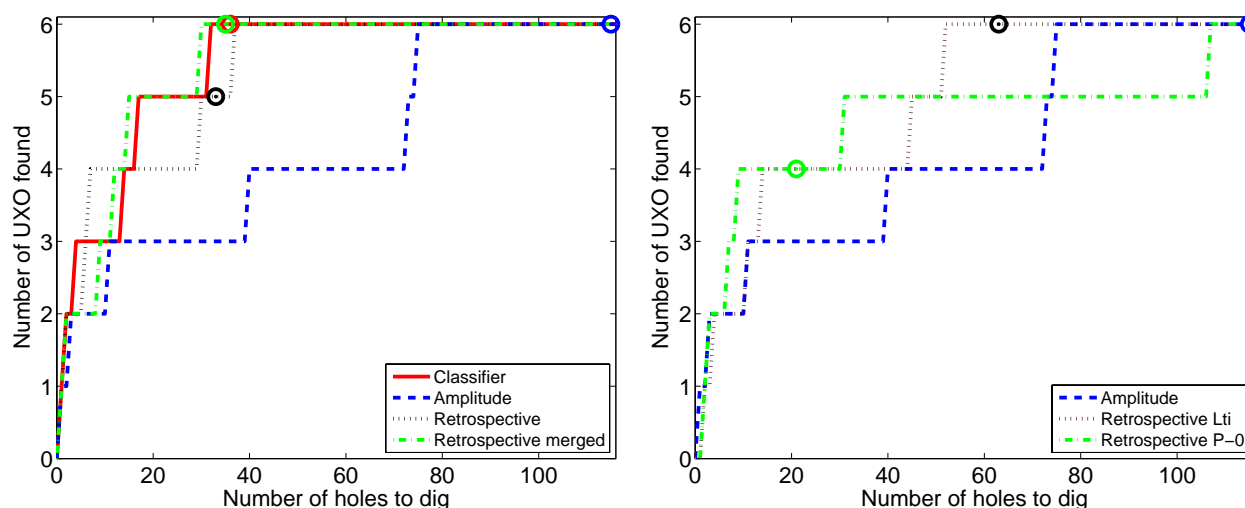


Figure 28. EM-63 ROC curves for 21-14 with ground-truth error corrected for one 37 mm target. Circles show the operating points for each diglist: (on left) “Retrospective” classifier includes time domain information β_1 parameter as an additional feature. “Retrospective merged” combines classifiers for low and high SNR amplitude anomalies; (on right) Comparison of retrospective Pasion-Oldenburg and Lti classifiers.

As digging proceeds, more ground-truth information becomes available and statistical analysis of features is better able to discern which features provide discrimination information for a given classification task. Figure 29(a), shows the contributions of polarization parameters to the first canonical eigenvector. This eigenvector was generated using the ground-truth obtained for Grid 19-14 and considers linear combinations of features which maximize the separation between 20

mm and 37 mm classes. The relative contribution of the decay parameter β_1 is significantly larger in Figure 29(a) than in Figure 20. The results of this canonical analysis are supported by a plot of $\log_{10}(k_1)$ and β_1 (Figure 29(b)). UXO of interest (37 mm projectiles) tend to have larger k values and slower decays (smaller β_1) than small UXO (i.e., 20 mm). However, Figure 29(b) suggests that these parameters provide very little discrimination between 37 mm and ubiquitous small arms (i.e., 50 caliber).

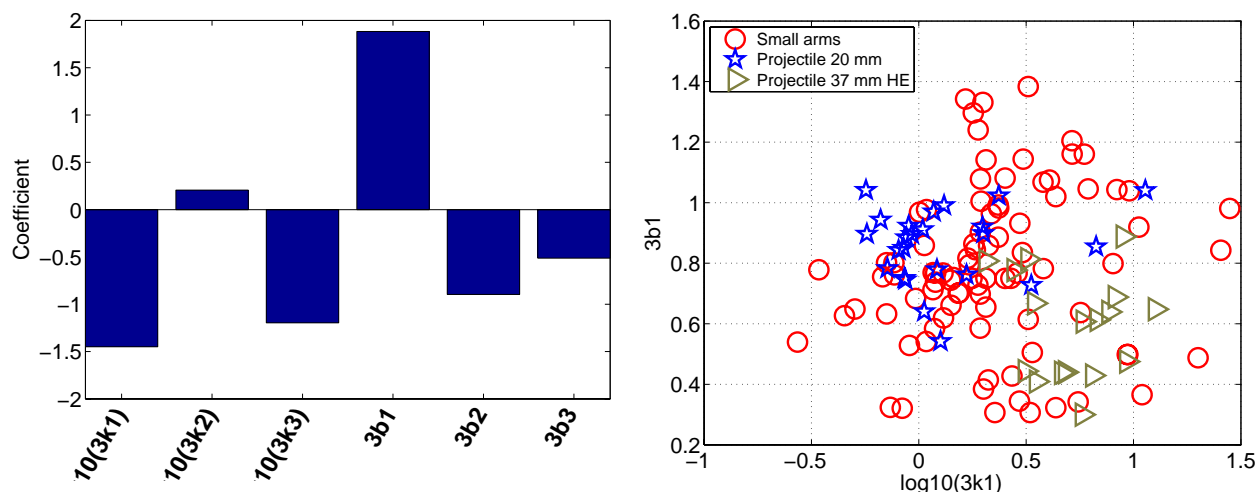


Figure 29. Analysis of features from 19-14: (on left) Contributions to first canonical eigenvector separating 20 mm and 37 mm; (on right)

Given the success of the discrimination strategy in the 20 mm Range-Fan for the first round, we decided to continue with our initial strategy for Grid 21-14. However, on the basis of the canonical analysis of 19-14 data, we retrospectively trained an SVM using $ksum$, β_1 and $misfit/amplitude^2$. The ROC for this classifier is shown in Figure 28(a). There is an initial improvement in retrospective classifier performance with respect to the original classifier, but ultimately there is no significant improvement in performance afforded by the inclusion of decay information. However, closer inspection of the ROCs indicates that three of the first four anomalies identified by both classifiers are high SNR anomalies (>50 mV maximum amplitude at the first time channel). The degradation of the retrospective classifier performance therefore occurs when identifying lower SNR anomalies (<50 mV amplitude), for which the decay parameters are unreliable. This suggests separate classification strategies for high amplitude and low amplitude anomalies. Figure 28(a) shows the resulting ROC when we incorporate decay information in discrimination of high amplitude anomalies and use the original classification strategy for low amplitude anomalies. Here the demarcation between high and low amplitude is chosen to be 50 mV. There is a slight improvement in overall performance with the merged ROC providing the lowest false alarm rate (FAR) for classifiers considered here.

Our submitted diglists for the EM-63 used Pasion-Oldenburg parameters for training classifiers. A question which arises retrospectively is how our choice of model affected classification performance. Figure 28(b) compares retrospective SVM classifiers trained using a P-O 3-dipole

model and an instantaneous amplitude (Lti) 3-dipole model applied to Grid 21-14. The P-O classifier is trained using all k and β parameters (a six-dimensional feature space), while the Lti model is trained using estimated polarizations for the first ten EM-63 time channels (a 30-dimensional feature space). The first ten time channels are used for the Lti model because these parameters were available for all feature vectors (polarizations are not estimated for low SNR channels). The performance of both the P-O and Lti classifiers is worse than the submitted classifier (shown in Figure 28(a)), indicating that our initial choice of a low-dimensional feature space was a good one. Both classifiers in Figure 28(b) have reasonable performance for the first four UXO targets found, but performance degrades considerably when trying to identify the last two UXO. Again, these correspond to low SNR targets for which parameter estimates may be unreliable. In particular, the final 37 mm identified with the P-O classifier has β values which are significantly different from other targets of the same class. This suggests that the explicit parameterization of the P-O model in terms of the decay constant β can degrade classification performance for low SNR targets. In particular, fitting noisy channels can produce highly unreliable estimates of β . In contrast, the Lti model has no explicit decay parameterization and so is less dependent on accurate estimation of a decay parameter which is sensitive to data at all time channels. Furthermore, elimination of noisy late time channels for the Lti model reduces the dimensionality of the feature space. In classification it is generally desirable to have a low-dimensional feature space to avoid the “curse of dimensionality”: the need for the training data size to grow exponentially for accurate estimation of class distribution parameters (Beran, 2005).

As an additional comparison, Figure 28(b) shows the ROC obtained when we use the P-O parameters to compute polarizations and then use the resulting smoothed Lti parameters (for the first ten time channels) to train the SVM classifier. This classifier provides a slight improvement over the P-O classifier, but again the final 37 mm target is difficult to find because its estimated Lti polarizations are significantly different from training 37 mm items.

4.3.5 EM-63 results on the Rocket Range

Results from the first round rocket range grids suggested that the sum of the (log-transformed) k parameters was not a robust indicator of target size for EM-63 data. In particular, if the secondary k s are small then the sum of k s can be quite small, despite a large k_1 value. Note that in this analysis we associate the largest k value with k_1 and assume k_1 corresponds to the axis of symmetry of the target, so that discrimination between prolate ($k_1 > k_2$) and oblate ($k_2 > k_1$) items is no longer possible with the three dipole model. This difficulty is illustrated in Figure 30 which highlights the estimated feature vector for an emplaced MK23 and shows the corresponding observed data for the first time channel. The k_1 value for target 3 is consistent with that of other MK23 targets in this grid. However, the k_2 value is quite small (with $k_2 = k_3$), so that resulting k_{sum} is significantly smaller than that of other MK23s. This example suggests that k_1 , the largest of the k parameters, is a more reliable feature for size-based discrimination and we therefore adopted this parameter for the remaining rocket range grids.

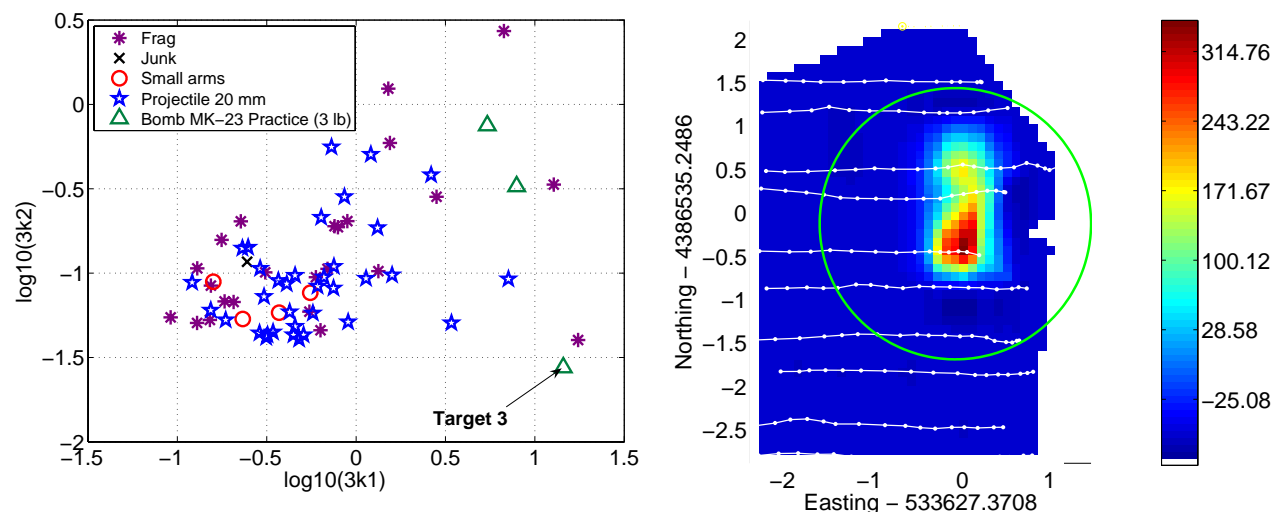


Figure 30. (on left) Feature data for Grid I12, EM-63. (on right) Observed data for Target 3 identified in the left panel.

Merged ROCs for the final round EM-63 classifiers in Grids I-13 and J-12 are shown in Figure 31. The ROC for amplitude thresholding is nearly ideal in this case, though classification still requires far fewer digs. Also shown in this figure is the ROC for a retrospective classifier trained with an additional feature β_I , incorporating time decay information. The retrospective classifier provides a slight improvement over the original classifier, but has some difficulty finding the final MK23 bomb in Grid I-13. This is a double-peaked anomaly which cannot be adequately fit with a single 3-dipole model (Figure 31), suggesting a multi-target scenario (though only a single target was identified in the groundtruth). This item occurs late in both the original classifier and retrospective classifier diglists because its *misfit/amplitude*² is relatively large.

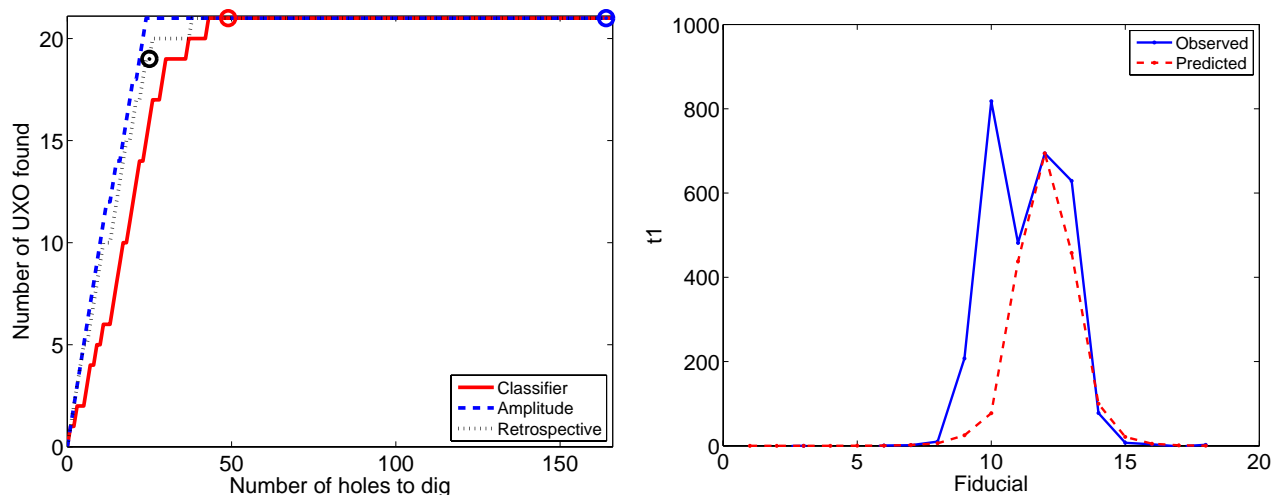


Figure 31. (on left) Merged ROC for Grids I-13 and J-12, EM-63. Retrospective classifier includes time decay information (β_I). (on right) Observed and predicted data for final UXO found in classifier diglists, Grid I-13.

4.4 Summary and discussion of classification results

Statistical classification results in the 20 mm Range-Fan were very positive, outperforming production amplitude thresholding and identifying a number of 37 mm targets which fell below production thresholds. Training and prediction were carried out in a feature space spanned by a measure of target size (the sum of the k parameters or the sum of the instantaneous polarization amplitudes at the first time channel) and a misfit measure weighted by anomaly amplitude (or its square). This latter feature provides a measure of how well we can model the observed data. Weighting by anomaly amplitude at the first time channel decreases the relative contribution of later (low SNR) time channels to the misfit. Inclusion of this feature in the parameter space codifies our preference, that, given two targets with identical size parameters, we choose to dig the target whose data is best fit with our model. This assumption relies on careful QC of data fits: if poor fits for UXO targets are accepted, then these targets will have lower priority in the classifier diglist. This difficulty was encountered for a number of MK23 targets in the rocket range, with retrospective QC indicating that the fits should not have been passed.

Performance of statistical classification in the rocket range was initially poor relative to that of amplitude thresholding. However, switching to $\text{misfit}/\text{amplitude}^2$ for subsequent grids produced comparable performance. Most MK23 were at shallow depths and so amplitude thresholding was a robust means of discrimination in this area. Of course, performance in the 20 mm Range-Fan indicates that classification would perform better had there been a deep target scenario.

Presently, the primary bottleneck in application of statistical classification to real datasets is the QC process. A classifier can only be as good as the features provided by inversion. This point is highlighted by the significant improvement in classification obtained for the 20 mm Range-Fan using retrospective features fit by an expert (see section 6.1).

Reliable QC requires visual inspection of all model fits to ensure that target masking isolates data of interest and that the fit adequately reproduces the data. Full automation of this process is likely neither possible nor desirable. However, automated output of a report summarizing all fit information will expedite QC (and reduce repetitive strain) in future demonstrations.

Another crucial requirement which arises from this demonstration is the need to characterize, as best as possible, the noise on the data and to exclude time channels with low SNR from inversion, especially in the EM-63. Explicit inclusion of time-domain information in a retrospective classifier was shown to slightly improve performance when restricted to high SNR targets. However, the explicit decay parameter in the P-O model can be unreliable when low SNR channels are included in the inversion.

Retrospective comparison of the P-O and instantaneous amplitude models suggested that classification performance can be significantly degraded by using the P-O model. However, the instantaneous amplitude model will likely also show a decrease in performance if polarizations are estimated for late-time, low SNR channels.

Estimation of noise and its propagation through to model uncertainties will also provide a rigorous way to quantify fit quality in classification. Aside from Monte Carlo simulations in Berans (2005), feature vector uncertainties have not been propagated through classifier training and prediction. This is a clear priority for subsequent demonstrations.

The difficulties encountered with noise in the EM-63 and the comparable performance for EM-63 and EM-61 array data indicate that the EM-63 may be best suited for cued-interrogation mode data acquisition. In this mode we can ensure sufficient data coverage and accurately measure sensor position and orientation, allowing the inclusion of time-domain information in a discrimination scheme. In contrast, the EM-61 array permits efficient acquisition of survey mode data and simple discrimination based upon target size.

Receiver operating points (“stop digging” criteria) were selected too early for a number of grids. These criteria were generally chosen as the SVM decision boundary, or in cases where there was overlap between classes, as the halfway point between the decision boundary and the support plane for the clutter class ($f = -0.5$ output by the SVM). Rigorous selection of stop-digging criteria by statistical means has been proposed by various authors (e.g., Zhang *et al.*, 2004) and will be investigated in future work. However, ensuring that all test data lie within the known training distributions is primarily a question of ensuring that parameter estimates are reliable. As outlined above, this depends upon improved QC and improved noise estimation.

UXO targets which are missed by a given classifier will lie in a region of overlap between UXO and clutter classes. It is these targets which are prime candidates for cued-interrogation. A reasonable strategy would be to dig targets based upon EM-61 array data until the overlapping region is reached. This region could be identified as the volume of the feature space between support planes, or by a flattening of the ROC observed thus far. Targets in the overlapping region could then be revisited in cued-interrogation mode and classification based upon cued-interrogation training data could then be applied to decide whether these targets should be dug. The cued-interrogation EM-63 training data could be generated with a test pit setup using targets dug from EM-61 array survey mode data.

This demonstration has provided a first test of classification capabilities in UXOLab and classification functionality has already been improved. Display capabilities have been upgraded and the code has been “modularized” to allow easy addition of features from new models. Capabilities for merging classes (e.g., 20 mm and 50 .caliber items are merged to become a small arms class) and displaying diglists and ROCs have also been written. A major development priority is to streamline classifier training and prediction so that the classification toolbox is friendlier to the novice user.

5 PERFORMANCE ASSESSMENT

Table 11 below shows the identified performance criteria along with an assessment of whether each performance objective was met for each sensor (EM-61 and EM-63) and each site (Rocket Range and 20 mm Range-Fan). Metrics for measuring performance are provided in Table 12, and further information on their calculation is provided in the text that follows.

Table 11. Performance assessment of the EM-61 and EM-63 surveys on the Rocket Range and 20 mm Range-Fan.

Type of Performance Objective	Primary Performance Criteria	Expected Performance (Metric)	Actual Performance (Objective Met?)			
			EM-61 RR	EM-61 20mm RF	EM-63 RR	EM-63 20mm RF
Qualitative	Terrain/Vegetation Restrictions	Operator acceptance for use at the site	Yes	Yes	Met	Met
	Ease of use (Hardware)	Operator and site geophysicist acceptance	Yes	Yes	Partly	Partly
Quantitative	Probability of Detection (Pd) of EM-63 sensor	\geq Pd for EM-61 towed array	NA	NA	Unknown	Unknown
	Probability of discrimination (Pdisc) with a 50% reduction in false-alarms	≥ 0.9	Yes	Yes	Yes	Yes
	False-alarm rate with PDisc = 1	> 25% reduction in false-alarms	Yes	Yes	Yes	Yes
	Location Accuracy of interpreted items	<0.2m	Yes	No	Yes	Yes
	Survey Rate for magnetometer system	1 hectare/day	Yes	Yes	Yes	Yes
	Survey Rate for EM-63 system	1/3 hectare/day	NA	NA	Yes	Yes
	Percent Site Coverage	>95%	Yes	Yes	Yes	Yes
	Processing Time (initial processing)	< 1 day per tile (1 acre)	Yes	Yes	Yes	Yes
	Processing Time (interpretation)	< 5 minutes operator time per anomaly	Unknown	Unknown	Unknown	Unknown
	Accuracy of inversion parameters	Within class variance of cooperative < single inversion	NA	NA	NA	NA

Table 12. Performance criteria and the metrics used for evaluation.

Performance Criteria	Description	Primary or Secondary
Probability of detection for EM-63	(# of EM-61 detected items detected with EM-63) / (# of EM-61 detected items)	<i>Primary</i>
Probability of Discrimination	(# of MEC items detected and recommended for excavation) / (# MEC items detected)	<i>Primary</i>
False alarm rate (FAR)	# of anomalies not corresponding to an ordnance item	<i>Primary</i>
Probability of False alarm (Pfa)	# false positives (i.e. declaration of ordnance) corresponding to clutter/# of opportunities for false positive	<i>Primary</i>
Geo-reference position accuracy	Distance to interpreted items	<i>Primary</i>
Terrain/vegetation restrictions	General qualitative observations on the suitability for the conditions encountered at the test-site	<i>Primary</i>
Survey Rate for magnetometer system	Hectares per day	<i>Secondary</i>
Survey Rate for EM-63 system	Hectares per day	<i>Secondary</i>
Percent Site Coverage	Percentage of area	<i>Secondary</i>
Processing Time (For Initial Processing)	Total minutes of operator time per tile	<i>Secondary</i>
Processing Time (Interpretation)	Total minutes of operator time per anomaly	<i>Secondary</i>
Ease of use (Hardware)	Number of people required to operate sensor and any support equipment. Skill level required by operators. Oversight required by site geophysicist, problems detected by QA officer	<i>Secondary</i>
Accuracy of inversion parameters	Comparison of spread in parameters for a given ordnance class for cooperative versus single inversion	<i>Secondary</i>

5.1 EM-61 Performance Assessment

5.1.1 Terrain/Vegetation Restrictions

Expected performance: Operator acceptance

Actual performance: Met

The EM-61 towed-array that was used for this discrimination study is almost identical to the system used for production survey work at FLBGR. The main differences are (i) five coils instead of three; (ii) use of RTS in-place of the GPS; and (iii) inclusion of crossbow. As such, the terrain/vegetation restrictions are the same as that of the production system. The production system has been used successfully at the site for the past three years and was designed to survey in areas with Yucca plants and moderate terrain. SNR requirements for discrimination are more stringent than for production work. This meant that in areas with significant densities of Yucca plants, the operator had to travel slowly to maintain data quality.

5.1.2 Ease of Use (Hardware)

Expected performance: Operator acceptance

Actual performance: Met

The terrain in the Rocket Range was undulating and selection of an optimal RTS base-station location proved challenging. There were a number of occasions during the first few days of surveying when a small section of a planned survey area proved to be inaccessible and a second RTS setup location was required. By modeling lines-of-sight in a GIS, we were able to minimize the number of RTS setups to cover an area.

Once or twice each day, the Crossbow IMU would undergo a large shock (e.g., as the array was pulled over a large Yucca plant) and become unstable. The operator would then need to stop, reset the unit and wait for it to stabilize. Typically this would cause approximately a 5 minute delay in the surveying.

Apart from the relatively minor issues with the RTS and IMU cited above, the EM-61 towed-array platform was easy to use and was well accepted by our operators.

5.1.3 Probability of Discrimination (P_{disc}) with a 50% Reduction in False Alarms

Expected performance: 90% or greater

Actual performance: Met

Table 13 summarizes the discrimination performance for each phase of the demonstration. We exclude the Phase II results at the Rocket Range as there was only one UXO item found. In addition, we only include anomalies where the inversion returned an acceptable feature vector

and list the “Failed fits” in a separate column. The table lists the false alarms, the number of UXO recovered and the Pd for the selected operating points of the amplitude and classification type methods. Note that we express our discrimination results using Pd (probability of detection) and not Pdisc (probability of discrimination). The difference is that Pdisc only considers performance on anomalies above the selected detection threshold, whereas Pd considers all anomalies for which ground-truth is available (and there is a valid fit to the data). This was done to emphasize that the classification method, at times, ranked low amplitude anomalies quite high in the priority list. We can draw the following conclusions regarding performance from the table:

- (i.) At the 20 mm Range Fan we tended to choose an operating point that was too aggressive and consequently would have missed a number of UXO. However, in each case these extra UXO were included not too much further down the diglist and all UXO were recovered with many fewer excavations than the amplitude based-method.
- (ii.) At the 20 mm Range-Fan the operating point of the amplitude based method was also too aggressive and a number of UXO would have been missed;
- (iii.) For each phase, we meet the objective that $P_{disc} > 0.9$ at the point where we dig 50% less false-alarms than the amplitude based method. In fact, we achieve $P_{disc} = 1$ for each phase;
- (iv.) For each phase, we meet the objective of at least a 25% reduction in false alarms (compared to the amplitude method) at the point where $P_{disc} = 1$. The smallest reduction is 64% with the Phase III rocket range results achieving a 90% reduction in false-alarms.

Table 13. EM-61 discrimination performance results at the 20 mm Range-Fan and Rocket Range.

Phase	# UXO (with valid fits)	AMPLITUDE			CLASSIFIER						
		False alarms (OP)	UXO (OP)	Pd (OP)	False alarms (OP)	UXO (OP)	Pd (OP)	Pd 50% false alarms	False alarms Pd=1	False alarm reduction Pd=1	Failed fits
Phase I 20mm RF	8	42	6	75%	6	6	75%	1	8	81%	28 (0)
Retrospective	8	42	6	75%	8	8	100%	1	8	81%	28 (0)
Phase II 20mm RF	9	59	8	89%	13	6	75%	1	21	64%	34 (1)
Phase I RR	6	84	6	150%	26	6	100%	1	26	69%	13 (2)
Retrospective	6	84	6	150%	24	6	100%	1	24	71%	13 (2)
Phase III RR	20	126	20	100%	36	20	100%	1	13	90%	68 (2)

For Phase I at both sites we show the actual (using *misfit/amplitude*) and retrospective performances (using *misfit/amplitude*²). The false-alarms, number of UXO and Pd at the operating points of the amplitude and classification methods are shown. The failed fits column lists the number of inversions with failed fits with the number of UXO shown in brackets. Pd at

the point with a 50% reduction in false alarms and the reduction in false alarms at $P_d = 1$ are also shown.

5.1.4 False-Alarm Rate with $P_{Disc} = 1$

Expected performance: Greater than 25% reduction in false alarms

Actual performance: Met

Performance met, as per Table 13 and the discussion in the previous section.

5.1.5 Location Accuracy of Interpreted Item

Expected performance: Better than 20 cm

Actual performance: No

On the 20 mm Range-Fan, the locations were biased by 3.3 cm East and 2.5 cm North, with root-mean-square (RMS) positional errors of 20 cm East/West and 26 cm North/South. Therefore, we did not meet the performance specification of 20 cm horizontal location accuracy. All 37 mm and most 20 mm and 50 caliber items were found within 30 cm of the EM-61 inversion location (Figure 32). For the 50 caliber bullets and 20 mm projectiles there is a definite flattening of the cumulative distribution curve at around 30 cm from the EM-61 location. Given the large numbers of these items in the 20 mm Range-Fan, it's possible that the validation crew was selecting a different item than the one modeled by the inversion. It is also possible that the poor positional performance on the 50 caliber bullets and 20 mm projectiles is due to lower SNR (see the retrospective analysis section for further discussion).

For the Rocket Range, the locations were biased by 6.9 cm East and 2.2 cm North with RMS errors of 24 cm East/West and 24 cm North/South. Therefore, we also did not quite meet the performance specification of 20 cm horizontal location accuracy (Figure 32). For the 29 UXO items with valid dipole fits, all but two have predicted locations within 25 cm of the ground-truth locations. The two exceptions include a fit in I-12 that should have been failed and a deep recovered model that was fit to an MK-23 in Grid J-13. Between 75-85% of shrapnel, 20 mm and junk items were recovered within 40 cm of the ground-truth location. For the small-arms, only just over 60% are recovered within 40 cm. This may be due to low SNR, but is almost certainly related to a ground-truth problem, whereby the incorrect object is associated with an anomaly.

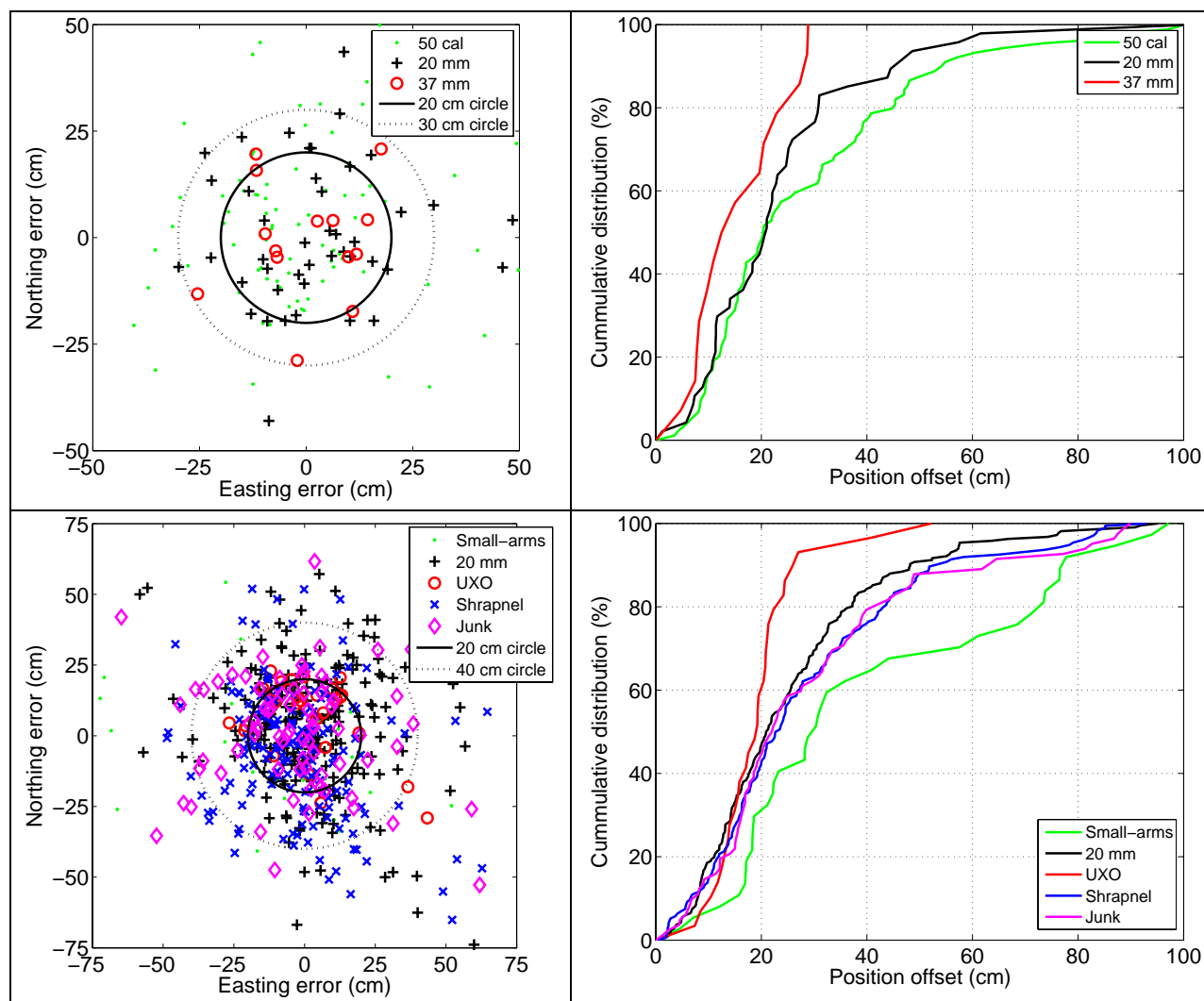


Figure 32. Difference in the ground-truth and inverted EM-61 locations. On the top row, we show the results for the 20 mm Range Fan, split into different object types and include a cumulative distribution curve. Positions were bias adjusted by -4 cm East and -1 cm North.

5.1.6 Percent Site Coverage

Expected performance: Greater than 95%

Actual performance: Met

The EM-61 towed array was able to cover 100% of the eight Rocket Range grids and two 20 mm Range-Fan grids. Within the 45 acre area surveyed as part of the congressional interest project, there were two inaccessible gullies that comprised about 1.5% of the total area. Note that in Figure 10 there are a number of generally linear areas without data coverage. These represent lines of data that were failed during data QC in October 2005. Inclement weather resulted in an early finish to the field season that year before we had a chance to perform recollects. These

were only completed in 2006 and were done with the production 3-coil system. Therefore, those data are not included in the analysis here.

5.1.7 Processing Time (Initial Processing)

Expected performance: Less than one day per grid (1 acre each)

Actual performance: Met

The initial processing for the 45 grids in the Rocket Range (including merging, filtering and target picking) took an analyst five days to complete, with around two days of QC. Thus, the processing proceeds at a rate of about 7 grids per day. This was a bit less than the 10 grids per day we usually achieve with the equivalent system in a production mode at the same site. We attribute some of the difference to unfamiliarity of the data processors with the 5-coil system and the RTS.

5.1.8 Processing Time (Interpretation)

Expected performance: Less than 5 minutes per anomaly.

Actual performance: Unknown

During this first demonstration, we slightly changed our procedures (different masking) and configuration (e.g., different constraints) a number of times, and therefore inverted each anomaly multiple times. In addition, we also made some code changes to both the inversion routines and the statistical classification. This meant that it was difficult to determine how much time was spent interpreting each anomaly, and how much time was spent improving our procedures. With a stable procedure, we believe that the 5 minute/anomaly number is feasible and will endeavor to be more consistent with our time-keeping practices for the next demonstration.

The most time-intensive tasks in interpretation are the visual QC of each and every anomaly along with a second layer of QC when the anomaly is re-inverted with a new mask and/or inversion parameters. Potential multiple anomalies were the most time-consuming of all, as we typically performed a fit with both single and paired targets and compared results.

5.2 EM-63 Performance Assessment

5.2.1 Terrain/Vegetation Restrictions

Expected performance: Operator acceptance

Actual performance: Met

The EM-63 was able to survey all areas on both the Rocket Range and the 20 mm Range-Fan. The cart is relatively bulky and heavy, and some difficulties were experienced when one or both wheels had to be pushed over many of the Yucca plants. The second operator would often need

to help pull the cart over these obstacles. The system would not be suitable for steeper open terrain or areas with more extensive occurrences of Yucca (the standard production EM-61 cart is much more maneuverable and could access more difficult terrain).

5.2.2 Ease of Use (Hardware)

Expected performance: Operator acceptance

Actual performance: Partly

Refer to the EM-61 section for comments on the ease of use of the RTS and IMU. As per the comments on terrain/vegetation, the EM-63 cart can be difficult to maneuver. Due to the low sampling rate and long measurement time, good data can only be collected when the cart is pushed slowly over the survey area (about $\frac{1}{2}$ of normal walking speed, around 2 km/hr). The requirement to collect background measurements every two-hours further reduces the rate of data collection. This slow rate of survey means that it's unlikely that this system would be deemed acceptable for production use at the site.

5.2.3 Probability of Discrimination (Pdisc) with a 50% Reduction in False-Alarms:

Expected performance: 90% or greater

Actual performance: Met

Table 14 lists the performance results for the EM-63 at the 20 mm Range-Fan and the Rocket Rage. This performance metric is met for all phases. The results for Grid 21-14 include two 37 mm projectiles at 40 cm that were below the production threshold. The Pd of the discrimination method includes these two projectiles and indicates that 110 false alarms are required to achieve Pd=1. This is the same number as the number of false-alarms with the production method. The difference is that the production method only recovers 6 UXO at that threshold compared to 8 for classification. To recover those same 6 UXO, the classification method only requires 36 false alarms (thus PDisc = 1 at the point where there is a 50% reduction in false-alarms).

Table 14. EM-63 discrimination performance results at both sites.

Phase	# UXO (valid fits)	AMPLITUDE			CLASSIFIER						
		False alarms (OP)	UXO (OP)	Pd (OP)	False alarms (OP)	UXO (OP)	Pd (OP)	Pd 50% false alarms	False alarms Pd=1	False alarm reduction Pd=1	Failed fits
Phase I 20mm RF	7	107	7	100%	37	7	100%	1	11	90%	28 (1)
Phase II 20mm RF	8	110	6	75%	37	6	75%	0.75	110	0%	55 (2)
Phase I RR	6	84	6	100%	26	6	100%	1	26	69%	37 (2)
Retrospective	6	84	6	100%	24	6	100%	1	24	71%	37 (2)
Phase III RR	21	143	21	100%	28	21	100%	1	22	85%	77 (1)

For Phase I at the Rocket Range we show the actual (using *misfit/amplitude*) and retrospective performances (using *misfit/amplitude*²). For a description of the columns refer to Table 13.

5.2.4 False-Alarm Rate with PDisc = 1

Expected performance: Greater than 25% reduction in false alarms

Actual performance: Met

As per the previous section and Table 14, this performance metric is met for each phase of classification at both the 20 mm Range-Fan and the Rocket Range.

5.2.5 Location Accuracy of interpreted item:

Expected performance: Better than 20 cm

Actual performance: Not met

At the 20 mm Range Fan the inverted locations are biased by 3.1 cm East and 1.1 cm South with RMS errors of 16 cm East/West and 16 cm North/South. At the Rocket Range the bias is 2.8 cm East and 0.8 cm North, with RMS errors of 20 cm East/West and 19 cm North/South. A visual comparison of the inverted and ground-truth locations for the 20 mm Range-Fan and Rocket Range is provided in Figure 33. At the 20 mm Range-Fan, around 85% of 37 mm projectiles are within 20 cm of the inverted location. For 20 mm projectiles and 50 caliber bullets around 75% and 60% are found within 20 cm. This degradation in performance for the smaller items could be due to lower SNR or to incorrect validation information (or both).

5.2.6 Percent Site Coverage:

Expected performance: Greater than 95%

Actual performance: Met

The EM-63 cart system was able to survey ~100% of the area on each of the eight grids on the Rocket Range and the two grids on the 20 mm Range Fans.

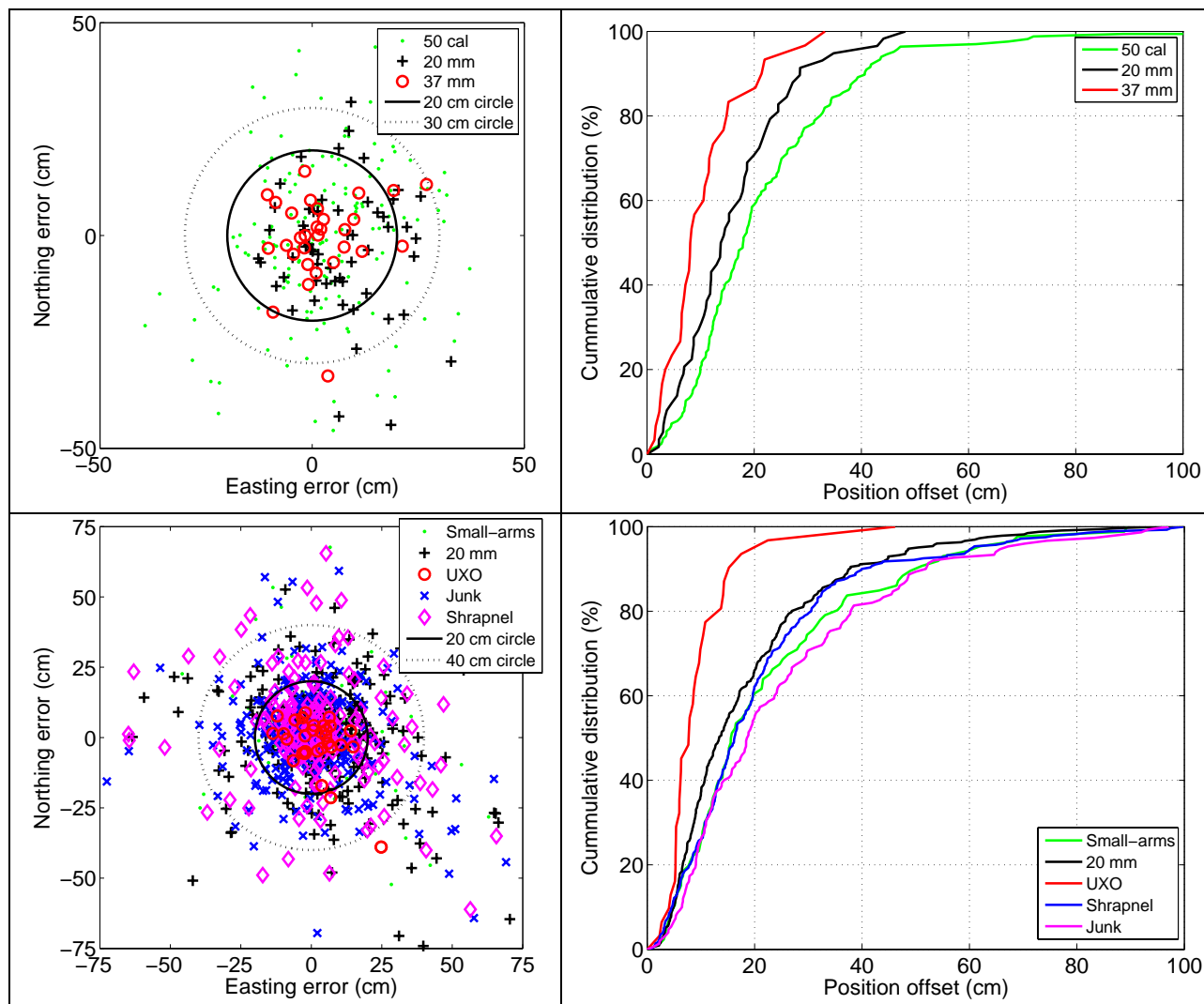


Figure 33. Difference in the ground-truth and inverted EM-63 locations. On the top row, we show the results for the 20 mm Range-Fan, split into different object types and include a cumulative distribution curve. No bias adjustments were applied to the positions.

5.2.7 Processing Time (Initial Processing):

Expected performance: Less than one day per tile

Actual performance: Met

To merge, process, target pick and QC a tile (1 acre) requires between 4-8 hours depending on the number of survey events that make up the tile and the number of calibration events performed each day.

5.2.8 Processing Time (Interpretation):

Expected performance: Less than 5 minutes per anomaly.

Actual performance: Unknown

See the EM-61 comments. An extra complexity with the EM-63 is the need to review both the spatial and temporal behavior of the data and model. In addition, there is about an 8-fold increase in the computational time required to conduct each inversion.

5.2.9 Probability of Detection (Pd) of EM-63 Sensor

Expected performance: Same as EM-61

Actual performance: Unknown

The production amplitude thresholds used for the EM-61 and EM-63 were significantly above the noise floors of both instruments. As such, there were many “detected” targets in each grid that were not selected for validation. Therefore, the validation data are insufficient to determine if one system outperforms the other in terms of detection. This detectability information is probably best obtained from the standardized UXO test-sites.

5.2.10 Accuracy of Inversion Parameters

Expected performance: Cooperative better than single inversion

Actual performance: NA

As we did not perform cooperative inversions, we cannot draw any conclusions regarding this performance metric.

5.2.11 Survey Rate for EM-63 system

Expected performance: Greater than 1/3 hectare per day

Actual performance: Yes

Table 15 lists the number of survey days spent collecting data on each grid, listed in the order the grids were surveyed. Note that there were a number of equipment and mobilization issues in the first week and a half, which explains the longer survey times for the earlier grids. Without that grid, we did achieve an average production rate of 1/3 hectare per day, which increased to about 0.4 hectare per day for the last five grids. Each field day is presumed to comprise a survey window of 8 hours from 8 AM to 4 PM.

Table 15. List of the number of field days required to collect EM-63 data at the Rocket Range and 20 mm Range-Fan.

Grid	J-13	J-12	I12	I-13	19-14	21-14	K-15	L-15	L-14	L-13	Average	Ha/Day
Days	2.5	1.75	1.25	1	2	1	1	1	1	1	1.35	0.30

6 RETROSPECTIVE AND ADDITIONAL ANALYSIS OF PERFORMANCE

In this section of the report, we conduct a retrospective analysis of our performance and also discuss a number of improvements we could make to improve our discrimination strategy.

6.1 Retrospective analysis of the EM-63 results at the 20 mm Range-Fan

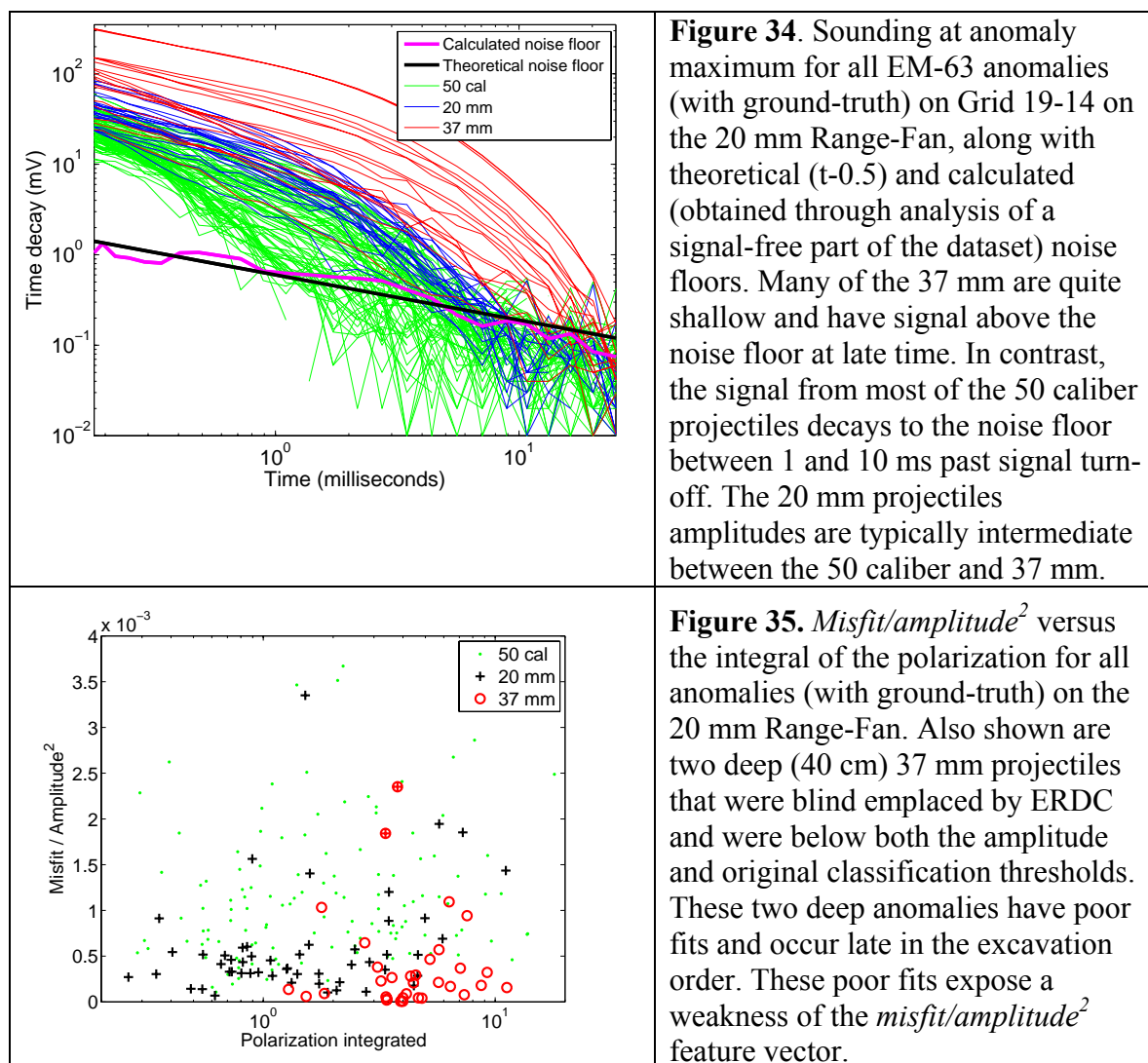
At the 20 mm Range-Fan, EM-63 statistical classification using a combination of a size-based and a misfit based feature vector was relatively effective at discriminating 37 mm projectiles from less hazardous 20 mm projectiles and 50 caliber bullets. In this section we conduct a retrospective analysis concentrating on the following three aspects of the results:

- (i.) Why was the *misfit/amplitude*² feature vector effective for this demonstration and what limitations are there in using it in other situations?
- (ii.) Why was it more difficult to tell the difference between 50 caliber bullets and 37 mm projectiles, than it was for 20 and 37 mm projectiles?
- (iii.) Why were feature vectors based on time-decay information not able to improve the EM-63 discrimination performance?

Why was the *misfit/amplitude*² feature vector effective for this demonstration and what limitations are there in using it in other situations?

As shown in the results section, the *misfit/amplitude*² feature requires vigilant quality control to prevent false negatives. The metric provides a relatively effective measure of the goodness of fit when the region to invert is of appropriate size, there are no other anomalies in the field of view and the best-fitting model is recovered by the inversion process (as opposed to convergence to a sub-optimal model). As the burial depth increases, the anomaly amplitude decreases and hence the *misfit/amplitude*² metric will tend to increase. Thus, deeper anomalies will be more likely to be classified as non-UXO. Figure 34 plots the sounding at the anomaly maximum for each of the items with ground-truth in the 19-14 grid along with theoretical and calculated ($t^{-0.5}$; Munkholm & Auken, 1996) noise floors. Most of the 37 mm projectiles have large amplitudes due to their large size and proximity to the surface. The 20 mm and 50 caliber projectiles, on the other hand, tend to have smaller amplitudes. Thus, one reason for the success of the *misfit/amplitude*² metric is that the 37 mm projectiles were relatively shallow. This is an *extrinsic* attribute and one that should clearly be less relied upon than intrinsic attributes like the object's polarizability. In fact, when two deep (40 cm) emplaced 37 mm projectiles² are included in the analysis (Figure 35), the classification performance degrades significantly. On the 21-14 grid, these two deep items are ranked as priorities 74 and 115 by the classifier (the next worst case anomaly is at rank 39).

² They were blind SEED items emplaced by ERDC and ranked below the excavation cut-offs in both the production and discrimination diglists



To conclude, the $misfit/amplitude^2$ feature vector is likely to be effective when the anomalies of interest have significant signal above the system noise-level. There is an increased chance of a misclassification as the burial depth approaches the maximum detection depth.

Why was it more difficult to tell the difference between 50 caliber bullets and 37 projectiles, than it was for 20 and 37 mm projectiles?

The validation results for 19-14 and 21-14 indicated that it was more difficult to tell the difference between 37 mms and 50 caliber items than between 37 mms and 20 mms. This at first glance is counterintuitive as the 20 mm projectiles are significantly larger than the 50 caliber projectiles and tend to have larger amplitudes as we saw in Figure 35. However, by inspection of Figure 36 we see that there are a large number of 50 caliber bullets with polarization integral comparable to that of the 37 mm projectiles. In contrast, there are relatively few 20 mm

projectiles with such a large polarization integral. To understand the reason for this discrepancy, we start with a comparison of the SNR of the different items (Figure 36). This is defined as:

$$SNR = \frac{\sum_{n=1}^N d_n^2 - N\sigma^2}{N\sigma^2}, \quad (16)$$

when there are N data points, d_n , and the noise has a variance of σ^2 . The 50 caliber projectiles tend to have lower SNR than the 20 and 37 mm projectiles. Additionally, fewer time-channels were used to constrain the 50 caliber inversions due to very low SNR at later times.

Next we compare the estimated depths of each item with the actual depth recorded during anomaly validation (Figure 37). There is generally very poor agreement with most fits pushed deep, particularly for the 50 caliber projectiles. Most of them are found within 10 cm of the surface yet their fitted depths are spread-out down to 80 cm. There is a clear correlation, for each object type, between the polarization and the error in depth recovery (Figure 37b). As the depth error increases the recovered polarization increases and results in four-orders-of-magnitude difference in the estimated size of the 50 caliber items. The error in depth recovery for the 20 mm projectiles is less and results in “only” two-orders-of-magnitude difference in the size. Consequently, the size distribution of the 50 caliber bullets overlaps more with the 37 mm projectiles, than is the case for the 20 and 37 mm projectiles. Thus, it is the lower SNR of the 50 caliber bullets and a consequent inability to constrain the object depth that makes 50 caliber bullets harder to discriminate than 20 mm projectiles.

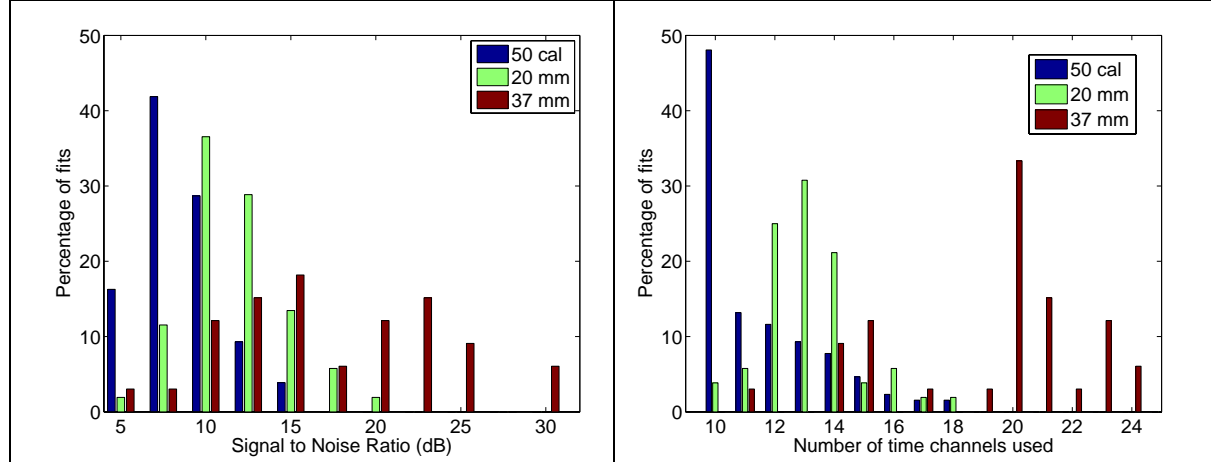


Figure 36. Left panel: Histogram of SNRs for 50 caliber, 20 mm and 37 mm anomalies on the 20 mm Range-Fan. The SNR is estimated via the data within the mask defined around each anomaly. The 37 mm anomalies typically have the highest SNR which is one reason that causes the *misfit/amplitude*² criterion to work well at this site. Right panel: Histogram showing the number of time-channels used in the inversion for the three different ground-truth items.

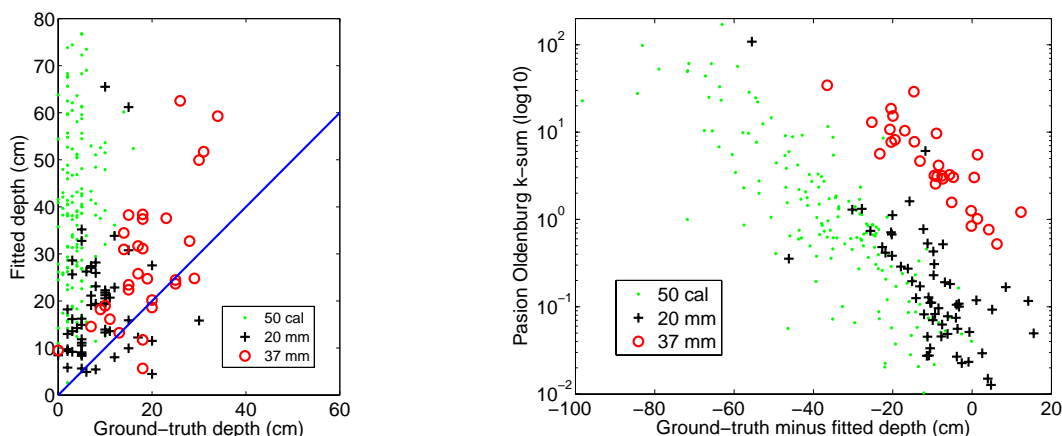
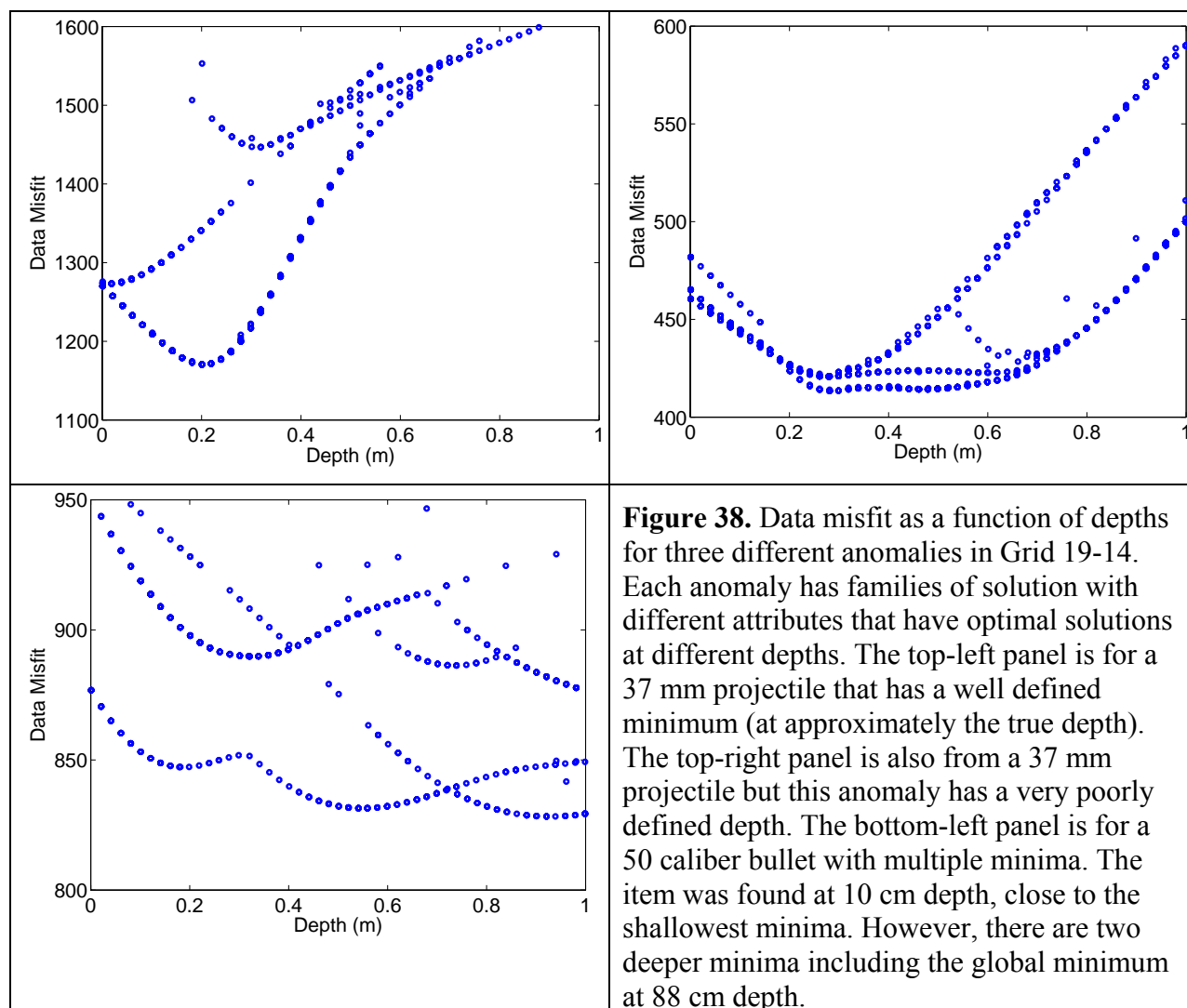


Figure 37. Fitted versus actual depth for 50 caliber, 20 mm and 37 mm projectiles at the 20 mm Range-Fan (on left). Error in depth estimation versus the estimated object size (on right).

In an attempt to understand this tendency to push items deep, we investigated the change in the least-squares misfit as a function of depth (Figure 38). We solved for the best fitting polarization model with the depth fixed at a certain value (red curves in Figure 5). We also constrained the polarization parameters using various criteria (e.g., $k_1 > k_2$, then $k_1 < k_2$), with the depth again fixed, and then solved for the best fitting model (blue circles in Figure 38). This produced different “families of solutions” that had optimal depths at (typically) different values. On a high SNR 37 mm anomaly there was a single well-defined minimum, with depth, at about the correct value of 20 cm (Figure 38a). In that case, the recovered polarization parameters provided a good estimate of the object’s size. On another 37 mm anomaly with lower SNR (Figure 38b), the misfit surface initially decreased with depth, then remained at about the same level from 25 to 55 cm depth, before increasing again. For this anomaly there is very poor depth resolution and a large uncertainty in both the recovered depth and polarization parameters. For a 50 caliber bullet with low SNR, the misfit surface has three local minima as the depth is increased (Figure 38c). The last and deepest minima has the minimum misfit and hence would be selected as the solution of the inverse problem. Thus, it appears that either a flat misfit surface, or multiple local minima with depth, creates ambiguity in the dipole depth and hence in the object’s estimated size. We do not, as of yet, know why the deeper solutions tend to fit better than the shallower solutions.



Why were feature vectors based on time-decay information not able to improve the EM-63 discrimination performance?

The EM-63 measures twenty-six time-gates from 180 μ s out to 25 ms compared to four time-gates from 216 μ s to 1 ms for the EM-61. Despite this ability to measure later in time, the performance of the EM-63 was no better than that of the EM-61. In fact, no feature vectors based on the time-decay properties of the polarizability were found to provide useful discrimination ability. From inspection of Figure 35, it is apparent that this is likely due to the faster rate of decay of the signal compared to the noise. There is very little signal at later time to constrain the decay of the polarizability, especially for the 50 caliber bullets where over half of the inversions were carried out using less than 12 of the available 26 time-channels (Figure 37). Consider the polarizability curves obtained by fitting test-stand data over 20 and 37 mm projectiles (Table 16). The axial-polarization of the 20 mm has a late-time decay constants of 3.4 compared to 4.25 ms,

for the 37 mm. From Figure 35, there should still be some signal out to 5-10 ms and therefore there should be enough information to discriminate 20 mm from 37 mm projectiles. On closer inspection of the recovered feature vectors, we found that many of the solutions for the 50 caliber bullets and 20 mm projectiles lay on the user-defined constraints. We therefore, re-inverted all the anomalies using a two-dipole model, and then selected the polarization curve with the largest integrated polarization,

$$L_i = \int_{.18}^{25} k_i t^{-\beta_i} \exp(-t / \tau_i) dt \quad (17)$$

where the time is in milliseconds. Note that during the re-inversion, we did not replace the Pasion-Oldenburg model with a power-law decay as we tended to do during the main demonstration.

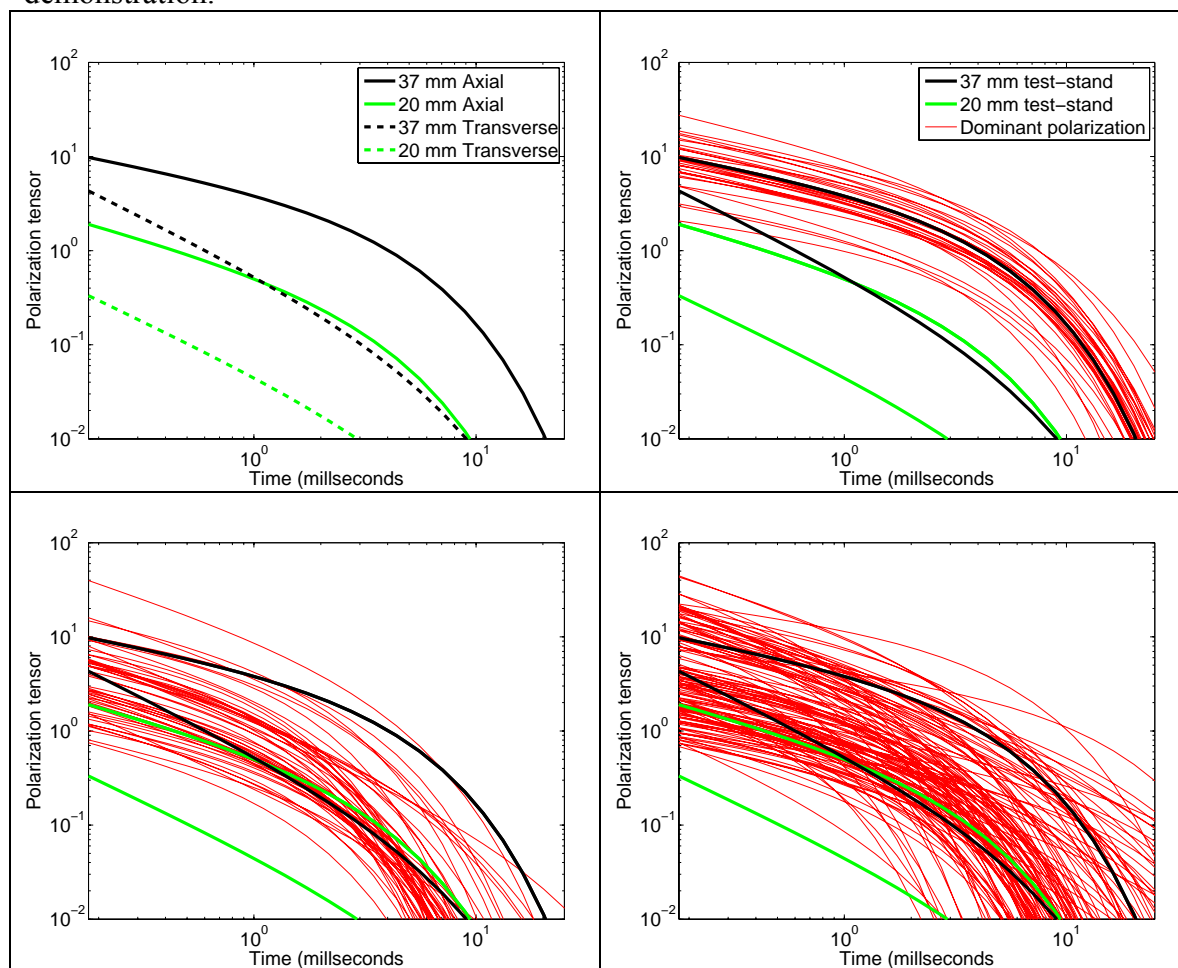


Figure 39. Polarization tensors recovered by inversion for all anomalies on the 20 mm Range-Fan. Top left panel: Shows the axial and transverse polarization tensors fit to test-stand data over 20 and 37 mm projectiles. Top right panel: Dominant polarizations for 37 mm projectiles. Bottom left panel: Dominant polarization for 20 mm projectiles; Bottom right panel: Dominant polarizations for 50 caliber projectiles.

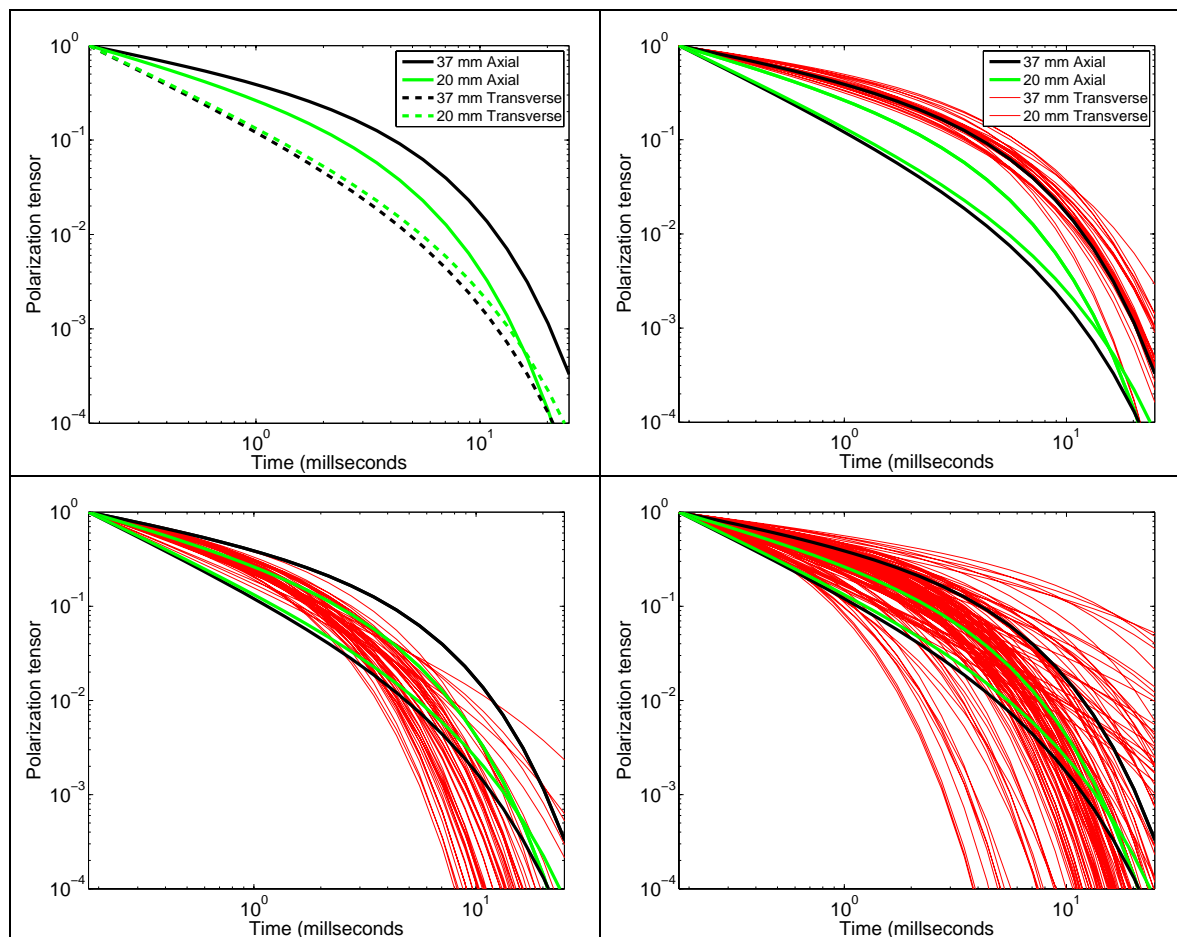


Figure 40. Normalized polarization tensors recovered by inversion for all anomalies on the 20 mm Range-Fan. Top left panel: Shows the axial and transverse polarization tensors fit to test-stand data over 20 and 37 mm projectiles. Top right panel: Dominant polarizations for 37 mm projectiles. Bottom left panel: Dominant polarization for 20 mm projectiles; Bottom right panel: Dominant polarizations for 50 caliber projectiles.

Table 16. Polarization parameters recovered from test-stand data collected over 20 and 37 mm projectiles.

	k_1	β_1	γ_1	k_2	β_2	γ_2
37 mm	4.84	0.44	4.25	0.62	1.15	5.67
20 mm	0.67	0.64	3.4	0.0519	1.1	6.28

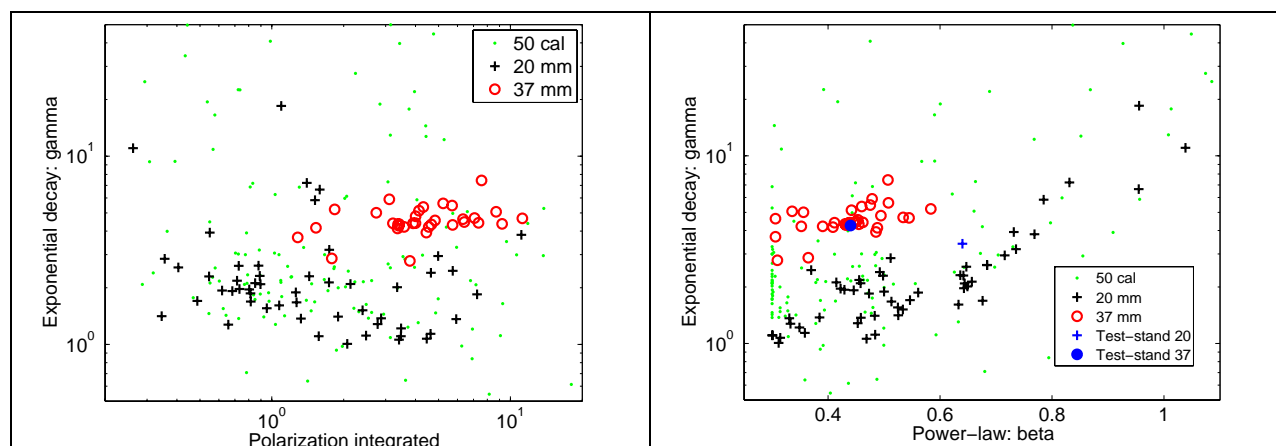


Figure 41. Recovered Pasion-Oldenburg parameters of the dominant polarization tensor for grids 19-14 and 21-14 in the 20 mm Range-Fan. Left panel: Gamma parameter version the integral of the polarization (indicative of size). Right panel: Gamma versus beta parameters.

In Figure 39, we plot the polarization curves obtained by inversion of test-stand data over 20 and 37 mm projectiles along with the dominant polarization curves for the 50 caliber bullets and the 20 and 37 mm projectiles. As noted in previous discussions there is considerable variation in the magnitude of the recovered polarizations due to the inability of the data to constrain the object depth (especially for the 50 caliber projectiles). However, when the polarization curves are normalized (by dividing by the magnitude of the polarization of the first-time-channel), there is much better separation between the different objects (Figure 40). The dominant polarizations for the 37 mm projectiles tend to follow the axial-polarization curve for the test-stand 37 mm. There is more spread at later times than one would expect due to the lower SNR at late time. For the 20 mm projectiles, the dominant polarization generally lies between the transverse and axial-polarizations from the test-stand 20 mm, out to about 5 ms. At later times, the SNR is too low to further constrain the polarization curve. For the 50 caliber bullets, the normalized polarization curves show a wide variation. Clearly, the SNR is not sufficient to constrain the decay parameters of these small objects.

When we now investigate a feature space comprising object size (integrated polarization) and two decay characteristics (beta and gamma), we can see that there is much better separation between the three classes (Figure 41). The beta and gamma parameters for the 37 mm projectiles cluster around the values obtained from the test-stand. For the 20 mm projectiles the beta/gamma parameters do not agree well with the test-stand values, but tend to be different to the 37 mm projectiles. For the 50 caliber bullets, there is a large spread in the beta/gamma parameters, with some overlap with the distribution of values for the 37 mm projectiles. If we now use all the feature vectors from 19-14 as training data and apply the resulting SVM classifier to 21-14 we obtain the ROC curve in Figure 42. At the operating point where all 37 mm projectiles are recovered, there are 21 items, comprising seven 37 mm projectiles, three 20 mm projectiles and

eleven 50 caliber bullets. The seven 37 mm projectiles include the two deep (40 cm) items emplaced by ERDC that were previously ranked as anomalies 74 and 115. In addition, the classification scheme does not have to rely on the problematic *misfit/amplitude*² feature vector. An alternative classifier trained on the ratio of the first polarization at the first and twentieth time channels, and the integral of the first polarization is also shown in Figure 42. The performance of both classifiers is comparable.

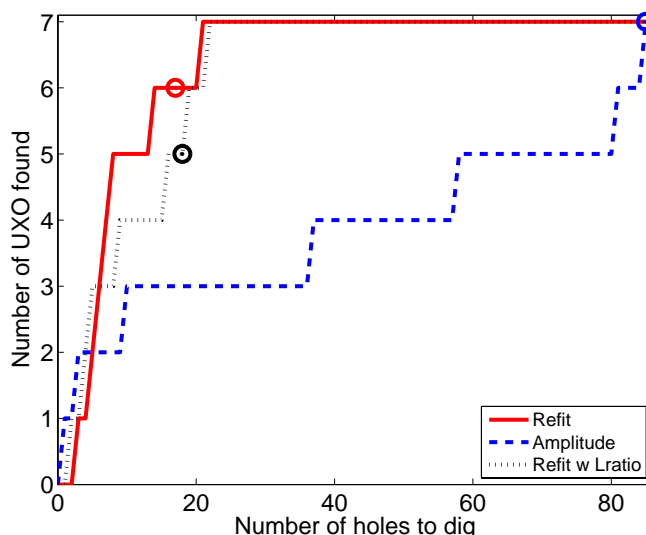


Figure 42. ROC curve for Grid 21-14 using a classifier trained on beta, gamma and polarization integral feature-vectors obtained from Grid 19-14. Classifier operating point at $f=0$ of the SVM classifier are shown as circles.

There was just enough SNR to obtain a relatively sensible estimate of the decay-rate of the dominant polarization curve of the 20 and 37 mm projectiles, but not for the 50 caliber bullets. This dominant polarization curve contained more discriminatory information than a data-based feature such as the decay characteristics of the sounding with maximum amplitude (Figure 43). The normalized soundings for the 20 and 37 mm projectiles generally fell between the axial- and transverse polarization curves obtained from test-stand data. This is to be expected as any sounding is comprised of a linear combination of the axial and transverse polarizations. However, depending on the degree of excitation of each component, there was considerable overlap in the decay characteristics of the different objects. This means that the dipole model (or at least some model) is an essential tool for discrimination at this site.

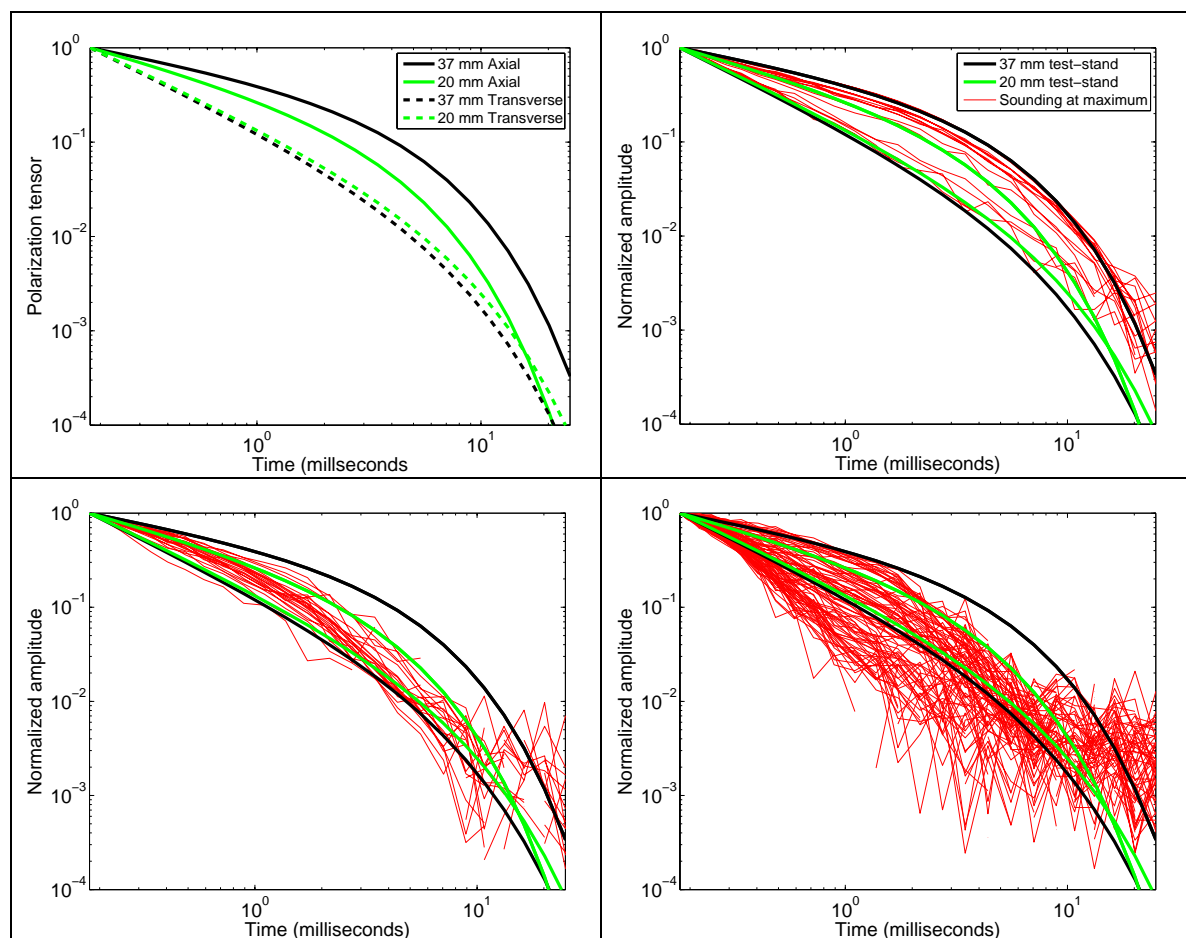
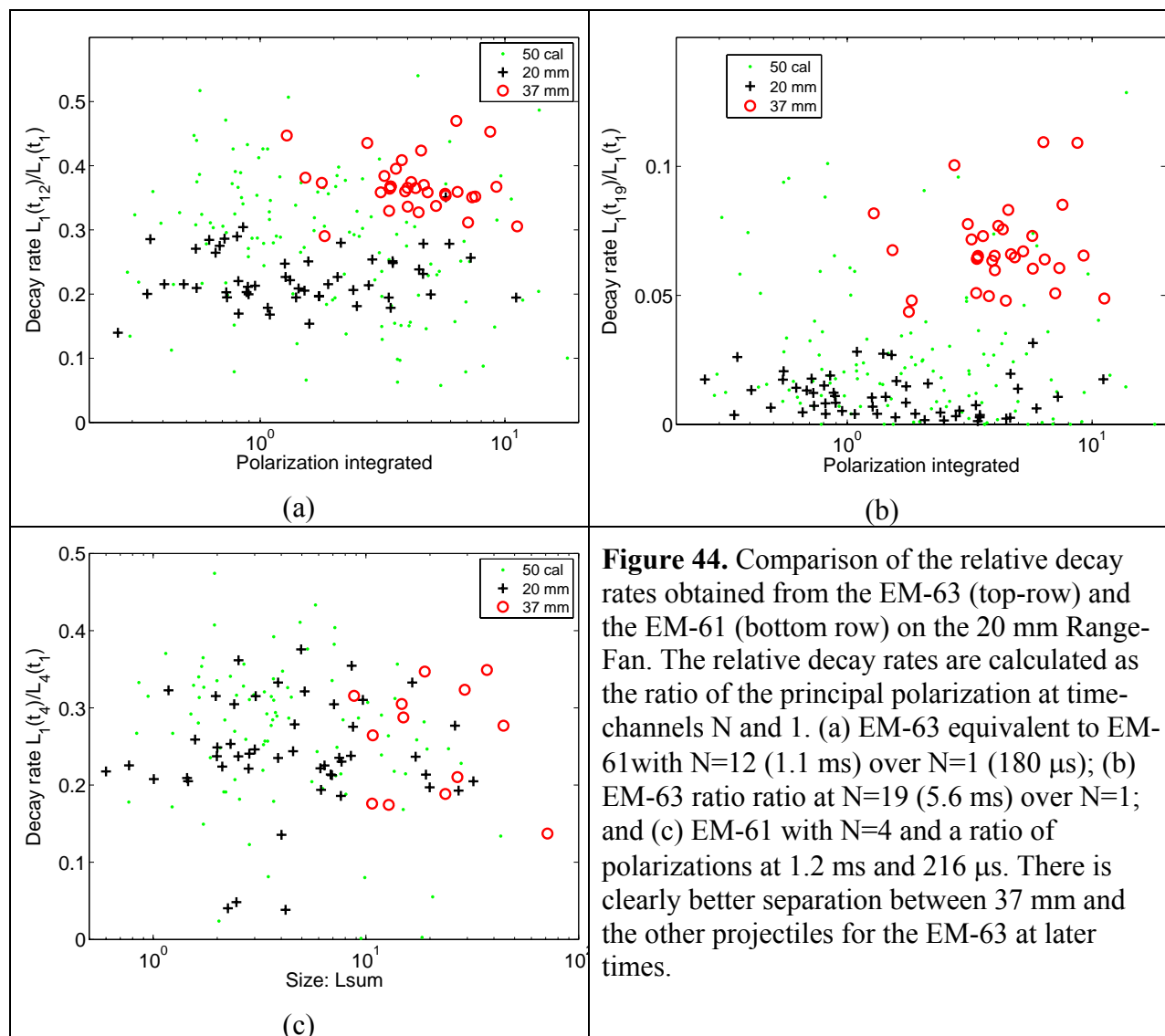


Figure 43. Normalized decay curves for the sounding with maximum amplitude on Grid 19-14. Top left panel: Shows the axial and transverse polarization tensors fit to test-stand data over 20 and 37 mm projectiles. Top right panel: Sounding at anomaly maximum for 37 mm projectiles. Bottom left panel: Sounding at anomaly maximum for 20 mm projectiles; Bottom right panel: Sounding at anomaly maximum for 50 caliber projectiles.

6.2 Retrospective analysis of the EM-61 data at the 20 mm Range-Fan

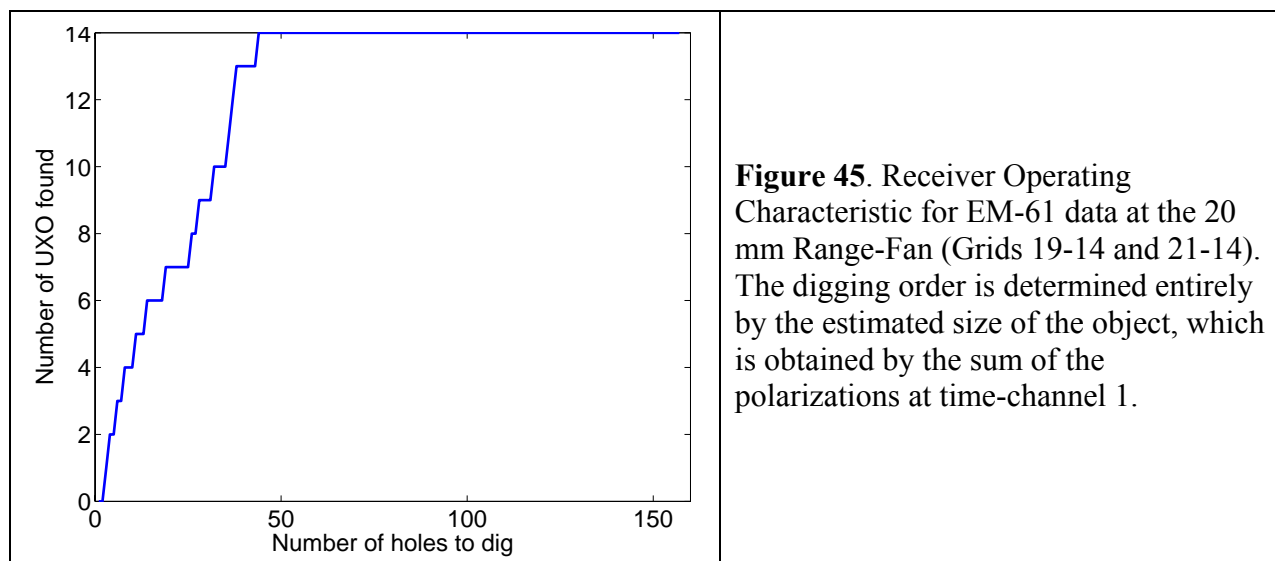
For the EM-61 data on the 20 mm Range-Fan, we performed an equivalent retrospective analysis as was done with the EM-63. We concentrated, in particular, on the decay rate information that could be extracted from the EM-61. The fourth and final time-channel in the EM-61 is at 1.2 ms which is a lot earlier than the 25 ms range of the EM-63. Without information at later times it might prove difficult to distinguish 20 mm from 37 mm projectiles. In Figure 44, we plot the ratio of the principal polarization at time-channels 4 and 1 against the sum of the polarizations at time-channel 1. By inspection of the plot it is evident that the decay-rate information does not improve the separation between the three different classes of items. We also show the equivalent plot for the EM-63 (time-channel 12 is at 1.1 ms) along with a plot that includes later time

information (out to time-channel 19 at 5.6 ms). For the EM-63, there appears to be good separation between 20 mm and 37 mm even at 1.1 ms. As a parameterized version of the polarizability was used for the EM-63, the later time-channels are most likely contributing information to constrain the polarization at earlier times. Thus, we cannot conclude that if the EM-63 stopped at 1.1 ms that it would produce better results than the EM-61.



The retrospective analysis indicates that time-decay information does not enhance the discrimination ability of the EM-61 in this scenario. Therefore, if we wanted to avoid using the *misfit/amplitude*² feature, we would need to rely solely on the size of the polarizability. If we prioritize the diglist by size and dig the largest items first we obtain the ROC curve in Figure 45. This curve includes data from both 19-14 and 21-14 and indicates that 30 false-alarms (twenty-

one 50 caliber bullets, nine 20 mm projectiles) would be required to find all fourteen 37 mm projectiles. This is just one more false-alarm than the SVM classification that used size and $\text{misfit}/\text{amplitude}^2$, and is significantly less than the 93 false-alarms at $\text{Pd}=1$ for the amplitude based method.



6.3 Additional analysis of the EM-61 results on the Rocket Range

In the performance assessment section we analyzed the accuracy of the location estimates for the Rocket Range data (Figure 33). We found that many items were recovered with 40 cm or more positional error. In Figure 46a, we extend that comparison to fitted depths, where it is evident that there is a relatively poor agreement. As with the 20 mm Range-Fan results, there is a tendency to push items deep and return a larger estimate of the polarization tensor (Figure 46b). However, even with this uncertainty in the size, the MK-23s are relatively well separated from the other items on the basis of the first time-channel of the principle polarization, $L_1(t_1)$. If we were to order our diglist such that the largest items are dug first, we return the ROC curve in Figure 47 and find that 40 false-alarms are required to recover 29 MK-23s. This ROC curve includes all anomalies with valid inversion fits on all eight Rocket Range grids. The performance of the size-based method is similar to amplitude-based thresholding, which required 50 false-alarms to recover the same 29 MK-23s. To put those results in context recall that almost all MK-23s were found close to the surface (< 10 cm) and hence had relatively large amplitudes and high SNR. Any deeper UXO would have lower amplitudes and SNR and we expect that the size-based method would start to more significantly outperform the amplitude based method. Of course, due to the absence of any deep items we can only speculate that this would be the case.

The performance of the size-based metric is good, thus obviating the need to use the $\text{misfit}/\text{amplitude}^2$ feature vector. However, keep in mind that during the demonstration, we were

searching for all UXO with calibers of 37 mm and greater. At that time, we did not know that we would only encounter the larger items and hence could rely entirely on a size-based metric. The red circles on Figure 47 show the items closest to $L_1(t_1) = 10.25$ and 3 (holes 132 and 347 respectively) which corresponds to the geometric mean and smallest polarizations recovered for a 37 mm at the 20 mm Range-Fan. Thus, if we needed to include items as small as a 37 mm, we would be forced to dig many more holes at the Rocket Range.

The training data available during the demonstration did not support the use of a shape- or time-based feature vector in the classification. Retrospectively, we plot a *shape feature*, $L_2(t_1)/L_1(t_1)$ against object size (Figure 48a). Most of the MK-23s occur between $L_2(t_1)/L_1(t_1) = 0.08$ to 0.22 which would separate them from a lot of the 20 mm, shrapnel and junk which tend to have larger values. However, there are a number of MK-23s with $L_2(t_1)/L_1(t_1) = 0.25$ to 0.4 and when these are included, it is evident that shape is not a good discriminator. A *time-decay feature*, $L_1(t_4)/L_1(t_1)$, holds more promise with all but one MK-23 within the range 0.14 to 0.2 (Figure 48b). The time-decay fit to the one outlier is shown in Figure 49. It is apparent that the data at time-channel 1 have been under-fit and consequently the ratio $L_1(t_4)/L_1(t_1)$ is over-inflated. By modifying the relative weights on time-channels 1 to 4 and refitting, the time-decay ratio decreases from 0.26 to 0.19 , which is within the range of the other MK-23s (the feature vectors of the two fits are joined by the red line in Figure 48b). The 20 mm projectiles tend to have $L_1(t_4)/L_1(t_1) > 0.2$, while a lot of the junk have $L_1(t_4)/L_1(t_1) < 0.1$, indicating a rapid decay of signal. Many of the items identified as junk are wire or thin-walled scrap and hence we expect such a rapid time-decay (e.g., Bell *et al.*, 2001).

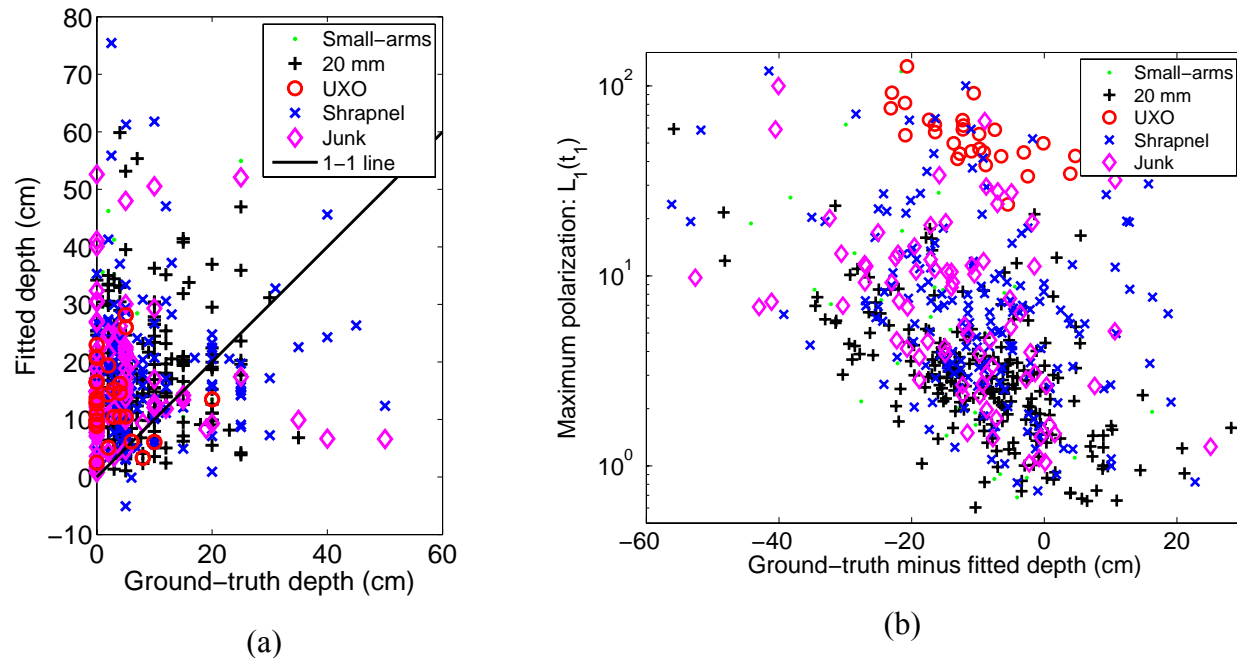


Figure 46. (a) Fitted EM-61 dipole depth versus actual depth for different objects at the Rocket Range. (b) Error in depth estimation versus the estimated object size. Calculations were made assuming the EM-61 coils were 25 cm above the ground-surface.

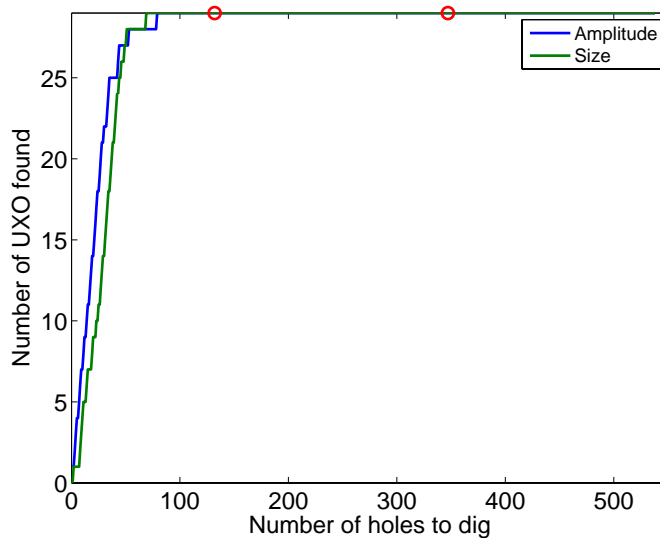


Figure 47. EM-61 ROC curves for amplitude and size-based methods applied to all eight grids at the Rocket Range. The red circles are for the items with $L_1(t_1) = 10.25$ and 3 (holes 132 and 347 respectively) which corresponds to the geometric mean and smallest polarizations recovered for a 37 mm at the 20 mm Range-Fan.

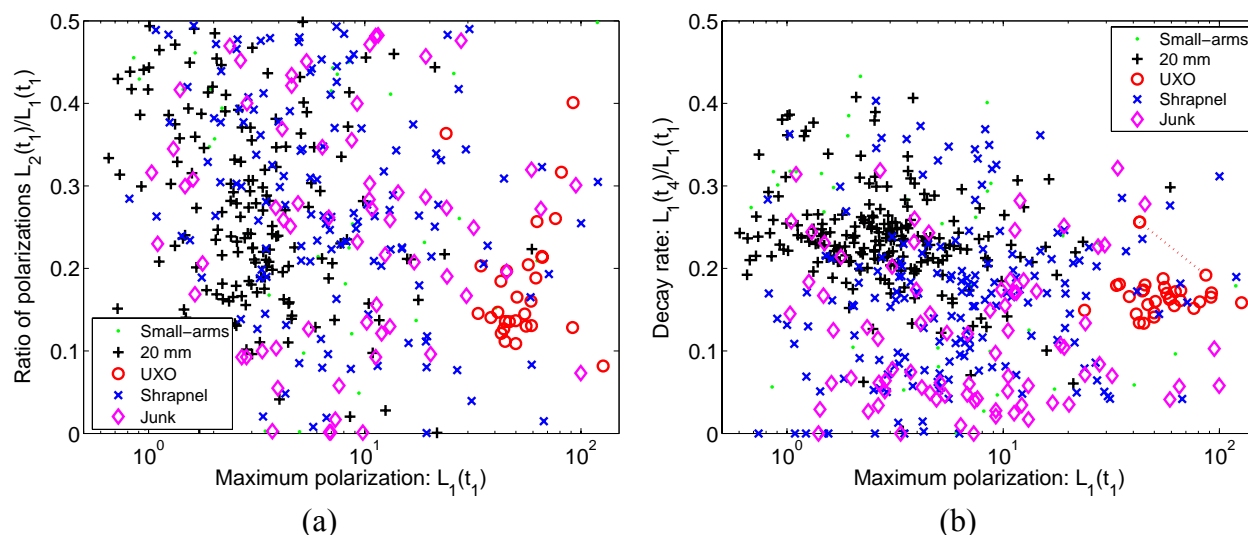


Figure 48. Comparison of instantaneous polarization parameters recovered from EM-61 data at all eight rocket range grids. (a) A plot of a *shape feature*, $L_2(t_1)/L_1(t_1)$ against a size feature, $L_1(t_1)$. (b) Plot of a *time-decay feature* $L_1(t_4)/L_1(t_1)$ against the same size feature, $L_1(t_1)$. The MK-23 bombs tend to be large and have decay parameter between 0.15 and 0.2. See text for a discussion of the two MK-23 features joined by the dashed red line.

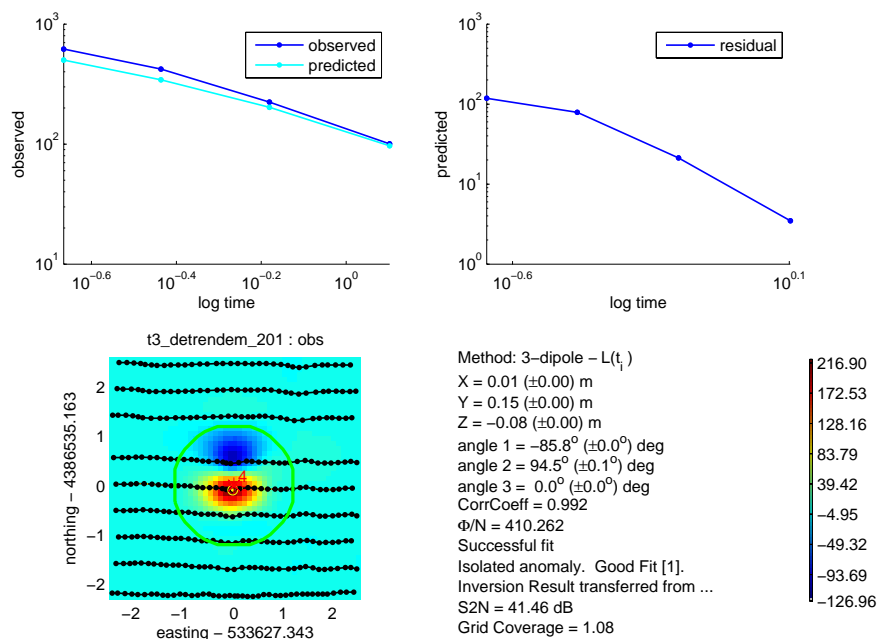


Figure 49. Original inversion fit to the MK-23 item identified in Figure 15 and discussed in the text. Plan view of the data at time-channel-3 and a comparison of the modeled and actual sounding at the anomaly maximum. The model returns a low-estimate of the response at time-channel-1 which causes an overestimate of the ratio $L_1(t_4)/L_1(t_1)$.

6.4 Additional analysis of the EM-63 results on the Rocket Range

Our final retrospective analysis concerns the EM-63 feature vectors and ground-truth at the Rocket Range. Figure 50a plots the fitted and ground truth depths for different classes of items on the Rocket Range, while Figure 50b plots object size (principal polarization k_1) versus the error in depth recovery. Results are similar to those presented previously for both the EM-61 and EM-63 at the 20 mm Range-Fan and Rocket Range (for the EM-61). That is, there is a tendency to push the items deep which results in an over-inflated estimate of the size of the polarization tensor.

Figure 51a plots a *shape-based* feature, k_2/k_1 , against the size of the principal polarization (k_1). Most of the MK-23s have values between 0 and 0.2 but there are a two items with values around 0.4. There is therefore considerable overlap with the other object types and this feature is unlikely to improve the classification performance. Figure 51b plots an *asymmetry* feature, $(k_3 - k_2)/k_2$, against the size of the principal polarization (k_1). Two-dipole models have asymmetry values of zero and 22 of the MK-23s were fit with the two-dipole model. There are seven MK-23s that were fit as three-dipole models and some of these have very large asymmetry values. From Figure 51 we conclude that shape or asymmetry features would not aid discrimination ability. It's clearly not a signal-to-noise issue that precludes the use of these features as almost all the rounds with large shape or asymmetry metric were at or near the surface. This suggests that it is either positional error or near-field non-dipolar effects that are the cause of the problems with these features.

We next investigate the usefulness of time-decay parameters derived from the EM-63 (Figure 52). We select the largest polarization, and then investigate the decay characteristics of the corresponding Pasion-Oldenburg decay curve: $L_I(N)/L_I(1)$. In Figure 52a, we plot the EM-61 equivalent decay characteristic, $N=12$ (1.1 ms) over $N=1$ (180 μ s). There is some separation between the MK-23s and the other items. This separation is more significant at both $N=15$ (2.17 ms) and $N=19$ (5.6 ms) as shown in Figures 52b and c. From the EM-61 decay characteristics on the Rocket Range, there were a significant number of objects identified as junk that had more rapid decay than the MK-23s (Figure 48b). This is not evident from the EM-63 decay, potentially because of the Pasion-Oldenburg re-parameterization process we used (see Section 4.3.5 on feature extraction). During this process, the exponential term was removed from the analysis and a power law-model was fit to the data, generally with only the first 10 to 15 time-channels used to constrain the fit. With the missing exponential decay, the polarization models don't decay rapidly at late-time and we return an over-inflated estimate of the relative decay rate. This suggests that some of the class separation evident in Figure 52 may be due to the characteristics of the model used, rather than something intrinsic in the data. To determine if this were the case we re-inverted each anomaly solving for both a power-law (beta) and exponential (gamma) term.

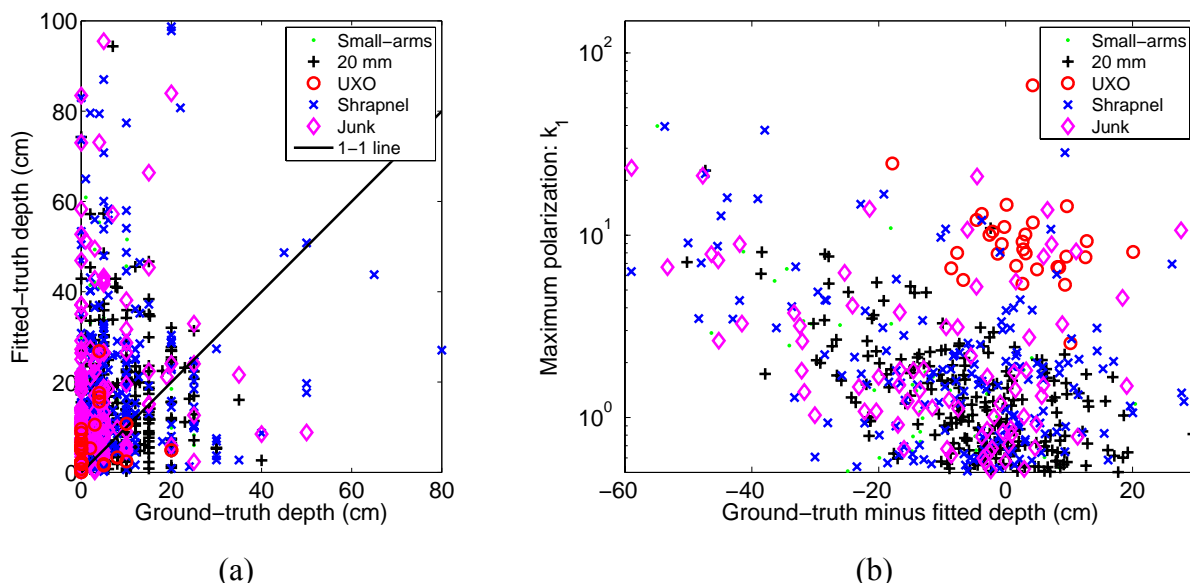


Figure 50. (a) Fitted EM-63-dipole depth versus actual depth for different objects at the Rocket Range. (b) Error in depth estimation versus the estimated object size. Calculations were made assuming the EM-63 coils were 20 cm above the ground-surface.

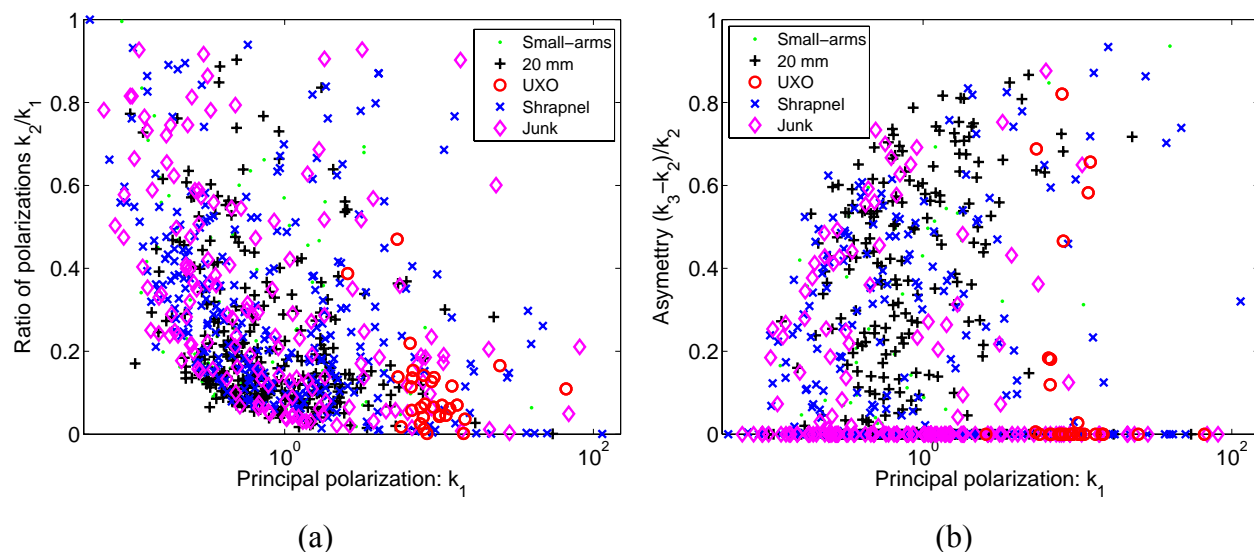
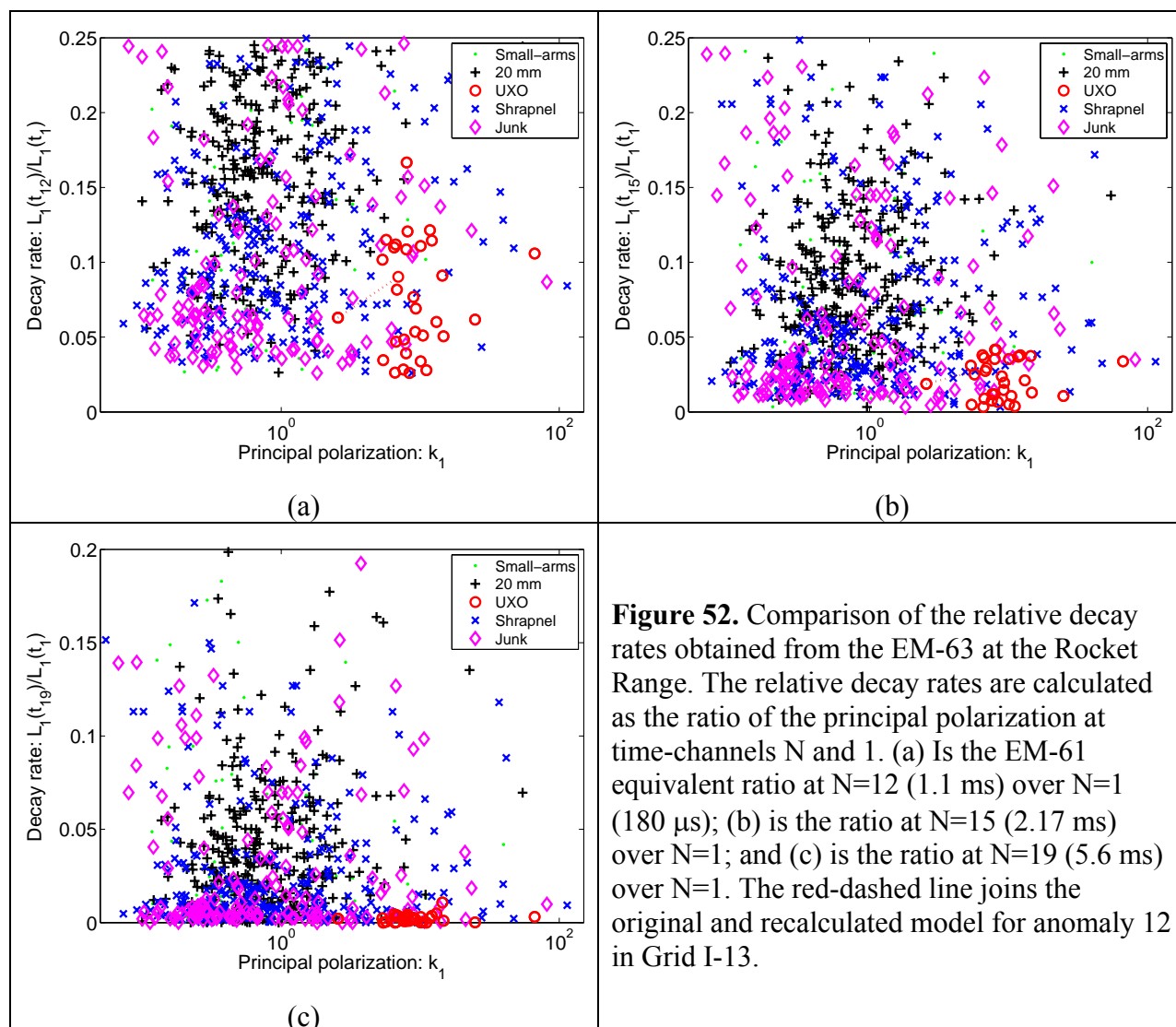


Figure 51. Comparison of polarization parameters recovered from EM-63 data on the Rocket Range. (a) Plot of a *shape-based* feature, k_2/k_1 , against size of the principal polarization (k_1); and (b) Plot of an *asymmetry* feature, $(k_3 - k_2)/k_2$, against size of the principal polarization (k_1). For two-dipole models this asymmetry parameter is zero.



For the re-inversions of the EM-63 Rocket Range data, we also included a %error in the covariance matrix (see next section) and fit only three-dipole models (so didn't have to decide whether to select a two-dipole model). We used the same masks that had been set-up for the original inversions. Figure 53a plots the principal polarizations curves for each of the 904 items that were re-inverted. There is some spread in the magnitudes of the polarization curves for the MK-23, but they have similar time-decay characteristics as revealed in a plot of normalized polarization curves (Figure 53b). The time-decay properties of the MK-23s are significantly different than the majority of the small arms, 20 mm, shrapnel and junk items. For the secondary (transverse) polarizations the time-decays of all the MK-23 are in close agreement out to about 2 ms. At around that time, or a bit later, seven of the curves diverge from the mean behavior.

The general consistency of the time-decay properties of the MK-23 polarizations indicates that they would provide useful information to aid discrimination (Figure 54). A plot of the ratio $L_1(t_{12})/L_1(t_1)$ versus $L_1(t_1)$ reveals that the MK-23s are tightly clustered with minimal overlap with the other items. This feature space would yield excellent classification performance as would the related feature space consisting of $L_1(t_1)$, β_1 and γ_1 (Figure 21c). The β_1 and γ_1 combination provides similar information to the $L_1(t_{12})/L_1(t_1)$ feature.

There is also reasonable separation between the MK-23s and the other items in a feature space comprised of the secondary polarizations (Figure 54b). However, the separation is not as consistent as that provided by the primary polarizations, and the secondary polarizations are unlikely to contribute extra information of value to the classification. Note, that for a measure of object size, we used $L_1(t_1)$ for the primary polarization and $L_2(t_1)$ for the secondary polarization. We found that these metrics provided better separation between MK-23s and other items than the Pasion-Oldenburg, k_1 and k_2 parameters (this was not the case in the original inversions due the re-parameterization procedure we used).

In Figure 55, we show the ROC curve for Grids I-13, J-12, K-15, L-13, L-14 and L-15 that would result by applying a Support Vector Machine Classifier trained with $L_1(t_1)$, β_1 and γ_1 on Grids I12 and J-13. We do not use the feature vectors from the test-plot data as these are not consistent with those obtained in grids I12 and J-13. The SVM classifier requires 28 false alarms to recover 24 MK-23s found in the other six grids. The classifier performance is similar to that of an amplitude based ranking scheme, where 37 false-alarms are required, and it is significantly better than a ranking based on size alone (91 false-alarms). As mentioned in previous sections, the amplitude based ranking performs so well because all the MK-23s are at or near the surface. The advantage of the SVM method is that we can stop digging much sooner. The point with $f=0$ (feature vector mid-way between clutter and UXO support planes) is at hole 223 (199 false-alarms) compared to the 725 excavations required by amplitude (701 false-alarms).

Finally, closer inspection of Figure 54a reveals that the majority of 20 mm projectiles are clustered in a similar region of feature space. There are a number of outliers, which could be due to low SNR or other data quality issues. However the plot does demonstrate that there are some regions of feature space that are free of 20 mm projectiles and that there is some chance a discrimination strategy designed for finding 20 mm projectiles could be developed (if such projectiles were deemed dangerous enough).

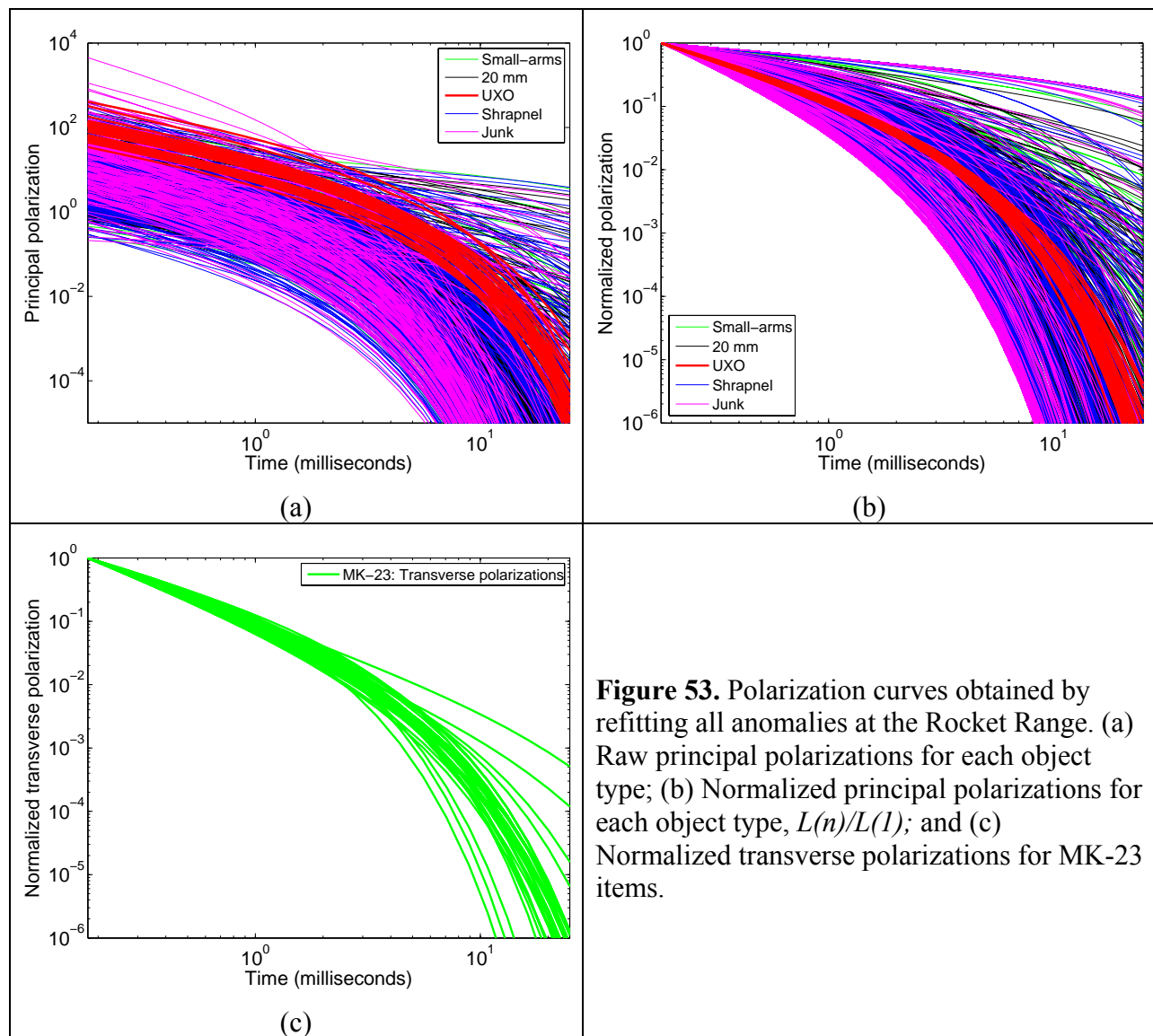


Figure 53. Polarization curves obtained by refitting all anomalies at the Rocket Range. (a) Raw principal polarizations for each object type; (b) Normalized principal polarizations for each object type, $L(n)/L(1)$; and (c) Normalized transverse polarizations for MK-23 items.

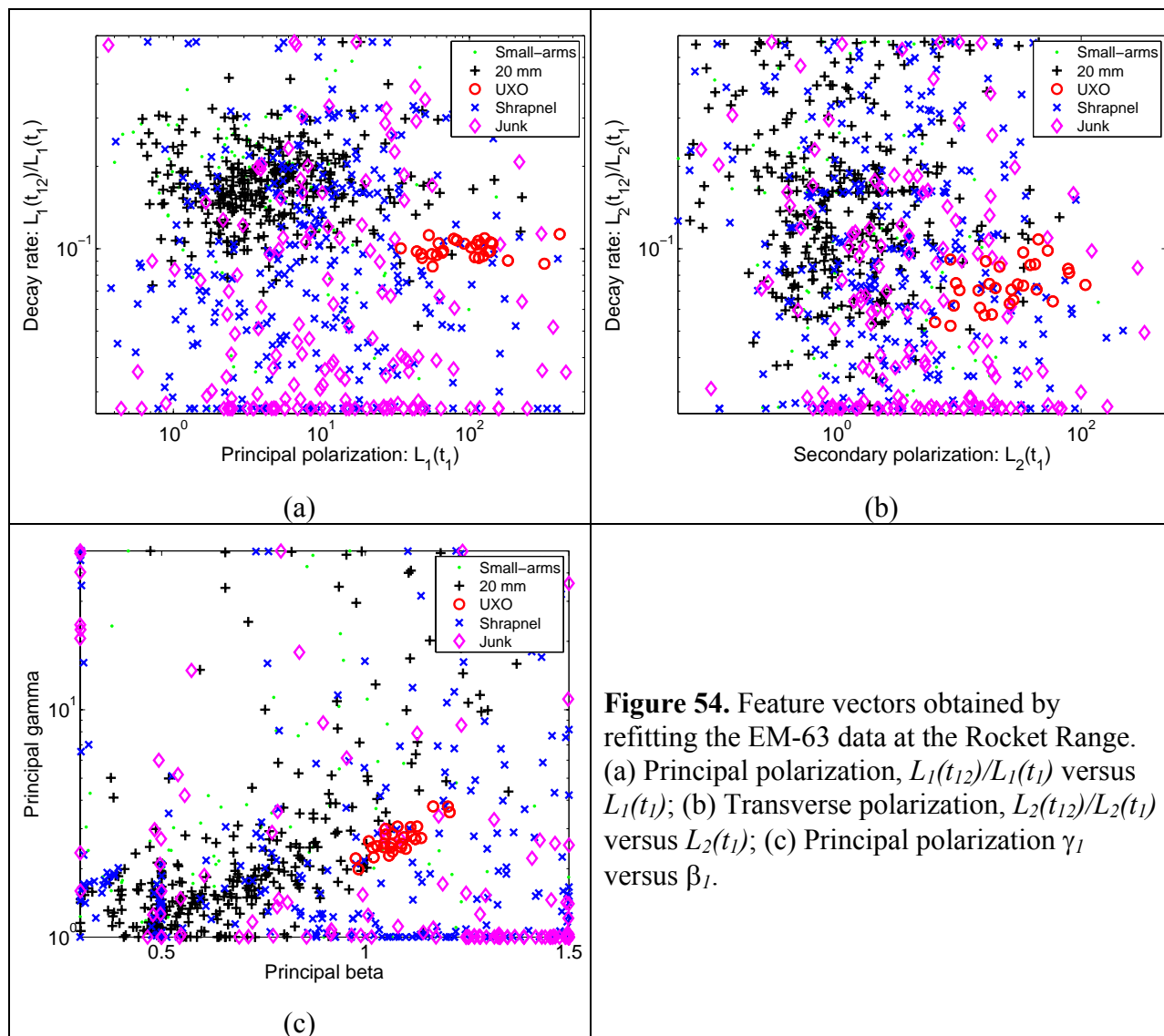


Figure 54. Feature vectors obtained by refitting the EM-63 data at the Rocket Range. (a) Principal polarization, $L_1(t_{12})/L_1(t_1)$ versus $L_1(t_1)$; (b) Transverse polarization, $L_2(t_{12})/L_2(t_1)$ versus $L_2(t_1)$; (c) Principal polarization γ_1 versus β_1 .

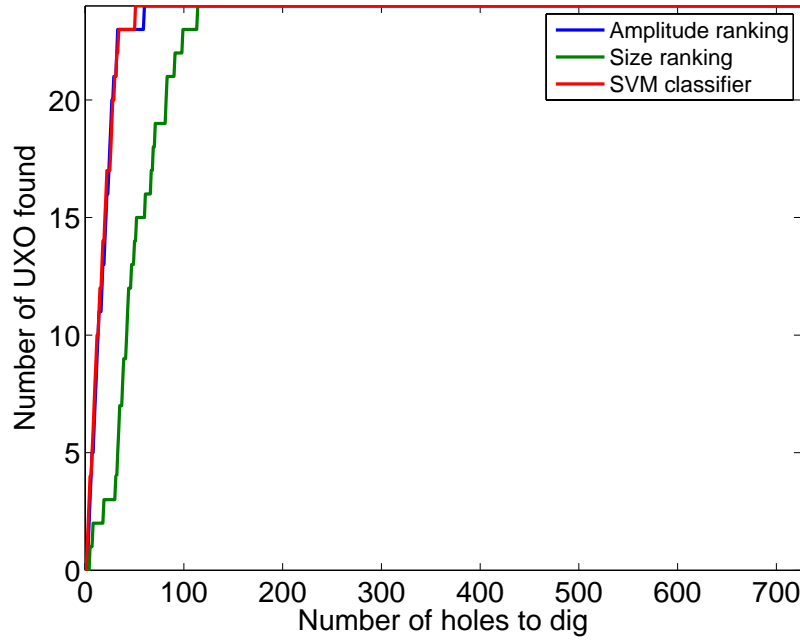


Figure 55. ROC curve for re-inverted EM-63 data on Phase II and III grids (I-13, J-12, K-15, L-13, L-14 and L-15) at the Rocket Range. The SVM classifier was obtained by training with feature-vectors $L_I(t_I)$, β_I and γ_I on the Phase I Grid, I12 and J-13. ROCs based on amplitude and size, $L_I(t_I)$, are also shown.

6.5 Investigation of data covariance matrix

Feature extraction involved solving the constrained least-squares optimization problem of Equation 6. Data errors play an important role in the feature extraction and are incorporated through the data covariance matrix. We assumed independently distributed Gaussian errors, and use the following data covariance matrix:

$$[V_d^{-1/2}]_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ \frac{1}{\delta_i + \varepsilon_i} & \text{if } i = j \end{cases} \quad (18)$$

where δ_i is a percentage of the i^{th} datum:

$$\delta_i = \%error \times [d_{obs}]_i \quad (19)$$

and ε_i is a base level error that is present in the i^{th} datum in the absence of a target. For this discrimination study we assumed that the $\%error$ was zero and used a base-error that varied with

time-channel (e.g., Figure 34). Thus, the error at each spatial location was assumed to be identical. We here conduct some simulations to determine how significant the percentage error was likely to be for an EM-63.

Data errors can arise from modeling inaccuracy, sensor noise and cultural/geological sources. Modeling errors include any discrepancy between the approximate forward mapping and an exact forward model. There are, essentially, two sources for discrepancies. The first is any inaccuracy in the functional form of the forward mapping. When inverting for the dipole polarization, higher order multi-poles act as correlated noise that has the potential to bias the recovered parameters. The second source of modeling errors is due to uncertainties in the modeling parameters that are not included in the model vector \mathbf{m} . Uncertainties in sensor positioning and orientation fall under this category. The data uncertainty resulting from these errors is approximately proportional to the signal strength. Our simulations show that Gaussian errors in the position and orientation can produce non-Gaussian, position dependent data errors (Figures 56 and 57). When simulating data collected over an 81 mm mortar (with polarization parameters determined from test-stand data) at depth of 30 cm, we found small position errors (2 cm standard deviation) and orientation errors (2 degree standard deviation in pitch, roll, and yaw) result in a nearly Gaussian error that is approximately 15 percent of the data value (Figures 58, 59 and 60). A %error of 15% is significant and would likely have exerted a strong influence on the feature vectors extracted for this study. For future discrimination study data processing, we will define an appropriate percent error.

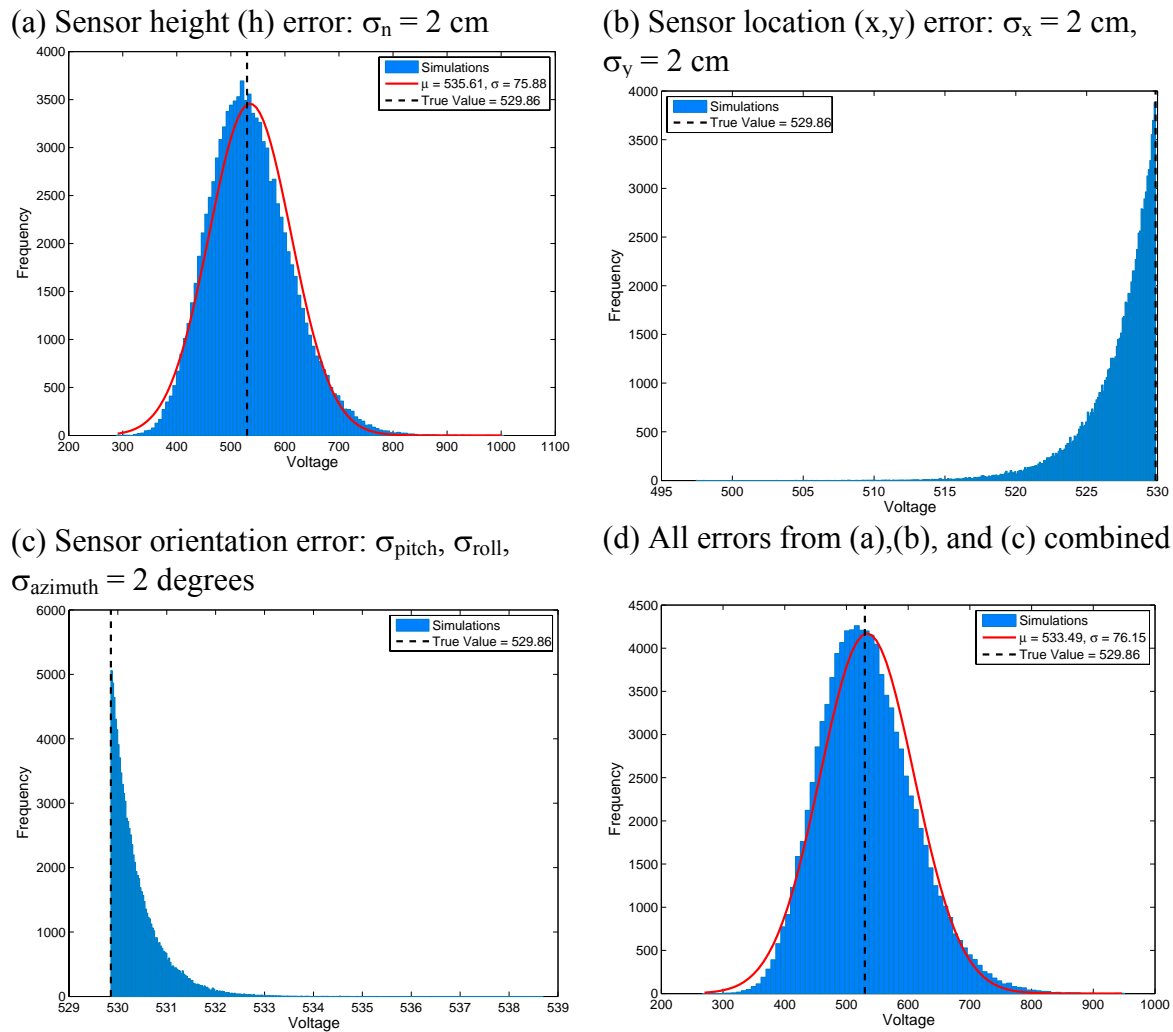
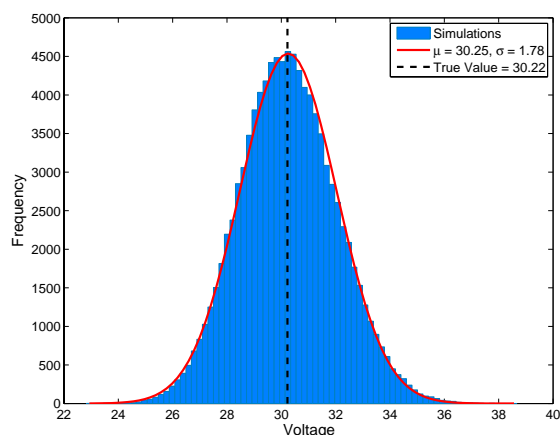
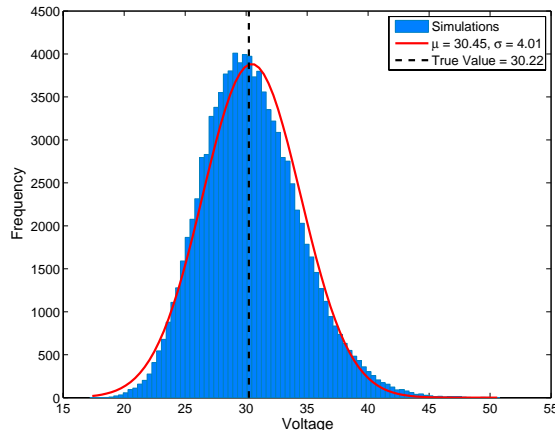


Figure 56. Error simulation results. Geonics EM-63 first time channel data from a vertical 81 mm target is simulated. The sensor is positioned at $(x,y,z) = (0,0,0.3)$ m, and the 81 mm mortar is located at $(0,0,-0.3)$ m. When there are no errors in sensor positioning and orientation, the sensor would record a value of 529.86 mV. In (a) – (c), we consider contributions of sensor height, location and orientation errors to the data spread. The asymmetry of (b) and (c) is due to the geometry of the problem. The maximum value measured by horizontal loop over a vertical target is directly over the target. In these simulations, any small changes in the orientation or position will produce a decrease in the signal strength. As a result, the most likely value for the data is the true value, and all other data will be lower in magnitude. This produces a one-sided distribution; (d) shows the result when including all errors. Directly over a target, the error in height contributes most to the variation in the signal value.

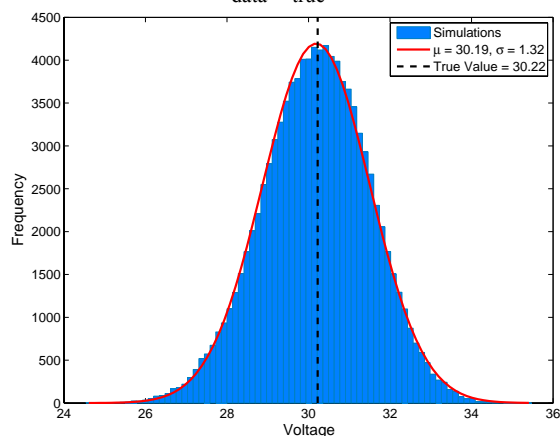
(a) Sensor height (h) error: $\sigma_n = 2$ cm. Best fit normal distribution has $\sigma_{\text{data}}/d_{\text{true}} \times 100 = 5.9\%$



(b) Sensor location (x,y) error: $\sigma_x = 2$ cm, $\sigma_y = 2$ cm. Best fit normal distribution has $\sigma_{\text{data}}/d_{\text{true}} \times 100 = 13\%$



(c) Sensor orientation error: $\sigma_{\text{pitch}}, \sigma_{\text{roll}}, \sigma_{\text{azimuth}} = 2$ degrees. Best fit normal distribution has $\sigma_{\text{data}}/d_{\text{true}} \times 100 = 4.4\%$



(d) All errors from (a),(b), and (c) combined. Best fit normal distribution has $\sigma_{\text{data}}/d_{\text{true}} \times 100 = 16\%$

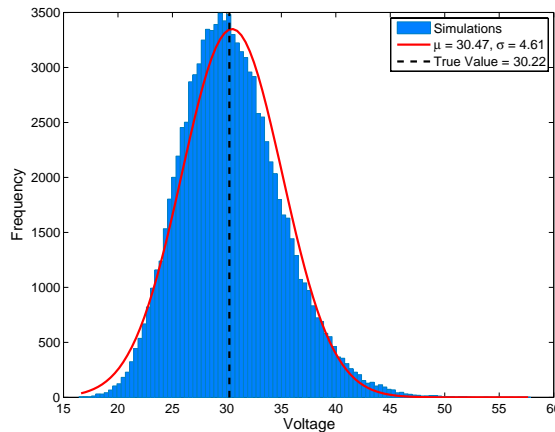


Figure 57. Error simulation results. Geonics EM-63 first time channel data from a vertical 81 mm target is simulated. The sensor is positioned at $(x,y,z) = (0.5,0.5,0.3)$ m, and the 81 mm mortar is located at $(0,0,-0.3)$ m. When there are no errors in sensor positioning and orientation, the sensor would measure 30.22 mV. In (a) – (c), we consider contributions of sensor height, location and orientation errors to the data spread. (d) shows the result when including all errors. There errors are nearly Normal.

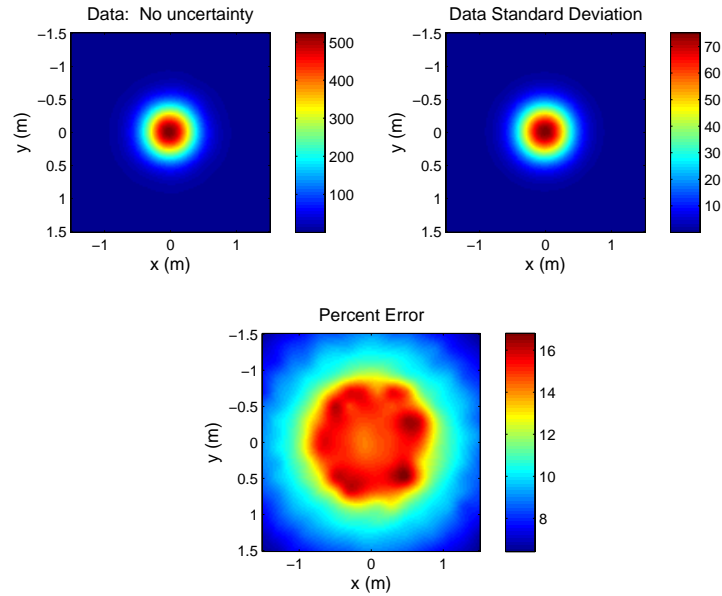


Figure 58. Percent error over a vertical 81 mm mortar. Sensor height and location uncertainties are Gaussian with $\sigma = 2$ cm. Sensor orientation has Gaussian errors of 2 degrees. The percent error due to sensor positioning errors varies from approximately 5 to 18 mV.

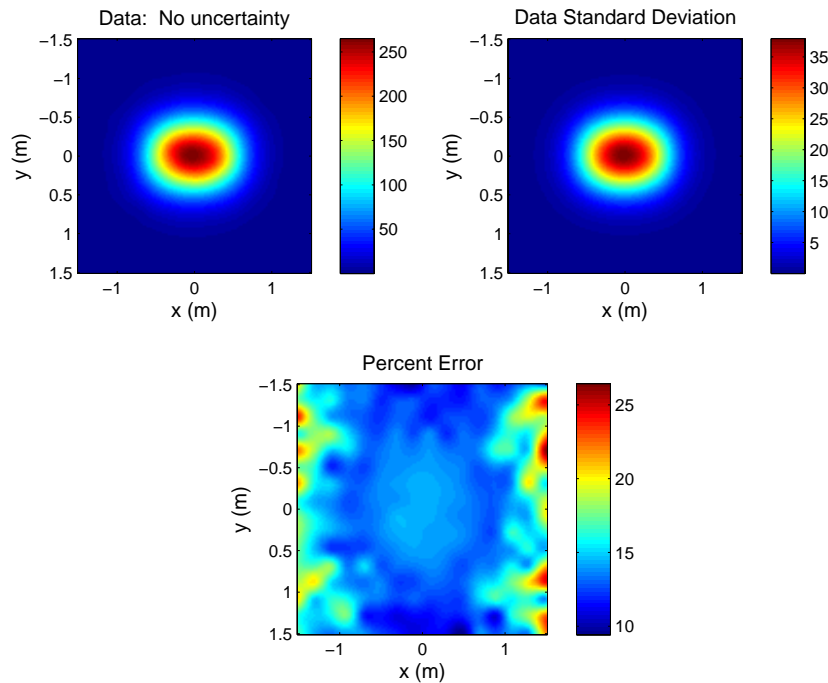


Figure 59. Percent error over a horizontal 81 mm mortar. The same sensor positioning uncertainties of Figure 58 are used. The percent error due to sensor positioning errors varies from approximately 10 to 25 mV.

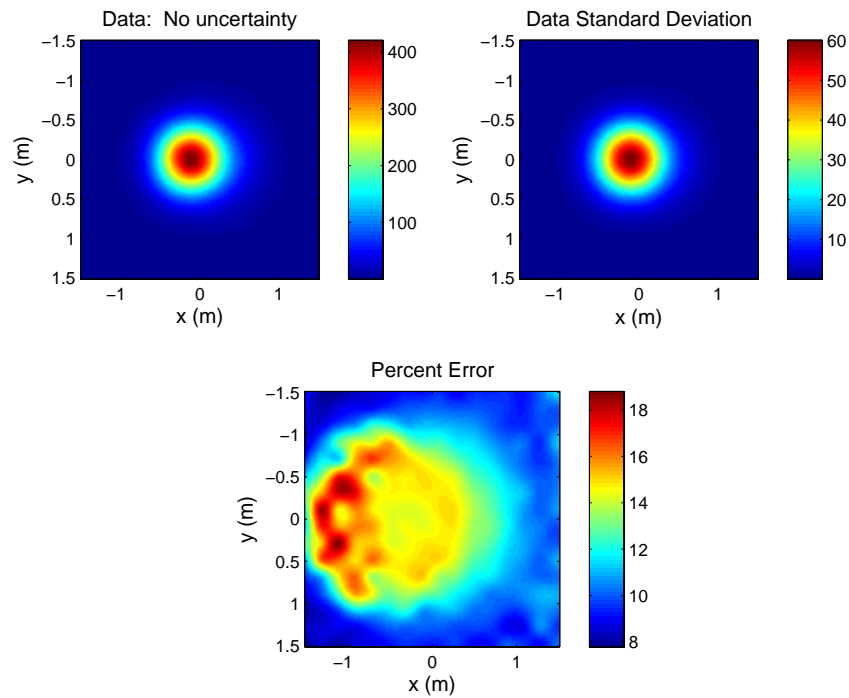


Figure 60. Percent error over an 81 mm mortar at a 45 degree dip. The same sensor positioning uncertainties of Figure 58 are used. The percent error due to sensor positioning errors varies from approximately 5 to 18 mV.

7 COST ASSESSMENT

Discrimination will likely only gain widespread acceptance if there is a decrease in the costs of clearance (relative to 100% detection), without a significant increase in the risk. The additional costs of data collection, processing and interpretation need to be offset by a reduction in the number of holes excavated. In this section of the report we make a comparison of the expected costs of discrimination based methods (EM-61 and EM-63) against a standard production survey with the EM-61. The costs for a standard detection mode survey include:

- (i) Site preparation (e.g., vegetation clearance, surface sweep);
- (ii) Geophysical survey;
- (iii) Data processing and interpretation;
- (iv) QA/QC;
- (v) Anomaly relocation (as appropriate); and
- (vi) Anomaly excavation and documentation.

The costs for the discrimination strategy include the above, minus the costs for anomaly relocation and excavation for those anomalies not slated for excavation, plus the costs for

- (i) Additional geophysical surveying (as required);
- (ii) Additional data processing and interpretation; and
- (iii) Additional documentation and QC.

For this demonstration we were able to accurately track the time (and hence costs) required for geophysical survey and initial processing. For the feature extraction and classification, we made some adjustments to the underlying inversion methods and trialed a number of different inversion strategies (e.g., noise-levels, masks etc) before settling on our final approach. Therefore, for these tasks we can only make estimates of reasonable times required for processing/interpretation.

For the cost comparison we estimate the fully burdened costs to the government for the geophysical and excavation phases of the work. We make the following assumptions:

- (i.) Mobilization/demobilization
 - a. The survey requires a 2000 km mobilization from Denver;
 - b. Mobilization includes 1 day preparation, 1 day set-up on site and 1 day for a testplot survey (and associated processing);
 - c. Demobilization includes 1 day of packing up on site and 1 day of organization back at home base.
- (ii.) The survey area is 100 acres;
- (iii.) There are 100 anomalies per acre;

- (iv.) Anomaly reacquisition proceeds at 150 items per day and requires a three person crew (two field technicians, one UXO technician);
- (v.) It costs \$100 to excavate each anomaly;
- (vi.) EOD escort rate and per-diem is that same as the field-technicians;
- (vii.) There is a management cost of 10% of for the geophysical portion of the work (the excavation portion is embedded within the cost per anomaly);
- (viii.) Daily rates are fully burdened and assume a 10 hour working day;
- (ix.) Processing requires a Geophysicist II level analyst while interpretation is 50% at the same level and 50% at the next highest level;
- (x.) Per-diem and hotel are assumed to total \$150 per person per-day, with 1.4 days of per-diem/hotel per day in the field (5 day week, 2 days on the weekend);
- (xi.) Equipment charge includes all equipment, vehicles (assumed 1 rental vehicle required) consumables and supplies;
- (xii.) Anomaly reacquisition costs are calculated assuming 2 field-technicians and an EOD tech can reacquire 150 anomalies per-day;
- (xiii.) Additional assumptions on each survey method are provided in Table 17.

Table 17. Assumptions used to compare the different survey methods. The excavation reduction numbers were estimated from the performance at the Rocket Range.

Attribute	EM-61 production	EM-61 discrimination	EM-63 discrimination
Survey Rate	10 acres per day	8 acres per day	1 per day
Equipment	5 EM-61, 1 GPS, tow-vehicle/sled, 2 vehicles	5 EM-61, 1 RTS, 1 IMU, tow-vehicle/sled, 2 vehicles	1 EM-63, 1 RTS, 1 IMU, 2 vehicles
Field personnel	3 field crew & 1 escort	3 field crew & 1 escort	2 field crew & 1 escort
Processing Time per acre	0.15 days	0.2 days	0.25
Processing Time per anomaly	0	5 minutes	10 minutes
Detected holes excavated	100%	75%	75%

The cost comparison using the stated assumptions is provided in Table 18. For the geophysical portion of the work, a standard production survey costs around \$876 per acre, compared to an EM-61 discrimination survey cost of \$1,755 per acre and an EM-63 survey cost of \$4,832 per acre. With a 25% reduction in false-alarms, the validation costs are reduced from \$11,890 per acre to \$8,910 per acre. This reduction is enough to make the EM-61 discrimination survey

more cost-effective than the production survey. Due to the slow-rate of survey and the length of time required for interpretation, the EM-63 discrimination method is more costly. For the costs to be less than those for EM-61 production, the excavations would have to be reduced by about 35% (35 anomalies per acre). In comparison, EM-61 discrimination has a break-even point of 8%, or just 8 anomalies per acre. Thus according to this cost model, the discrimination efficiency of the EM-61 does not have to be very high for cost-savings to be realized.

Table 18. Comparison of the costs for the different modes of survey using the assumptions in Table 17, and in the bullet points immediately before that table.

System	EM-61 production			EM-61 discrimination			EM-63 Discrimination		
Geophysics	Daily Rate	# Days	Cost	Daily Rate	# Days	Cost	Daily Rate	# Days	Cost
Mobilization/Demobilization	NA	NA	\$32,543	NA	NA	\$34,125	NA	NA	\$19,793
Survey (Equipment/consumables)	\$1,250	10	\$12,500.88	\$1,250	12.5	\$15,626	\$743	100	\$74,300
Survey (Labor)	\$1,834	10	\$18,336.00	\$1,834	12.5	\$22,920	\$1,375	100	\$137,520
Hotel and per-diem	\$600	14	\$8,400.00	\$600	17.5	\$10,500	\$450	140	\$63,000
Processing (Initial)	\$463	15	\$6,942.00	\$463	18.75	\$8,678	\$463	25	\$11,570
Processing (Discrimination)	\$628	0	\$0.00	\$628	104.167	\$65,427	\$628	208.33	\$130,854
Deliverables/Maps	\$463	2	\$925.60	\$463	5	\$2,314	\$463	5	\$2,314
Management (10%)	NA	NA	\$7,964.72	NA	NA	\$15,959	NA	NA	\$43,935
Total (GEOPHYSICS)			\$87,612			\$175,549			\$483,287
Validation	Unit Rate	# Holes	Cost	Unit Rate	# Holes	Cost	Unit Rate	# Holes	Cost
Reacquisition	\$19	10000	\$188,217.60	\$19	7500	\$141,163	\$19	7500	\$141,163
Validation	\$100	10000	\$1,000,000.00	\$100	7500	\$750,000	\$100	7500	\$750,000
Total (VALIDATION)			\$1,188,217.60		15000	\$891,163		15000	\$891,163
Total			\$1,275,829.56			\$1,066,712			\$1,374,450

Obviously the cost analysis is dependent on the values of a number of parameters that could vary widely from the numbers quoted here. These include: the cost per anomaly to excavate; the number of anomalies per acre; the time required to interpret each anomaly; and the reduction in the number of excavations due to discrimination (percentage of holes to dig). Table 19 provides a comparison of how the cost varies for different numbers of anomalies per acre, and different percentages of holes to dig. As the number of anomalies per acre decreases, the discrimination efficiency of the EM-63 must improve substantially for it to be cost-competitive to the production survey (due to the greater percentage cost of the geophysics compared to the validation). On the other hand, the EM-61 remains cost competitive with the standard production survey with very small reductions in the number of holes to dig.

In table 20, we investigate the influence of the time-required to process each anomaly on the costs for the EM-61 and EM-63 surveys (assuming 100 acres and 100 anomalies/acre). As the time for processing increases there is only a modest increase in the price. According to our

price-assumptions, each extra minute of interpretation time only increases the cost by \$14,395. This is the equivalent of excavating an extra 144 anomalies.

Table 19. Comparison of the cost of survey for the different methods with percentage of holes to dig and different numbers of anomalies per acre. The smallest reduction in holes to dig that produces a cost less than the standard production method is marked in green.

Number holes	200			50		
Holes to dig	EM-61 Prod	EM-61 Disc	EM-63 Disc	EM-61 Prod	EM-61 Disc	EM-63 Disc
100%	\$2,464K	\$2,624K	\$3,033K	\$681K	\$734K	\$1,005K
90%		\$2,386K	\$2,766K		\$674K	\$946K
75%		\$2,030K	\$2,410K		\$585K	\$857K
66%		\$1,816K	\$2,196K		\$532K	\$803K
50%		\$1,436K	\$1,815K		\$560K	\$708K
25%		\$842K	\$1,221K		\$288K	\$560K

Table 20. Comparison of the cost of survey for the different methods with different % of holes to dig and different amounts of time required for interpretation of each anomaly. The smallest reduction in holes to dig that produces a cost less than the standard production

Time/anomaly	1		5		10	
Holes to dig	EM-61	EM-63	EM-61	EM-63	EM-61	EM-63
100%	\$1,306K	\$1,542K	\$1,363K	\$1,600K	\$1,435K	\$1,671K
90%	\$1,187K	\$1,423K	\$1,244K	\$1,480K	\$1,317K	\$1,552K
75%	\$1,009K	\$1,244K	\$1,067K	\$1,302K	\$1,138K	\$1,374K
66%	\$902K	\$1,138K	\$958K	\$1,196K	\$1,031K	\$1,267K
50%	\$712K	\$947K	\$770K	\$1,005K	\$842K	\$1,077K
25%	\$415K	\$651K	\$473K	\$708K	\$545K	\$780K

In summary, EM-61 discrimination will be more cost-effective than standard EM-61 production surveying if discrimination can eliminate the need to excavate 4 out of every 100 anomalies. If excavations can be reduced by as much as 25% then there will be an approximate cost saving of 20%. EM-61 discrimination costs compare favorably to production surveying at both higher and lower anomaly density. For a 100 acre survey with 100 anomalies/acre, an EM-63 discrimination survey is only likely to be more cost-effective if it can achieve a 35% reduction in false-alarms. As the anomaly density per-acre decreases, this requirement increases sharply due to the need to offset the high costs of surveying with the EM-63.

8 IMPLEMENTATION ISSUES

8.1 Environmental Checklist

No permits were needed for this demonstration.

8.2 Other Regulatory Issues

Initially we intended to engage the regulatory community through a review of the technology by Mr. Jeffrey Swanson, the site regulator from the State of Colorado Department of Health and Environment. We have not proceeded with this plan due to the upcoming ESTCP Discrimination Study which is aimed at engaging the regulatory community so that discrimination becomes an accepted practice.

8.3 End-User Issues

The regulators and stakeholders will need to buy-in for these discrimination methodologies to be accepted on site and fulfill the promise they hold for reducing remediation costs by decreasing the number of excavations required on a remediation site. The main issue related to acceptance of this methodology will be, we believe, transparency in the decision making process and usefulness of the software in facilitating decision making.

The main advantages of the technology are a potential reduction in the number of non-hazardous items that need to be excavated, thus reducing the costs of UXO remediation. There are two key aspects to the demonstrated technology (i) hardware and (ii) software. On the hardware side, we are concentrating on the demonstration of COTS sensors like the EM-61, EM-63 and Cs-vapor magnetometer. On the software side, UXOLab, as used for this demonstration, is currently available through the University of British Columbia. This software is largely a research focused code but has been used successfully by Sky Research for production work at FLBGR and at National Guard sites in Montana.

8.4 Advantages and Limitations of the Technology

The main advantage of the technology is a potential reduction in the number of non-hazardous items that need to be excavated, thus reducing the costs of UXO remediation. There are two key aspects to the demonstrated technology (i) hardware and (ii) software. On the hardware side, we are concentrating on the demonstration of COTS sensors like the EM-61, EM-63 and cesium (Cs) vapor magnetometer. As each of these instruments measure only one component of a vector field, a measurement at a single location provides limited information. As a consequence, relatively dense measurements in two-dimensions are required for accurate recovery of relevant target parameters. These measurements must be very precisely positioned and oriented for discrimination to be successful. SERDP/ESTCP are sponsoring the development of several next

generation sensors with multi-component receivers. These newer sensors have the potential to significantly improve the estimation of target parameters using a much lower density of measurements. Over the next few years, these sensors may replace the EM-61/EM-63 and Cs vapor magnetometers. However, there will still be large volumes of data collected and processed with the older sensors, and there is no guarantee that any of the new sensors will be rugged and flexible enough for the diverse environments of the many hundreds of MEC contaminated sites in the country.

On the software side, advantages of UXOLab and the algorithms within the package include:

- The software contains all the functionality required to process raw geophysical data, detect anomalous regions, and perform geophysical inversion and discrimination.
- UXOLab contains algorithms for inverting magnetic and TEM datasets both separately and cooperatively using a number of different polarization tensor formulations.
- Has an extensive set of algorithms for rule-based and statistical classification.
- UXOLab has been configured in a modular fashion, so that as new sensor technologies become available (e.g. new TEM systems with multi-component receivers etc), the inversion functionality will be immediately available to those new sensor systems.

Thus, the UXOLab software is powerful and flexible. Due to its modularity and extensive functionality, the software will provide an excellent environment for data processing and interpretation for the next generation of sensors.

9 SUMMARY AND RECOMMENDATIONS FOR THE NEXT DEPLOYMENT

Section 4.4 summarized the discrimination performance of our classification methodology as applied during the demonstration process (that is before we had access to any ground-truth for the next phase). In Section 6, we conducted retrospective analysis of the EM-61 and EM-63 data on both the 20 mm Range-Fan and the Rocket Range. Here, we compile a list significant conclusions from the demonstration process and the retrospective analysis and then discuss recommendations for the next demonstration.

9.1 List of Significant Conclusions

9.1.1 Discrimination Performance at 20 mm Range-Fan

On the 20 mm Range-Fan the goal was to discriminate hazardous 37 mm projectiles from less hazardous 20 mm projectiles and 50 caliber bullets. For both the EM-63 and EM-61 two feature vectors were used to guide the Support Vector Machine statistical classification. One was based on the size of the polarization tensor, the other on the quality of the fit. Feature vectors based on ratios of polarization tensor components (shape) or on the time-decay properties of the polarization did not improve classification performance (except during retrospective analysis, see section 9.1.8). For both Phases I and II and for both instruments the SVM classification method outperformed a ranking based on amplitude. The last detected UXO was ranked quite high in each of the classification diglists and digging to that point would have resulted in a 60-90% reduction in the number of false-alarms. This operating point is of course unknown prior to digging. We found that using a stop-digging criteria of $f=0$ (mid-way between UXO and clutter class support planes), was too aggressive and more excavations were typically required for full recovery of detected UXO. Due to lower SNR, more 50 caliber bullets than 20 mm projectiles were ranked high by the SVM classifier. Both the amplitude and SVM methods performed quite poorly on two deep (40 cm) emplaced 37 mm projectiles, exposing a potential weakness of the *misfit/amplitude*² metric (see section 9.1.4 for more discussion). Retrospective analysis revealed that thresholding on the size of the polarization tensor alone would have yielded good discrimination performance.

Conclusion: Discrimination of 37 mm projectiles from 20 mm projectiles and small-arms is feasible at the 20 mm Range-Fan using an EM-61 towed array and a diglist ranking scheme based on the size of the polarization tensor.

9.1.2 Discrimination Performance at the Rocket Range

The goal at the Rocket Range was to find all munitions with calibers equaling or exceeding 37 mm in diameter. On the eight grids used for this demonstration, all but one UXO item was a MK-23 practice bomb, and these were all found relatively close to the surface. This resulted in an amplitude based ranking performing very well. The SVM classifier used essentially the same feature vectors as the 20 mm Range-Fan. Again, shape and time-based features of the

polarizability did not aid classification performance. After modifying the goodness-of-fit metric from *misfit/amplitude* to *misfit/amplitude*² the performance of the SVM classifier and the amplitude ranked methods were comparable. A significant advantage of the SVM method is that digging can stop sooner. We found reductions in false-alarm rates of 70-90% were achievable. Use of the *misfit/amplitude*² metric introduced a significant additional QC burden into the discrimination procedure. Retrospective analysis, where digging order is determined by entirely by size, results in good discrimination performance with the potential for minimal false-alarms.

Conclusion: Discrimination of MK-23 projectiles appears to be feasible at the Rocket Range when using an EM-61 towed array and a diglist ranking scheme based on the size of the polarization tensor. Performance against other munitions found in the area is unknown as none of these were encountered during the demonstration (except for a single 2.5 inch rocket).

9.1.3 Training Data Requirements

For the first phase of classification, only limited training data were available to help select feature vectors and decide on a classification strategy. Excavation of these phase I grids provided a richer set of training data, particularly over objects in the clutter class. Minor modifications to the feature vectors were made for the second phase of classification but only one UXO was recovered during the subsequent digging phase. For the third-phase, the classifier used the same training data and feature vectors as the phase II validations. A single phase of excavations appeared to provide sufficient training data for both the applied and retrospective classification strategies.

Conclusion: Excavation of one grid at the 20 mm Range-Fan and two grids at the Rocket Range provided sufficient information to build a classification strategy.

9.1.4 Disadvantages of the *misfit/amplitude*² Feature Vector

The *misfit/amplitude*² feature was effective against shallow 37 mm projectiles and MK-23 practice bombs. There were two significant disadvantages of the method. The first relates to the additional QC burden and the consequences of any errors in that QC process. As we found at the Rocket Range, if poor inversion fits are accepted, a UXO can be ranked quite low by the SVM classifier. The second disadvantage concerns deep UXO such as the two 37 mm projectiles emplaced in the 20 mm Range-Fan. The fits to these deep rounds are poor and they are hence ranked very low by the SVM classifier.

Recommendation: Do not use the *misfit/amplitude*² feature vector unless the goal is to find UXO in a depth region where the SNR will be high.

9.1.5 Depth and Size Ambiguity Caused by Low SNR (and Positioning Inaccuracies)

At the 20 mm Range-Fan it was found that 50 caliber bullets caused more false-alarms than 20 mm projectiles, even though they are significantly smaller. Retrospective analysis revealed that

this was caused by a lower SNR³ on the 50 caliber bullets. There was insufficient SNR to constrain the depth of the item and inversion solutions tended to be pushed deep due to either flat-objective functions or the presence of multiple locally optimal solutions. Consequently, size estimates of 50 caliber bullets obtained from the k-sum of EM-63 data varied across four-orders of magnitude and tended to be larger than their size would suggest. For the larger 20 and 37 mm projectiles, size estimates varied by around two-orders of magnitude, but there was less overlap between the two classes (than between 50 caliber bullets and 20 mm projectiles). The relatively poor depth performance on shallow, high SNR MK-23 practice bombs at the Rocket Range indicates that positional errors (and potentially un-modeled dipole components) also cause uncertainty in the object depth (and hence in the object size).

Conclusion: Depth and size are poorly constrained when estimated from single component sensor data obtained with currently available positional precision. However, size estimates may still provide useful information to prioritize digging order.

9.1.6 The Role of Cooperative Inversion

The amplitude of the polarization tensor components can be a powerful discrimination metric (it provides a measure of object size). However, the reliability of polarization tensor amplitudes are limited by the accuracy of the depth estimates which in turn are limited by low SNR, positional accuracy and number of relevant observations. Depth and position constraints from magnetic data (if they are accurate) have the potential to improve estimates of polarization amplitudes obtained from EM data. Due to positional problems with the magnetometer survey data, we were unable to demonstrate the cooperative inversion procedure at FLBGR. On the 20 mm Range-Fan, the applicability of cooperative inversion would have been limited (even with precisely positioned magnetic data), due to the poor detection performance of the magnetometer data on 50 caliber bullets. Magnetics could perhaps contribute by rejecting polarization solutions to small objects that are pushed deep during the inversion process.

Conclusion: Cooperative inversion has the potential to improve estimates of polarization tensor amplitudes recovered from EM-data. The applicability of the method is likely limited for discrimination of small objects such as the 50 caliber bullets and the 20 and 37 mm projectiles at the 20 mm Range-Fan.

9.1.7 Shape and Asymmetry as Discriminators

Shape (ratio of secondary and primary polarizations) and asymmetry (difference in secondary and tertiary polarizations) based feature vectors did not improve discrimination performance at the Rocket Range or 20 mm Range-Fan. For many of the MK-23s the shape ratios estimated from both EM-61 and EM-63 data are clustered. However, there are a number of MK-23s with significantly different polarization ratios and, consequently, the shape based metric cannot be relied upon for discrimination. SNR, positional uncertainty and insufficient “looks” at the object

³ Positioning error and sparse data coverage also likely contributed to our inability to constrain inversion parameters.

all contribute to the inability to consistently constrain the different polarization components.

Conclusion: SNR, positional accuracy and sampling diversity of the tested systems was insufficient to consistently constrain all three components of the polarizability.

9.1.8 Time-Decay Metrics for Discrimination

During the demonstration, feature-vectors derived from the time-decay properties of the polarization tensor were not used to aid discrimination performance of either instrument. Retrospective analysis revealed that time-decay properties of the principal polarization tensor could have been used to distinguish MK-23 practice bombs from other items on the Rocket Range (for both instruments). On the 20 mm Range-Fan, the time range of the EM-63 is long enough that the slower decay rate of the 37 mm could have been distinguished from 20 mm projectiles. In contrast, the EM-61 did not sample late enough in time to aid discrimination. The noise-floor decays as $1/t^{0.5}$ while signal falls off more rapidly. This means that the accuracy of time-decay parameters extracted from low SNR anomalies is generally limited (e.g the 50 caliber bullets on the 20 mm Range-Fan).

Conclusion: For both the EM-61 and EM-63, the time-decay properties of the principal polarization of the MK-23 practice-bombs were well-constrained and significantly improved discrimination performance. Only the EM-63 sampled late enough in time to distinguish 20 mm from 37 mm projectiles.

9.1.9 EM-61 to EM-63 Comparison

The EM-61 and EM-63 discrimination methods tested during the demonstration both used size and goodness-of-fit related feature vectors. Performance on both the Rocket Range and 20 mm Range-Fan were comparable, except the EM-63 data took about 5-8 times longer to collect and required more processing time and effort. In addition, the EM-61 is currently the instrument of choice for production EM-surveying. Unless the time-decay information from the EM-63 is used for discrimination, the EM-61 is clearly the preferred instrument. However, even if the time-decay information from the EM-63 is useful, the slow speed of survey means that it's unlikely the EM-63 is suitable for use in the full-survey mode.

Conclusion: Speed of survey, ease of use and reliability make the EM-61 more suitable for discrimination methods based on size-based metrics. The EM-63 is better suited for a cued-interrogation mode where it has the potential to constrain the time-decay properties of the polarization over a wider time-range.

9.2 Comments and Recommendations Concerning Feature Extraction

The previous section addressed the performance of the inversion procedures and the conclusions that can be drawn from the results obtained. While UXOLab had previously been used for analysis of EM-61 and EM-63 data at APG/YPG, this demonstration represented the first true test of its capabilities on a live-site with a large number of detected anomalies. The volume of

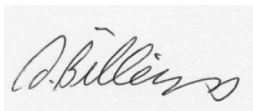
data collected and the need to fit both EM-61 and EM-63 anomalies meant that a number of less experienced UXOLab users had to be called upon to assist with the feature extraction. The following represents a summary of lessons learned during implementation of the discrimination process:

- (i.) ***Selection of data to invert:*** The observation locations to invert are determined by a user-specified mask around each anomaly. These masks are usually estimated through automated procedures that also deal with the case of overlapping anomalies. The masks can significantly influence the inversion results, either through inclusion of signal from adjacent anomalies, or by their impact on the estimated signal-to-noise ratio at each channel. Adjusting masks turned out to be one of the most difficult and time-consuming tasks during the inversion process. Better procedures for automated masking are required and we have begun to develop some potential solutions.
- (ii.) ***Estimation of covariance matrix and temporal data to invert:*** The covariance matrix was assumed to be constant for each survey grid and varied only as a function of time-channel. For some grids, there was considerable spatial variation in the sensor noise floor. The covariance matrix determines the relative weight placed on each observation location and on each time-channel. Together with the mask, it is used to estimate a signal-to-noise ratio at each time-channel and determines which of them are used to constrain the inversion. For the next demonstration we plan to
 - a. Allow the covariance matrix to vary with observation location;
 - b. Include a percentage error term in the covariance matrix;
- (iii.) ***Overlapping anomalies:*** Visual examination of spatial anomaly pattern can be misleading for determining the number of objects. A horizontal target with a dominant axial polarization (as is the case with rod-like UXO) can lead to an anomaly with a two peaks. The peak separation is a function of the target depth and transmitter loop size. During interpretation we needed to decide if the anomaly was best fit with a single target, or a pair of targets. This was time-consuming because we had to invert the item as a single object, then as a pair of targets where the each masked portion of the anomaly with a single dipole model. Better tools for determining whether to fit as one or two items are required within UXOLab.
- (iv.) ***Reparameterization of Pasion-Oldenburg model:*** After recovering a Pasion-Oldenburg model, we examined the variance of the γ parameter. If it were high, we replaced the P-O model with a power-law model $L(t) = kt^{-\beta}$. Retrospective analysis revealed that this process significantly limited our ability to use time-based feature vectors in the discrimination and could have been avoided if we used lower constraints on time-decay parameters.

10 POINTS OF CONTACT

Table 21. Points of Contact

Point of Contact Name	Organization Name/ Address	Phone/Fax/Email	Role in Project
Dr. Stephen Billings	Sky Research	541-552-5185 (phone) 604-827-3221 (fax) Stephen.billings@skyresearch.com	Principal Investigator
Jerry Hodgson	USACE-Omaha	402-221-7709 (phone) Jerry.L.Hodgson@nwo02.usace.army.mil	Federal Advocate
Stacey Kingsbury	Sky Research	540-961-9132 (phone) 661-457-1936 (fax) Stacey.kingsbury@skyresearch.com	Program Manager
Donald Yule	USACE-ERDC	601 634-2964 (phone) Don.E.Yule@erdc.usace.army.mil	Contracting Officer's Representative



Stephen Billings
March 22 2007

11 REFERENCES

- Bell, T., 2005, Geo-location Requirements for UXO Discrimination: Paper prepared for UXO Location Workshop, Annapolis, May 2005.
- Bell, T. H., B. J. Barrow, and J. T. Miller, "Subsurface discrimination using electromagnetic induction sensors", *IEEE Transactions on Geoscience and Remote Sensing*, volume: 39, issue: 6, pages:1286 – 1293, June 2001.
- Beran, L. S., 2005, Classification of unexploded ordnance. Master's thesis, University of British Columbia.
- Beran, L., Billings, S. D. & Oldenburg, D. W., 2004, A Comparison of Classification Algorithms for UXO Discrimination, proceedings from the UXO Forum, St. Louis, March 9-12, 2004.
- Billings, S. D., L. R. Pasion, and D. W. Oldenburg, "Inversion of magnetics for UXO discrimination and identification," in Proc. 2002 UXO Forum, Orlando, FL, Sep 2002a.
- Billings, S. D., L. R. Pasion, and D. W. Oldenburg, "Discrimination and identification of UXO by geophysical inversion. Phase II: Inversion of total-field magnetics," U.S. Army Corps of Engineers, Engineer Research and Development Center, Tech. Rep. ERDC/GSL TR-02-16, 2002b.
- Billings, S.D., and F. Hermann, "Automatic detection of position and depth of potential UXO using a continuous wavelet transform", in Proceedings of SPIE Vol. 5089, *Detection and Remediation Technologies for Mines and Minelike Targets VII*, pp 1012-1022, 2003.
- Billings, S.D., 2004, Discrimination and classification of buried unexploded ordnance using magnetometry: *IEEE Transactions of Geoscience and Remote Sensing*, 42, 1241 - 1251.
- Collins, L., Zhang, Y., Li, J., Wang, H., Carin, L., Hart, S., Rose-Phersson, S., Nelson H. & McDonald, J. R., 2001, A comparison of the performance of statistical and fuzzy algorithms for unexploded ordnance detection: *IEEE Transactions on Fuzzy Systems*, 9, 17-30.
- Carin, L., Zhang, Y. & Liao, X., 2004, Detection of Buried UXO via Active Selection of Labeled Data, proceedings from the UXO Forum, St. Louis, March 9-12, 2004.
- Foley, J., Mehl, R., Fonda, R. Glenn, T., and M. Blair, Sensor Orientation Effects on UXO Geophysical Target Discrimination, SERDP & ESTCP Symposium & Workshop, Washington DC, Dec 2-5, 2004
- Grimm, R. E., 2003, Triaxial modeling and target classification of multi-channel, multicomponent EM data for UXO discrimination: *Journal of Environmental and Engineering Geophysics*, 8, 239-250.
- Hart, S.J. et al. (2001) Using Physics-based modeler Outputs to Train Probabilistic Neural Networks for Unexploded Ordnance (UXO) Classification in Magnetometry Surveys. *IEEE Trans. Geosci. Remote Sensing* 39, 797-804.

- Hwang, J. N., Choi, J.J., Oh, S. & Marks, R. J., 1991, Query-Based Learning Applied to Partially Trained Multilayer Perceptrons. *IEEE Trans. Neural Networks* 2, 131-136.
- Lewis D. D. & Gale W.A., 1994, A Sequential Algorithm for Training Text Classifiers., proceedings from 17th ACM International Conference on Research and Development in Information Retrieval.
- Munkholm, M.S. & Auken, E., 1996, Electro-magnetic noise contamination on transient electromagnetic soundings in culturally disturbed environments: *Journal of Environmental and Engineering Geophysics*, 1,119-127.
- Nelson, H. H., Bell, T. H., McDonald, J. R. & Barrow, B., 2003, Advanced MTADS classification for detection and discrimination of UXO, Technical report, Naval Research Laboratory.
- Pasion, L. & Oldenburg, D. 2001, A Discrimination Algorithm for UXO Using Time Domain Electromagnetics: *Journal of Engineering and Environmental Geophysics*, 28, no. 2, pages 91-102.
- Pasion, L. R., Billings, S. D. & Oldenburg, D. W., Joint and Cooperative Inversion of Magnetics and Electromagnetic Data for the Characterization of UXO Discrimination Problems, *The Symposium on the Application Geophysics to Engineering and Environmental Problems (SAGEEP)*, San Antonio, Texas, U.S.A., Feb 6-10, 2003
- Smith, J. T., H. F. Morrison and A. Becker, 2004a, Parametric Forms and the Inductive Response of a Permeable Conducting Sphere, *J. Environmental and Engineering Geophysics*, 9, 213-216.
- Smith, J. T., H. F. Morrison and A. Becker, 2004b, Resolution Depths for Some Transmitter-Receiver Configurations, *IEEE Transactions on Geoscience and Remote Sensing*, 42(6), pp. 1215-1221.
- Zhang, Y., Collins, L. M., & Carin, L., 2003a, Model-based statistical signal processing for UXO discrimination: Performance results from the JPG-V demonstration: *Proceedings of SPIE Volume 5089*, 1116-1126.
- Zhang, Y., Collins, L., Yu, H., Baum, C. E. & Carin, L., 2003b, Sensing of Unexploded Ordnance with Magnetometer and Induction Data: Theory and Signal Processing. *IEEE Trans. Geosci. Remote Sensing*, 41, 1005-1015.
- Zhang, Z., Liao, X. & Carin, L., 2004, Detection of buried targets via active selection of labeled data: Application to sensing subsurface UXO. *IEEE Transactions on Geoscience and Remote Sensing*: 42, 2535–2543\.

APPENDIX A

Magnetic Positioning Problems

In this appendix, we discuss positional problems with the magnetometer data that prevented their use for cooperative inversion. We also apply the magnetic remanence metric to 37 mm projectiles at the 20 mm Range-Fan (Grid 19-14) and MK-23 practice bombs at the Rocket Range (Grids I-12, I-13, J-12, J-13).

1. Positioning Issues with Magnetometer Data

Magnetometer data were collected on all eight Rocket Range and both 20 mm Range-Fan grids. The surveys were conducted with a man-portable quad-sensor magnetometer array. Primary positioning was provided by a Leica TPS 1206 RTS with a Crossbow IMU used to make minor adjustments to the primary positions. Despite our best efforts with the processing we were unable to produce magnetic data with positional accuracy acceptable for cooperative inversion. For instance Figure A-1 shows the data and model fit to an emplaced 37 mm projectile on Grid 19-14 in the 20 mm Range-Fan. The model and data match well for the series of four lines directly over the anomaly. However, on the return transect, the magnetic positive lies about 44 cm further to the south than what it should if the dipole model is correct. This represents a 22 cm bias in the North-South location (the bias contributes to the error in both transect directions). If we adjust parameters such as magnetometer to RTS time-difference, or the lever arms between sensors and RTS, we can get a better match to this dipole. However, this will then introduce positional error into other parts of the survey. We could not find an acceptable set of parameters that produced well positioned data throughout the whole survey. We believe that there are three reasons for the poor positional performance:

- (i.) The pole with the RTS prism was approximately 1.75 meters above the sensors and was at right-angles to the sensor frame. When walking the frame is typically held at an angle of 15 degrees so that the sensors are the correct height above the ground. This placed the RTS prism about 45 cm in-front of the sensors. This was a deviation from our standard operating procedures, where the RTS (or GPS) antenna is always configured so that it is directly above the sensors when surveying (the field-crew and QA officer did not properly understand the standard operating procedure [SOP]). To correct for the RTS/sensor location difference, we had to rely heavily on the pitch and roll values returned by the IMU. As per (ii) below, these were not always reliable and hence considerable positional error was introduced;
- (ii.) The IMU was on the rear of the sensor array about 75 cm from the center of gravity of the system. As data were collected the gait of the operator caused a lot of up/down, backwards/forwards motion of the IMU. These accelerations caused a degradation in the accuracy of the IMU pitch and roll, with the error growing over time. Zero-velocity updates were used to minimize these errors, but were not collected often enough given the non-ideal position of the IMU. The front of the array, directly above the sensors would have provided a more suitable location (presuming the IMU did not interfere with the magnetometer readings); and

- (iii.) During data collection, the array tended to sway back-and-forth due to the gait of the operator as he/she walked across the survey area. This caused the velocity of the RTS prism to fluctuate and created difficulties for the tracking mechanism in the Leica base-station (see example data stream in Table A-1). It would track the prism, then lose it temporarily, then overshoot the prism location as it tried to regain lock. The tracking problems translated to along-line positional errors on the order of tens of centimeters. To mitigate this problem, the operator needs to carefully control the motion of the quad-sensor array while they walk. This is difficult to do, but can be learned through sufficient practice.

Table A-1. Sample of RTS data showing the time and distance increments between successive measurements. At points 14 to 17, the distance increments vary widely causing a large change in apparent velocity. We believe that these fluctuations are caused by back-and-forth movements of the RTS prism that create problems for the tracking algorithm in the Leica base-station.

Point	Easting (m)	Northing (m)	Time increment (s)	Distance Increment (cm)
1	527560.7	4384800	0.183	26.2
2	527560.75	4384800.2	0.185	6.5
3	527560.76	4384800.3	0.174	26.1
4	527560.74	4384800.6	0.181	25.0
5	527560.72	4384800.8	0.140	16.1
6	527560.72	4384801	0.183	14.1
7	527560.73	4384801.1	0.187	18.0
8	527560.76	4384801.3	0.154	28.9
9	527560.79	4384801.6	0.226	25.4
10	527560.78	4384801.8	0.180	13.1
11	527560.78	4384802	0.138	25.9
12	527560.76	4384802.2	0.222	30.2
13	527560.75	4384802.5	0.180	17.1
14	527560.79	4384802.7	0.188	42.5
15	527560.87	4384803.1	0.116	0.0
16	527560.87	4384803.1	0.154	10.1
17	527560.88	4384803.2	0.191	32.7
18	527560.89	4384803.5	0.175	5.7

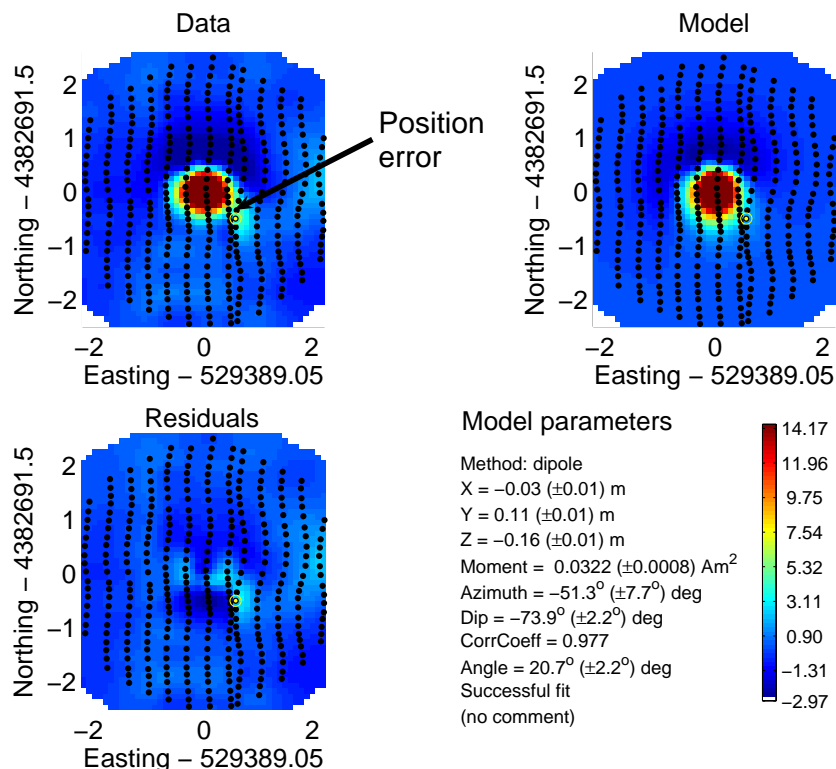


Figure A-1: Data, dipole model, residuals and model parameters for a 37 mm projectile on Grid 19-14 (anomaly 1 in Figure A-2). Observation locations are shown as black dots. On the series of 4 lines (one transect) directly over the dipole, there is a good match between model and data. For the return transect to the East, the fit is poor due to a North-South positional error in the data.

2. Analysis of Grid 19-14 in the 20 mm Range-Fan

With the above positional issues in mind, we conducted an analysis on Grid 19-14 on the 20 mm Range-Fan to determine if the magnetometer data would be usable in a cooperative inversion process. Figure A-2 shows a total field image with the locations of nineteen 37 mm projectiles and twenty-four 20 mm projectiles shown. All the 37 mm projectiles produce obvious anomalies, while many of the 20 mm are not detected by the magnetometer. As per the retrospective analysis, the big challenge with the EM sensors was the ability to constrain the position and depth of the object. Uncertainty in depth translates to uncertainty in size, and our classification results showed that size is a powerful discrimination metric. Magnetometers are not particularly sensitive to small objects and we would not expect the cooperative inversion process to aid the feature extraction for the smaller 20 mm projectiles, or the 50 caliber bullets. We inverted all nineteen 37 mm projectiles to determine if the magnetic data could be used to constrain the EM-fits to those objects.

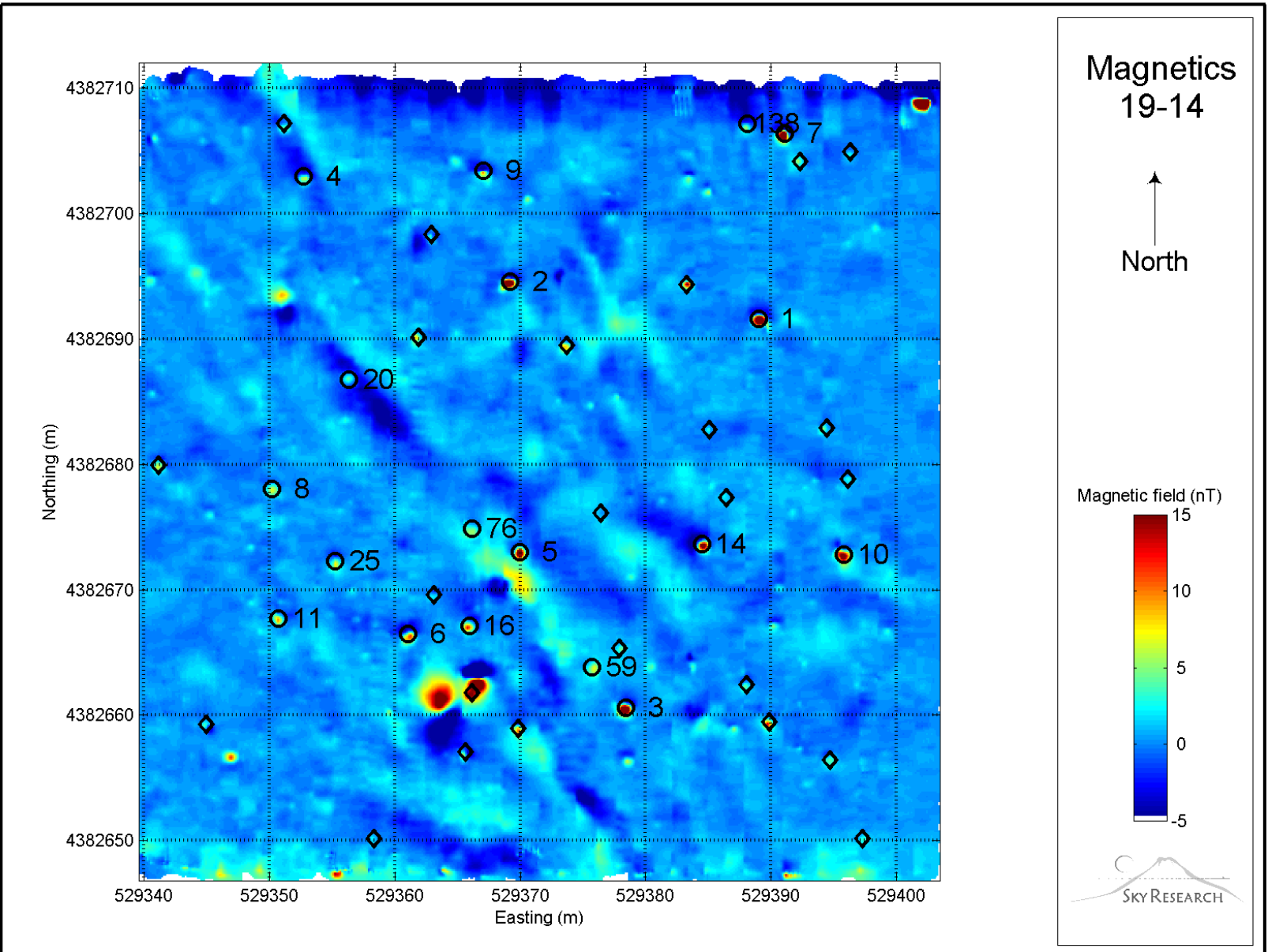


Figure A-2: Image of the total magnetic field on Grid 19-14. The locations of 37 mm rounds detected by the EM-63 are labeled and shown as circles, 20 mm rounds as diamonds.

Figure A-3 compares the positional and depth errors of the EM-63 and magnetometer fits to the 37 mm projectiles on 19-14. As the cumulative distribution plots show, the magnetometer positions and depths are as bad as or worse than the EM-63 positions and depths. Thus, we conclude that cooperative inversion would not help to constrain the polarization parameters of the 37 mm projectiles.

Figure A-4 compares the recovered moments of the 37 mm projectiles to *dipole feasibility curves* (Billings, 2004) for 20 mm and 37 mm projectiles and 60 mm mortars. The dipole feasibility curves represent the moments a given object can produce through induced magnetization alone. The recovered moments for the 37 mm projectiles tend to cluster around the feasibility curve for the 37 mm projectiles and would produce relatively low estimates of remanent magnetization (Figure A-4(b)). This is a discrimination metric that we have used successfully for 76 mm projectiles and larger in Montana. The results shown here demonstrate that remanence may prove useful for projectiles as small as 37 mm in caliber.

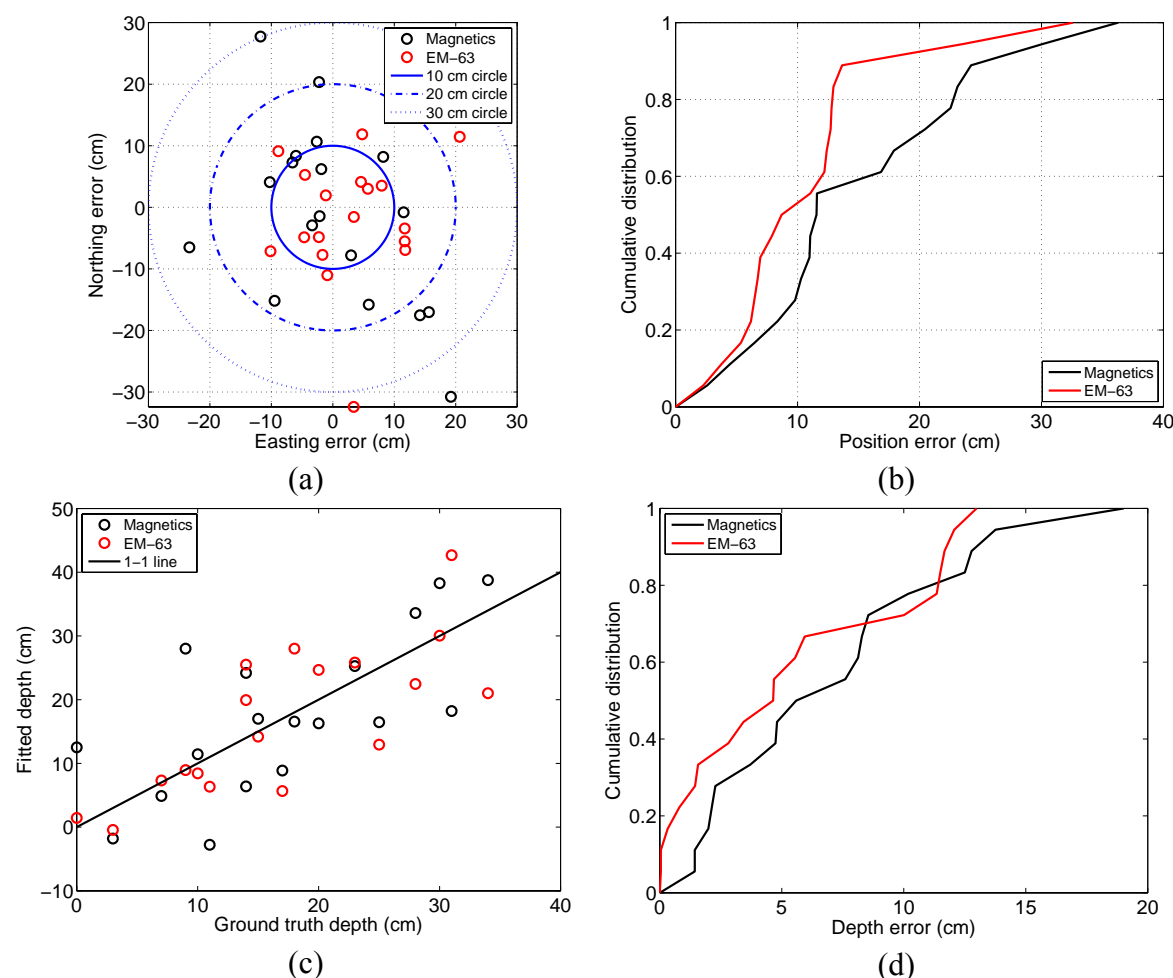


Figure A-3. Comparison of ground-truth and magnetometer and EM-63 fitted positions and depths of 37 mm projectiles on Grid 19-14: (a) Error in position; (b) Cumulative distribution of position error; (c) Fitted versus ground-truth depths; and (d) Cumulative distribution of depth errors.

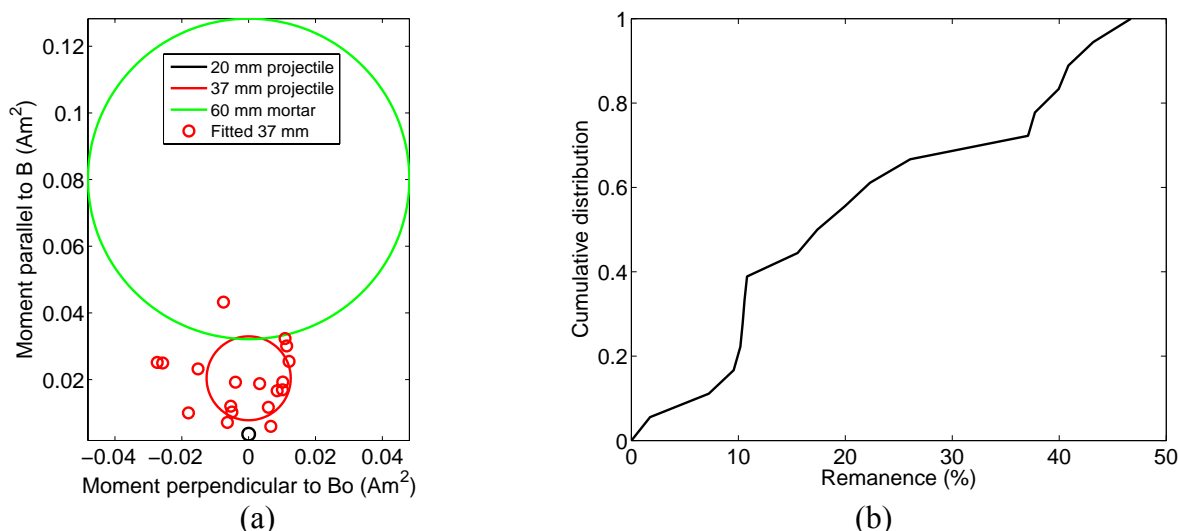


Figure A-4. Application of magnetic remanence metric to Grid 19-14: (a) Plot of recovered moments of 37 mm projectiles against dipole feasibility curves for 20 and 37 mm projectiles and 60 mm mortars; (b) Cumulative distribution of the remanence metric.

3. Analysis of MK-23 Rounds on the Rocket Range

An additional complication with the Rocket Range data was that the RTS coordinate system was set-up using poorly positioned survey control. Instead of using the accurate locations established by the on-site surveyor, the field crew mistakenly used the roughly positioned grid markers to perform their set-up. This problem was compounded by the field-crew neglecting to check positional accuracy on an independent third point (part of the SOP). The incorrect set-up was not detected by the on-site QA officer and had to be rectified post-survey by identifying the grid corners and translating and rotating those until they closely matched the true locations. Easting values had to be translated by 3.871 meters East and Northing values by 1.953 meters South. In addition, the locations had to be rotated by 1.8 degrees counter-clockwise around the origin of Grid I12.

Figure A-5 shows the resulting detrended magnetic field image for Grids I12, I-13, J-12 and J-13 with locations of MK-23 rounds marked. There are some geological features present within the region, but these do not effect the detection of the MK-23 rounds. All of them have relatively well defined dipole signatures. Figure A-6 compares the depths returned by inversion of the magnetometer data with depths predicted by the EM-63 inversions. Around 60% of the magnetometer inversions are within 5 cm of the true depths which is slightly better than the performance of the EM-63. The largest error is 17 cm compared to over 25 cm for the EM-63. The results demonstrate that the magnetometer data may be able to provide useful constraints to the EM-63 data. The problem with using the magnetometer inversions as constraints in a cooperative inversion process is that the Easting and Northing locations of the magnetics do not agree very well with the ground truth (probably due to the RTS correction procedure we used).

The magnetic data are positioned well enough that estimates of the magnetic dipole moment can be used to estimate the magnetic remanence metric (Figure A-7). The recovered dipole moments for the MK-23s tend to cluster around the lower half of the dipole feasibility curve for a MK-23 practice bomb (obtained by a spheroid approximation to the shape and size of the item). This is consistent with the ground-truth observation that most MK-23s were found within 30 degrees of horizontal (thus the semi-major axis of the item is at a large angle to the Earth's magnetic field). All Mk-23s have a predicted remanence metric of less than 50%, with 18 of the 29 rounds having a remanence of 20% or less. As with the 37 mm rounds in the 20 mm Range-Fan, this analysis indicates that remanence would be a very useful discrimination diagnostic.

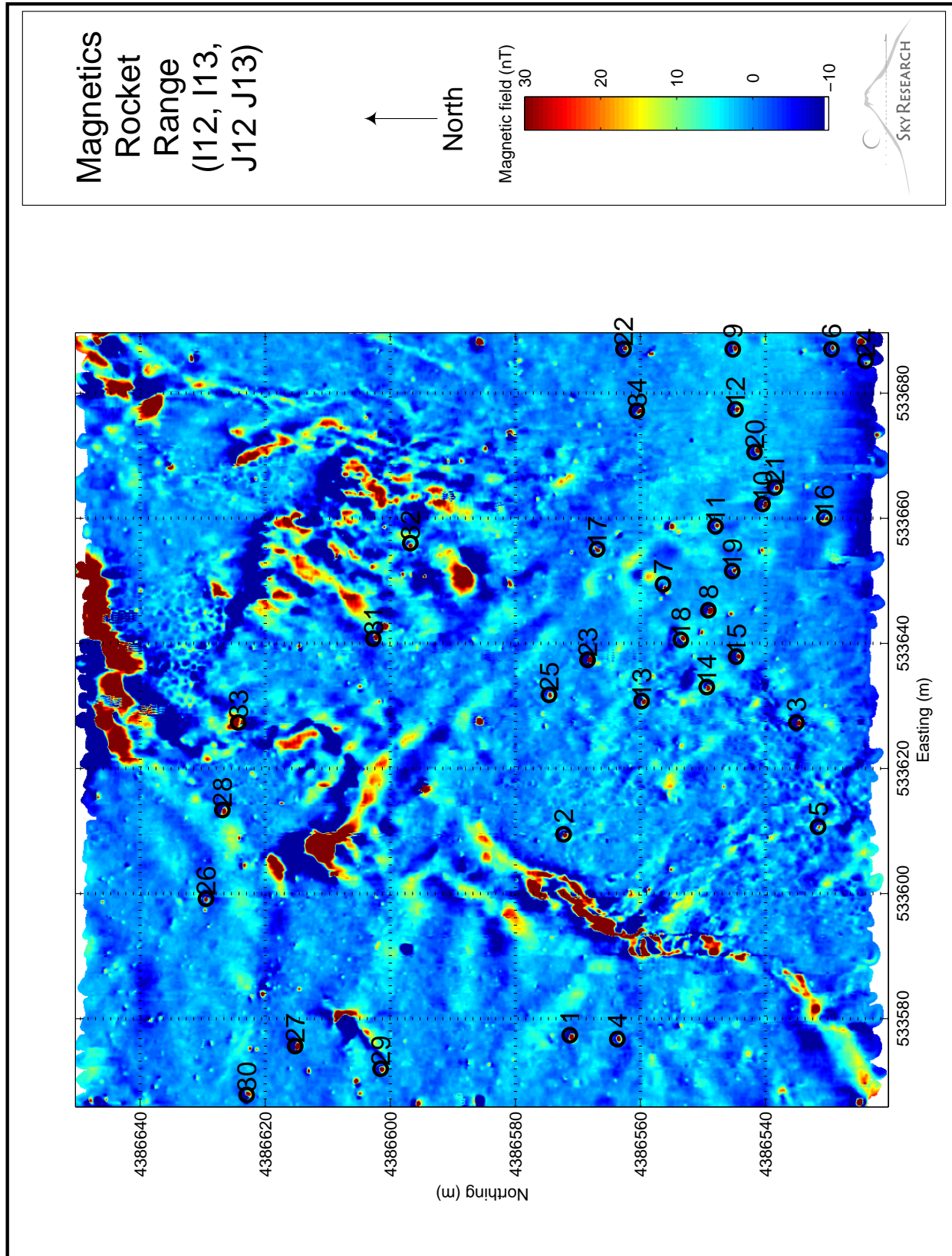


Figure A-5. Image of the total magnetic field on Rocket Range Grids I12, I-13, J-12 and J-13. The locations of MK-23 practice bombs detected by the EM-63 are labeled and shown as circles.

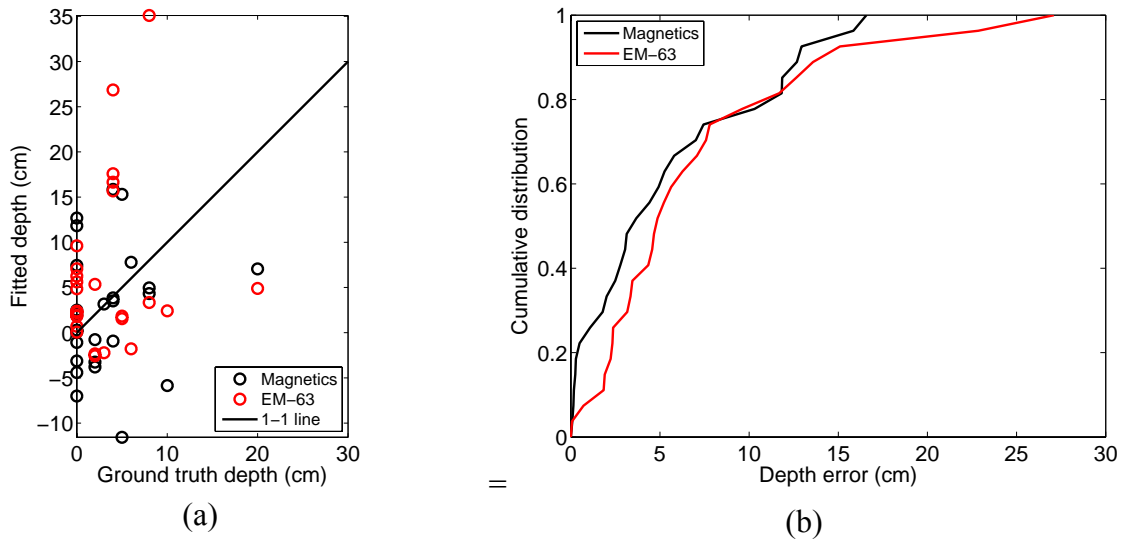


Figure A-6. Comparison of ground-truth and magnetometer and EM-63 fitted depths of MK-23 practice bombs at the Rocket Range: (a) Fitted versus ground-truth depths; and (c) Cumulative distribution of depth errors.

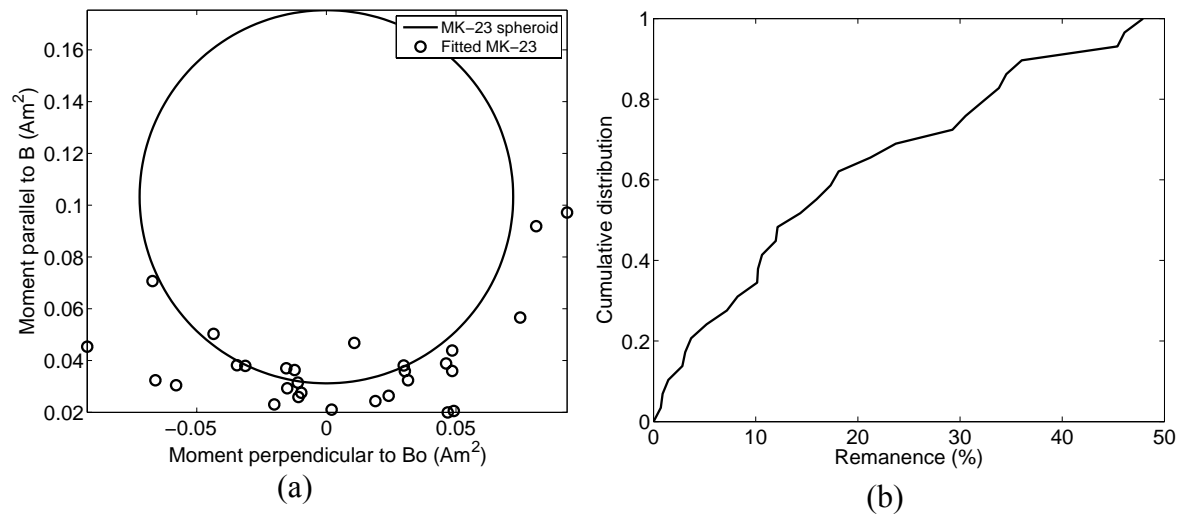


Figure A-7. Application of magnetic remanence metric to MK-23 rounds at the Rocket Range: (a) Plot of recovered moments of MK-23 practice bombs against the dipole feasibility curve for an MK-23 practice bomb; (b) Cumulative distribution of the remanence metric.

APPENDIX B

Selection of Excavation Threshold

Another important component of the demonstration is the criterion used to select the anomalies to excavate to test the discrimination performance. This threshold needs to be essentially independent from the classification so we can conduct unbiased tests of performance. In the demonstration plan we indicated that we would chose an amplitude threshold that was equivalent to what would be chosen in a production UXO clearance. In reality we have chosen the threshold to be slightly lower than that used for production work, to determine if there are anomalies below the production threshold that would be found by discrimination.

The analysis presented in this appendix was submitted to the Program Office at the same time as the interpretations for the Phase I grids.

1. EM-61 Array Threshold

Production work at the Rocket Range is conducted under the assumption that the smallest items of concern are a 3-pound MK-23 practice bomb or a 57 mm HEAT M307 projectile. For production work at the site, excavation thresholds were established to detect these items down to a maximum depth of 1 foot (30 cm). On the FLBGR test plot there are no 57 mm projectiles and a number of MK-23 at depths ranging from a little over a half-foot (18 cm) to 4 feet (1.2 m). Unfortunately there is no MK-23 at 1 foot. For the regular production work that conducted on site, a hole was dug an MK-23 was emplaced at a one foot depth. For the SKY 3-coil EM-61 array (three 1 m by 0.5 m coils arranged to make an effective transmitter of 3 m x 0.5 m in dimension), the MK-23 produced a maximum response of just over 15 mV in the third-time-channel of the center coil. Based on this test and the amplitudes at the 0.58 and 2 feet depths, the production threshold was set at 15 mV.

The array deployed for this demonstration comprised five 1 m by 0.5 m EM-61 coils arranged side by side such that they approximate a 2.5 m by 1 m transmitter. As the transmitter current for the 3-coil and 5-coil cases are approximately the same, the 5-coil transmitter will have a moment that is $(2.5 \times 1)/(3 \times 0.5) = 5/3$ larger. This means that the amplitudes for the 5-coil case should be around $5/3$ times larger⁴. To verify that this is indeed the case, we plot the maximum amplitudes on the FLBGR test-plot for the 3-coil and 5-coil cases, where both datasets were collected east-west (Figure B-1). As expected there is some scatter in the amplitudes due to geometry differences but overall the amplitudes follow the expected $5/3$ ratio. Indeed least squares regression predicts a best fitting line with a ratio = 1.689 (compared to $5/3=1.667$). Thus an equivalent production threshold for the 5-coil EM-61 is $5/3 \times 15 \text{ mV} = 25 \text{ mV}$.

⁴ The maximum amplitude will depend on the specifics of the sensor to target geometry, which sensor passes closes to the object, differences in the spatial characteristics of the primary field etc.

For the first set of grids we have used a threshold of 15 mV which is a little lower than the production equivalent of 25 mV. This was selected to determine if there are munitions of concern just below the production threshold that could be correctly classified as UXO. This would support the case for discrimination where all available features are used to make excavation decisions, not just a simplistic criterion based on amplitude.

For the 20 mm Range-Fan, the main item of concern is the 37 mm projectile and these are expected to be cleared to a depth of $\frac{1}{2}$ foot (15.2 cm). The challenge is to recover all 37 mm projectiles and reject a lot of the 20 mm projectiles, 50 cal bullets and shrapnel at the site. In 2005 during the EM-61 survey of the 20 mm Range-Fan, six 37 mm projectiles were emplaced in Grid 20-14 at depths ranging from 7.6 cm to 22.9 cm. At the 15.2 cm depth, the amplitude in the third time-channel was at least 50 mV (Table B-2). At the maximum emplacement depth of 22.9 cm the amplitude was 33 mV or greater. If we use the same 15 mV threshold as the Rocket Range, then we should recover all 37 mms at $\frac{3}{4}$ of a foot (22.9 cm) and may get some 37 mms at deeper depths.

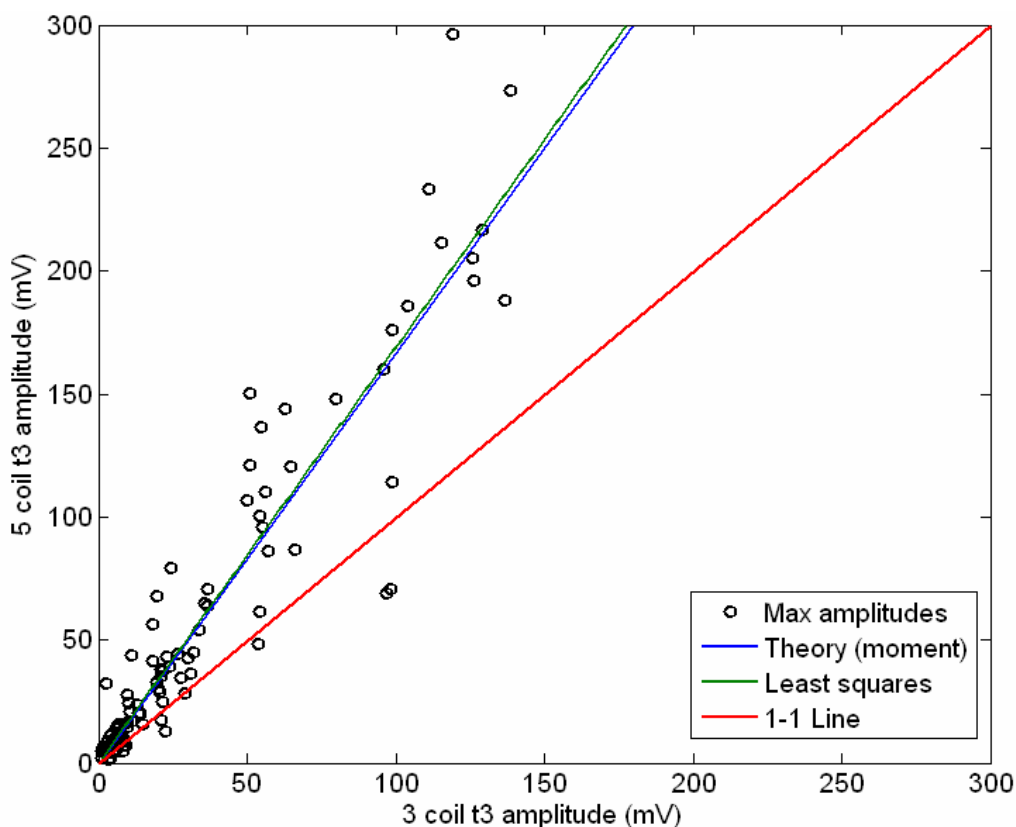


Figure B-1. Comparison of the amplitudes on the test-plot of time channel 3 of the 5-coil SKY-array to the 3-coil array used for production work. The 5-coil amplitudes are (on average) $5/3$ times larger than the 3-coil array. This difference is due to the larger moment of the 5-coil array.

Table B-1. Number of anomalies in each of the 4 tiles with amplitudes > 25 mV, between 15 and 25 mV and > 15 mV.

Grid	> 25 mV	15 to 25 mV	Total > 15 mV
I-12	24	30	54
J-13	25	31	56
19-14	15	50	65
20-14	25	81	106

Table B-2. Maximum amplitudes in time-channel 3 for the six 37 mm projectiles emplaced in Grid 20-14.

Name	Easting	Northing	Depth (cm)	Azimuth	Dip	t3 (mV)
1	529406.37	4382695.29	7.6	137	86	135.9
2	529409.21	4382669.25	7.6	-69	3	80.3
3	529409.48	4382692.07	15.2	-147	85	106.0
4	529439.13	4382702.78	15.2	20	8	50.2
5	529413.23	4382692.93	22.9	-56	83	41.8
6	529437.34	4382700.65	22.9	50	3	33.1

2. Threshold for the EM-63

On the test-plot, the maximum amplitude in time-channel 1 of the EM-63 is 63 mV for the one MK-23 at 0.58 foot depth, and between 5-6 mV for the two separate MK-23s at a depth of two feet. It is difficult to extrapolate the expected amplitude value at the required one-foot depth from these results as the observed amplitude depends on a number of extraneous factors (closest point of approach, orientation of the item etc). We consider the EM-61 as the primary detection device and for the Rocket Range we just selected a value for the EM-63 that produced a comparable number of targets. For the first two-grids (I12 and J-13), we settled on a value of 15 mV. Once we determine what items are recovered in the lower amplitude range, we may increase the threshold to prevent unnecessary excavations of small items.

The main item of interest in the 20 mm Range-Fan is the 37 mm projectile. As no representatives of these were emplaced in the test-plot, Sky Research emplaced 19 rounds in grids 19-14 and 21-14 prior to survey with the EM-63. The rounds were emplaced at depths ranging from 7.6 cm (1/4 foot) to 30.5 cm (1 foot). The depths of burial, location, orientation and maximum amplitude in channel t₁ are shown in Table B-3. For a detection depth of ½ foot the minimum amplitude is 93 mV. At one foot, the amplitude ranges from 28.6 mV up to 73.6 mV, with only one item below 35 mV. We elected to set our threshold at 35 mV which requires excavation of 42 holes in 19-14 and 82 holes in 21-14. Lowering the threshold to 25 mV (Table B-4) would have required the excavation of an additional 214 items (which we feel is logistically infeasible). As a QC

measure we have elected to excavate 10% of those anomalies in the 25-35 mV range (in addition, there are a number of anomalies that were above the EM-61 threshold but below the EM-63 threshold that will be excavated).

Table B-3. Maximum amplitude at time-channel 1 of the EM-63 over the 19 emplaced 37 mm projectiles in the 20 mm Range-Fan. Only one anomaly (marked in pink) has an amplitude less than 35 mV.

Label	Northing	Easting	Depth (cm)	Azi	Dip	t1 (mV)
1	4382702.87	529352.94	7.6	-131	0	150.3
2	4382703.34	529367.24	7.62	-86	30	101.9
3	4382706.41	529391.06	7.62	-171	45	113.8
4	4382694.50	529369.13	7.62	-173	60	308.1
5	4382691.61	529389.25	7.62	-20	90	314.7
6	4382678.05	529350.26	15.2	-130	0	110.6
7	4382672.94	529395.75	15.24	-165	30	93.3
8	4382673.10	529369.83	15.24	-73	45	150.3
9	4382666.37	529361.04	15.24	-1	60	135.1
10	4382660.78	529378.20	15.24	169	90	151.1
11	4382705.51	529488.54	22.9	-126	0	45.8
12	4382706.35	529506.01	22.86	12	30	65.9
13	4382691.06	529475.76	22.86	6	45	55.5
14	4382689.97	529494.14	22.86	136	60	111.1
15	4382668.50	529480.86	22.86	-131	90	133.1
16	4382666.66	529514.11	30.5	-86	0	28.6
17	4382656.89	529487.12	30.48	-171	30	45.6
18	4382656.47	529497.57	30.48	-173	45	73.6
19	4382661.33	529524.30	30.48	-20	60	36.0

Table B-4. Number of anomalies in each grid at different amplitude thresholds.

Grid	> 35 mV	25 to 35 mV	15 to 25 mV	Total
I-12	29	11	38	78
J-13	25	20	49	94
19-14	42	82 (8)	NA	124 (50)
21-14	82	132 (13)	NA	214 (95)

3. Total Number Anomalies Recommended for Excavation

The amplitude based dig-sheets for the EM-61 and EM-63 were merged using the criterion that the anomalies were from the same source if they were within 1 meter of each other. There were then a number of anomalies that were only above the threshold in one dataset or the other, and all of these were selected for excavation.

For I12 all anomalies classified as potential UXO by the statistical classifier were in the amplitude dig-sheet (88 anomalies). For J-13, there were 13 items classified as UXO that were left off the 102 anomalies in the amplitude dig-sheet. 7 of these were added as QC picks to that diglist. For the 19-14 dig-sheet there were a total of 81 anomalies, but only 17 of 78 anomalies identified as potential UXO (by the EM-63) were on that dig-sheet. This is evident in Figure B-2 below, which shows that there is very little correspondence between high amplitude anomalies on the production diglist and anomalies ranked as potential UXO. Consequently, we decided to add all the extra anomalies classified as UXO to the dig-sheet to determine if there are UXO that we would miss using the amplitude criteria alone.

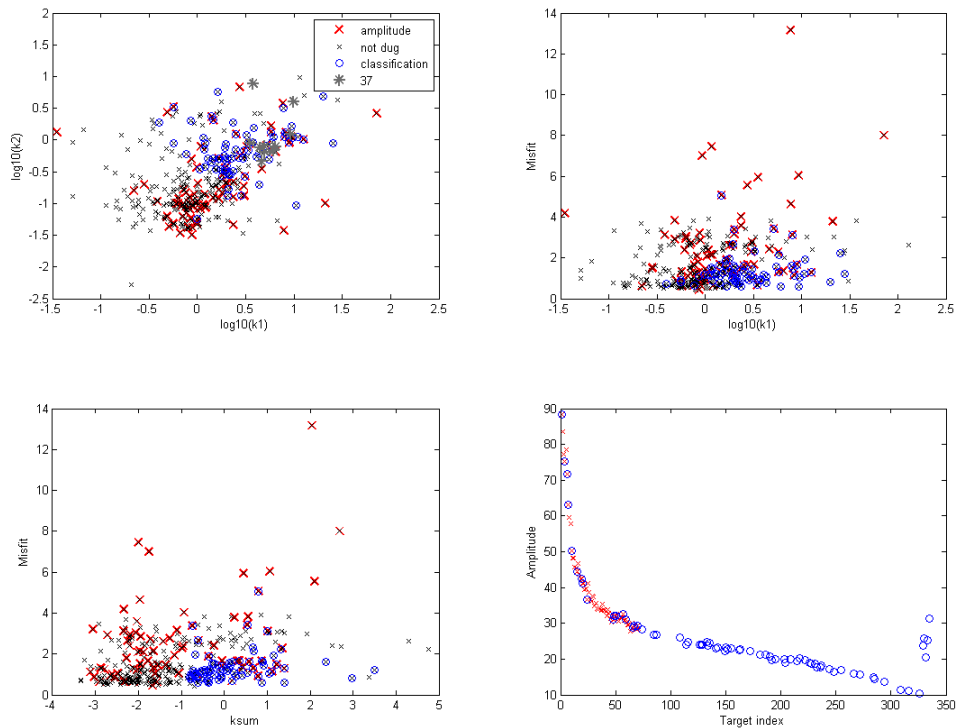


Figure B-2. Comparison of the anomalies on the amplitude only diglist for 19-14 with those anomalies classified as potential UXO via statistical classification. There is little overlap between the two dig-sheets.

APPENDIX C

Table C-1: List of Items Buried in the FLBGR Test Plot

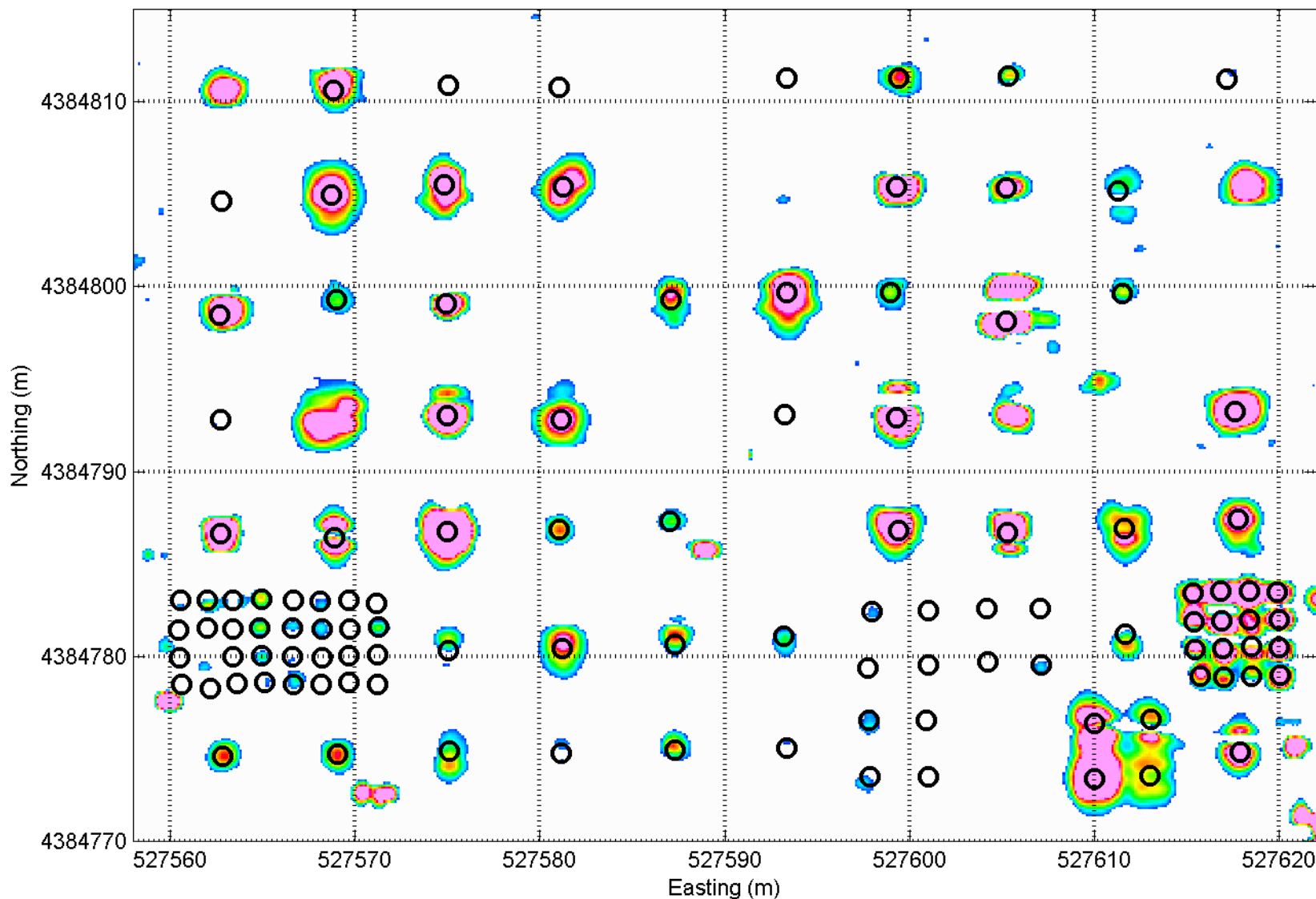
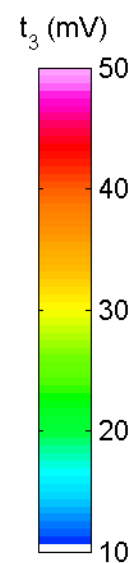
Cell	Item(s)	Az	Dip	Weight (lb)	Depth (ft)	East (m)	North (m)
1	Barbed/Telecom Wire	0	0	5	0 to 1	527562.90	4384810.68
2	M69 (4)	0	0	17	3	527568.87	4384810.61
3	Frag (small, misc)	0	0	10	2	527575.03	4384810.88
4	MK23 (1)	0	60	3	3	527581.03	4384810.76
6	OE Scrap	0	0	22	3	527593.04	4384811.12
7	M69 (1)	90	0	4	2	527599.01	4384811.21
8	MK23 (1)	0	0	3	0.58	527605.04	4384811.18
9	M38	90	45		2	527611.10	4384811.17
10	OE Scrap	90	45	30	4	527617.14	4384811.23
11	SCAR (1)	0	30	6	3	527562.77	4384804.64
12	Rusted PIGs (2) - 3' apart	90	45	68	3	527568.96	4384804.61
13	Non-OE Scrap	0	0	15	1	527575.12	4384804.85
14	SCAR (4)	45	15	35	2	527581.06	4384804.76
16	M69	-90	0		1.5	527593.12	4384804.90
17	Barbed Wire & Frag	0	30	12	1	527599.10	4384805.08
18	OE Scrap	0	0	15	1	527605.07	4384805.05
19	SCAR (4)	0	0	35	4	527611.28	4384805.17
20	SCAR (1) & SCAR (2)	45	45	24	4 & 2	527617.38	4384805.01
21	SCAR (4)	0	60	135	3	527562.71	4384798.49
22	OE Scrap	90	90	30	3	527568.90	4384798.73
23	MK23 (3)	0	60	9	0.58	527574.99	4384798.70
25	M69 (1)	0	0	2	2	527587.09	4384798.63
26	Frag - Large	0	0	30	3	527593.18	4384798.93
27	SCAR (2)	90	45	16	3	527599.15	4384798.96
28	OE Scrap	0	0	40	1	527605.16	4384799.02
29	OE Scrap	0	60	30	2	527611.31	4384798.98
31	MK23 (2)	30	0	7	2	527562.77	4384792.46
32	PIG (1) & M69 (1) - 5' apart	0	0	44	4 & 2	527568.89	4384792.67
33	Frag - Large	0	45	33	1	527574.80	4384792.54
34	PIG (1)	0	30	52	4	527580.90	4384792.81
36	MK23 (1)	0	0	3	2	527593.33	4384792.81
37	M69 (1) & MK23 (2)	0	0	20	1	527599.30	4384792.90
38	2.25" Rocket SCAR (2)	0	0, 90		3	527605.24	4384792.80
40	SCAR (1) & 100-lb bomb (1)	90	0	16	2	527617.31	4384792.89
41	Frag - Large	0	0	15	1	527562.77	4384786.45
42	SCAR (1)	0	0	10	2	527568.92	4384786.45
43	100-lb bomb	0	0	23	1	527575.08	4384786.57

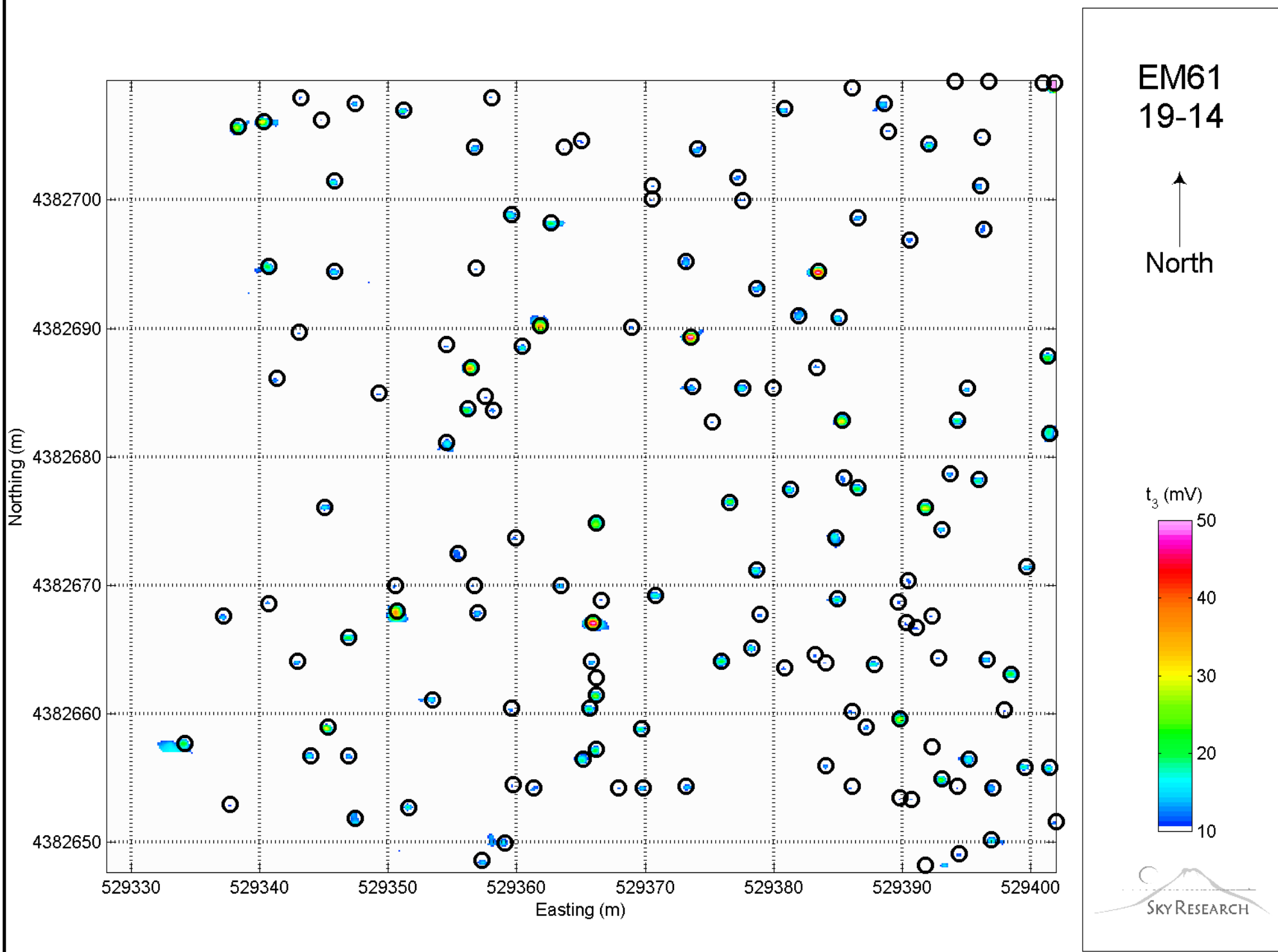
Cell	Item(s)	Az	Dip	Weight (lb)	Depth (ft)	East (m)	North (m)
44	2.75" Warhead (1)	0	60	6	2	527580.80	4384786.57
45	100-lb bomb (1) & MK23 (3)	0	0	35	3	527586.93	4384786.75
47	Scrap	0	45	48	3	527599.42	4384786.83
48	2.75" Warhead (3)	0	0	12	1	527605.27	4384786.71
49	SCAR (2) & M69 (2)	0	0	24	3	527611.31	4384786.86
50	SCAR (2) & 100-lb bomb (1)	0	0	26	2	527617.34	4384786.73
51.1	20mm (1)	0	-30	0.23	0.25	527560.55	4384783.02
51.2	20mm (1)	0	60	0.23	0.25	527562.05	4384783.04
51.3	20mm (1)	0	0	0.23	0.25	527563.39	4384783.04
51.4	20mm (1)	0	90	0.23	0.25	527564.94	4384783.12
51.5	20mm (1)	0	-30	0.23	0.5	527560.45	4384781.44
51.6	20mm (1)	0	60	0.23	0.5	527562.04	4384781.52
51.7	20mm (1)	0	0	0.23	0.5	527563.41	4384781.48
51.8	20mm (1)	0	90	0.23	0.5	527564.87	4384781.54
51.9	20mm (1)	0	-30	0.23	0.75	527560.54	4384779.94
51.1	20mm (1)	0	-30	0.23	0.25	527560.55	4384783.02
51.11	20mm (1)	0	0	0.23	0.75	527563.43	4384779.99
51.12	20mm (1)	0	90	0.23	0.75	527564.99	4384780.04
51.13	20mm (1)	0	-30	0.23	1	527560.62	4384778.48
51.14	20mm (1)	0	60	0.23	1	527562.08	4384778.60
51.15	20mm (1)	0	0	0.23	1	527563.61	4384778.53
51.16	20mm (1)	0	90	0.23	1	527565.10	4384778.61
52.1	Misc Fuzes (1)	0	-30	0.09	0.25	527566.68	4384783.02
52.2	Misc Fuzes (1)	0	60	0.39	0.25	527568.14	4384783.01
52.3	Misc Fuzes (1)	0	0	0.10	0.25	527569.69	4384783.03
52.4	Misc Fuzes (1)	0	90	0.17	0.25	527571.19	4384782.87
52.5	Misc Fuzes (1)	0	-30	0.6	0.5	527566.61	4384781.53
52.6	Misc Fuzes (1)	0	60	0.5	0.5	527568.22	4384781.52
52.7	Misc Fuzes (1)	0	0	0.7	0.5	527569.75	4384781.51
52.8	Misc Fuzes (1)	0	90	0.5	0.5	527571.27	4384781.62
52.9	Misc Fuzes (1)	0	-30	0.6	0.75	527566.63	4384779.97
52.10	Misc Fuzes (1)	0	60	0.35	0.75	527568.22	4384779.96
52.11	Misc Fuzes (1)	0	0	0.7	0.75	527569.68	4384780.05
52.12	Misc Fuzes (1)	0	90	0.5	0.75	527571.23	4384780.10
52.13	Misc Fuzes (1)	0	-30	0.5	1	527566.68	4384778.48
52.14	Misc Fuzes (1)	0	60	0.5	1	527568.18	4384778.50
52.15	Misc Fuzes (1)	0	0	0.40	1	527569.67	4384778.62
52.16	Misc Fuzes (1)	0	90	0.5	1	527571.25	4384778.51
53	2.25" Rocket - SCAR	0	-30	5.9	3	527575.04	4384780.35
54	M38	0	-30	8.7	3	527581.20	4384780.41
55	M69	0	-30	4.1	3	527587.29	4384780.65

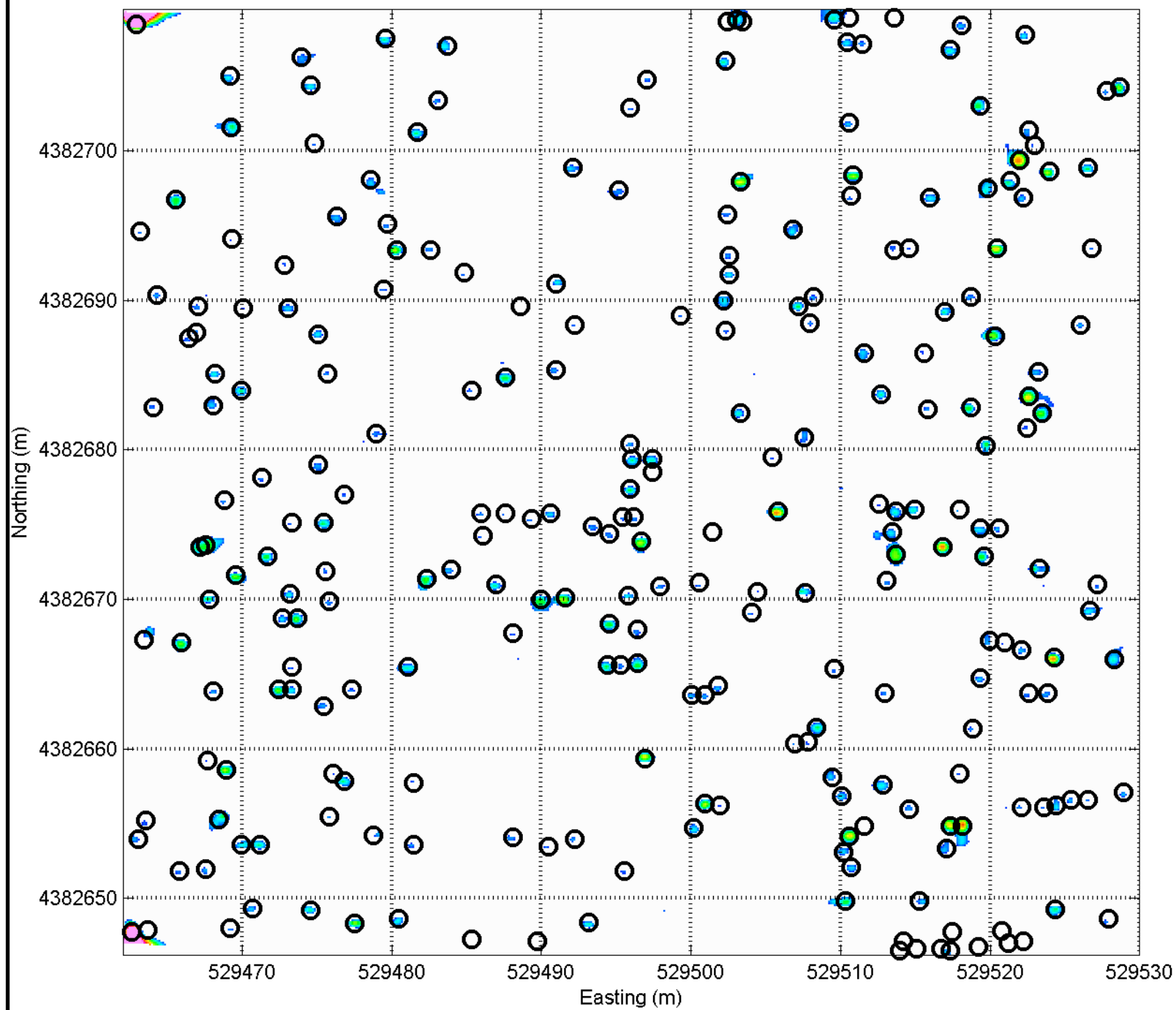
Cell	Item(s)	Az	Dip	Weight (lb)	Depth (ft)	East (m)	North (m)
56	M38	0	60	6.9	3	527593.21	4384781.08
57.1	M50 (1)	0	-30	1.5	1	527597.97	4384782.45
57.2	M50 (1)	0	60	1.5	1	527600.98	4384782.50
57.3	MK23 (1)	0	-30	2.25	2	527597.74	4384779.41
57.4	MK23 (1)	0	60	2.25	2	527601.00	4384779.57
58.1	M50 (1)	0	0	1.5	2	527604.15	4384782.60
58.2	M50 (1)	0	90	1.5	2	527607.05	4384782.62
58.3	MK23 (1)	0	0	2.25	4	527604.23	4384779.71
58.4	MK23 (1)	0	90	2.25	4	527607.12	4384779.54
59	M69	0	60	4.1	3	527611.64	4384781.19
60.1	37mm (1.5" x 6" Pipe)	0	-30	1.3	0.25	527615.34	4384783.45
60.2	37mm (1.5" x 6" Pipe)	0	60	1.3	0.25	527616.83	4384783.54
60.3	37mm (1.5" x 6" Pipe)	0	0	1.3	0.25	527618.36	4384783.56
60.4	37mm (1.5" x 6" Pipe)	0	90	1.3	0.25	527619.88	4384783.49
60.5	37mm (1.5" x 6" Pipe)	0	-30	1.3	0.5	527615.36	4384781.90
60.6	37mm (1.5" x 6" Pipe)	0	60	1.3	0.5	527616.89	4384781.92
60.7	37mm (1.5" x 6" Pipe)	0	0	1.3	0.5	527618.38	4384781.97
60.8	37mm (1.5" x 6" Pipe)	0	90	1.3	0.5	527619.96	4384782.00
60.9	37mm (1.5" x 6" Pipe)	0	-30	1.3	0.75	527615.44	4384780.41
60.10	37mm (1.5" x 6" Pipe)	0	60	1.3	0.75	527616.94	4384780.43
60.11	37mm (1.5" x 6" Pipe)	0	0	1.3	0.75	527618.46	4384780.57
60.12	37mm (1.5" x 6" Pipe)	0	90	1.3	0.75	527619.99	4384780.47
60.13	37mm (1.5" x 6" Pipe)	0	-30	1.3	1	527615.71	4384778.94
60.14	37mm (1.5" x 6" Pipe)	0	60	1.3	1	527616.96	4384778.87
60.15	37mm (1.5" x 6" Pipe)	0	0	1.3	1	527618.48	4384778.96
60.16	37mm (1.5" x 6" Pipe)	0	90	1.3	1	527620.04	4384778.98
61	M38	0	0	8.5	3	527562.85	4384774.63
62	M38	0	90	5.8	3	527569.04	4384774.75
63	M69	0	0	4.2	3	527575.13	4384774.90
64	M69	0	90	4.3	3	527581.17	4384774.80
65	2.75" Rocket Warhead	0	-30	5.6	2	527587.29	4384774.95
66	2.75" Rocket Warhead	0	0	5.7	2.5	527593.36	4384775.04
67.1	M50 (2); Separation: 2' N-S	0	0	2.25	1	527597.78	4384776.54
67.2	M50 (2); Separation: 2' N-S	0	0	2.25	2	527600.92	4384776.56
67.3	M50 (2); Separation: 4' N-S	0	0	2.25	1	527597.86	4384773.49
67.4	M50 (2); Separation: 4' N-S	0	0	2.25	2	527601.00	4384773.51
69.1	M69 (2); Separation: 2' E-W	0	0	4.2	2	527609.97	4384776.38
69.2	M69 (2); Separation: 2' E-W	0	0	4.2	3	527613.05	4384776.61
69.3	M69 (2); Separation: 4' E-W	0	0	4.2	2	527610.08	4384773.55
69.4	M69 (2); Separation: 4' E-W	0	0	4.2	3	527613.00	4384773.56
70	Surface Frag & M69	0	0	6.4	2	527617.88	4384774.82

Appendix D
Images of EM-61 and EM-63 Data

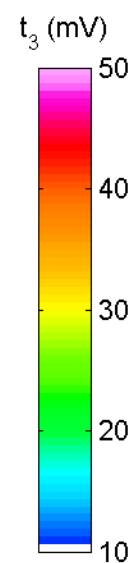
EM61
Testplot

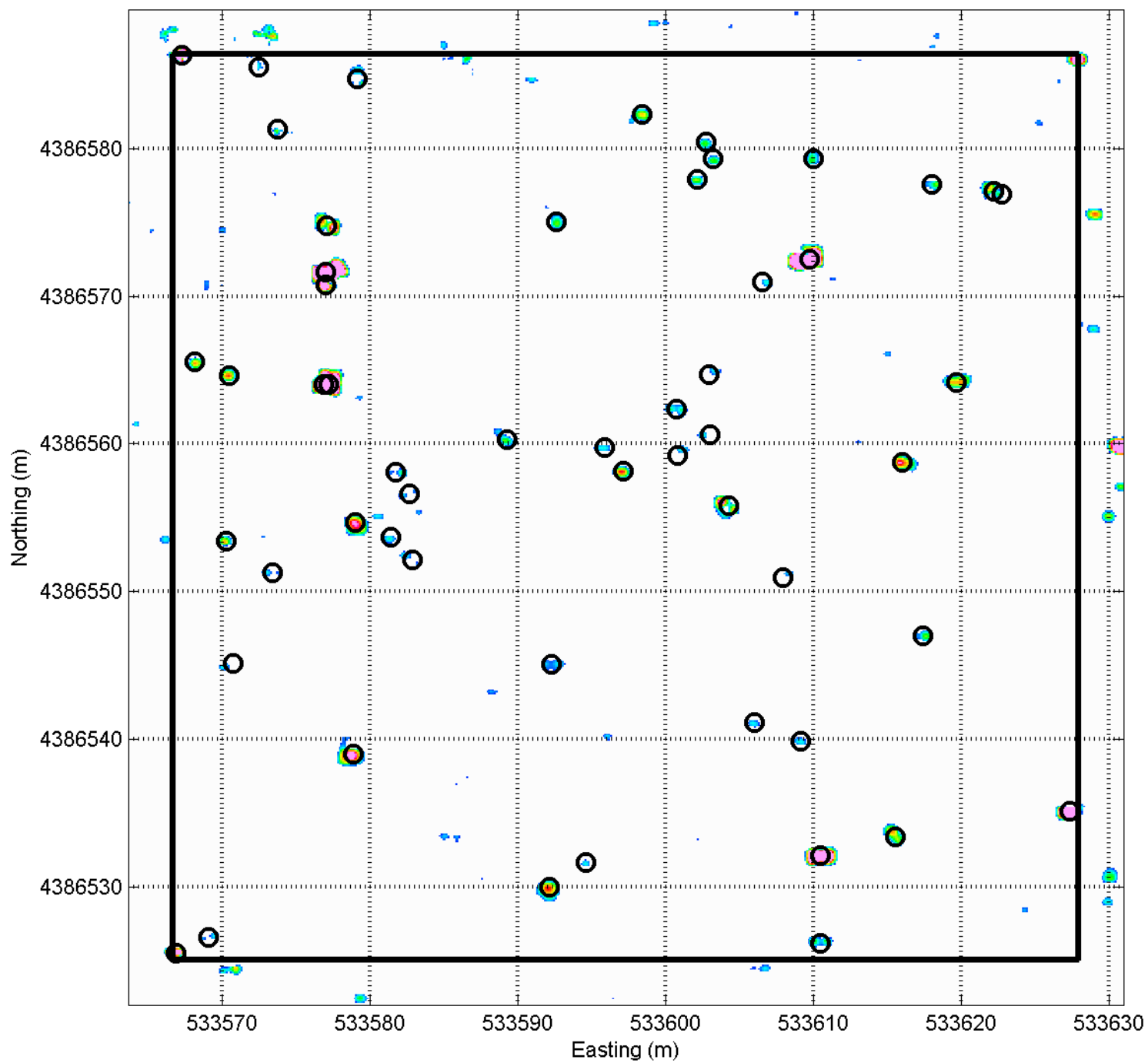




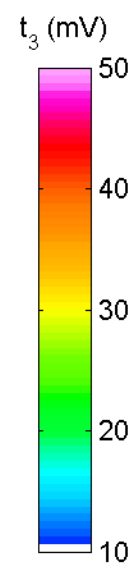


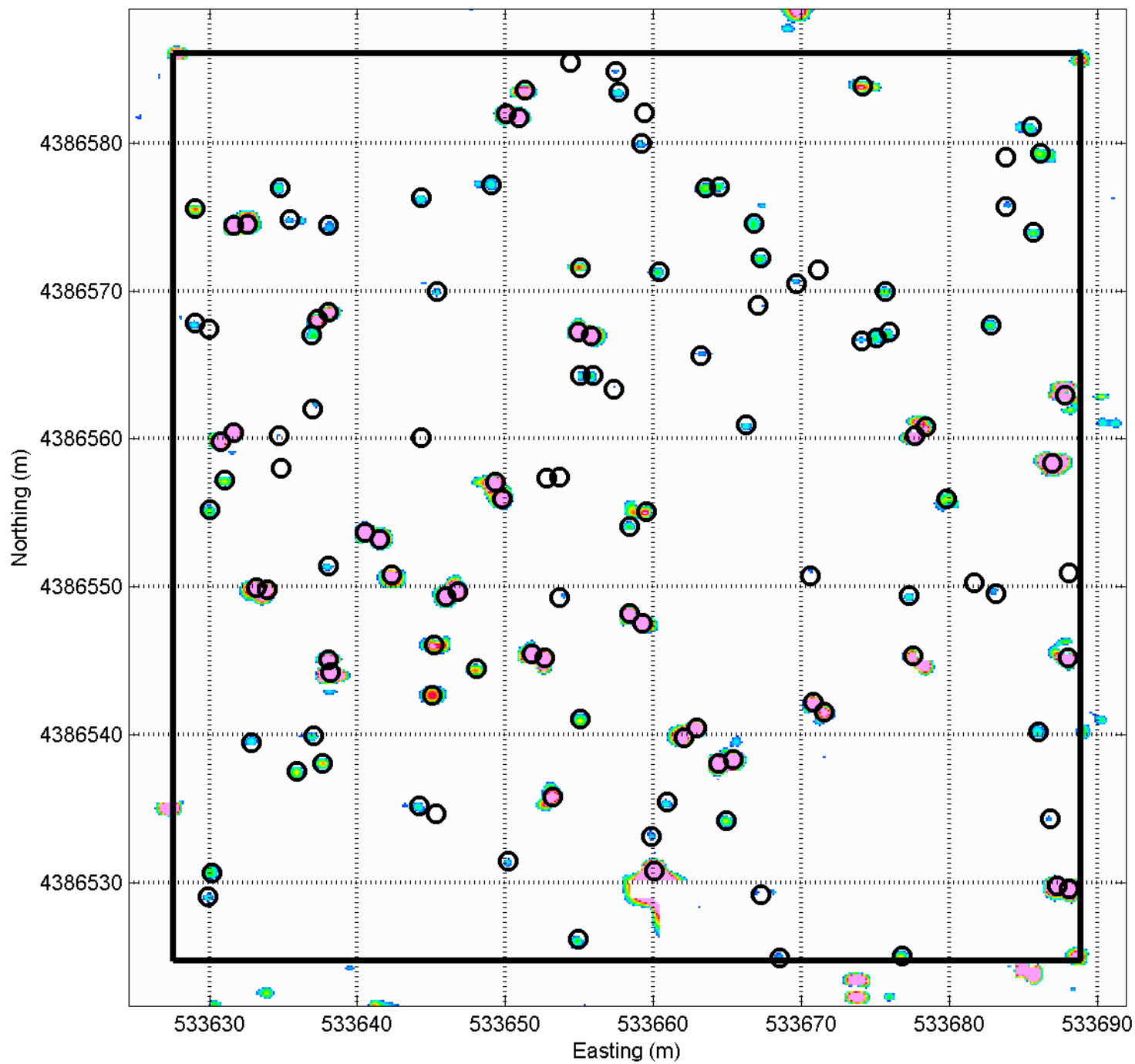
EM61
21-14



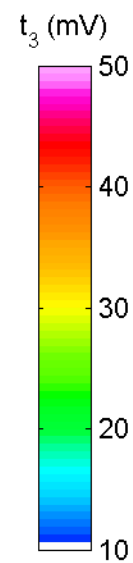


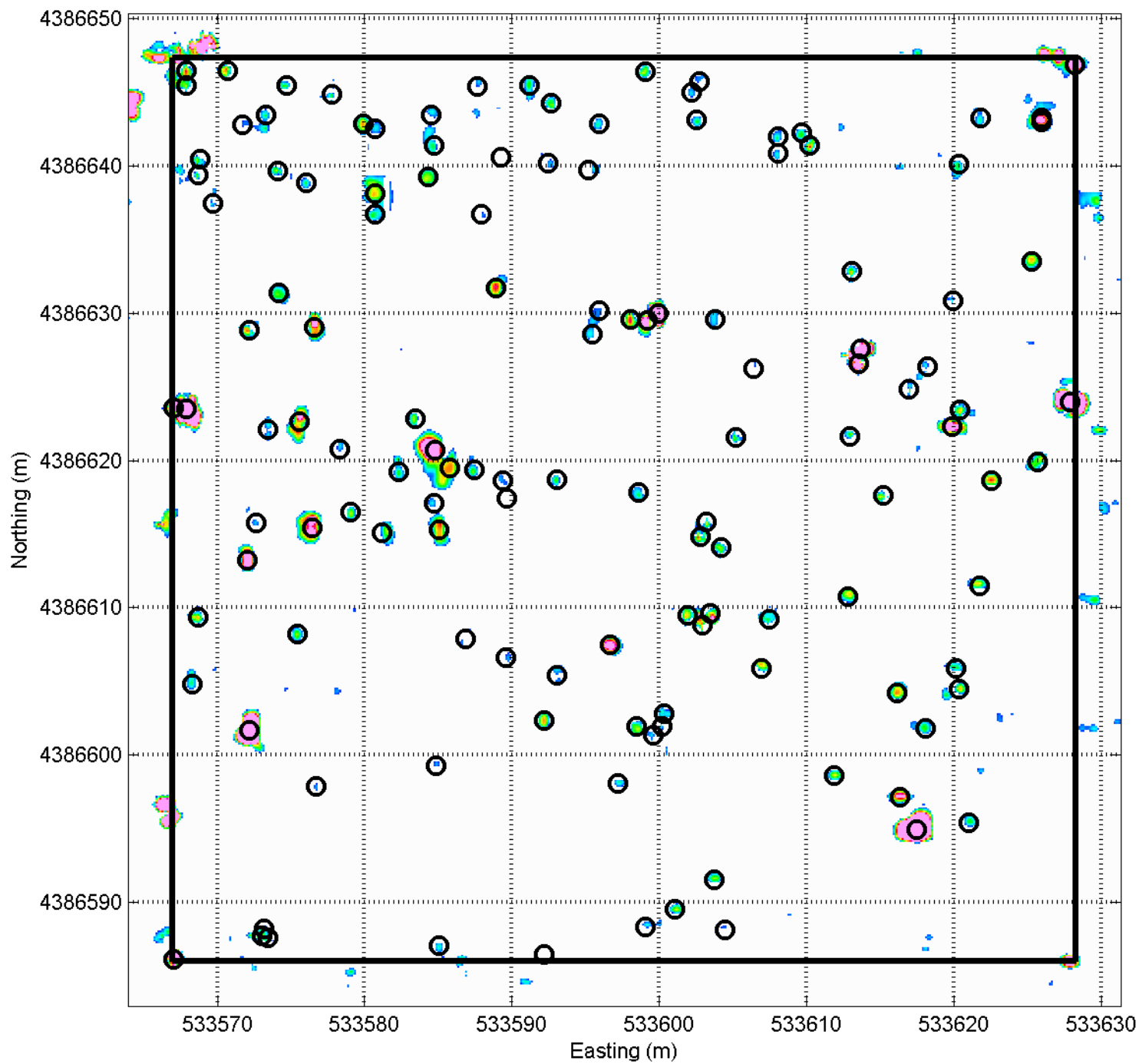
EM61
I-12





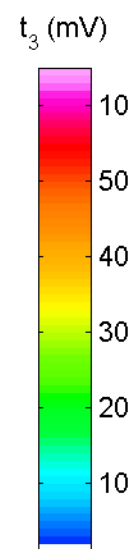
EM61
I-13

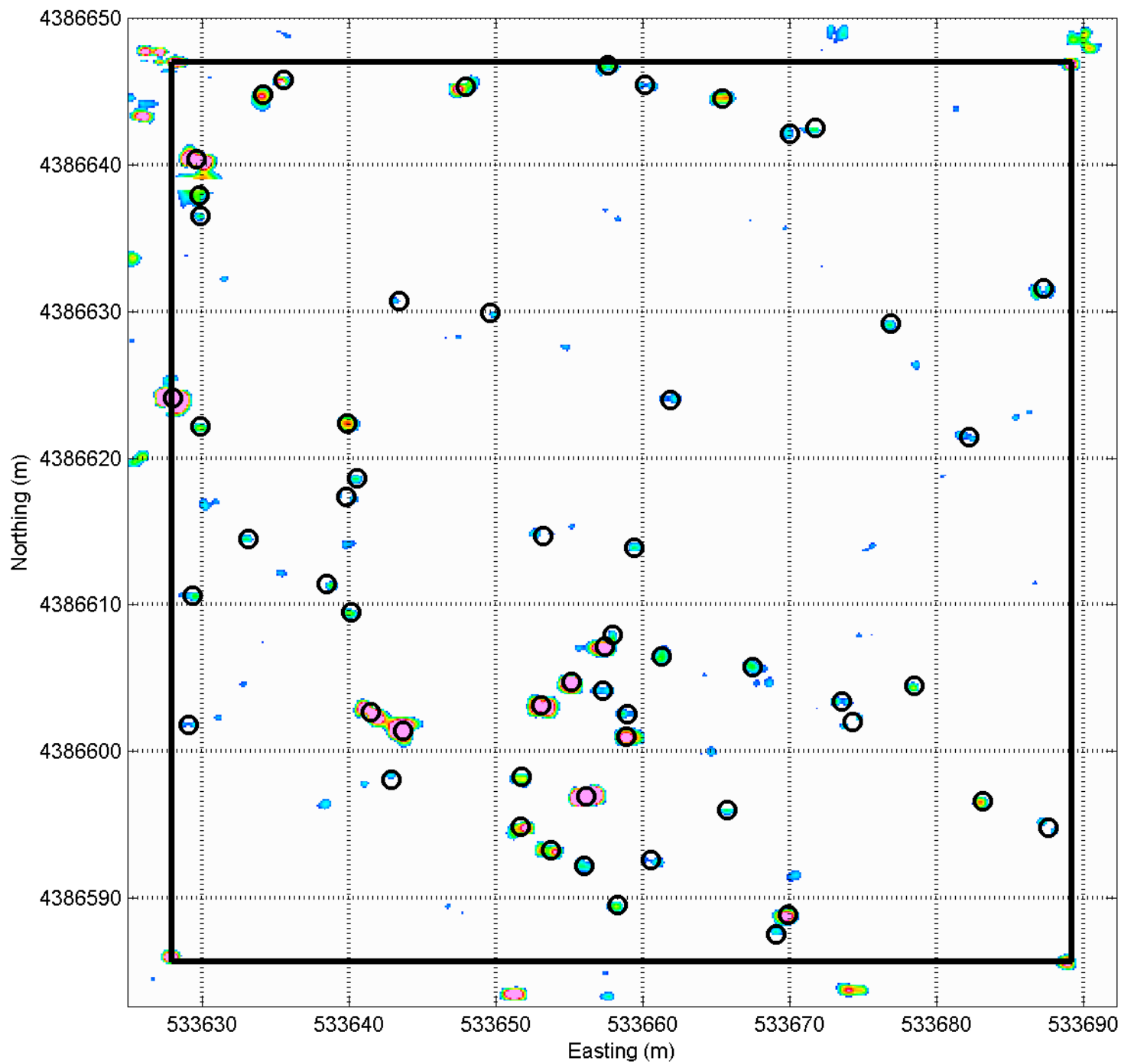




EM61
J-12

North
↑

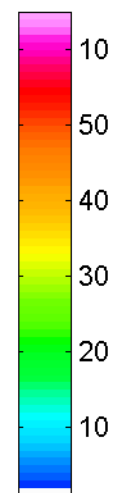


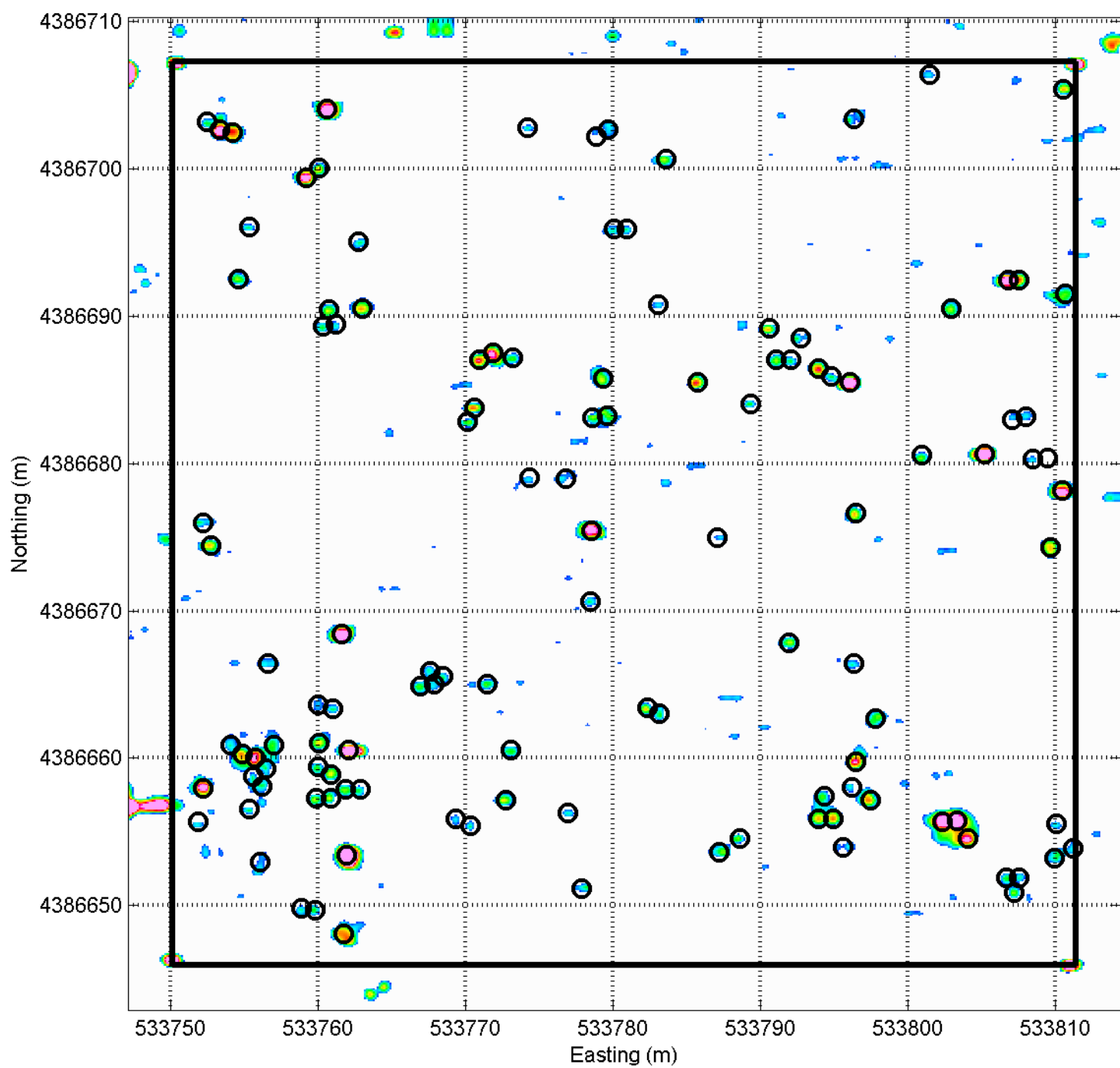


EM61
J-13

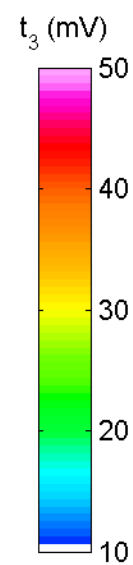


t_3 (mV)



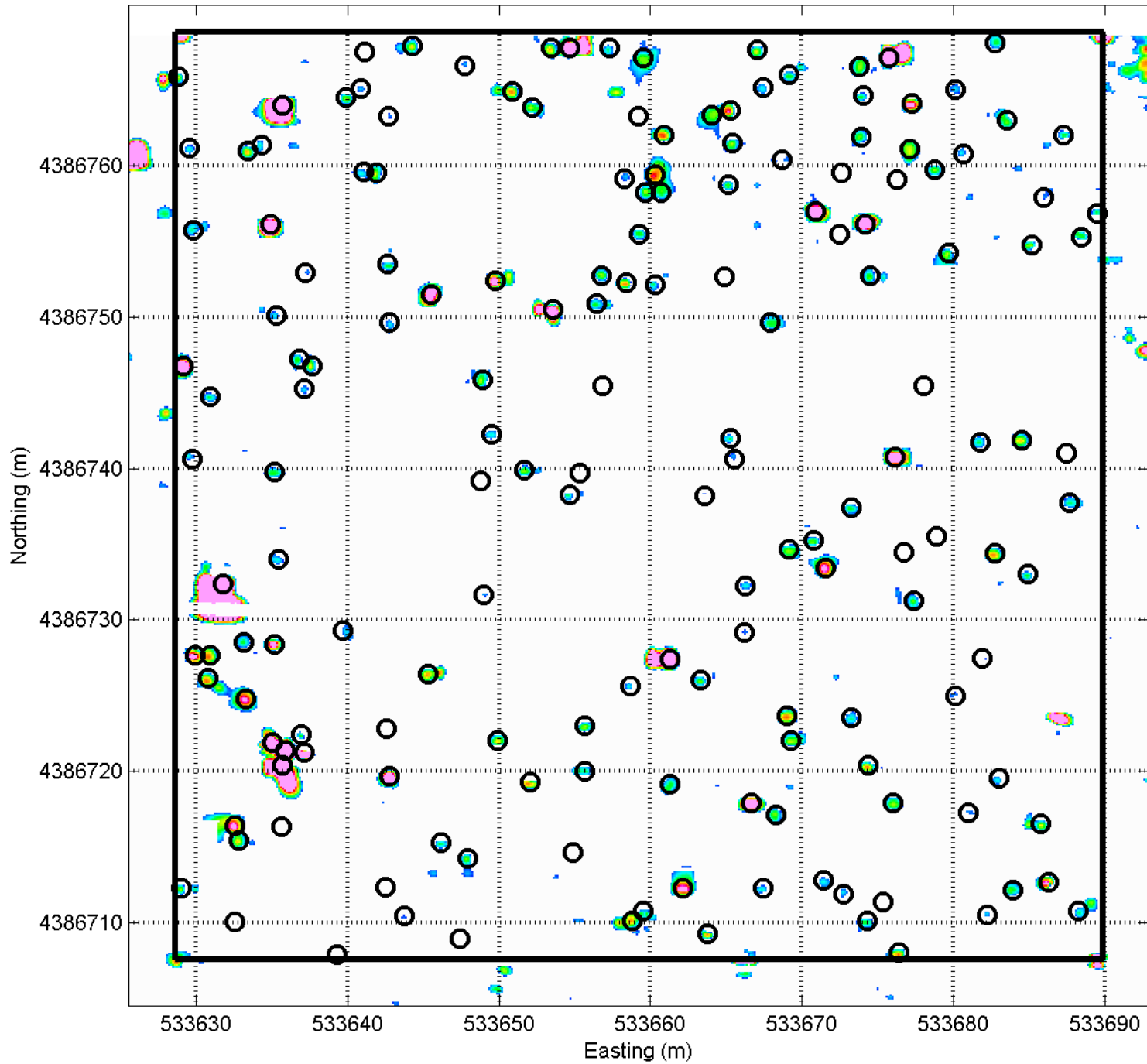


EM61
K-15



EM61
L-13

North
↑

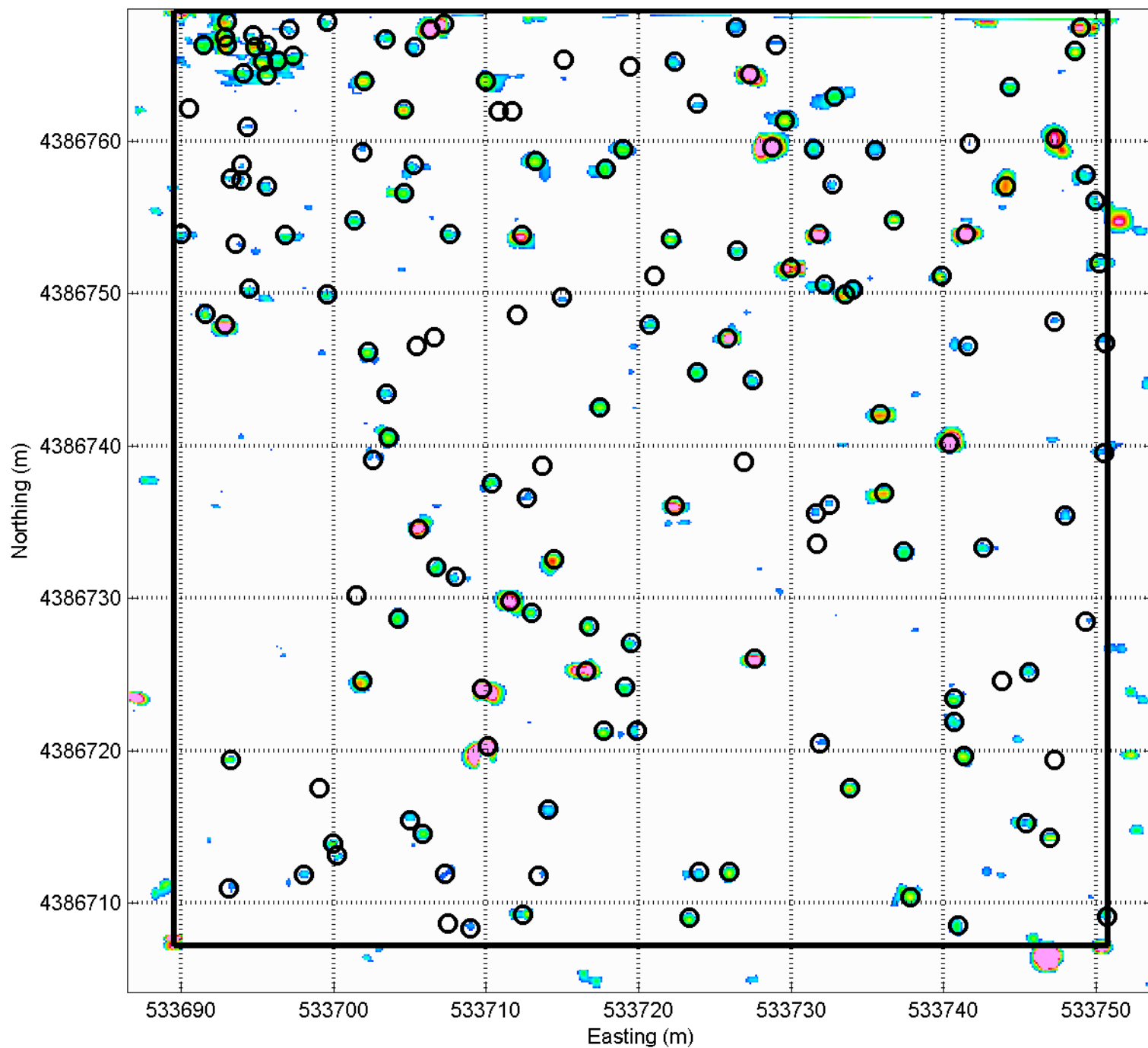
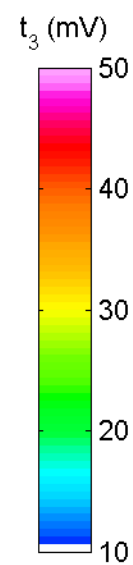


t_3 (mV)
50
40
30
20
10

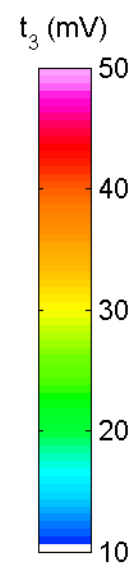
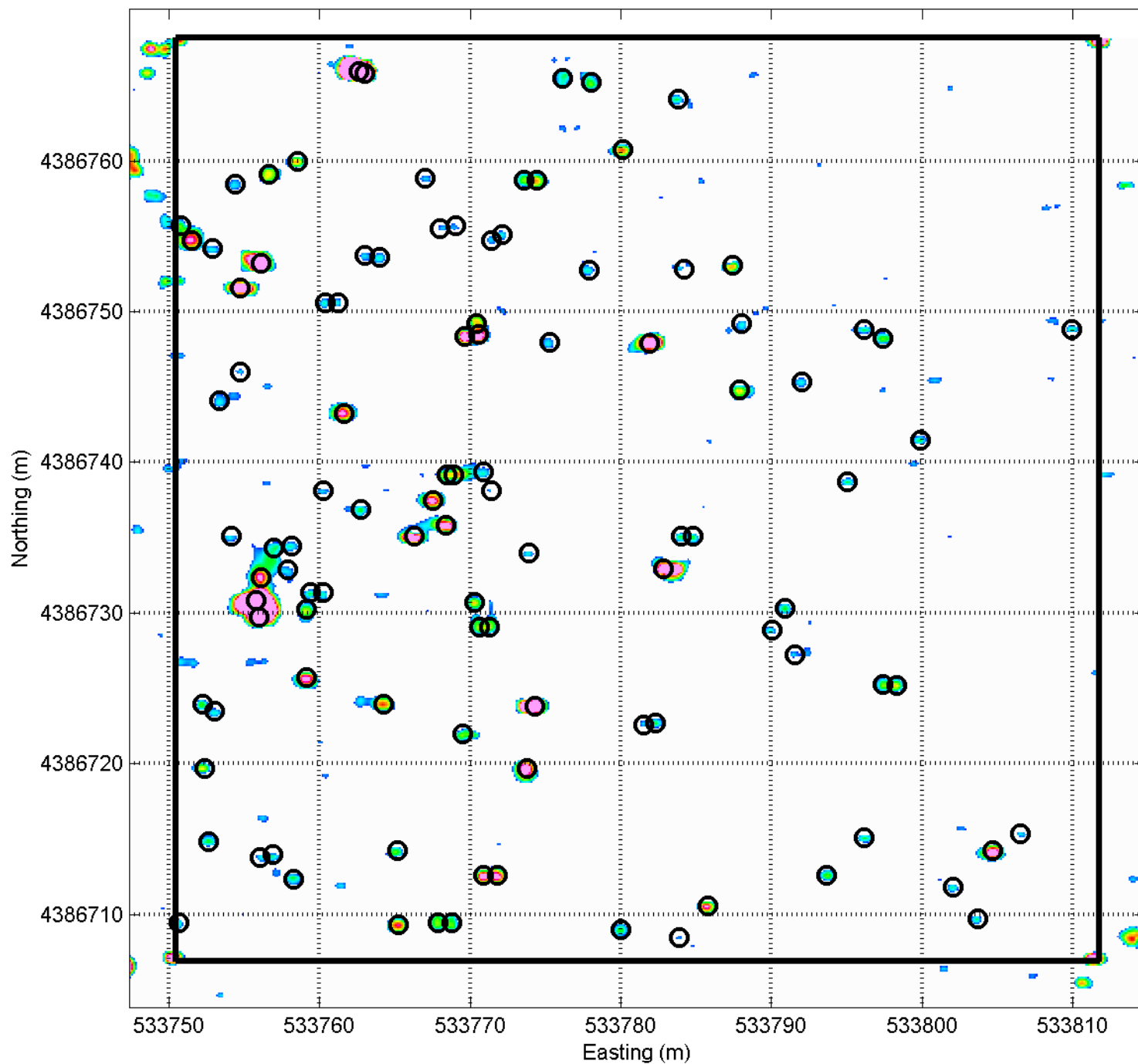


EM61
L-14

North
↑

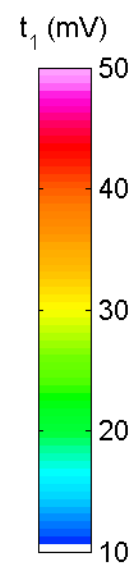
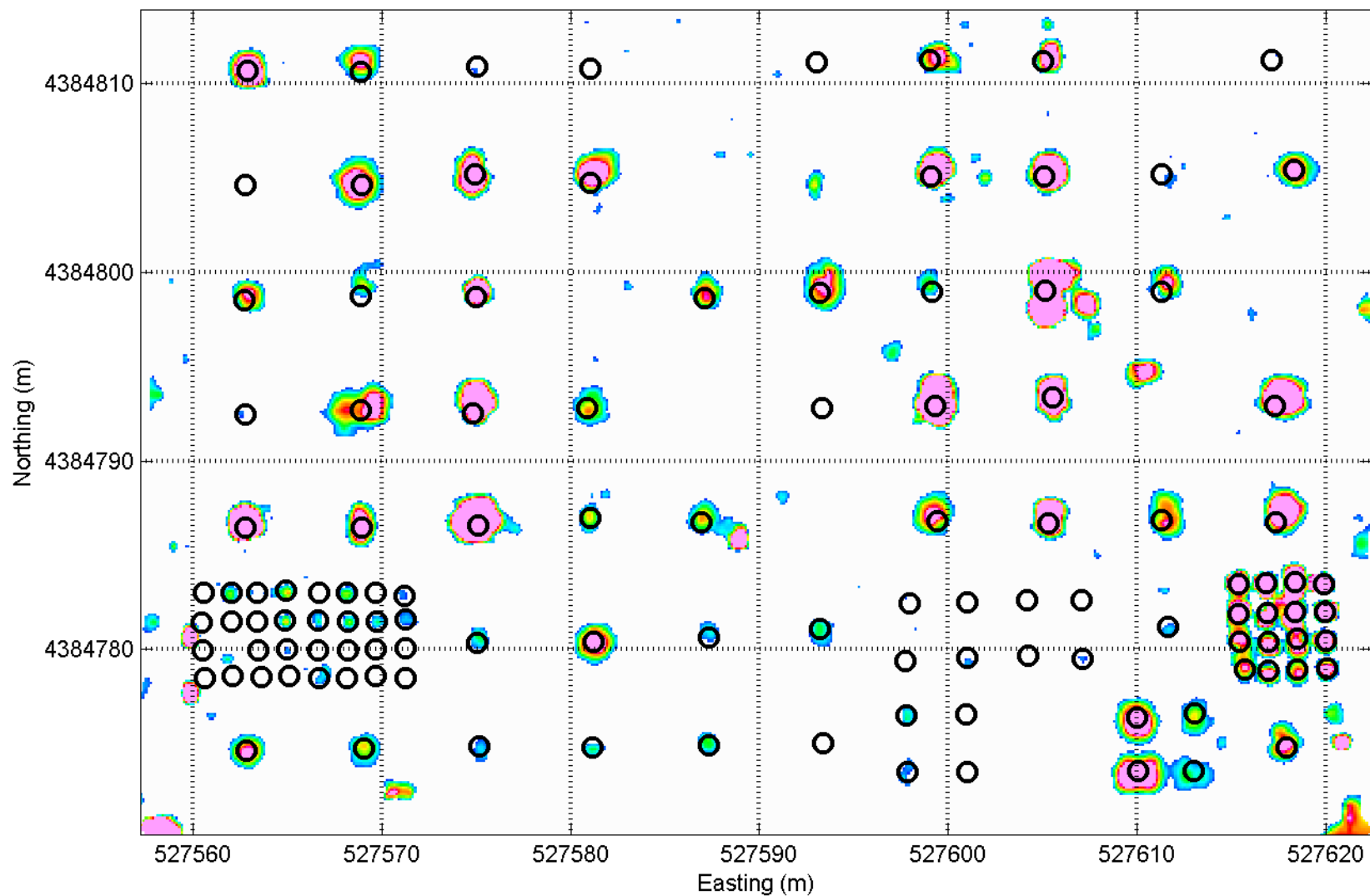


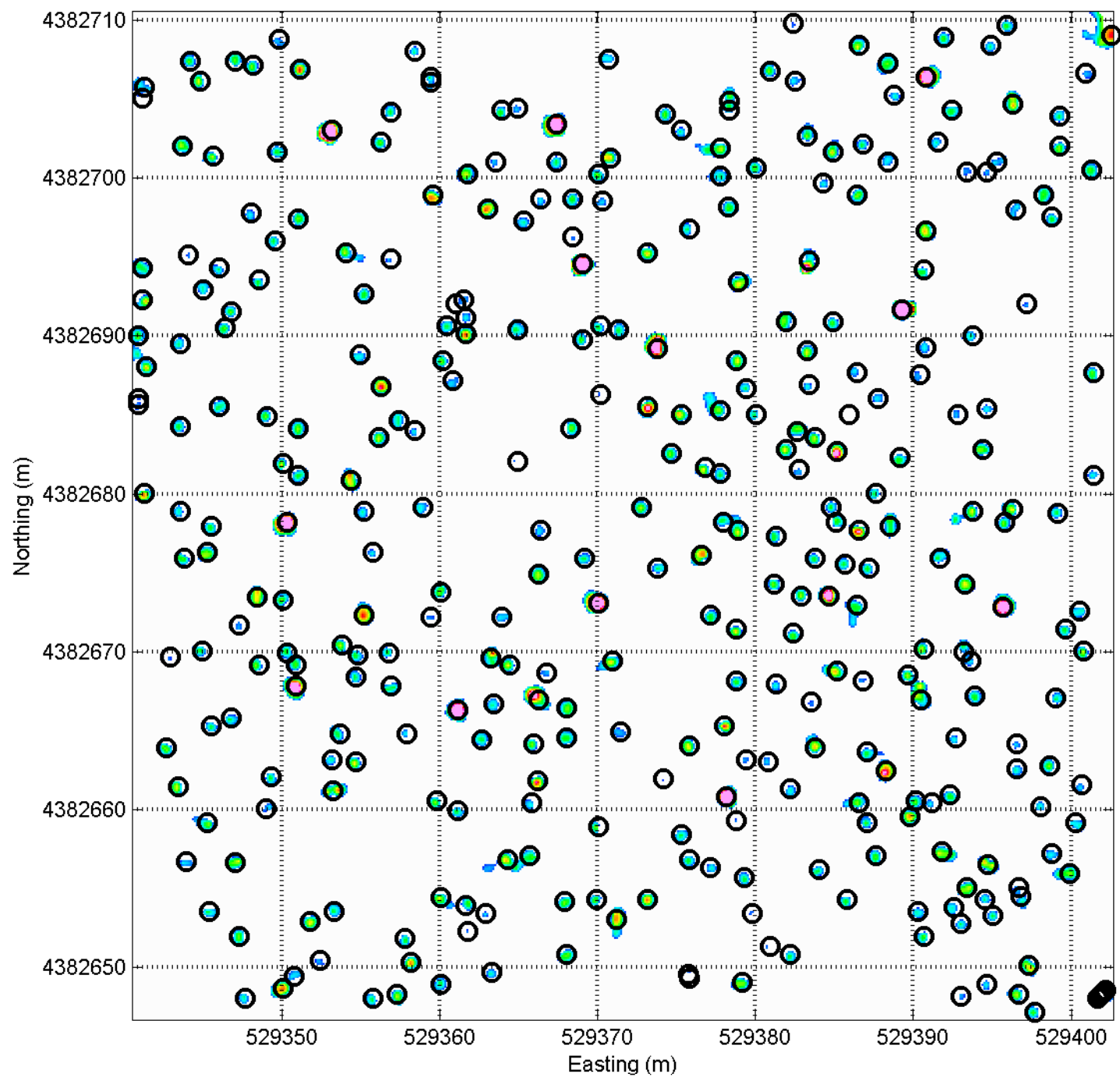
EM61
L-15



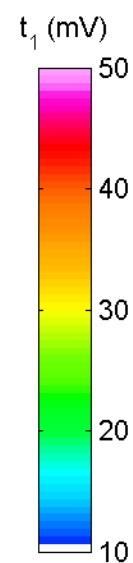
EM63 Testplot

North
↑

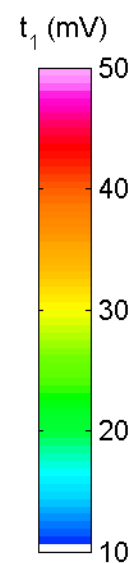
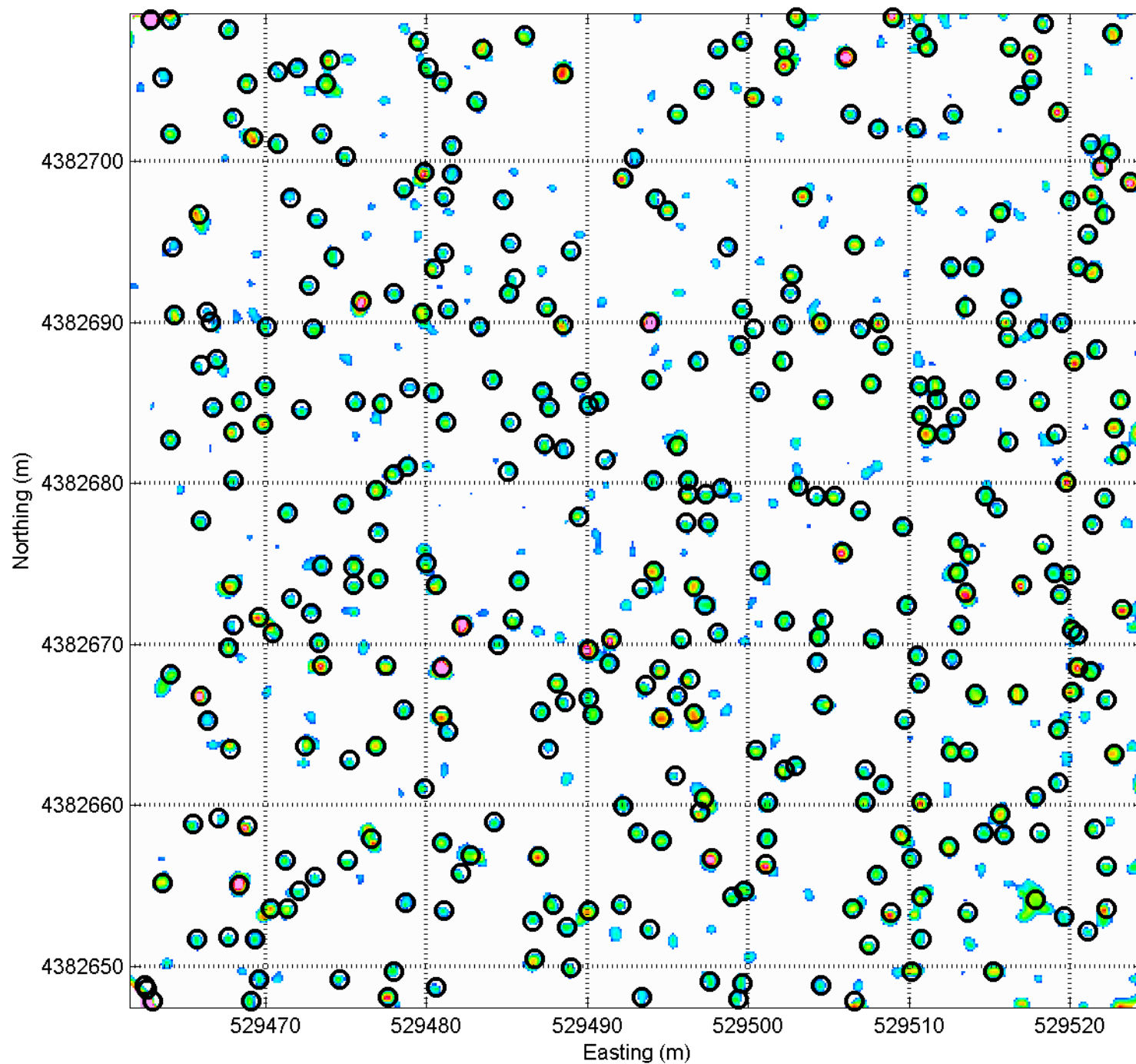


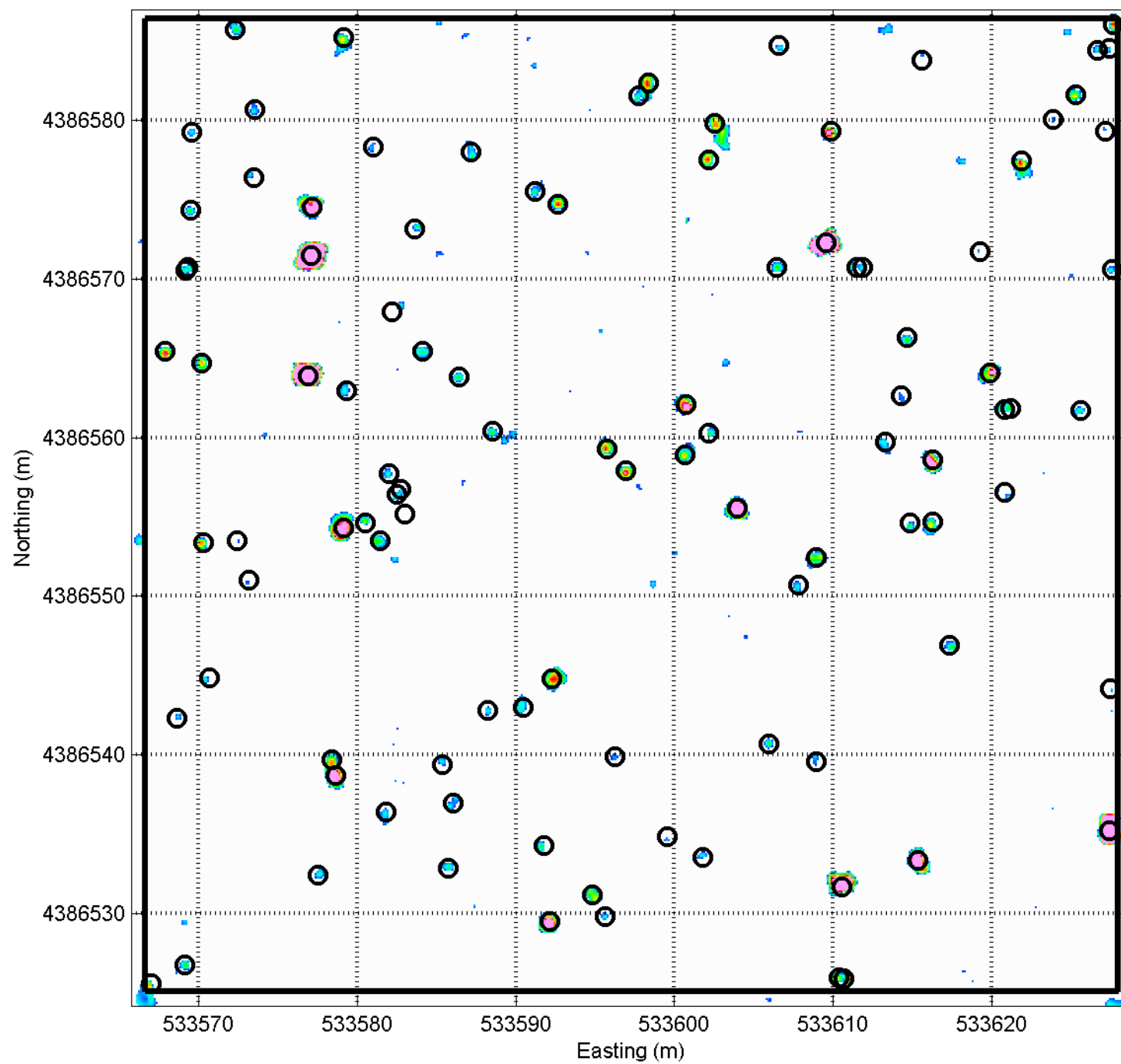


EM63
19-14



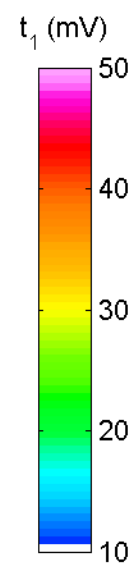
EM63
21-14





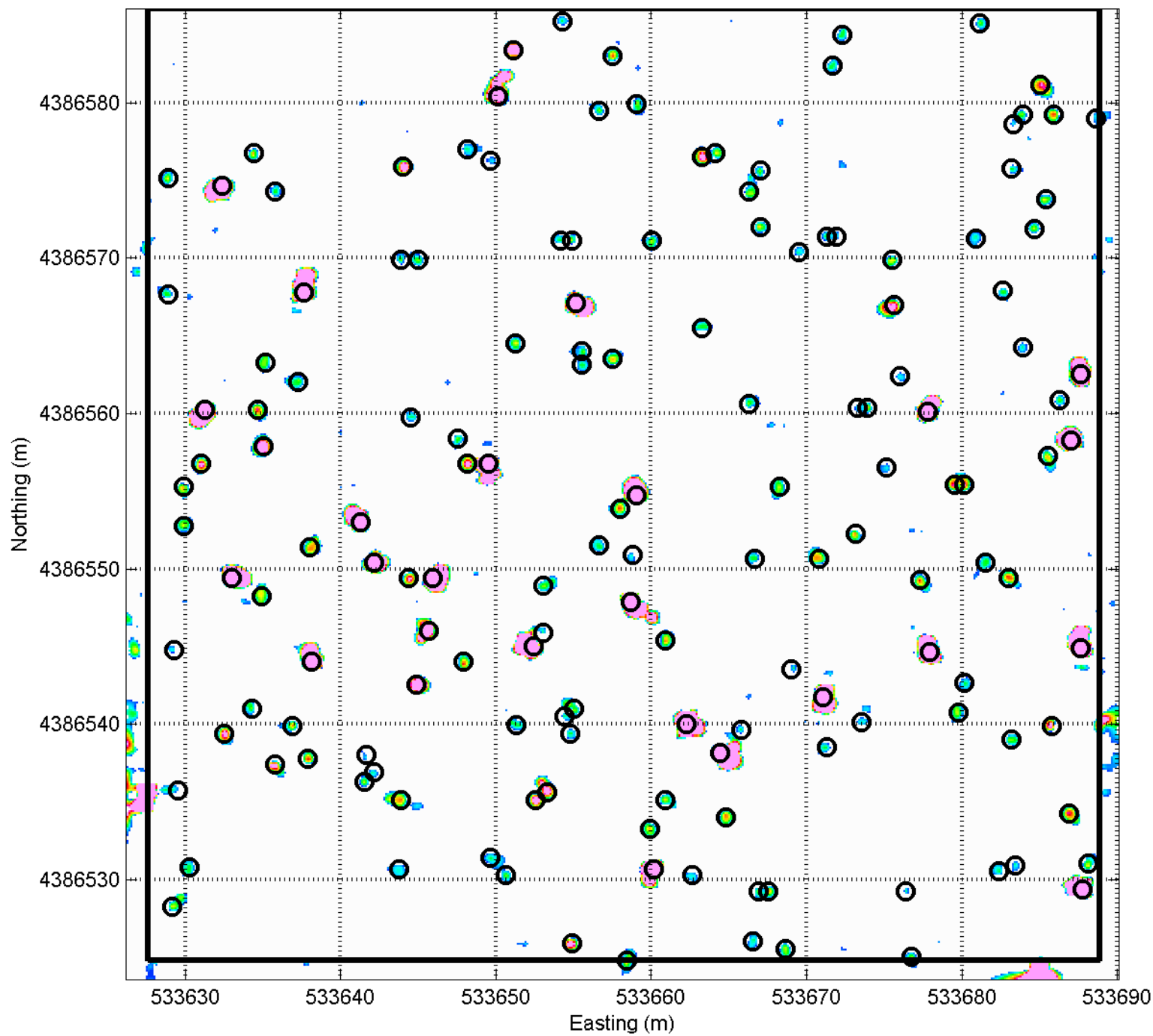
EM63
I-12

↑
North



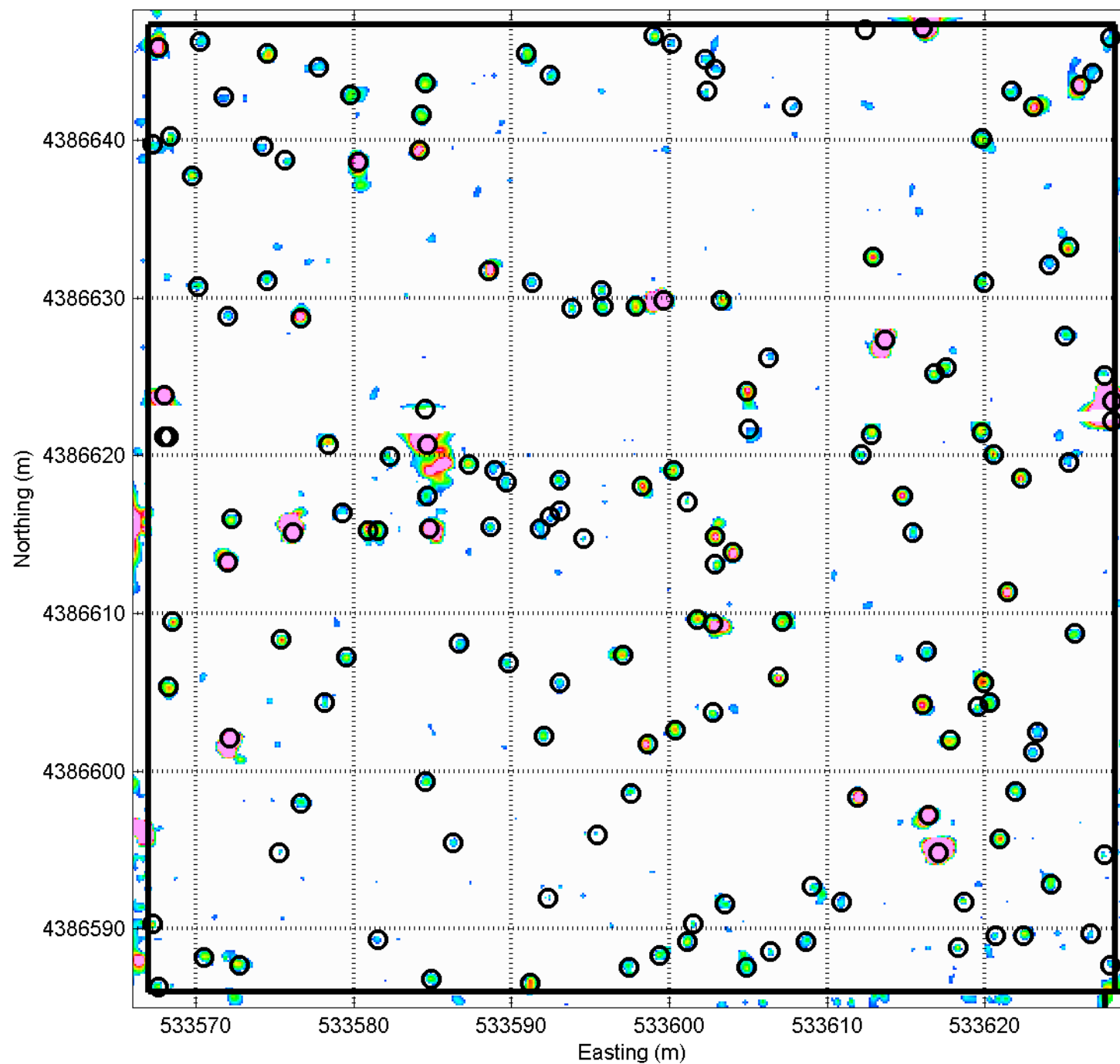
EM63
I-13

North
↑



EM63
J-12

North
↑



t_1 (mV)

50

40

30

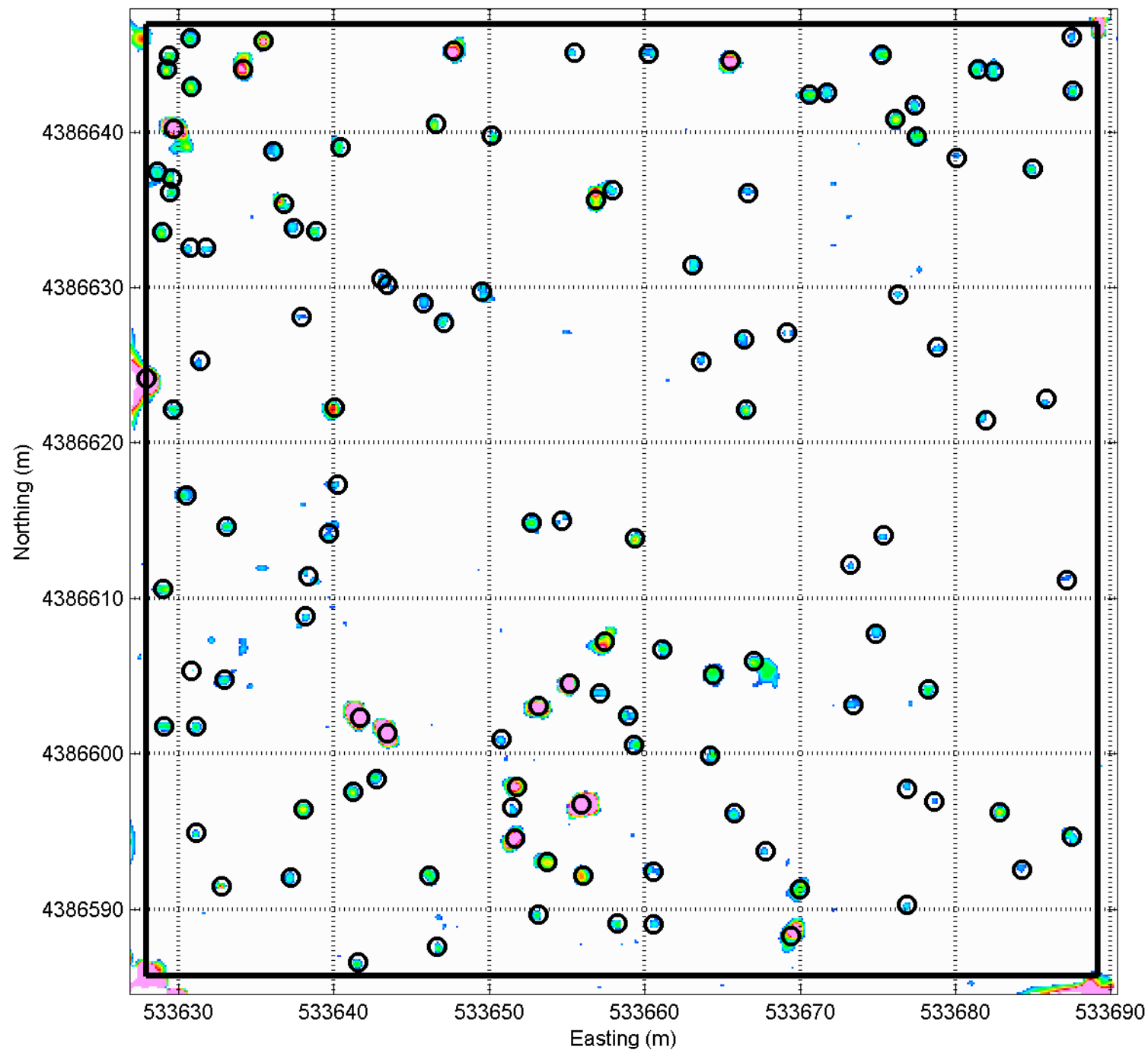
20

10



EM63
J-13

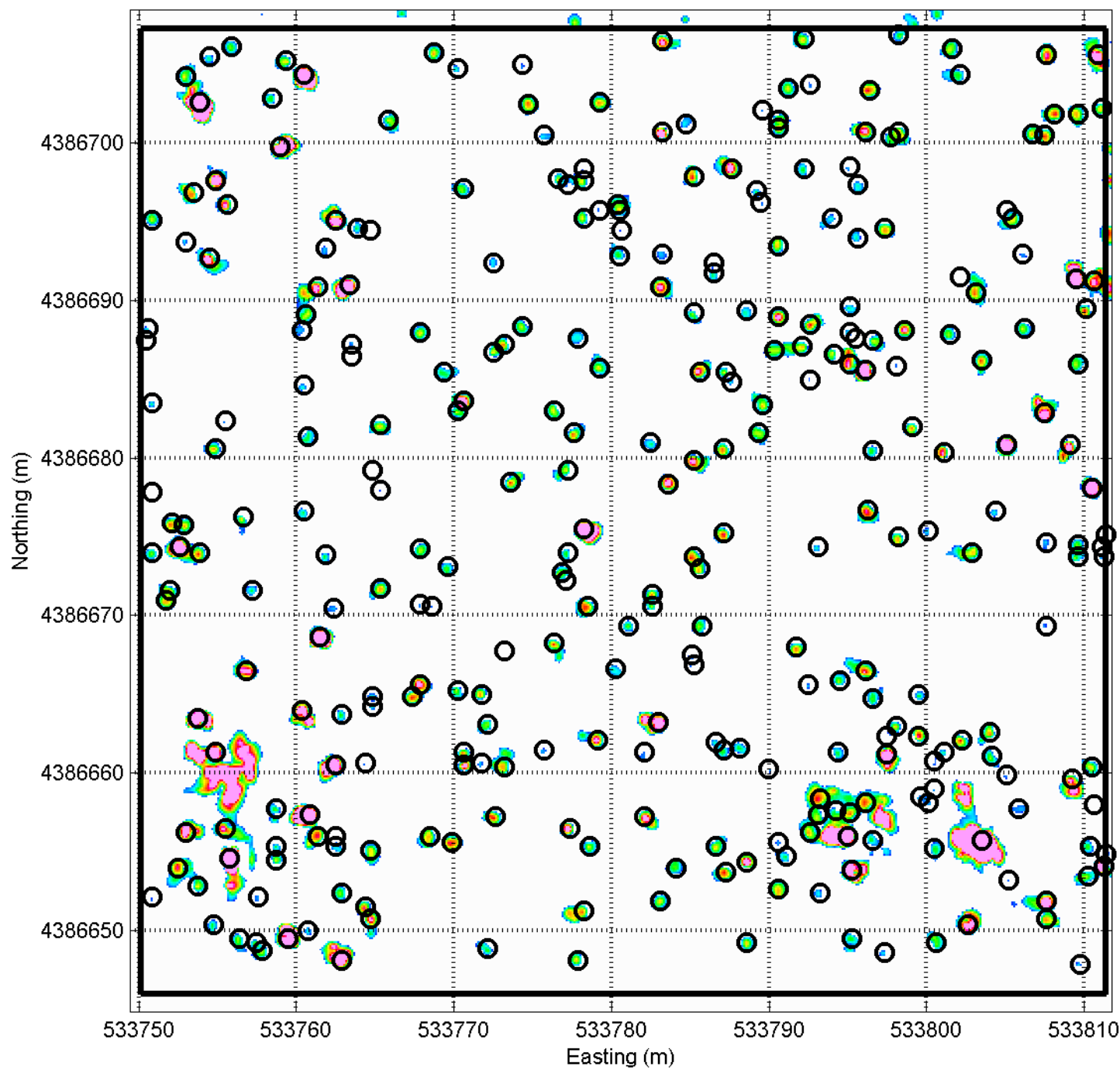
North
↑



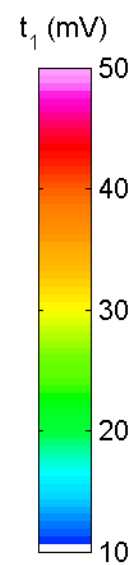
t_1 (mV)

50
40
30
20
10



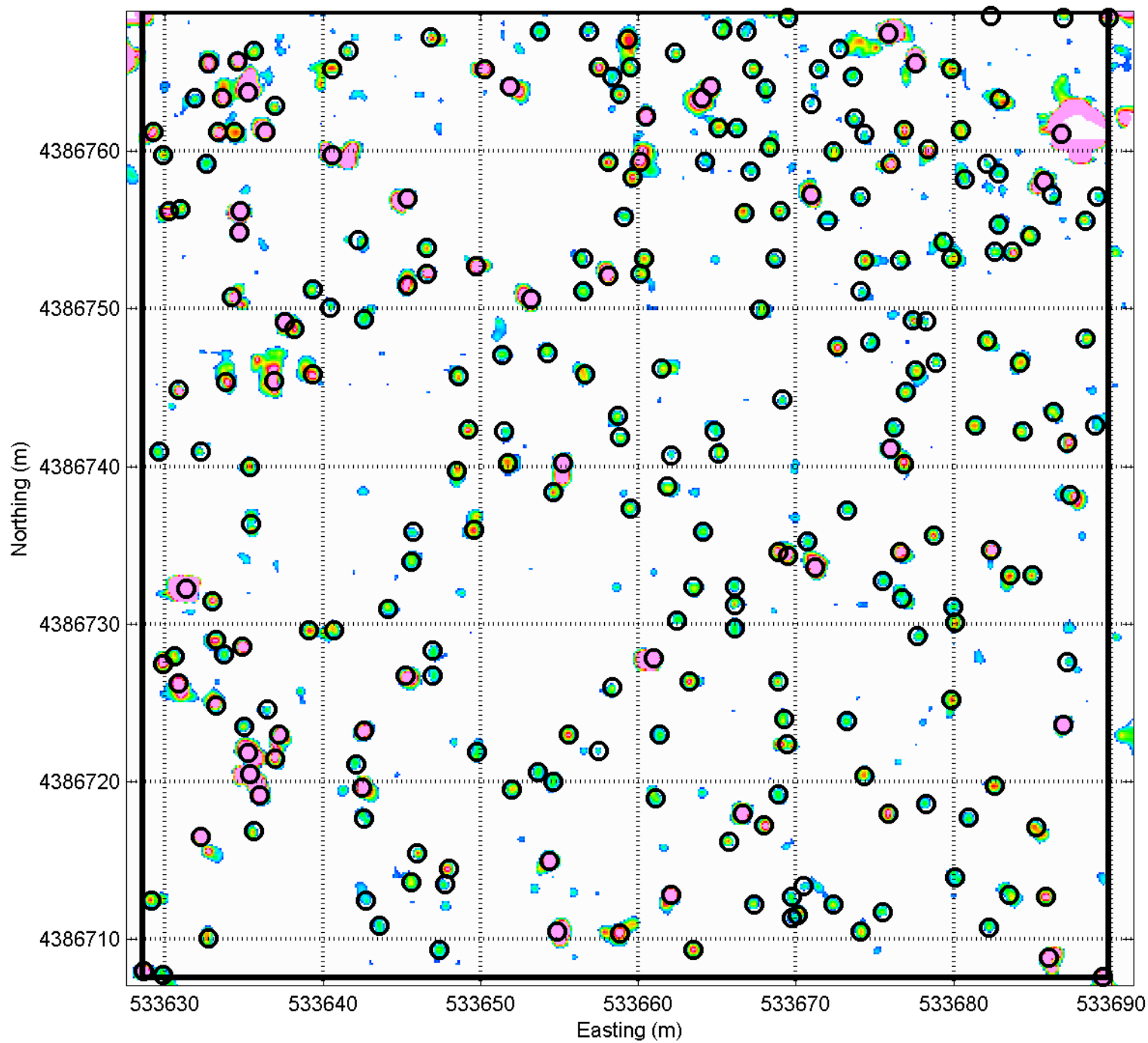


EM63
K-15



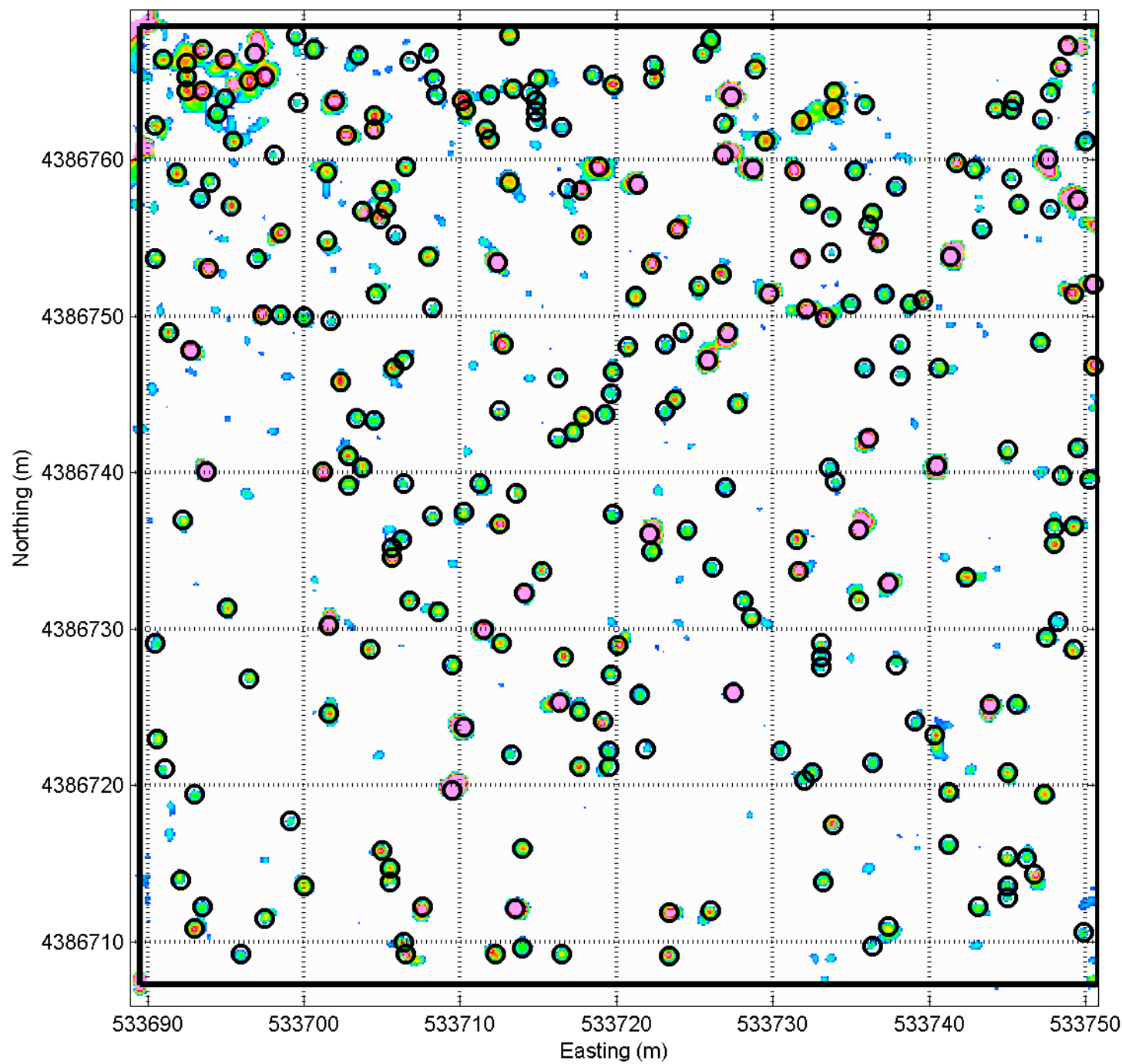
EM63
L-13

North
↑



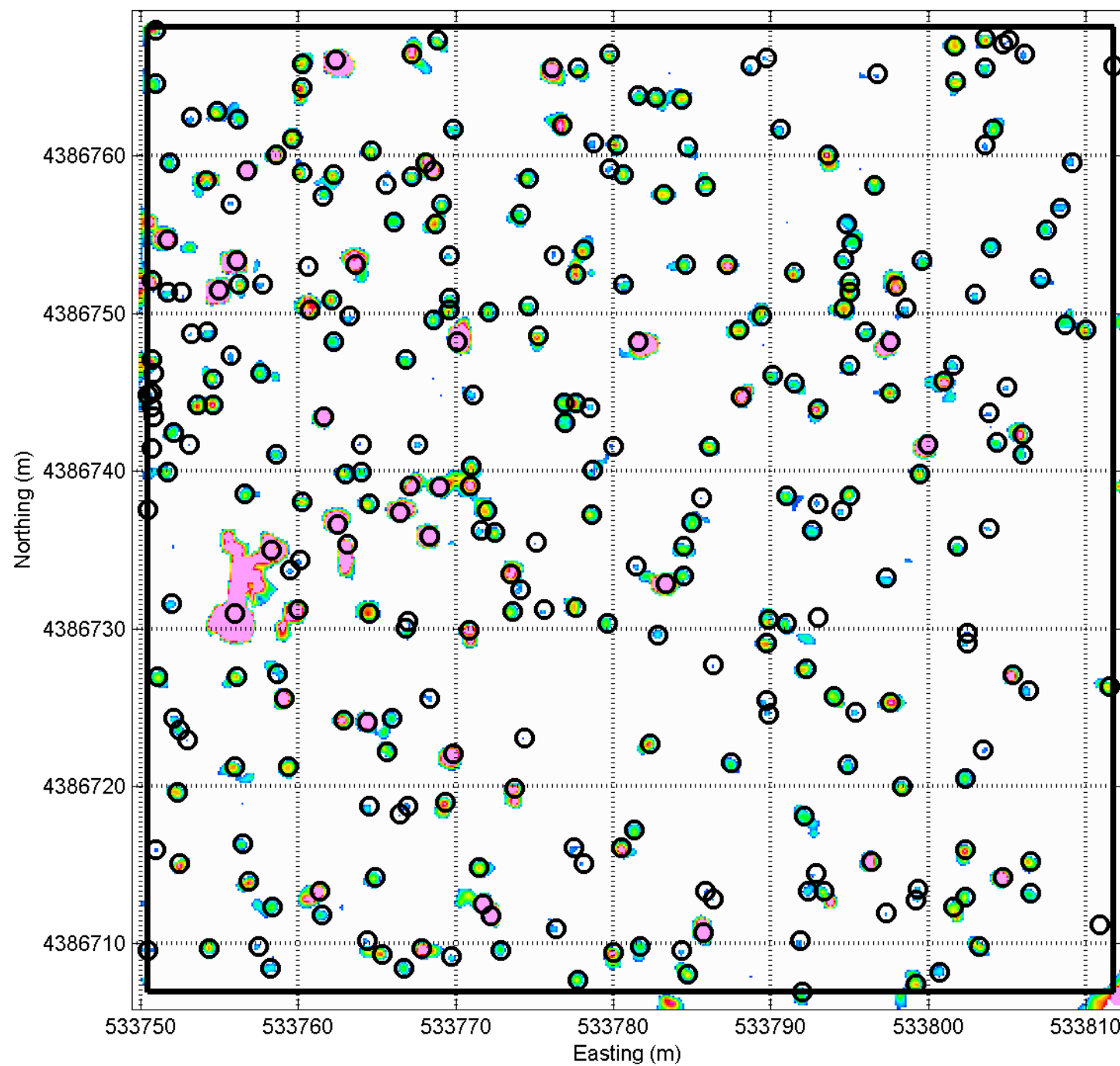
EM63
L-14

North
↑



EM63
L-15

North
↑



t_1 (mV)

50
40
30
20
10

