FINAL REPORT

Statistical and Adaptive Signal Processing for UXO Discrimination for Next-Generation Sensor Data

SERDP Project MM-1442

SEPTEMBER 2009

Leslie M. Collins
Duke University

This document has been approved for public release.

Strategic Environmental Research and Development Program

Table of Contents

I. Background 1

II. Objective 1

III. Summary of Major Accomplishments 1

Modeling 1

Statistical Signal Processing and Feature Selection 2

Active Learning and Kernel-Based Algorithms 2

IV. Scientific Progress and Accomplishments 3

Modeling 3

Sensor Modeling 3

Target Modeling 5

USGS ALLTEM Sensor-Target Response Model 5

LBL BUD Sensor-Target Response Model 6

Statistical Signal Processing and Feature Selection 8

Multi-axis Sensor Simulations 9

Multi-axis Sensor Simulations Summary 11

Inversions with Uncertainty 12

Inversions with Uncertainty Summary 14

Alternate Inversion Techniques 14

Alternate Inversion Techniques Summary 21

Data/Measurement Selection 22

Data/Measurement Selection Summary 27

Sensor Management 28

Sensor Management with Simple Feature Model 29

Sensor Management with Full Observation Model 30

Camp Sibert Discrimination Study Results 32
Sensor Management Summary 37
Feature Selection 38
Feature Selection Summary 43
Fisher Information for Model Inversion 43
Fisher Information for Model Inversion Summary 46
USGS ALLTEM Test Stand Data Processing 46
USGS ALLTEM Test Stand Data Processing Summary 48
LBL BUD YPG Data Processing 48
LBL BUD Camp Sibert Discrimination Study Data Processing 49
LBL BUD Camp Sibert Discrimination Study Data Processing Summary 56
EMI Dipole Model Inversion Modifications 56
Motivations for Considering EMI Dipole Model Inversion Modifications 57
EMI Dipole Model Inversion Modifications 60
Modified Model Inversion Parameter Estimates 64
Feature Selection and Classifier Design 67
LBL BUD AEM Sensor Camp Sibert Discrimination Study Results 67
EMI Dipole Model Inversion Modifications Summary and Discussion 69
Overlapping Signatures 70
Overlapping Signatures Summary 73
Active Learning and Kernel-Based Algorithms 73
Semi-Supervised Active Learning 74
Semi-Supervised Active Learning Summary 79
Improved Concept Drift 80
Improved Concept Drift Summary 84

Multi-Task Learning 85

Multi-Task Learning Summary 89

V. Project Summary and Future Work 90

VI. Technology Transfer 91

VII. Manuscripts Published/Submitted and Presentations Given 91

Manuscripts 91

Presentations 91

VIII. References 93

I. Background

To address the limitations associated with current sensors, SERDP and ESTCP have been supporting efforts to develop a new generation of UXO sensors that produce data streams of multi-axis vector or gradiometric measurements, for which optimal processing has not yet been carefully considered or developed. Here, we outline a research program where our goal was to address this processing gap, employing a synergistic use of advanced phenomenological-modeling and signal-processing algorithms. The focus of the research was: (i) exploitation and refinement of Duke's existing phenomenological models to accurately predict the underlying target signatures for the new sensor modalities, with the goal of pinpointing physical parameters that can be utilized within the signal processing architecture; (ii) development of physics-based statistical signal processing approaches applicable to the problem in which vector data is available from such sensors; (iii) development of the theory of optimal experiments to guide the optimal deployment of sensor modalities; and (iv) development of active learning and kernel-based algorithms that yield target classification based upon all available data (not based on individual feature vectors from isolated targets), as well as allowing data collected across multiple sites to be integrated within the classifier.

II. Objective

The objective of this project was to develop algorithms that provide UXO classification capability for multi-axis sensors that are superior to single-axis solutions. To achieve our goals, we leveraged and extended our previous SERDP-sponsored research. Under SERDP support, both exact and approximate phenomenological models have been developed for single-axis EMI and magnetometer sensors. These models have been used successfully within the context of statistical signal processing algorithms. The processing algorithms have demonstrated improved discrimination of UXO from clutter, for isolated and overlapping target signatures. We have also developed an information-theoretic framework based on the theory of optimal experiments, with which we have achieved near-optimal performance without *a priori* training data; the algorithm is trained adaptively as the data is acquired. As in our prior SERDP projects, we did not propose development of new sensors, but rather focused on optimal processing of data from next-generation sensors, based on insights from advanced models.

III. Summary of Major Accomplishments

Modeling

- Developed an electromagnetic model for data measured by the LBL BUD AEM system.
- Developed an electromagnetic model for data measured by the USGS ALLTEM system in the ZZM polarization, which consists of the single z-directed transmit coil circumscribing the middle of the cube and the two large, 1-meter, z-directed receive coils located at the top and bottom of the cube.
- Implemented a magnetization tensor target model assuming the target is represented by a parametric model for three orthogonal and unique magnetic dipoles.
- Implemented a magnetization tensor target model assuming the target is not represented by a parametric model.

Statistical Signal Processing and Feature Selection

- Performed simulations of multi-axis sensors which provided support for the premise that multi-axis sensors could improve target parameter estimates, and thereby improve discrimination of UXO from clutter.
- Proposed a Bayesian model inversion algorithm which is robust to sensor position uncertainty.
- Evaluated multiple alternate model inversion techniques to improve model parameter estimates.
- Evaluated the effects of preferentially selecting informative measurements for model inversion, as well as methods for determining the most informative measurements.
- Developed an information-theoretic sensor management framework, and demonstrated on real data that it enabled very similar detection performance to an unmanaged procedure with substantially fewer observations.
- Proposed a new feature selection algorithm (parallel-sequential – ParSe) and demonstrated that it appears to provide better feature section performance than either sequential forward or sequential backward searches, and evaluated the effect of the choice of performance metric which is optimized by the feature selection.
- Proposed a novel objective function (Fisher information measure) for model inversion, and demonstrated that this approach in conjunction with conventional least squares objective function can provide improved model parameter estimates.
- Processed USGS ALLTEM Test Stand data automatically, with no human intervention or oversight.
- Processed LBL BUD YPG data automatically, with no human intervention or oversight.
- Processed LBL BUD Camp Sibert Discrimination Study data automatically, with no human intervention or oversight.
- Identified shortcomings in the current EMI dipole inversion process, and modified the inversion process to overcome the shortcomings and improve performance.
- Evaluated matching pursuits as a method to provide robust performance in the case of overlapping target signatures.

Active Learning and Kernel-Based Algorithms

- Investigated semi-supervised active learning to improve discrimination of UXO from clutter.
- Improved previously developed concept drift algorithm to provide better performance on data collected across multiple sites.
- Developed multi-task learning classification algorithm, which merges ideas from semi-supervised and multi-task learning for multiple sets of sensor data collected with the same sensor.

IV. Scientific Progress and Accomplishments

Modeling

The source and receiver coils of the USGS ALLTEM and LBL BUD AEM sensors have been rigorously modeled, wherein the size, shape, and orientation of the coils have been accounted for; dipole approximations for the sensor coils were not assumed as part of this analysis. In addition to rigorously modeling the spatial properties of the sensor coils, the temporal properties of the source have also been rigorously modeled.

Sensor Modeling

The transmit coils for the sensors under consideration are large enough to preclude modeling the coil with a single magnetic dipole. To address this modeling task rigorously, two approaches are considered for modeling the sensor transmit coils: the Biot-Savart Law, and the surface equivalence theorem, or Huygens' Principle. Both of these approaches model the transmit coil current using the superposition of smaller current segments or loops. Using the surface equivalence theorem, or Huygens' Principle, the sensor coils are modeled using a superposition of magnetic dipoles. This technique approximates the sensor transmit coil with a grid of individual magnetic dipoles, as illustrated in Figure 1 with a general rectangular transmit coil. Each of the dipoles has a position in space and an orientation. The orientation is orthogonal to the plane of the sensor coil, and is aligned so that current is flowing in the proper direction. The dipole currents inside the sensor coil cancel each other, leaving only the current on the outside, which serves as the approximation to the coil current.

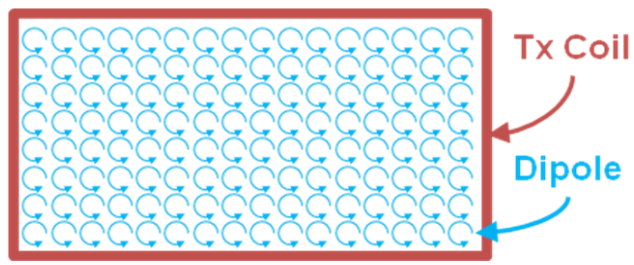

Figure 1. Example sensor coil model in which superposition of magnetic dipoles is used to approximate the current in the coil (surface equivalence theorem, or Huygens' Principle).

The representation of the LBL BUD sensor using this approach is shown in Figure 2. The location of the dots represents the centers of the magnetic dipoles used to represent the coils using superposition of discrete magnetic dipoles, or Huygens' Principle. Each transmit coil orientation is shown in a different color (red is x-directed, green is y-directed, and blue is z-directed), and the receive coils are represented by black dots. The receive coils are circular loops, and therefore are appropriately modeled by magnetic dipoles. Although this model completely represents the LBL BUD AEM sensor, it is important to note that the sensor model is rotated 45° relative to (x,y)-coordinates. While this rotation does not affect the model inversions to estimate

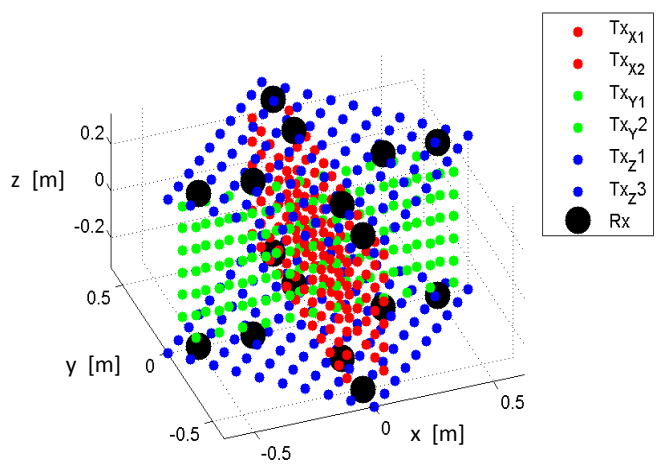

Figure 2. LBL BUD AEM sensor as modeled using Huygens' Principle for the transmit coils and a dipole for the receive coils. Each dot represents the location of a magnetic dipole.

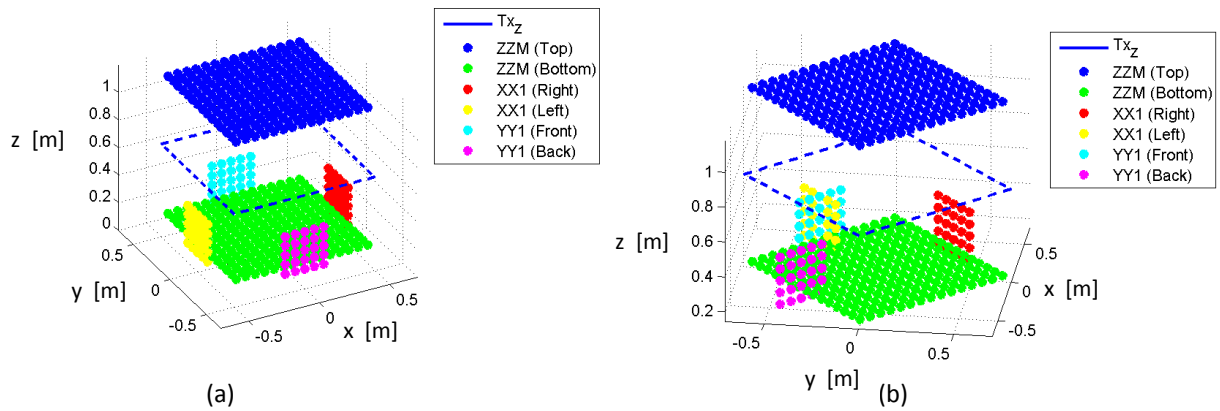

Figure 3. USGS ALLTEM sensor as modeled using the Biot-Savart Law for transmit coils and Huygens' Principle for receive coils for cart rotation of (a) 0° and (b) 45° relative to Northing/Easting. Each line segment represents a current segment, and each dot represents the location of a magnetic dipole.

target parameters, it is important to recognize that the estimated target orientation parameters may have a bias of 45° as a result of the rotation in the sensor model.

Using the Biot-Savart Law, the transmit coil loop is approximated as a series of current segments, and the total incident magnetic field is modeled as the sum of the contribution from each of the individual current segments. The implementation is a discretization of the contour integral over the transmit coil current loop (Biot-Savart Law),

$$B = \frac{\mu_0}{4\pi} \oint_{C'} \frac{dl' \times \vec{R}}{R^3},$$

where dl' represents a current segment and $\vec{R}$ is the vector between the current source and the point at which the magnetic field is calculated.

The representation of the USGS ALLTEM sensor using this approach is shown in Figure 3. The transmit coils are represented by dashed lines, with each line segment representing a current element utilized in the discretized approximation to the Biot-Savart Law. The receive coils are modeled using a superposition of magnetic dipoles, or Huygens' Principle, and the locations of the magnetic dipoles employed for this model are represented by colored dots, with each receive coil plotted in a unique color. As can been seen in Figure 3, at the completion of this effort the model is still incomplete. Although the USGS ALLTEM sensor uses three transmit coils and 18 receive coils to measure 19 individual polarizations, only the z-directed transmit coil and the two large 1m receive coils on the top and bottom of the sensor are fully modeled. Thus, only the ZZM polarization can be modeled. The x- and y-directed transmit coils are not included in this model, and none of the smaller receive coils are fully modeled. Four of the vertically oriented small receive coils have been partially incorporated into the model; their locations and orientations are correct when the cart is not rotated relative to Northing/Easting (Figure 3a), but incorrect when the cart is rotated (Figure 3b). These limitations in the USGS ALLTEM model were discovered after the project ended. We have since been working to correct and complete this model, but the progress included in this report is limited to the status of the project when this funded effort ended. Questions regarding the current status of this model can be directed to Stacy Tantum or Leslie Collins.

Target Modeling

A target sensed by an EMI sensor may be modeled using a magnetization tensor, which takes the form

$$M(t)=\begin{bmatrix} f_1(t) & 0 & 0 \\ 0 & f_2(t) & 0 \\ 0 & 0 & f_3(t) \end{bmatrix} \quad \text{or} \quad M(\omega)=\begin{bmatrix} f_1(\omega) & 0 & 0 \\ 0 & f_2(\omega) & 0 \\ 0 & 0 & f_3(\omega) \end{bmatrix},$$

for time-domain or frequency-domain sensors, respectively. The functions on the diagonal, $f_n(t)$ or $f_n(\omega)$, may be uniquely specified for each sensor. When the target is assumed to be composed of magnetic dipoles, the diagonal terms are given by

$$f_n(t)=C_n+\sum_k M_{nk}e^{-\omega_{nk}t} \quad \text{or} \quad f_n(\omega)=C_n(0)+\sum_k \frac{\omega M_{nk}}{\omega - j\omega_{nk}},$$

for time-domain or frequency-domain sensors, respectively. The constant term C_n is non-zero only for ferrous targets. Typically, the first term of the sum dominates the target response, and therefore only the first term in the sum is necessary to sufficiently model the target response. This is equivalent to assuming the target may be represented by three mutually orthogonal magnetic dipoles. When the target is assumed to be rotationally symmetric (a body-of-revolution, or BOR), then $f_1(t)=f_2(t)$. The application of this general target model to targets sensed by the USGS ALLTEM sensor and LBL BUD AEM sensor is discussed below.

USGS ALLTEM Sensor-Target Response Model

The ALLTEM system is a multi-coil sensor, which emits a repetitive triangle pulse. The coils are large enough that they undermine the use of a single dipole to model for the source and receiver coils, so a Biot-Savart electromagnetic analysis of the source coils and Huygens' Principle for the receiver coils have been employed. In these analyses, the size, shape and orientation of the coils have been accounted for rigorously. In addition, the temporal properties of the source have also been rigorously modeled. These sensor parameters have been used to compute the time-evolving properties of the magnetic fields incident on the target. These fields are then used as the excitation for the magnetic dipole target model described above, which is then used to compute the time-evolving induced magnetic fields that are received at the sensing coils (rigorously modeled).

The target model for the USGS ALLTEM sensor is based on the magnetization tensor model previously introduced. For this sensor, the functions on the diagonal are $f_n(t)=C_n+M_n e^{-\omega_n t}$. When the target is assumed to be a BOR, then $f_1(t)=f_2(t)$. When the BOR assumption is not made, then each of the three functions may be unique.

With these models in place, we may now invert for the target model parameters, using a least-squares fitting procedure. This analysis yields the EMI dipole parameters of general targets, as sensed by the ALLTEM system. In addition, the analysis takes into account all data collected by the ALLTEM system, at multiple locations. The original implementation of the dipole model assumed the target is a body-of-revolution (BOR), and as such two sets of target parameters were extracted from the data. An example inversion for an 81mm shell at 0° inclination (Test Stand data) using the USGS ALLTEM sensor model with the BOR target assumption is shown below in Figure 4. The top left image is the measured data for the ZZM polarization as a function of space, the top right image is the model fit for the ZZM polarization

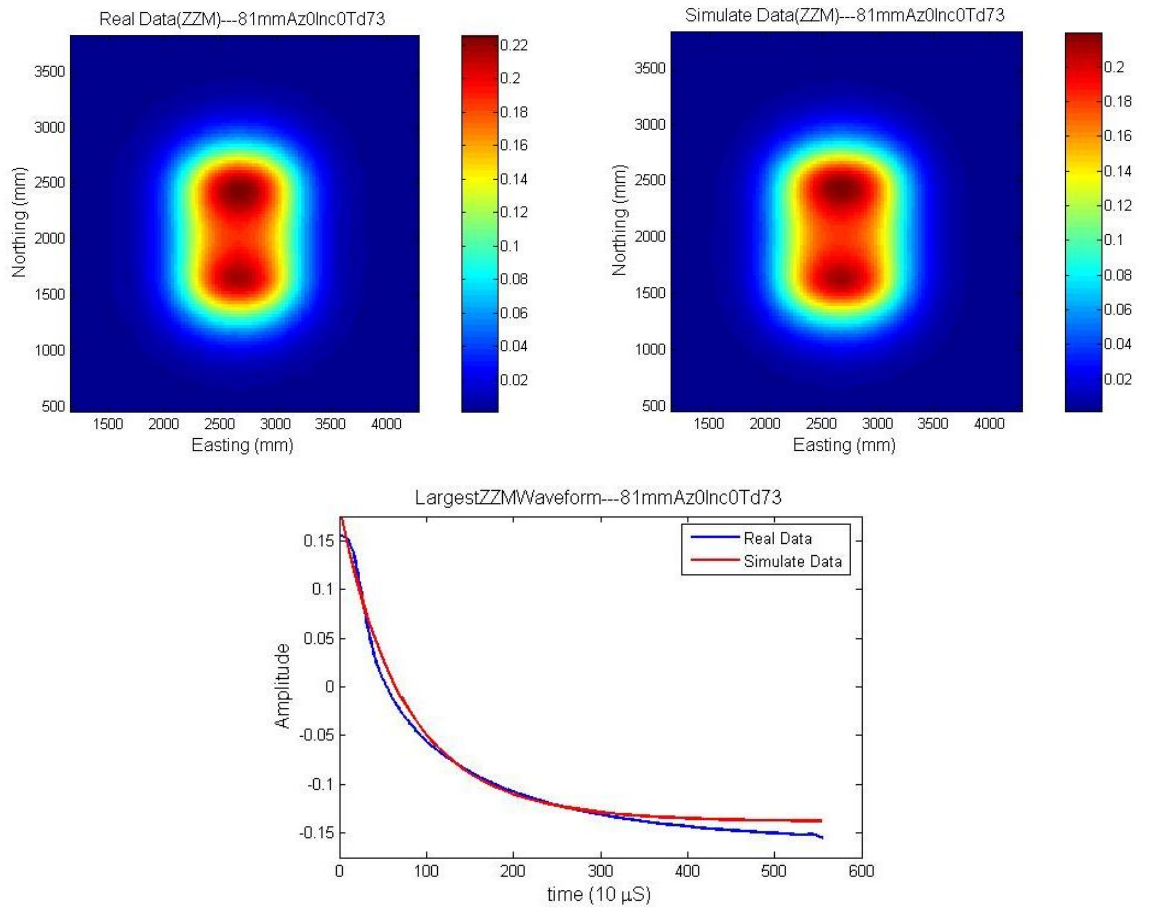

Figure 4. Example model inversion for an 81mm shell at 0° inclination (Test Stand data) using the USGS ALLTEM sensor model with the BOR target assumption. The top left image is the measured data for the ZZM polarization as a function of space, the top right image is the model fit for the ZZM polarization as a function of space, and the bottom plot shows measured and modeled time-domain responses for the measurement with the largest ZZM polarization response.

as a function of space, and the bottom plot shows measured and modeled time-domain responses for the measurement with the largest ZZM polarization response. This example illustrates the generally good agreement between the measured data and the modeled data using these sensor and target models.

The BOR assumption was later removed, and the resulting model inversions provided three sets of target parameters. The analysis, with and without the BOR assumptions, has been applied to all measured test stand data provided to date, with successful dipole-model inversions for a large class of UXO. The framework is now poised to analyze classification performance (based on the extracted EMI-model features), when provided with field data for UXO and non-UXO. Results from the dipole model analysis of the test stand data are discussed below.

LBL BUD Sensor-Target Response Model

A rigorous model of the LBL BUD AEM system has also been developed, including a full analysis of the details of all transmitting and receiving coils and the true physical dimensions of the coils. A single dipole approximation for the sensor transmit coils was not assumed as part of this analysis. This model

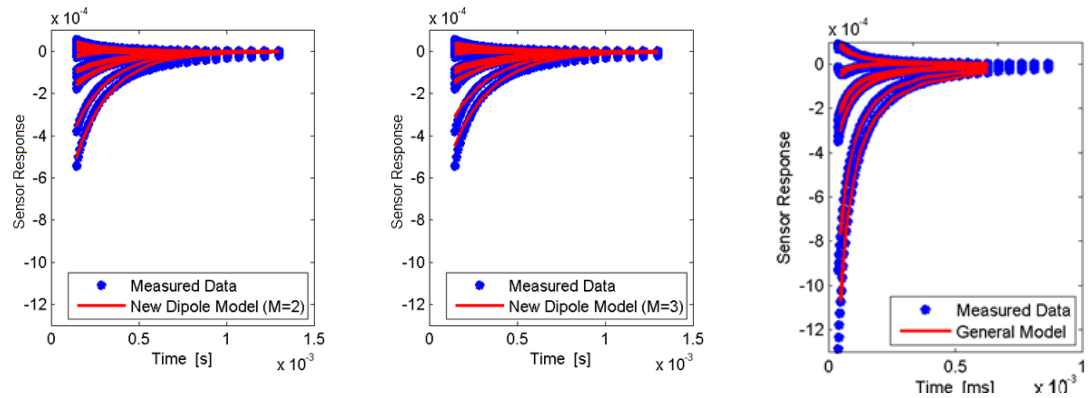

Figure 5. Example model inversions for Camp Sibert UXO target SE2-48 using the LBL BUD sensor model and magnetization tensor target models. Each plot shows the measured data and the modeled data. The left plot presents results for the magnetic dipole target model with the BOR target assumption $f_1(t) = f_2(t)$, the center plot is for the magnetic dipole target model without the BOR target assumption $f_1(t) \neq f_2(t)$, and the right plot is for the general (non-parametric) target model.

was successfully applied to measured BUD data collected by LBL in the YPG field test and the Camp Sibert discrimination study, wherein features were extracted from the measured data. In this analysis the target was modeled using the previously described wideband EMI magnetic dipole target model and the properties of the sensor were analyzed using the new LBL BUD AEM sensor model.

The target model for the LBL BUD AEM sensor is based on the magnetization tensor model described above. For this sensor, the functions on the diagonal are $f_n(t) = M_n \omega_n e^{-\omega_n t}$, and when the target is assumed to be a BOR, then $f_1(t) = f_2(t)$. When the BOR assumption is not made, then each of the three functions may have unique parameters. For the general model, each of the three functions $f_n(t)$ are arbitrary (non-parametric) time-domain functions. This formulation is a modification of the Berkeley target model.

The fits of the data were generally of high quality, and in the detection and classification work based on the extracted features encouraging separation was observed between UXO and non-UXO items. The magnetic dipole model for the targets originally included a body-of-revolution (BOR) assumption, which resulted in two sets of estimated object parameters for a target. The BOR assumption was later removed, so instead of extracting two sets of object parameters, three were extracted. We have also analyzed the BUD sensor data via a modification of the Berkeley model. Example inversions for all three target models applied to measured data for Camp Sibert UXO target SE2-48 are shown in Figure 5. Each plot shows the measured data and the modeled data. The left plot presents results for the magnetic dipole target model with the BOR target assumption $f_1(t) = f_2(t)$, the center plot is for the magnetic dipole target model without the BOR target assumption $f_1(t) \neq f_2(t)$, and the right plot is for the general (non-parametric) target model. Each model provides progressively better fits to the data, with the general magnetization tensor model providing the best fit to the data.

The inversion quality from the dipole model and modified Berkeley model are comparable, with the latter simpler in many ways. Both models yielded relatively good classification performance. Results from the analysis of the YPG calibration lane and Camp Sibert discrimination study data are discussed below.

Statistical Signal Processing and Feature Selection

Progress in the area of statistical signal processing proceeded in two stages: exploratory research in areas that were anticipated to be beneficial once the multi-axis sensor models and multi-axis data became available, and processing of field-collected LBL BUD AEM and test stand USGS ALLTEM multi-axis sensor data.

Initial efforts focused on simulations which could be completed without the benefit of multi-axis sensor models or sensor data. Specifically, simulations were performed to explore the degree to which multi-axis sensors may improve target parameter estimates and a Bayesian approach to mitigating sensor position errors was investigated. Simulations results indicate that, as expected, multi-axis sensors provide data that enables better estimation of target parameters (lower variability in the parameter estimates) and more robust discrimination performance, particularly when there is uncertainty in the sensor position. Additional simulations investigated explicitly addressing sensor position uncertainty within the parameter estimation algorithm. The Bayesian approach to mitigating sensor position uncertainty presented here also provided improved parameter estimates and more robust discrimination performance. Although it may not be necessary to mitigate sensor position uncertainty for sensors which measure multiple responses at a single sensor position, it could prove useful for sensors which take measurements at multiple sensor positions, or future applications in which multi-axis sensors measure the subsurface response at multiple sensor positions.

Additional simulations investigated alternate inversion techniques such as exploiting established single-axis inversion algorithms to improve multi-axis inversions and constraining multi-axis inversions to improve performance. The study undertaken here suggests that single-axis inversion algorithms may not perform as well when they are simply extended to multi-axis data. Furthermore, constraining the multi-axis inversion algorithms, such as through including an estimate of target depth, has the potential to improve inversion performance.

Another avenue of exploratory research involved data/measurement selection and sensor management. Data/measurement selection involves selecting a subset of the measured responses to use for model inversions and target classification. In this approach, the potential utility of a particular measurement is determined after it has been recorded. Thus, this approach does not affect the data collection process; it merely aims to select the most useful of all the measurements for target classification. Simulation results suggest that a principled approach to selecting a subset of the measured data for model inversion has the potential to improve discrimination performance. An alternate approach to this problem is to employ sensor management to predict which measurements would be most beneficial for target classification. Thus, this approach impacts the data collection process; it aims to guide the selection of the next sensor and measurement location so as to maximize the additional information the next measurement will provide. Simulations exploring this approach also demonstrate that it has the potential to improve target discrimination: discrimination performance is slightly improved while the number of measurements needed to achieve that performance is dramatically reduced. In summary, data/measurement selection attempts to determine the most useful of all the recorded measurements after data collection has been completed, while sensor management attempts to adaptively select during the data collection process only those measurements that will be most useful.

Finally, model-based processing of multi-axis sensors has the potential to produce a large number of features. Since many classification techniques degrade when the number of variables utilized for classification becomes too large, feature selection is often employed to reduce the feature set to a subset of salient features for classification. Many feature selection algorithms trade-off efficiency for

accuracy. A novel feature selection algorithm which attempts to strike a balance between efficiency and accuracy has been investigated here, and applied to features derived from the field-collected multi-axis sensor data. Initial results indicate this approach shows promise

Once multi-axis sensor data became available (LBL BUD YPG Calibration Lane in November 2006, USGS ALLTEM Test Stand data in July 2007, and LBL BUD Camp Sibert data in April 2008) algorithm development specifically for those sensors was undertaken. Initially, algorithm development focused on aspects which were not dependent on a sensor model, as the sensor models were not fully available until some time later. Relatively late into the program, it was discovered that the multi-axis inversions using the full sensor models were not as accurate as they could be. The model inversion process has since been improved, and a subset of the analyses have been repeated with the modified, improved, inversion process. The inversion process modifications and resulting changes to previously reported results are described in detail.

Multi-axis Sensor Simulations

Typical early generation electromagnetic induction (EMI) systems used for UXO detection and discrimination have two co-located coils – one for transmitting the electromagnetic field, and one for receiving the field resulting from currents induced in the subsurface object. Historically, these coils have been located so that the axes of both coils are perpendicular to the ground. Recent studies have suggested that adding additional transmit and/or receive coils in other orientations can improve sensitivity and discrimination performance by exciting the target more completely and measuring the full vector field instead of just the z-component. The goal of this preliminary study was to assess the level of performance gain using simulated data, but realistic field scenarios and uncertainties.

It has been established that UXO can be adequately modeled with a magnetic dipole target model, as described previously. The imaginary resonant frequencies of the EMI resonant modes are a function of target material parameters. Thus, these features can be used after the data is inverted using the model for signal processing and classification. Imagine a cylinder coordinate system with the target's symmetry axis as z, the frequency-dependent moment can be expressed as

$$M(\omega) = \hat{z}\hat{z}\left[m_z(0) + \sum_k \frac{\omega m_{zk}}{\omega - j\omega_{zk}} \right] + \left(\hat{x}\hat{x} + \hat{y}\hat{y}\right)\left[m_p(0) + \sum_i \frac{\omega m_{pi}}{\omega - j\omega_{pi}} \right]$$

The six target moment parameters are $m_z(0)$, $m_p(0)$, m_{zk}, m_{pi}, ω_{zk}, and ω_{pi},, where $m_z(0)$ and $m_p(0)$, account for the dipole moments contributed by ferrous targets. In practice, assuming a single term for the remnant magnetization for both the *x*- and *y*-axes ($m_p(0)$) has worked very well, though this assumption could readily be relaxed so there are two terms, one for each axis. The parameters ω_{zk} and ω_{pi},, are resonant frequencies, which are determined by the target geometry and material properties. Generally the first term in the sum, which is the principle dipole moment along each coordinate axis, is all that is needed to provide an accurate representation of the measured data from UXO and clutter.

In our initial simulations, we consider a system with one transmitter at a fixed orientation and three receiver coils along three perpendicular axes: x, y, and z. We consider a simulated target whose resonant frequencies along horizontal and perpendicular directions are 463 Hz and 168 Hz. These parameters were estimated from 81mm projectile field data (AETC database courtesy of Jonathan Miller). The target is buried 0.5 meters deep. The signal energy, defined as the integral of the squared signal, due to the target is a function of the target/sensor orientation, and generally averages about 6×10^7. To minimize the effect of the target orientation on conclusions based on this simulation, we use

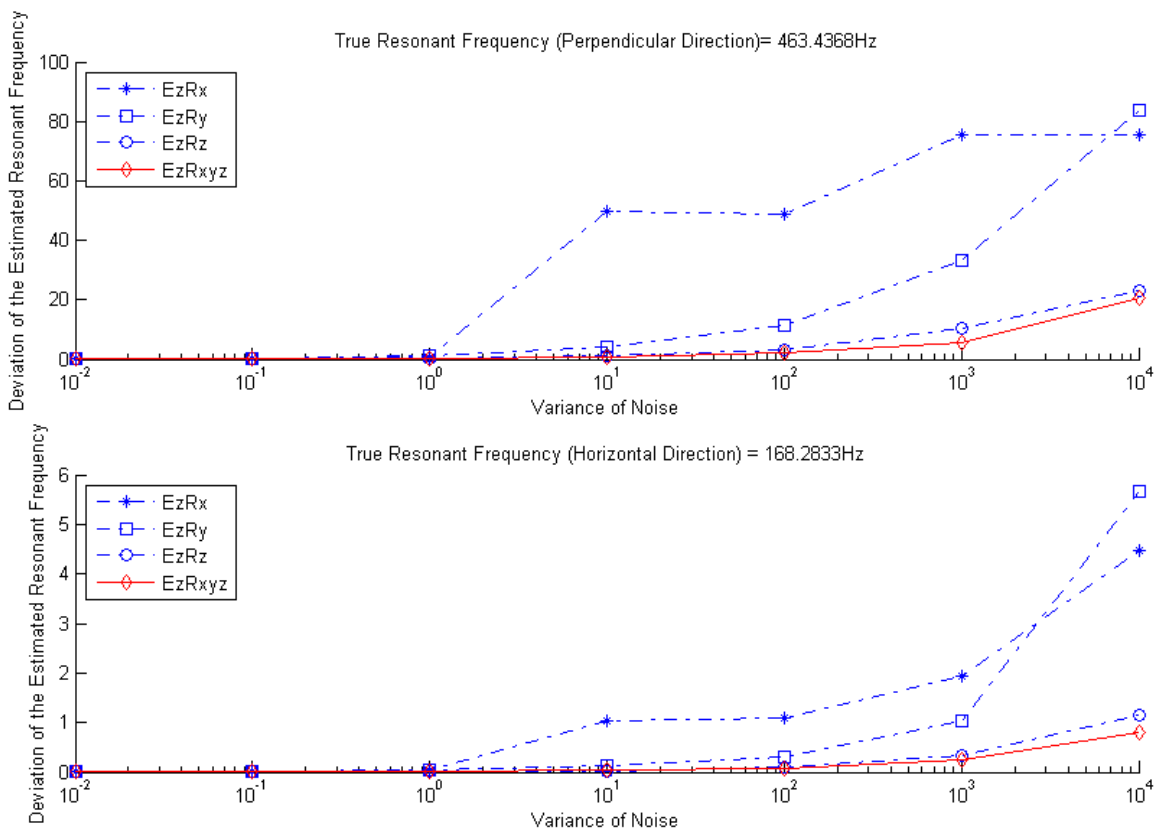

Figure 6. Standard deviation of estimated target resonant frequency as a function of noise variance for four different simulated system configurations. Ez indicates an excitation coil whose axis is perpendicular to the ground (z dimension), Rx, Ry, Rz denote receive coils in the x, y, and z dimensions, respectively, and Rxyz denotes three receive coils, one in each of the dimensions. The signal energy, defined as the integral of the squared signal, due to the target is a function of the target/sensor orientation, and generally averages about 6×10^7. The SNR corresponding to the noise variance ranges from about 35dB to about 95dB.

a uniform distribution for the target orientation. Using the magnetic dipole target model, we calculate the electromagnetic field measured from the target when the receiver coil is located in the three different orientations and add Gaussian noise to the simulated field. Using our standard inversion algorithms, we obtain the estimated moments, the position and the orientation of the buried target from the simulated noisy field.

One mechanism by which to compare the performance of various system configurations is to consider the mean and standard deviation of the estimated moments/resonant frequencies as a function of the level of the Gaussian noise. Figure 6 shows the standard deviation of the estimated resonant frequencies for the simulated 81 mm target. The standard deviation of the estimates using a three-axis system is the lowest, illustrating that the three-axis receive coil provides better performance with increasing noise variance than any of the single axis systems.

In order to further quantify performance gain, the simulated 81 mm projectile was classified against a simulated clutter field where the clutter moments were distributed uniformly. Each simulated clutter target was generated by randomly drawing its resonant frequencies from uniform distributions bounded by 500 and 1000 for ω_z and 700 and 1500 for ω_p. In addition, the clutter target orientation parameters

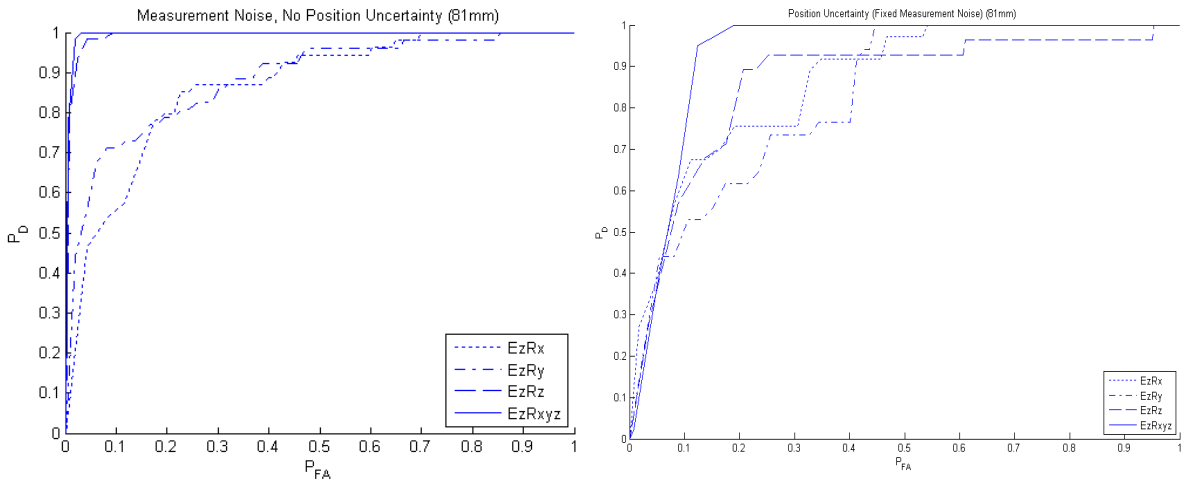

Figure 7. ROC curves plotting probability of detection versus probability of false alarm for simulated single- and multi-axis sensors when the sensor position is precisely known (left panel) and is uncertain (right panel). Measurement noise is included in both cases.

were randomly drawn from uniform distributions whose extents were the entire range of allowable values. A Bayesian classifier was used to discriminate the UXO object from the clutter using the estimated moments. Testing and training of the classifier were performed separately. Figure 7 illustrates the classification performance achieved from three single-axis and one multi-axis system when the sensor position is precisely known (left panel) and when it is uncertain (right panel). Some performance gain is obtained for the three axis system when the sensor position is precisely known. In the more realistic case of uncertain sensor position, however, substantially more performance gain is observed in the case of the multi-axis system.

Multi-axis Sensor Simulations Summary

This series of multi-axis sensor simulations provided support for the premise that multi-axis sensors could improve target parameter estimates, and thereby improve discrimination of UXO from clutter. The standard deviation on the target parameter estimates is lower for the simulated multi-axis system than for any of the individual simulated single-axis systems. More precise target parameter estimates provides greater confidence in the target identification. In addition, detection performance for the simulated multi-axis system is better than detection performance for any of the individual simulated single-axis systems, particularly when the spatial locations of the sensor measurements are subject to error.

These simulations focused on simulated systems in which there was a single transmitter and three orthogonal receivers. Existing multi-axis systems use either transmitters in multiple orientations with receivers in a single orientation (LBL BUD AEM system) or transmitters in multiple orientations with receivers in multiple orientations (USGS ALLTEM system). Further multi-axis sensor simulations that are consistent with the geometries of existing multi-axis sensors could provide insight from a signal processing perspective as to what those sensors are exploiting in the sensing phenomenology to provide improved performance over single-axis systems, as well as if future revisions to the sensor systems may improve performance even more.

Inversions with Uncertainty

Phenomenological modeling coupled with statistical signal processing has significantly improved capabilities for discriminating UXO from benign clutter using EMI sensor data. However, a potential limitation of this coupled approach for UXO detection is the model inversion has been shown to be sensitive to uncertainty in the measurement positions; measurement position uncertainty leads to more variable parameter estimates. While improved GPS helps mitigate the problem, it does not solve the problem. Thus, the problem of sensor measurement position uncertainty must be explicitly considered and addressed in order to overcome the discrimination performance degradations it causes. We are proposing a Bayesian approach to mitigating the effects of sensor position. The method specifically acknowledges that uncertainty in the sensor positions exists, and incorporates this knowledge into the processing algorithm to find the maximum likelihood feature estimates by integrating over the uncertain measurement positions.

Generally, the conventional approach to model inversion is to employ a least squares method to minimize the error between the measured data, $\mathbf{r}$, and the model prediction, $\mathbf{s}$. Suppose the measured data can be predicted with a forward model $\mathbf{s}(\Omega,\gamma,\rho,k)$, where Ω represents the parameters to be estimated for subsequent target discrimination, γ, represents then parameters to be estimated as part of the model but are irrelevant for subsequent discrimination (nuisance parameters), ρ represents model parameters which are assumed to be known, and k represents the signal dependence on time, frequency, space, etc. While this approach is frequently applied in many parameter estimation applications, it implicitly assumes that all fixed parameters (ρ) are accurate. If the parameters ρ are not known with certainty, the incorrect assumption that they are known may degrade inversion performance.

In our Bayesian approach, we instead take as the estimate of the desired parameters, Ω, that value of those parameters which maximizes the probability density function of the desired parameters given the measured data, $\mathbf{r}$

$$f(\Omega \mid \mathbf{r}) = \frac{\int_{\rho}\int_{\gamma} f(\mathbf{r} \mid \Omega,\rho,\gamma) f(\gamma) f(\rho) d\gamma d\rho f(\Omega)}{f(\mathbf{r})}.$$

Thus, the integration over the uncertainty in the nuisance parameters eliminates the need to estimate those parameters and the integration over the assumed known parameters mitigates the reliance on their accuracy.

The simulation study considered here is the discrimination of a single known UXO target from axially symmetric clutter. The frequency-domain sensor data are measured at seven frequencies: 90, 150, 330, 930, 2790, 8190, and 20010 Hz. The selected UXO poles are Ω = [1000 2000] Hz, and the clutter poles are assumed to be uniformly distributed from 500 to 1500 Hz for ω_p and 1500 to 2500 Hz for ω_z. The other target parameters are selected from uniform distributions.

The proposed Bayesian model inversion technique is evaluated through numerical simulations, with the magnetic dipole target model previously described employed to simulate the sensor response to both UXO and clutter targets. This model characterizes the target in terms of poles, or imaginary resonant frequencies, which are the features utilized for UXO/clutter discrimination. The characteristic poles are estimated from the simulated sensor data via the conventional nonlinear-least-squares (NLS) approach and the proposed Bayesian model inversion technique. The root mean square (RMS) error of the pole estimates and discrimination performance are evaluated for a realistic range of sensor position

uncertainties. The results show that the Bayesian approach to model inversion is an effective technique for reducing the standard deviations of the target parameter estimates, and improving UXO/clutter discrimination in the presence of sensor measurement position uncertainty.

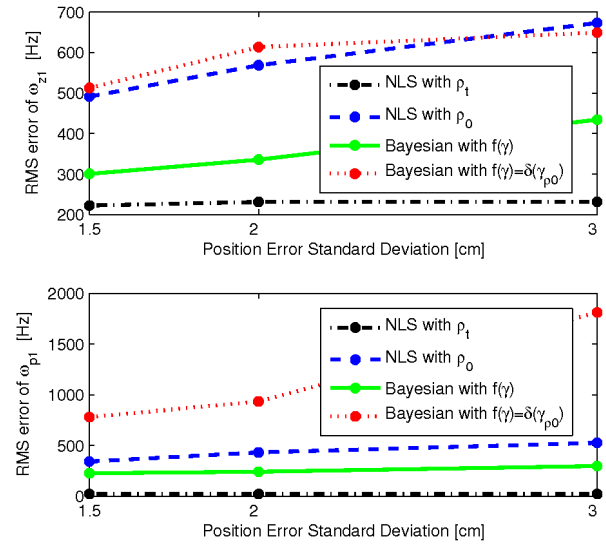

Figure 8. RMS error of pole estimates for the various inversion approaches. The nominal pole are ω_z =2000Hz and ω_p =1000Hz.

The error in the estimated features as a function of the standard deviation of the position error are shown in Figure 8. The black curves are found assuming the true positions are known with conventional (least squares) estimation, the blue curves use the assumed (nominal) positions in the conventional inversion, the green curves utilized Monte Carlo integration in the inversion process, and reduce the error over the conventional (blue) approach. The red curves also utilize Monte Carlo integration in the inversion process, but incorrectly assume knowledge of some parameters is true, and demonstrate the performance degradation that can result from incorrect prior knowledge.

The RMS errors of the pole estimates show that larger sensor position uncertainty leads to larger errors. In addition, they demonstrate the inadequacy of utilizing nominal sensor positions with the NLS algorithm; errors in the assumed sensor positions substantially increase the RMS error in the pole estimates. They also show that applying an appropriate Bayesian model inversion technique [Bayesian with $f(\gamma)$] substantially reduces the RMS error compared to a conventional approach that utilizes nominal sensor positions [NLS with ρ_0], although the estimates do not achieve the upper bound in performance (or lower bound in estimation error) obtained when the true sensor position is known [NLS with ρ_t]. Further, comparison of the two Bayesian techniques, integrating over the uncertainty [Bayesian with $f(\gamma)$] and estimating the uncertain parameters [Bayesian with $f(\gamma) = \delta(\gamma_{\rho 0})$], demonstrates that the integration approach provides superior performance. This occurs because the estimate of g obtained using the NLS algorithm with ρ=ρ_0 is rather poor, and assuming incorrect values for γ in the Bayesian algorithm seriously degrades estimation performance.

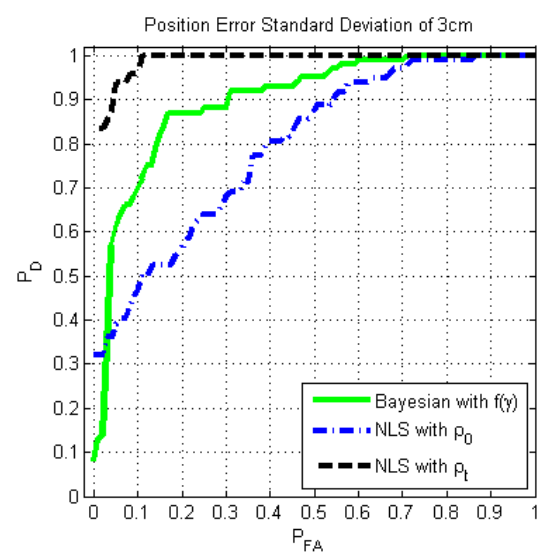

Figure 9. ROC curves illustrating improved performance with the proposed Bayesian approach to mitigate sensor position uncertainty.

A similar trend is evident in the ROCs shown in Figure 9; position uncertainty degrades discrimination performance when it is not

addressed by the inversion method. Integrating over the uncertainty, however, mitigates the sensor position errors and improves discrimination performance. The black curve again assumes true positions are known and provides an upper bound. Clear performance improvement in terms of the probability of making a correct decision (green vs. blue) is observed when using our proposed approach. With respect to unnecessary digs saved (P_{FA} at P_D=1), however, the performance improvement is modest.

Inversions with Uncertainty Summary

It has been shown that sensor position errors degrade the performance of algorithms which employ phenomenological models. The performance degradation caused by the sensor position errors cannot be overcome by simply increasing the sensor data SNR; the position uncertainty must be considered explicitly within the signal-processing strategy. Addressing the sensor position errors with the Bayesian inversion framework proposed here effectively mitigates the sensor position errors and improves both feature estimation and target detection performance.

Alternate Inversion Techniques

As part of this project, we are concerned with determining how best to extract features from more complex data sets. There are many aspects of the inversion process to be considered including a variety of error metrics, decision criteria, and inversion techniques. The most traditional approach for inversion relies on the Levenburg-Marquart technique for performing a non-linear least squares search of the feature space to find the best fit of the features to the data. Generally, this problem has not been studied in depth, although several problems associated with multiple local minima when inverting data from the UXO application area have been noted. We first considered the degree to which multi-axis sensor data degraded results produced by the inversion algorithms originally developed for single-axis sensor data, then we investigated the extent to which constraining the multi-axis sensor data inversion by using inversion results from other sensing modalities or single-axis sensor data inversions could improve the multi-axis sensor data inversion as well as improvements to the implementation of the inversion algorithms for the multi-axis sensor data.

In our initial simulations, the traditional approach continued to be fairly robust, so no changes in error metric (L2 norm), decision criteria (lowest error), or inversion process (L-M using Matlab's lsqnonlin function) were made. However, it was noted that the multiple local minimum problem escalated for multi-axis data. As we moved towards processing real data, specifically the LBL BUD AEM data, we began to encounter serious issues with our inversion approach. There are several reasons that this might occur – one is that the underlying model no longer fits the data, the other is that the inversion process that worked marginally on single axis data was simply not appropriate for more complicated data sets. At the time this study was completed, the LBL BUD AEM sensor model was still under development.

We undertook a study to ascertain how significantly the inversion process was degraded with multi-axis data, and whether prior knowledge on some of the parameters that are estimated in the inversion could improve the inversion results. The first thing that we learned is that for the multi-axis sensor, many more inversions do not converge (partially due to the multiple minima in the search space) for the multi-axis simulations than the single-axis simulations. Table I shows the randomization of the parameters used in our initial study of the inversion problem. Figure 10 shows pseudo-histograms of the number of iterations it takes for the inversion to converge to an answer – multi-axis results are shown on the left, single-axis results are shown on the right. Clearly, many more inversions do not converge (partially due to the multiple minima in the search space) for the multi-axis simulations than the single-axis simulations.

theta	[0, pi]
phi	[0, pi]
depth	[.25, .75]
dx	[-.25, .25]
dy	[-.25. .25]
C1	N(0,5)
M1	[5, 35]
W1	10^[1.5, 2.5]
C2	N(0,5)
M2	[15, 45]
W2	10^[2.9, 3.4]

Table I. Model parameters used in the inversion study.

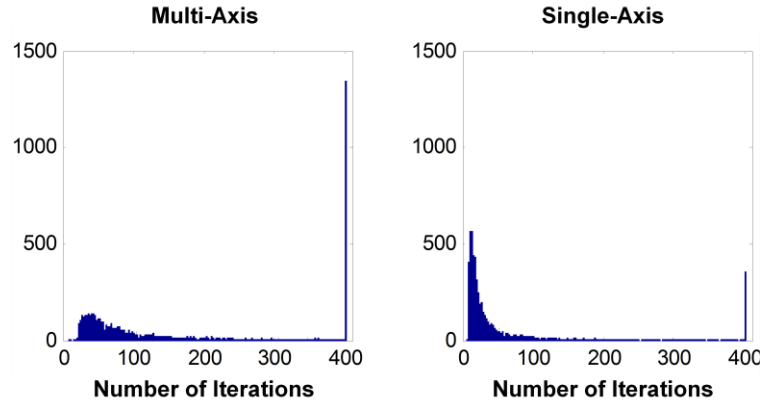

Figure 10. Pseudo-histogram of number of iterations required for the inversion of simulated UXO data to converge. The maximum number of iterations is limited to 400.

One of the issues associated with the difficulty of this problem is the number of parameters that we are trying to extract from a given set of data. The magnetic dipole target model requires 11 parameters, while the AETC model requires 3-5. [Interestingly, in a side study of the AETC model, these inversion problems did not occur – but discrimination could not be effected on the AETC GEM-3 data base with the extracted features.]. In an effort to understand the data, we set up a simulation in which a subset of the parameters was assumed known, and the impact of this knowledge was considered in light of the inversion performance. While it is impossible to know any of the parameters exactly, it is conceivable that estimates could be obtained from a different sensor modality (e.g. magnetometer). Our main question with this study was whether we should attempt to use a single-axis inversion to obtain estimates of the parameters, and which parameters would be the most useful or helpful. The cases considered are listed in Table II.

There are two metrics we considered with regards to the inversion 'success' – number of iterations to converge and error between the model fit and the data. We tabulated (1) the percentage of inversions that required fewer than 100 iterations and (2) the percentage of inversions requiring more than 400 iterations, and these results are shown in Table III.

	Parameters known	Parameters to find via lsqnonlin and emif2
case 1	All position (theta, phi, depth, dx, dy)	6
case 2	All dipole (C1, C2, M1, M2, W1, W2)	5
case 3	theta	10
case 4	phi	10
case 5	depth	10
case 6	dx	10
case 7	dy	10
case 8	C1	10
case 9	M1	10
case 10	W1	10
case 11	C2	10
case 12	M2	10
case 13	W2	10

Table II. Cases considered/parameters known for the follow-on

	iters < 100	iters > 400
case1	53.1%	20.6%
case2	84.0%	5.4%
case3	2.9%	78.2%
case4	1.7%	84.5%
case5	27.5%	29.7%
case6	1.5%	82.5%
case7	1.5%	83.5%
case8	1.3%	80.7%
case9	3.7%	73.1%
case10	1.8%	75.6%
case11	1.5%	84.3%
case12	5.5%	66.0%
case13	2.2%	79.7%

Table III. Summary of number of iterations per case.

One interesting point can be noted when looking at cases 3-13 is that the parameter that most substantially decreases the number of iterations required for inversion is depth. Considering the dipole parameters, the resonant frequency gains (M) appear to help more than the resonant frequencies (W). However, these numbers are only related to the time to converge, not accuracy of the inversion. Speeding up the inversion process is only useful if the resulting inversions are accurate.

Table IV shows calculations of the dipole parameter accuracy for each of the 13 cases. Four parameters are shown – m1, w1, m2, and w2. Under each parameter are two columns listing the percentage of inversions (out of the total 6500) where the error was (left) less than 1% of the true parameter value or (right) greater than 25% of the true parameter value. Since the 4 parameters in the table are all dipole parameters, case 2 (given all 6 dipole parameters) has the true values and is listed as zero error. There is also a single case (case 9, 10, 12, 13) for each parameter for which error is zero; these are the case where the true parameter value was provided to the inversion. Ideally, for each parameter we want the value in the left column (small error) to be big and the right column (large error) to be small.

From Table IV, we can see that providing depth information (case 5) is still quite helpful for both fast and accurate inversions. In general, giving one parameter of a dipole (c, m or w) helps with finding the other 2 parameters. However, information is not helpful across the dipoles (e.g. knowing c1 does not help find c2, m2, or w2). Based on just a visual analysis, the depth parameter again appears to be useful (always 70% to 80% of inversions with less than 1% error).

These results are consistent with 'prevailing knowledge', and may suggest that to improve accuracy an estimate of depth from a magnetometer might be helpful. The results also suggest that using an inversion from a single axis version of a multi-axis system may be less prone to catastrophic errors (Figure 10), and that the estimates of the parameters obtained from that inversion might be used to constrain the multi-axis inversion and thus improve estimation errors (Table IV).

We continued this effort in order to assess if there are other tools in the Matlab toolbox library that can be leveraged to provide better inversion results. The 'lsqnonlin' function has several flags that can be set, and the impact these parameters have on inversion performance was assessed. This effort also

	m1		w1		m2		w2	
	error<1%	error>25%	error<1%	error>25%	error<1%	error>25%	error<1%	error>25%
case1	95.4%	1.5%	95.3%	1.4%	99.8%	0.1%	99.8%	0.1%
case2	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%	100.0%	0.0%
case3	53.4%	11.9%	53.9%	10.5%	65.4%	9.7%	84.4%	5.5%
case4	46.9%	16.3%	47.7%	10.9%	56.9%	10.0%	81.3%	3.3%
case5	73.3%	11.7%	73.0%	7.0%	82.7%	9.9%	82.9%	5.7%
case6	51.6%	14.8%	52.3%	10.3%	66.7%	9.2%	83.3%	4.8%
case7	51.6%	14.3%	52.8%	10.0%	66.5%	8.8%	84.1%	4.9%
case8	70.2%	8.2%	74.4%	12.1%	59.8%	9.3%	83.5%	10.7%
case9	100.0%	0.0%	72.0%	14.9%	62.3%	9.8%	81.9%	10.6%
case10	85.5%	3.6%	100.0%	0.0%	82.0%	3.1%	94.6%	2.2%
case11	47.9%	13.3%	48.7%	18.6%	64.6%	7.9%	82.3%	9.9%
case12	53.9%	14.3%	54.9%	20.2%	100.0%	0.0%	82.9%	12.5%
case13	64.3%	8.2%	65.8%	9.0%	73.8%	4.9%	100.0%	0.0%

Table IV. Summary of errors per case.

monitors the SIG/SAIC work in developing alternative sampling strategies for inversion and will leverage any advances made there. Progress included a thorough assessment of the inversion tools Matlab provides and a consensus on how flags should be set without the resampling approaches under investigation by SIG.

Other work in this general area (improving our ability to obtain robust estimates of the parameters from time-domain data) was initiated and has continued based on the issues that arose in processing the real data. Specifically, an alternate signal model which may be better suited for inversion, a sum of decaying exponentials signal model, has been investigated. The relationship between the sum of decaying exponentials signal model and the dipole model suggests that the sum of decaying exponentials signal model may lead to an inversion problem that is better-posed since it eliminates ambiguity between inversion parameters.

The dipole model is a weighted sum of the terms in the magnetization tensor, $f_n(t) = M_n \omega_n e^{-\omega_n t}$, where the weights on the individual terms are determined by the target orientation (inclination and azimuth relative to the sensor) and the transmitter/target/receiver orientation (target depth and horizontal offsets from the sensor center). Thus, given the target, transmitter, and receiver orientation parameters, the dipole target signal model without a BOR assumption is

$$s_D(t) = \sum_{n=1}^{3} A_{Dn} M_n \omega_n e^{-\omega_n t}$$

where A_{Dn}, a function of the target orientation and the transmitter/target/receiver orientation, is the weight on the nth term. The LBL BUD AEM sensor utilizes pairs of receiver coils, and records the difference between the two coils in a receiver pair. Denoting the signal on the top coil in a pair by $s_{DT}(t)$ and the signal on the bottom coil by $s_{DB}(t)$, the measured LBL BUD EM sensor response may be modeled as

$$\begin{aligned} s_{BUD}(t) &= s_{DT}(t) - s_{DB}(t) \\ &= \sum_{n=1}^{3} A_{DTn} M_n \omega_n e^{-\omega_n t} - \sum_{n=1}^{3} A_{DBn} M_n \omega_n e^{-\omega_n t} \\ &= \sum_{n=1}^{3} (A_{DTn} - A_{DBn}) M_n \omega_n e^{-\omega_n t} \\ s_{BUD}(t) &= \sum_{n=1}^{3} \tilde{A}_{Dn} e^{-\omega_n t} \end{aligned}$$

where $\tilde{A}_{Dn} = (A_{DTn} - A_{DBn}) M_n \omega_n$. Thus, the dipole target model for the LBL BUD sensor can be expressed in the form of a weighted sum of decaying exponentials, which suggests an alternate signal model of a weighted sum of decaying exponential signals.

To test the validity of the above analysis, synthetic data were generated using the magnetic dipole target model and LBL BUD AEM sensor model, and then decay rates were estimated using the weighted sum of decaying exponentials model. The target parameters selected for the synthetic data were: M_1=5, ω_1=2000, M_2=5, ω_2=5000, inclination $\theta=\pi/4$, azimuth $\phi=\pi/4$, depth=0.75m, sensor position=[0,0]m. The noise-free synthetic data is shown in Figure 11. Each panel shows the responses for the 8 receiver pairs for a single transmitter.

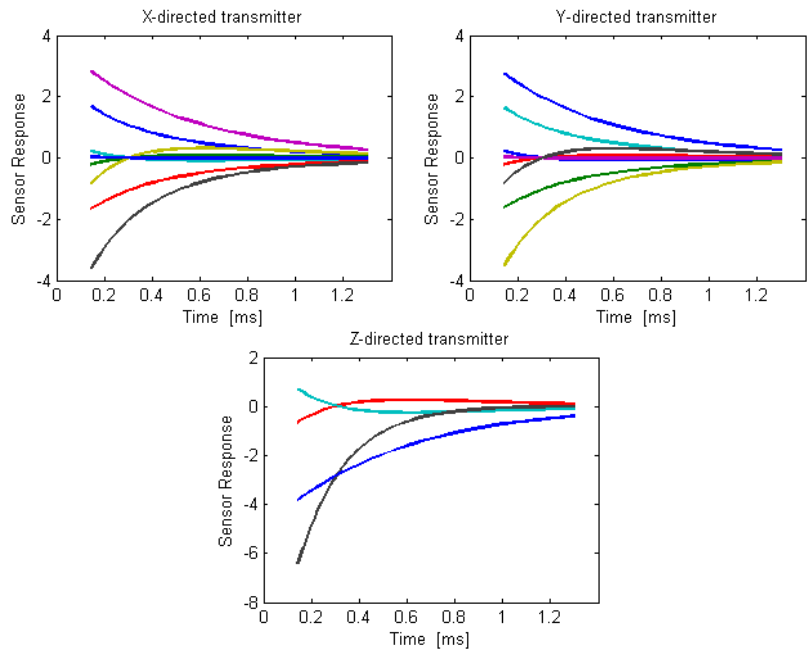

Figure 11. Synthetic LBL BUD AEM sensor data for a magnetic dipole target model with parameters M_1=5, ω_1=2000, M_2=5, ω_2=5000, inclination $\theta=\pi/4$, azimuth $\phi=\pi/4$, depth=0.75m, sensor position=[0,0]m.

The decay rates are estimated from the 4 strongest receivers for each transmit coil using the weighted sum of decaying exponentials signal model The decay rate estimates are shown and tabulated in **Error! Reference source not found.** and **Error! Reference source not found.**. **Error! Reference source not found.** plots the estimated decay rates (the true decay rates are α_1=2000 and α_2=5000) and **Error! Reference source not found.** tabulates the decay rate estimates and highlights those estimates with large error. The decay rates are estimated using a single transmit coil and receive coil pair (blue dots), using all 4 receiver pairs for a single transmit coil (red squares), and using all 4 receiver pairs for all 3 transmit coils (green diamond). In **Error! Reference source not found.**, the receiver number is color-coded to correspond to the color curve used to plot the synthetic data in Figure 11. It is interesting to note that the decay rate estimates obtained using data from a single receiver are prone to large errors, while decay rate estimates obtained using data from multiple receivers is not. This is consistent with the initial multi-axis simulations evaluating the quality of the parameter estimates for single-axis and multi-axis sensors. The simultaneous inversion of data from multiple channels improves parameter estimates in both cases.

The fundamental difference between the magnetic dipole and sum of decaying exponentials signal models is that the magnetic dipole model includes the transmitter/target/receiver orientation in the model, and therefore imposes constraints on the relative amplitudes of the received signals in the inversion process. Since this model maintains a distinction between the scaling constant due to the transmitter/target/receiver orientation, A_{Dn}, and the scaling constant due to the magnetization tensor, $M_n\omega_n$, there may be ambiguity in the inversion. The sum of decaying exponentials signal model, on the other hand, does not include transmitter/target/receiver orientation, and therefore does not impose constraints on the relative amplitudes of the received signals. Consequently, the scaling constants due to the transmitter/target/receiver orientation and the magnetization tensor are grouped

together into a single constant $\tilde{A}_{Dn}$ which removes that potential source of ambiguity from the inversion.

Since the applicability of the sum of decaying exponentials signal model has been established by demonstrating its relationship to the dipole model, further work focused on improving the numerical methods implemented for decay rate estimation via nonlinear least squares (NLS) methods. The specific factors that have been considered are: 1) parameterization of the signal model (decay rates, logarithm of the decay rates, or poles), 2) scaling of the data prior to applying standard NLS estimation algorithms, 3) selection of exit criteria for the NLS algorithms, 4) fitting the selected model to the data or the logarithm of the model to the logarithm of the data, 5) the choice of the NLS estimation algorithm (Levenberg-Marquardt or Gauss-Newton), and 6) the use of analytic solutions or numerical estimates of the Jacobian. The overriding goal has been to develop a "Decay Rate Estimation Toolbox" which will provide robust and accurate decay rate estimation for a wide variety of measured sensor data with minimal input from the user. It is anticipated that the lessons learned while improving the numerical methods for decay rate estimation will provide insight as to how the numerical methods for inverting other models might be improved as well.

The weighted sum of decaying exponentials signal model is linear in the amplitudes, or weights, for each decaying exponential but nonlinear in the decay rates. Thus, estimation of the model parameters (amplitudes and decay rates) can be approached using separable least squares. This approach utilizes a nonlinear technique, such as gradient descent, to estimate the decay rates, but a conventional linear method to estimate the amplitudes given the current estimate of the decay rates, and has the benefit of reducing the dimensionality of the space explored using nonlinear techniques by one half. In this approach, the quality of the amplitude estimates directly affects the performance of the nonlinear search for the decay rates which minimize the objective function. Within the literature, it has been suggested that utilizing less than 100% of the available data to estimate the amplitudes provides amplitude estimates with lower variance. Preliminary results from simulated data with a single pole suggest that the final data point to utilize for amplitude estimation is at a fixed point in time, regardless of the sampling rate or the length of the measured signal. Simulations are continuing to determine the relationship between the final data point and the decay rate(s) present in the signal. The results of this work have the potential to improve decay rate estimation by improving the linear amplitude estimation step so that the nonlinear search for decay rates which optimize the objective function is more effective and efficient.

Simulations continue to be ongoing to evaluate the effects of each of the factors under consideration. The CRLB for decay rate estimates will be utilized to assess the performance of each approach for decay rate estimation. It is anticipated that the parameterization of the signal model will impact decay rate estimation performance primarily by altering the shape of the error surface to make it better suited for numerical minimization so that the minimum can be found more quickly and accurately. This effect has been observed when a pole parameterization is used rather than a decay rate parameterization. It is anticipated that the logarithm of the decay rates will provide similar benefits as the pole parameterization without the need to select a scaling constant as the pole parameterization requires. It remains to be seen if the logarithmic parameterization will provide decay rates estimates with errors as low as those provided by the pole parameterization. The scaling of the data prior to applying the algorithms is an important consideration because it both affects the exit criteria for the NLS estimation algorithms and has numerical implications if the magnitude of the measured signals is very small. The current approach is to scale the measured signals so they have an average signal energy equal to 1. With this scaling choice, the desired residual (or energy in the error) at which the NLS estimation

algorithm will exit can be interpreted as a percentage of the average signal energy. Thus, the energy in the error is at most a pre-defined percentage of the energy in the measured signals. The selection of "linear space" or "logarithmic space" for estimating the decay rates will emphasize the early and late portions of the measurement differently. A logarithmic scale places more emphasis on the later part of the measured responses than the linear scale does. Though there are some suggestions in the literature that a logarithmic scale may improve decay rate estimation (under conditions of Poisson noise), it has not yet been determined if this will be the case for our problem. The selection of the NLS algorithm may be affected by the selection of the model parameterization. Since the different parameterizations are anticipated to produce error surfaces with different characteristics, they may not all be well-suited for the same NLS algorithm. The Jacobian for each of the proposed signal models has been calculated and implemented. Initial results indicate that numerical estimation of the Jacobian provides more stable and efficient decay rate estimations. This counter-intuitive result is not yet fully understood, though it is suspected that there may be numerical reasons for this. It is possible that very near the minimum of the errors surface the Jacobian becomes quite small, and the analytic solution becomes numerically unstable while the numerical solution is able to maintain numerical stability. Efforts at investigating this will continue.

The current implementation of the Decay Rate Estimation Toolbox has been applied to both the USGS ALLTEM data and the LBL BUD AEM data, and the results have been promising. Example inversions for USGS ALLTEM data and LBL BUD AEM data using the sum of decaying exponentials signal model are shown in Figure 12 and Figure 13. Figure 12 shows example inversions for an 81mm shell at 0° inclination (USGS ALLTEM Test Stand data), and Figure 13 shows example inversions for Camp Sibert UXO target SE2-48 (LBL BUD AEM data). In both figures, each plot shows the measured data and the modeled data, as well as the residual between the two below. The left plot presents results for a single decaying exponential signal model, the center plot is for two decaying exponentials, and the right plot is for three decaying exponentials. These figures illustrate that the measured data is well-modeled by a sum of decaying exponentials signal model, and the residuals are fairly low. Further, increasing the number of decaying exponentials in the model decreases the residual.

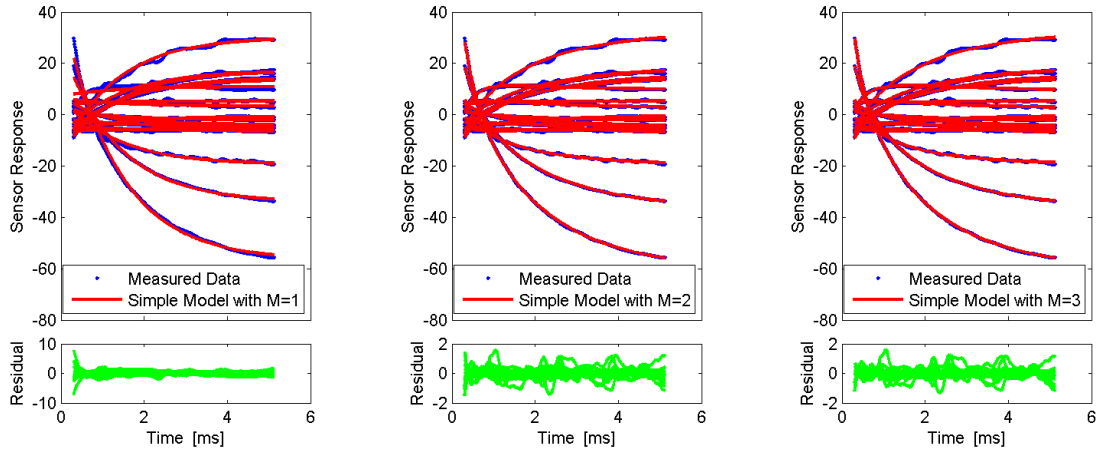

Figure 12. Example model inversion for an 81mm shell at 0° inclination (USGS ALLTEM Test Stand data) using the sum of decaying exponentials signal model. Each plot shows the measured data and the modeled data, as well as the residual between the two below. The left plot presents results for a single decaying exponential, the center plot is for two decaying exponentials, and the right plot is for three decaying exponentials

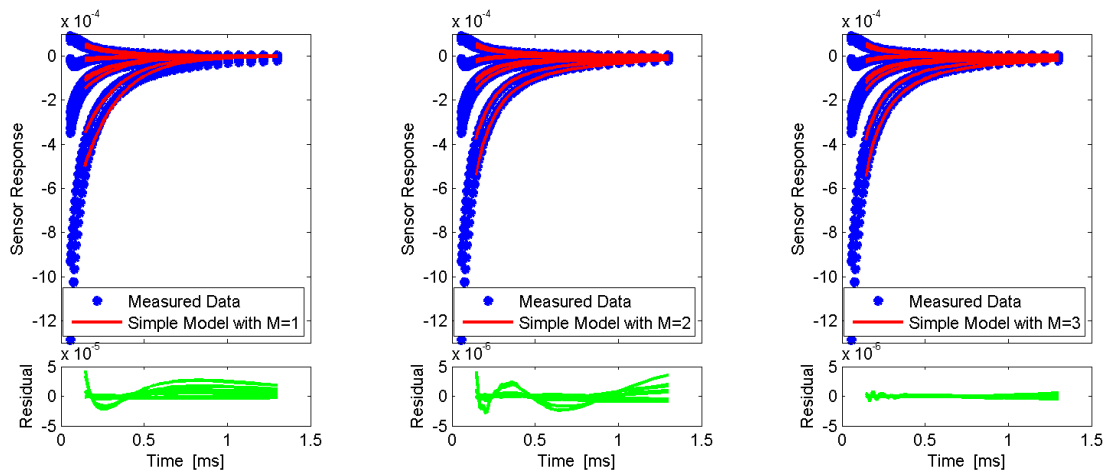

Figure 13. Example model inversions for Camp Sibert UXO target SE2-48 (LBL BUD data) using the sum of decaying exponentials signal model. Each plot shows the measured data and the modeled data, as well as the residual between the two below. The left plot presents results for a single decaying exponential, the center plot is for two decaying exponentials, and the right plot is for three decaying exponentials.

Alternate Inversion Techniques Summary

Model inversion results have been found to be influenced by the manner in which the inversion is implemented, both in the choice of parameters in standard least squares algorithms and choice of parameters to estimate in the inversion or assume known via an auxiliary sensor measurement. Simulations results showed that having a good estimate of the target depth a priori may be useful for improving target parameter estimates. They also showed that inversions for a single-axis from a multi-axis sensor tend to be less prone to catastrophic failure, and suggest an inversion strategy whereby the single-axis data are inverted first with those results utilized to constrain the multi-axis data inversion. With all these approaches and strategies, care must be taken to ensure the optimization parameters

within standard numerical optimization routines are set appropriately, as inappropriate parameter choices can adversely affect the inversion.

Data/Measurement Selection

A common approach to the UXO discrimination task is to utilize phenomenological models of the response of subsurface objects to sensor excitation. The fitted models can be used to generate features for UXO discrimination. One recent focus of our work has been an investigation of the effect of several properties associated with the measured data on the model inversion results. Measurement spatial density, the number of measurements, measurement energy, and measurement spatial coverage were considered to investigate any relationships between the measurement parameters and model fit error or UXO discrimination performance. The goal was to consider whether any results may provide evidence supporting the existence of an optimal subset of the measurements.

This study used data collected from the open field range at Yuma Proving Ground using the GEM-3 sensor on the MTADS system. A total of 169 anomalies were considered, consisting of 106 UXO and 63 non-UXO clutter. A phenomenological dipole model was fit for each anomaly to eighteen different data sets with various measurement densities and spatial areas, and the relationship between these properties of the data and the model fit error were analyzed. The fit error metric used in this study is the error between the data and the model fit normalized by the energy in the data,

$$\varepsilon = \frac{\sum_n \sum_\omega \left[\hat{Y}(n,\omega) - Y(n,\omega)\right]\left[\hat{Y}(n,\omega) - Y(n,\omega)\right]^*}{\sum_n \sum_\omega Y(n,\omega) Y(n,\omega)^*},$$

Where $\hat{Y}$ and Y are the modeled response given the estimated parameters and the measured data, respectively, and * denotes the complex conjugate. The normalization of the residual by the energy in the data ensures that the model fit error can be fairly compared for data with differing numbers of measurements as well as for data at different signal-to-noise ratios.

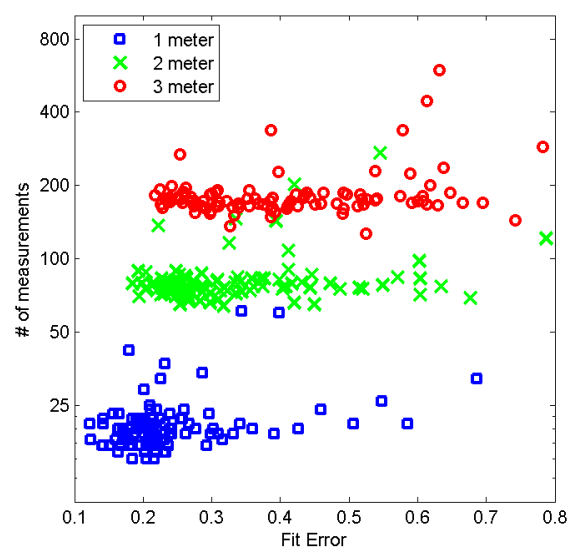

Figure 14. Model fit error versus number of measurements used for model fitting for three different polygon sizes: 1 meter diameter (blue), 2 meter diameter (green), and 3 meter diameter (red).

Model fit error versus the number of measurements utilized for the inversion is shown in Figure 14 for three different polygon sizes: 1 meter (blue), 2 meter (green), and 3 meter (red). The measurement density is relatively constant across the three polygon sizes, so as the polygon diameter increases the number of measurements also increases. For this reason, there is some ambiguity as to whether the trends suggested in this plot are due to changes in the number of

measurements or the polygon diameter. Nevertheless, it is interesting to observe that both the model fit error and the variability in the model fit error tend to increase as more measurements are utilized for the inversion, or as the polygon diameter increases. This suggests that some combination of fewer measurements and smaller diameter polygons can reduce the model fit error, and thereby improve the model inversions.

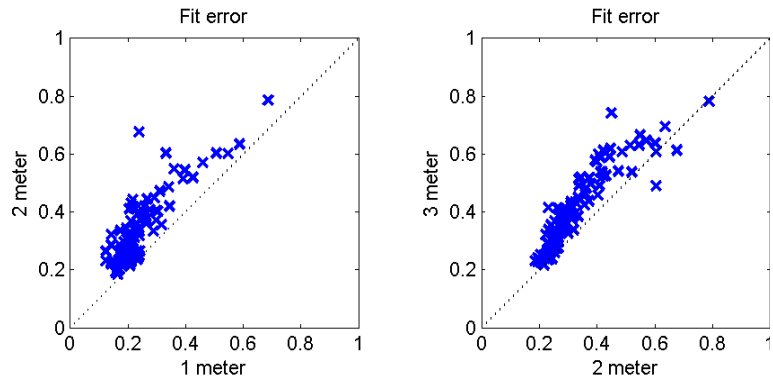

Figure 15. Scatter plots of the model fit error achieved with different diameter polygons. Left panel shows 1m polygon versus 2m polygon; right panel shows 2m polygon versus 3m polygon. Each point corresponds to one of the 106 targets. Points along the diagonal line have the same fit error using both polygon diameters.

Fit error with different polygon diameters and measurement densities have also been analyzed individually. Comparisons of the fit error for the three polygon diameters considered in this study are shown in Figure 15. Each panel plots the fit error for one polygon diameter versus the fit error for a second polygon diameter (2 meter versus 1 meter in the left panel and 3 meter versus 2 meter in the right panel). A point on the diagonal line indicates that the fit error is the same for both polygon sizes. A point below the diagonal line indicates that the fit error is smaller for the polygon diameter labeling the y-axis, while a point above the diagonal line indicates that the fit error is smaller for the polygon diameter labeling the x-axis. All but a few of the points plotted in these two graphs are above the diagonal line, indicating that in all but a few cases, reducing the polygon size resulted in a smaller fit error. Thus, smaller polygons appear to be favored for model inversion.

The effects of measurement density on fit error are shown in Figure 16 for each of the polygon diameters under consideration. For the 169 objects used in current study, the average number of measurements within the 1-meter polygon is 21.5 measurements, with standard deviation of 7.2. Using the 2-meter polygon, the average number of measurements for each object is 83.7 with standard deviation 26.0. The 3-meter polygon centered at each object location encloses an average of 187.1 measurements with standard deviation 56.1. Across all three polygon sizes the measurement density is quite consistent with an average over all 169 objects of 26.8 measurements per square meter. The standard deviation across objects with the same size polygon is likely driven by additional passes from different directions over some objects. The data for each target were inverted using all of the available measurements (100%) and a random draw of half the available measurements

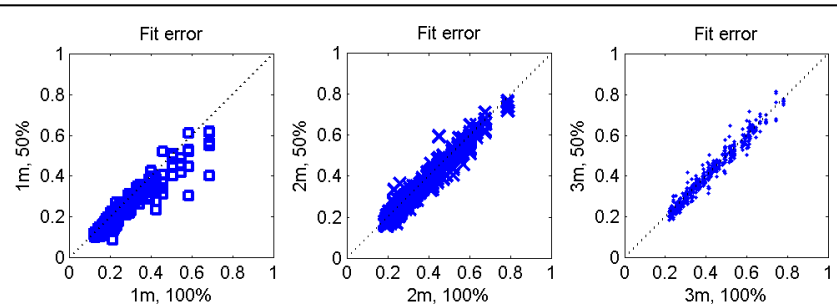

Figure 16. Scatter plot of fit error for two measurement densities: all available measurements (100%) on the x-axis and random draws of half of the measurements (50%) on the y-axis. Points along the diagonal line have the same fit error using either measurement density.

(50%), and the resulting model fits from these two inversions are plotted along the x-axis and y-axis, respectively. A point on the diagonal line indicates that both sets of measurements (100% and random 50%) resulted in the same model fit error. A point below the diagonal line indicates that the reduced (50%) measurement set provided a smaller fit error, and a point above the diagonal line indicates that the full (100%) data set provided a smaller fit error. For all three polygon diameters, the data are clustered tightly about the diagonal line, suggesting that data density (number of measurements) does not play a significant role in model fit error. In addition, these results taken together with the previous results (model fit error versus polygon size in Figure 15) suggest that polygon size plays a larger role than data density (number of measurements) in reducing the model fit errors plotted in Figure 14.

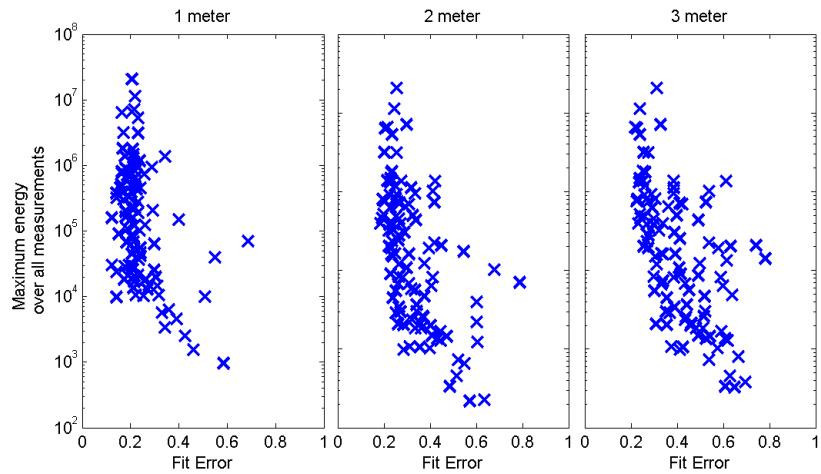

Figure 17. Scatter plot of fit error versus measurement energy for the 106 UXO targets. Subplots are for the 1m diameter polygon, 2m diameter polygon, and 3m diameter polygon.

Model fit error versus measurement energy is shown in Figure 17 for each of the three polygon diameters. For all three polygon diameters, there is a general trend toward an inverse relationship between the measurement error and the model fit.

In summary, analysis of the relationship between the properties of the data and the model fit error indicated that the number of measurements within a given spatial area (i.e. density) had little effect on model fit error; however, measurement energy was inversely related to model fit. This latter result is consistent with studies that have investigated the stability of parameter estimates under different measurement signal-to-noise ratios. Results of UXO discrimination testing performed on the eighteen data sets displayed a wide range of performance, indicating that discrimination performance can be affected significantly with the same targets, model, and discrimination strategy when using different data collection parameters. Additionally, the targets that were not correctly identified were inconsistent across different data sets. Thus, there may be an optimal data set for each individual anomaly that will provide improved discrimination performance.

The next stage of this study was to investigate how discrimination performance may be improved through the use of a strategy for determining which measurements are used in the model inversion for each anomaly. Two data selection strategies were considered. The first method is based on a previous application of the Theory of Optimal Experiments to subsurface sensing; however, for this work the technique was applied to post-processing of the data set rather than determining the optimal measurements during data collection. The second method attempted to minimize the effect of outlier measurements during the model inversion by removing them from the set of data. Both strategies require an initial inversion of the model using all available data. On average, the first method discarded 67% of the measurements for each anomaly whereas the second method discarded 28% of the measurements for each anomaly.

The first technique for selecting a subset of data points for backfitting the dipole model, termed MaxFI, utilized the gain in Fisher information [**1**]. The Fisher information gain (FIG) has been applied previously to subsurface sensing with EMI sensors. Liao and Carin [**2**] used an equivalent dipole model to perform adaptive EMI sensing, sequentially determining the optimal measurement locations using estimates of the model parameters. In this study, the same technique will be considered as a post-processing technique, as opposed to pre-processing, to determine which subset of measurements offer the largest gain in Fisher information based on the model parameter estimates calculated using all available data. The Fisher information gain is a metric calculated from the Fisher information matrix, which is inversely related to the variance of the parameter estimates. Thus, maximizing the Fisher information is equivalent to minimizing the variance of the model parameters. A rigorous development of the gain in Fisher information for the dipole model can be found in Liao and Carin [**2**]. Assume the EMI sensor measurements are Y = $f\left(p;\beta_{\mathrm{INT}},\beta_{\mathrm{EXT}}\right)$ + *Z*, where $f\left(p;\beta_{\mathrm{INT}},\beta_{\mathrm{EXT}}\right)$ is the dipole model response and *Z* is white Gaussian noise. The dipole model response is a function of a set of measurement parameters *p* = {ΔNorthing, ΔEasting, ω}, the set of intrinsic model parameters β_{INT} = {m_{p0}, m_{p1}, ω_{p1}, m_{z0}, m_{z1}, ω_{z1}}, the set of extrinsic model parameters β_{EXT} = {θ, Φ, depth}. The Fisher information matrix J is comprised of a sum of components J_n calculated for each of the N measurements

$$J=\sum_{n=1}^{N} J_n \tag{1}$$

where

$$J_n=\mathrm{Re}\left\{\left[\nabla_{\beta_{\mathrm{INT}}} f\left(p_n;\beta_{\mathrm{INT}},\beta_{\mathrm{EXT}}\right)\right]\left[\nabla_{\beta_{\mathrm{INT}}} f\left(p_n;\beta_{\mathrm{INT}},\beta_{\mathrm{EXT}}\right)\right]^H\right\} \tag{2}$$

with J_n evaluated at the nth measurement, and the superscript H denotes conjugate transpose.. The gain in Fisher information for the ith measurement, with parameters p_i, can be calculated via Eq. (3)

$$\mathrm{FIG}\left(i\right)=\log\left|I+F^H\left(\sum_{n=1,n\neq i}^{N}\frac{1}{\sigma_n^2}J_n\right)F\right| \tag{3}$$

where F is a six by two matrix

$$F=\left[\mathrm{Re}\left\{\nabla_{\beta_{\mathrm{INT}}} f\left(p_i;\beta_{\mathrm{INT}},\beta_{\mathrm{EXT}}\right)\right\},\mathrm{Im}\left\{\nabla_{\beta_{\mathrm{INT}}} f\left(p_i;\beta_{\mathrm{INT}},\beta_{\mathrm{EXT}}\right)\right\}\right]. \tag{4}$$

There are several differences between the implementation used here and that used by Liao and Carin. In Liao and Carin, they did not distinguish between intrinsic and extrinsic parameters; the Fisher information matrix J was an 11 by 11 matrix (over all 11 parameters of the dipole model). In this study, the information gain focused on the six intrinsic parameters used for generating features for UXO discrimination. Also, the goal in Liao and Carin was to develop an iterative sequential method for collecting sensor measurements by first estimating the model parameters from the available data, then determining the measurement parameters of the next data point to maximize the FIG, and re-estimating the parameters (through model inversion) using the added data. The current study performed the FIG calculations after collecting the data to determine a subset of the most informative data points.

The second technique, termed MinResidual, utilized the residuals between the measured data and the model fit estimated with all available data. The handling of outliers has been considered in the statistics community for model estimation [**3**]. Several methods for identifying influential outliers have been proposed for linear models, but for nonlinear models such as the dipole model, measuring outlier

influence is not as straightforward. Therefore, a first attempt at a data selection strategy using the residuals simply removes data points with high residual. The justification for a strategy that removes data points with large fit error is as follows. The numerical optimization to find the model parameters during backfitting is driven by the residuals. If a measurement has a large error after backfitting it is an indicator that the dipole model is not capable of describing that point. However, if the point has high leverage it could corrupt the numerical optimization by driving the optimization towards a point it cannot reach. Removing high-residual measurements will also reduce the fit error (albeit via manipulation), which may improve performance. The residual at each measurement was calculated using Eq. (5)

$$\varepsilon(n) = \frac{\sum_{\omega}\left(\hat{Y}(n,\omega) - Y(n,\omega)\right)\left(\hat{Y}(n,\omega) - Y(n,\omega)\right)^{*}}{\sum_{\omega} Y(n,\omega) Y(n,\omega)^{*}}. \tag{5}$$

The data selection strategies were performed on the results for the two-meter diameter polygons, which averaged 83.7 measurements per anomaly. The data subsets were selected by setting a threshold equal to the average of the Fisher information gain or residual, respectively, for each data selection method. By retaining only measurements with an above-average FIG, the number of measurements for each object ranged from 9% to 46% of the original measurements (33% average, or 27.6 measurements on average). Retaining only the measurements with below-average residual results in data densities of 40% to 94%, with an average retention of 72% of the available data. There are numerous other methods for using the FIG or residuals for data selection. Rather than setting a fixed threshold for removing data, measurements could be weighted based on either FIG or residual. The FIG and residual could also be combined in a data selection strategy; however, preliminary investigations of such techniques suggest that a conceptual relation between leverage in statistical model fitting and the Fisher information should be established to assist in the determination of what points may be beneficial.

The inversion results using the selected data sets had smaller fit errors for most UXO targets, as shown in Figure 18. This figure shows the change in fit error from performing the inversion using the full data set to performing the inversion using the a reduced data set found using the two data selection approaches previously described. Each data point is the change in fit error for a single target. The change in fit error using the Fischer Information Gain is plotted on the x-axis and the change in fit error using the MinResidual is plotted on the y-axis. In both cases, a positive change indicates a reduction in the fit error. For all but a few targets, the fit error is reduced using either measurement selection approach, as demonstrated by all but a few data points being

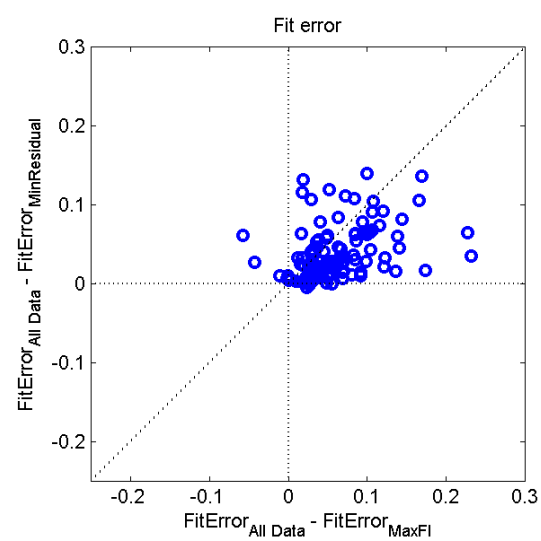

Figure 18. Scatter plot of the change in fit error using each of the two data selection strategies. Points along the diagonally line have an equal change in fit error using either data selection strategy. Positive values indicate a reduction in fit error.

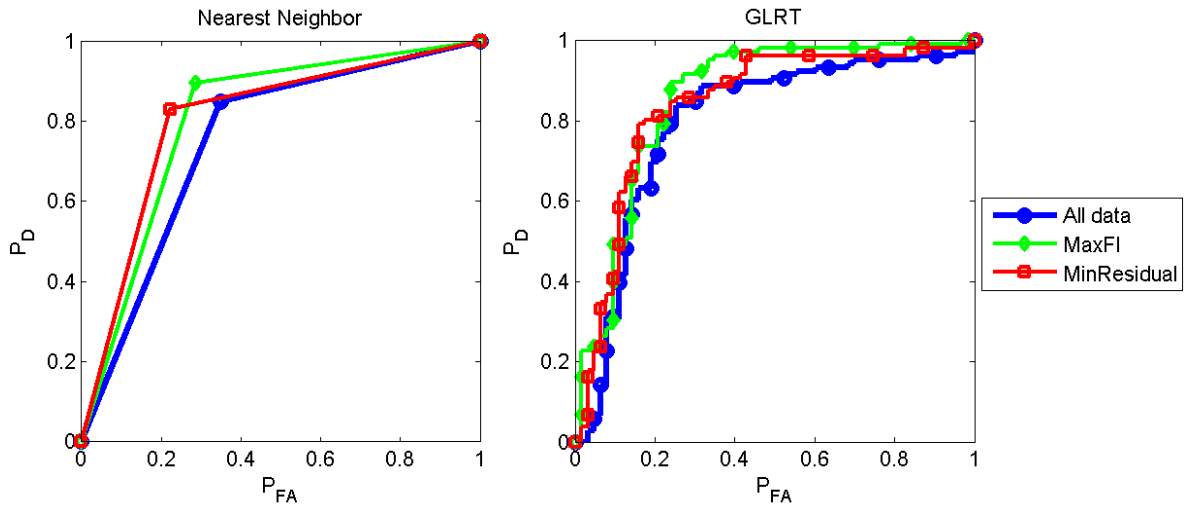

Figure 19. Unexploded ordnance discrimination performance using the data selection strategies. Results are shown for two classification algorithms, nearest neighbor (left) and generalized likelihood ratio test (right), and plotted as the probability of false alarm (P_{FA}) versus probability of detection (P_D).

located in the upper right quadrant of the graph.

ROC curves for UXO discrimination performance are shown in Figure 19 for two classifiers: a nearest neighbor classifier (left) and a generalized likelihood ratio test (right). For both classifiers, the MinResidual and MaxFI measurement selection approaches result in improved discrimination performance, with the MaxFI technique generally providing slightly better performance than the MinResidual technique. Overall, classification results were modestly improved, as shown in Figure 19, with the probability of classification error using the nearest neighbor classifier (left panel) decreasing from 35.8% for the baseline classifier using all the data to 27.4% for the classifier using maximum Fisher Information.

Data/Measurement Selection Summary

Analysis of the relationships between properties of the data, such as polygon size, measurement density and measurement energy, and the fit error indicate that selecting data with the goal of improving model inversions may provide better target parameter estimates, which should, in turn, improve detection performance. The outcomes of this study may lead to guidelines for determining a set of sufficient measurements given the data-dependency in model fitting and UXO discrimination performance. These outcomes also suggest a potential benefit from data selection strategies, based on the evidence supporting the existence of an optimal subset of the measurements, encouraging further investigation.

We have also extended the study of the effect of polygon size to include the magnetometer sensor. In this current study, the magnetometer sensor and model are being considered to illustrate the benefits of optimal selection of the polygon, with specific emphasis on the accuracy of the anomaly depth estimates. The magnetometer model was inverted for anomalies in the Camp Sibert data collection using polygon sizes ranging in diameter from 0.5 to 5 meters. The initializations of the model parameters were consistent across polygon size; the only variable was the polygon diameter. Results indicate substantial variability in the estimated anomaly depth as a function of polygon diameter. Depth estimates for clutter objects were more sensitive to polygon size than depth estimates for UXO. For a majority of the UXO, the depth estimates versus polygon size behave consistently; the few exceptions

require further examination. This result does indicate the potential benefit of active polygon selection in magnetometer data for target depth estimation. Future work will consider investigating model-based and empirical methods for polygon selection in both the magnetometer and EMI sensors.

Sensor Management

Automated sensor management is the process of autonomously tasking a suite of sensors in order to best fulfill a set of mission objectives in the presence of situationally varying resource constraints. The use of a sensor manager in a specific setting provides benefits such as assisting a human operator who would otherwise be overwhelmed by the number of tasking decisions that need to be made, removing a human operator from a dangerous application setting and thereby ensuring the safety of the operator, or facilitating the autonomous operation of a system or network of system. Sensor management has been studied extensively in recent years for a variety of applications ranging from robotic obstacle avoidance [**4**] to wireless network battery life management [**5**][**6**] to military command, control, communications, and intelligence (C3I) applications [**7**][**8**][**9**][**10**][**11**][**12**][**13**][**14**][**15**][**16**][**17**][**18**].

Previous work by the authors has developed a framework for sensor management for the static target detection problem in which *M* heterogeneous sensors search for *N* targets within a grid of *C* cells [**19**][**20**][**21**][**22**]. This framework is based on the work of Kastella [**23**] and functions by tasking the sensors to make a new observation that maximizes the expected information gain that is obtained, using the Kullback-Leibler divergence as a measure of information. This information-theoretic sensor management approach has been consistently demonstrated to outperform a direct search procedure in which the sensors sweep through the grid in a predefined search pattern. A variety of observation models have been considered in [**19**][**20**][**21**][**22**]; binary observations are considered in [**19**][**22**], non-binary observations are considered in [**20**], and correlated observations are considered in [**21**]. In addition, uncertainty modeling has been addressed in [**19**][**20**] in order to consider the effects that unknown sensor performance characteristics will have on the overall performance of the sensor manager. Correct modeling of the uncertainty present in the problem was demonstrated to be an important component of a robust sensor manager.

In many applications, different sensors are highly complementary in performing a desired task. For example, in the landmine detection application, electromagnetic induction (EMI) sensors and ground penetrating radar (GPR) sensors have been shown to be useful for detecting different types of mines. Fusion of these two sensing modalities results in increased performance over that obtained with only a single modality [**24**]. For this reason, modeling of the relationship between observations made by different sensors and how those observations impact classification is a crucial step. Earlier work in [**21**] incorporated the modeling of correlated sensor observations, but this work had several limitations. Firstly, the observations modeled were restricted to follow a multivariate normal density. In addition, it was possible to define the correlation parameters presented in [**21**] in such a way as to obtain an invalid covariance matrix. To address these limitations, a new framework for observation modeling is presented in this study that imposes no restrictions on the densities of the observed data and that furthermore models observations in the feature space corresponding to features generated from the observed data in a cell rather than a space with dimensionality equal to the number of observations that have been made in that cell. This modeling approach is motivated by the physics of static target detection problems such as landmine detection and UXO discrimination. In such applications, pattern classification is typically performed by generating a single feature (or group of features) for each sensor that combines the information obtained from all observations made with that sensor. Thus, a model is required for the features under both the “target present” and “no target present” hypotheses. A second model should then govern how the sensor features are determined from the individual sensor

observations and how new observations should update the densities for the feature values for that sensor. As in [**19**][**20**][**21**][**22**], the sensor manager will continue to maximize the expected information gain obtained with each new sensor measurement, and the Kullback-Leibler divergence will continue to be used as the measure of information.

Sensor Management with Simple Feature Model

The sensor management framework considered in this work uses M sensors to search for N targets in a grid of C cells. Each cell has a binary state, either "target" or "no target," corresponding to whether that cell contains a target; the cell state for cell c is denoted $S_c = 1$ or $S_c = 0$ to indicate "target" and "no target," respectively. Without loss of generality, it will be assumed that each of the M sensors produces a single feature in each cell. To create a model where a sensor produces multiple features, the dimensionality of the following feature model may be straightforwardly increased from M to the total number of features from all sensors. Let F_M represent all of the features and let f_m represent the feature that corresponds to sensor m. Using training data, a model is learned for the joint density of all features conditioned on each of the two possible states: $p(F_M \mid S = 0)$ and $p(F_M \mid S = 1)$. Note that here the notation $S = 1$ and $S = 0$ is used to denote the "target" and "no target" hypotheses in general and not in reference to a specific cell c.

Now let $F_{c,M}$ represent all features in cell c and let $f_{c,m}$ represent the feature that corresponds to sensor m in cell c. In this section, it will be assumed that the sensors directly observe the feature values $f_{c,m}$. The following section will introduce modeling for the more realistic case when sensor observations of the feature values are corrupted by noise. Let $F_{c,k}$ represent a vector of observed features, where k is a vector of indices indicating which of the M features have been observed; similarly, define $F_{c,k*}$ to be the vector of unobserved features, with $k*$ denoting the features that have not yet been observed. State probabilities may then be computed for each cell under the "target" and "no target" hypotheses conditioned on the features that have been observed in that cell using the traditional Bayesian framework:

$$\Pr\left(S_c = s \middle| F_{c,k}\right) = \frac{p\left(F_{c,k} \middle| S_c = s\right)\Pr\left(S_c = s\right)}{\sum_{t=0}^{1} p\left(F_{c,k} \middle| S_c = t\right)\Pr\left(S_c = t\right)}. \tag{1}$$

The densities for the observed features are simply the marginal densities for those features derived from the full joint density model for the features in cell c:

$$p\left(F_{c,k} \middle| S_c = s\right) = \int p\left(F_{c,M} \middle| S_c = s\right) dF_{c,k*}. \tag{2}$$

Note that now models are required for the joint densities of the features in cell c, $p(F_{c,M} \mid S_c = 0)$ and $p(F_{c,M} \mid S_c = 1)$. One reasonable choice for these densities is to use the full joint density models for the features, $p(F_M \mid S = 0)$ and $p(F_M \mid S = 1)$, that were produced from the training data. Other choices for the prior density are of course also possible. This work, however, will assume the use of $p(F_M \mid S = 0)$ and $p(F_M \mid S = 1)$ as the cell-specific joint densities.

As mentioned in Section 1, the sensor manager uses the Kullback-Leibler divergence as a measure of information. The Kullback-Leibler divergence is defined for discrete densities p and q as

$$D_{KL}\left(p(\cdot) \| q(\cdot)\right) = \sum_x p(x) \log\left(\frac{p(x)}{q(x)}\right). \tag{3}$$

The sensor manager maximizes the expected information gain that will be obtained with a new sensor observation. The information gain for a new sensor observation is defined to be the Kullback-Leibler divergence between the state probabilities in the cell after the new observation and the state probabilities in the cell before the new observation. Letting $f_{c,k+1}$ represent the new observed feature and $F_{c,k+1}$ represent the observed feature vector including the new feature, the expected information gain may be computed analytically as follows:

$$\Delta D = E\left\{D_{KL}\left(S_c\left(\cdot\middle|F_{c,k+1}\right)\middle\|S_c\left(\cdot\middle|F_{c,k}\right)\right)\right\} = \int D_{KL}\left(S_c\left(\cdot\middle|F_{c,k+1}\right)\middle\|S_c\left(\cdot\middle|F_{c,k}\right)\right) p\left(f_{c,k+1}\middle|F_{c,k}\right) df_{c,k+1}\,. \tag{4}$$

The predictive density for the new observed feature, $p(f_{c,k+1} \mid F_{c,k})$, may be written as

$$p\left(f_{c,k+1}\middle|F_{c,k}\right) = \sum_{s=0}^{1} p\left(f_{c,k+1}\middle|S_c = s, F_{c,k}\right)\Pr\left(S_c = s\middle|F_{c,k}\right). \tag{5}$$

Finally, the conditional densities in (5) may be computed from the full joint densities for the features as

$$p\left(f_{c,k+1}\middle|S_c = s, F_{c,k}\right) = \frac{\int p\left(F_{c,M}\middle|S_c = s\right) dF_{c,(k+1)^*}}{p\left(F_{c,k}\middle|S_c = s\right)}\,. \tag{6}$$

The expected information gain may be computed for each sensor that has not yet made an observation using (4). The sensor manager may then task the sensor suite to make an observation using the sensor that will maximize the expected information gain. Depending on the results of previous observations in a cell, it is possible that the cell state probabilities will not change substantially if additional information is incorporated. In such a case, making additional observations using new sensors will result in wasted time and resources. Consequently, a threshold, λ, may be determined and the sensor manager should be tasked to make a new observation only if $\Delta D > \lambda$.

Equations (2) and (6) specify how marginal and conditional densities may be determined from the full joint densities $p(F_{c,M} \mid S_c = 0)$ and $p(F_{c,M} \mid S_c = 1)$. Specific functional forms for the full joint densities allow the analytical derivation of the relevant marginal and conditional densities. For example, if a Gaussian mixture model (GMM) is used to model the full joint densities, the marginal densities and conditional densities of all possible feature combinations may be shown to be GMMs as well. The use of such functional forms avoids the need for numerical determination of the marginal and conditional densities.

Sensor Management with Full Observation Model

Consider each feature as having a true value for each of the cells under consideration. The previous section assumes that features are directly observed by the sensors without the presence of corrupting noise, meaning that the true feature value is observed in a single sensor observation. In realistic environments, however, noise will be present in the observed data, and as a consequence, noisy observations of this true feature value will be obtained. Such a perspective is readily motivated by considering a feature such as the maximum energy response of a buried object. Different maximum energy values will be observed with different passes over the object depending on the thermal noise of the sensor, the exact soil properties and sensor location, and other confounding factors. Even for features that require additional processing to generate, such as dipole model features or matched subspace detector features, different observations of the object will result in different feature outputs that may be considered to be noisy observations of a true feature value. In this work, the noise that corrupts sensor observations will be assumed to be additive white Gaussian noise.

Let $x_{c,k,m}$ represent the kth observation made in cell c with sensor m, let $X_{c,k,m}$ represent observations $x_{c,1,m}, \ldots, x_{c,k,m}$, and let $X_{c,M}$ represent all data collected in cell c. Furthermore, assume that the variance of the additive noise, σ^2, is known. The observation model for a new sensor observation is therefore

$$x_{c,i,m} \sim N\left(f_{c,m}, \sigma^2\right), \tag{7}$$

and the posterior density for the features given the sensor observations may be written using Bayes's rule as

$$p\left(F_{c,M} \middle| S_c = s, X_{c,M}\right) = \frac{p\left(X_{c,M} \middle| S_c = s, F_{c,M}\right) p\left(F_{c,M} \middle| S_c = s\right)}{\int p\left(X_{c,M} \middle| S_c = s, F_{c,M}\right) p\left(F_{c,M} \middle| S_c = s\right) dF_{c,M}} \tag{8}$$

The posterior state probabilities for a cell now depend on the observations $X_{c,M}$ that have been collected in the cell:

$$p\left(S_c = s \middle| X_{c,M}\right) = \frac{p\left(X_{c,M} \middle| S_c = s\right) \Pr\left(S_c = s\right)}{p\left(X_{c,M}\right)}, \tag{9}$$

which may be rewritten as

$$p\left(S_c = s \middle| X_{c,M}\right) = \frac{\int p\left(X_{c,M} \middle| S_c = s, F_{c,M}\right) \Pr\left(S_c = s \middle| F_{c,M}\right) p\left(F_{c,M}\right) dF_{c,M}}{p\left(X_{c,M}\right)}. \tag{10}$$

Given the feature values, the sensor observations are independent of the cell state, meaning that $p(X_{c,M} \mid S_c = s, F_{c,M}) = p(X_{c,M} \mid F_{c,M})$. Given the feature values, this independence holds because the feature values are the only variables driving the observed data, as shown in (7). The observed data is assumed to be normally distributed with means equal to the feature values and known variances. Thus, given the feature value, the sensor observation does not depend on the underlying cell state. Finally, the numerator may be rewritten using Bayes's rule to obtain a final form for the state probability:

$$p\left(S_c = s \middle| X_{c,M}\right) = \frac{\int p\left(X_{c,M} \middle| F_{c,M}\right) p\left(F_{c,M} \middle| S_c = s\right) \Pr\left(S_c = s\right) dF_{c,M}}{p\left(X_{c,M}\right)}. \tag{11}$$

The expected information gain obtained with a new sensor observation $x_{c,k+1,m}$ is determined similarly to the expected information gain in Section 2:

$$\Delta D = E\left\{D_{KL}\left(S_c\left(\cdot \middle| X_{c,M}^{new}\right) \middle\| S_c\left(\cdot \middle| X_{c,M}^{old}\right)\right)\right\} = \int D_{KL}\left(S_c\left(\cdot \middle| X_{c,M}^{new}\right) \middle\| S_c\left(\cdot \middle| X_{c,M}^{old}\right)\right) p\left(x_{c,k+1,m} \middle| X_{c,M}^{old}\right) dx_{c,k+1,m}, \tag{12}$$

where $X_{c,M}^{new}$ represents the sensor data after the new observation $x_{c,k+1,m}$ is collected and $X_{c,M}^{old}$ represents the sensor data before the new observation is collected. The predictive density, $p(x_{c,k+1,m} \mid X_{c,M})$, for a new observation given the data already collected by that sensor may be expanded to give

$$\begin{aligned} p\left(x_{c,k+1,m} \middle| X_{c,M}^{old}\right) &= \int p\left(x_{c,k+1,m} \middle| F_{c,M}, X_{c,M}^{old}\right) p\left(F_{c,M} \middle| X_{c,M}^{old}\right) dF_{c,M} \\ &= \int p\left(x_{c,k+1,m} \middle| F_{c,M}\right) \left(\sum_{s=0}^{1} p\left(F_{c,M} \middle| S_c = s, X_{c,M}^{old}\right) \Pr\left(S_c = s \middle| X_{c,M}^{old}\right)\right) dF_{c,M} \end{aligned}, \tag{13}$$

which may be evaluated using (7), (8), and (11). For specific functional forms of the densities in (13), the integral may be solved analytically. For example, using the observation model of (7) with posterior densities for the features in a cell that are GMMs results in a predictive density that is itself a normal density. As in Section 2, the sensor manager operates by greedily maximizing the expected information

gain that may be obtained with each new sensor observation. The expected information gain in (12) is computed for each sensor in each cell, and the sensor manager directs the sensor suite to make the observation that has the largest expected information gain. A threshold, λ, for the information gain may be determined so that a new observation is only made if $\Delta D > \lambda$.

Camp Sibert Discrimination Study Results

Having presented the new framework for observation modeling, an illustrative example is now presented that examines the performance of the sensor manager on real data from the UXO discrimination application. The dataset considered only contains a single observation of the feature for each object with each of the sensors, so the model from Section 2 is used in the following analysis. The dataset, collected at Camp Sibert, contains 59 targets and 183 clutter objects, each of which is assumed to occupy a cell, giving a total number of 242 cells in the grid under consideration. The proportions of targets and clutter are used to set the prior probabilities of the "target" and "no target" states. Observations are available using three different sensing modalities: a time-domain EMI sensor (EM61), a frequency-domain EMI sensor (GEM), and a magnetometer. Energy, defined as the integral of the squared sensor response, is used as the feature for the two EMI sensors, and estimated depth is used as the feature for the magnetometer. A scatter plot of the data in three dimensions is shown in Figure 20, along with each of the three different two-dimensional representations of the data. As may be seen from the figure, the data appears to be relatively separable, which would indicate the possibility that the use of a sensor manager could allow UXO to be correctly classified using fewer sensor observations than

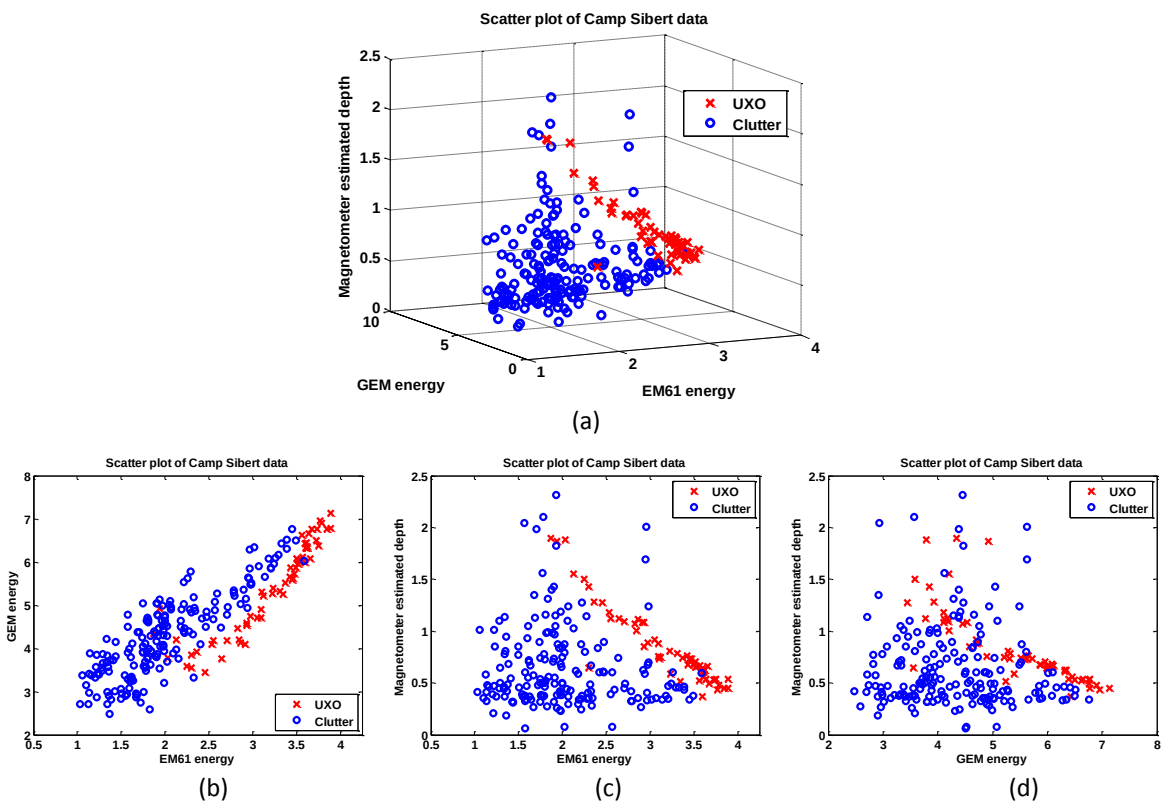

Figure 20. Scatter plots of UXO data that has been collected by an EM61 sensor, a GEM sensor, and a magnetometer. The full, three-dimensional dataset is plotted in (a), while (b), (c), and (d) show the three different two-dimensional scatter plots of the data.

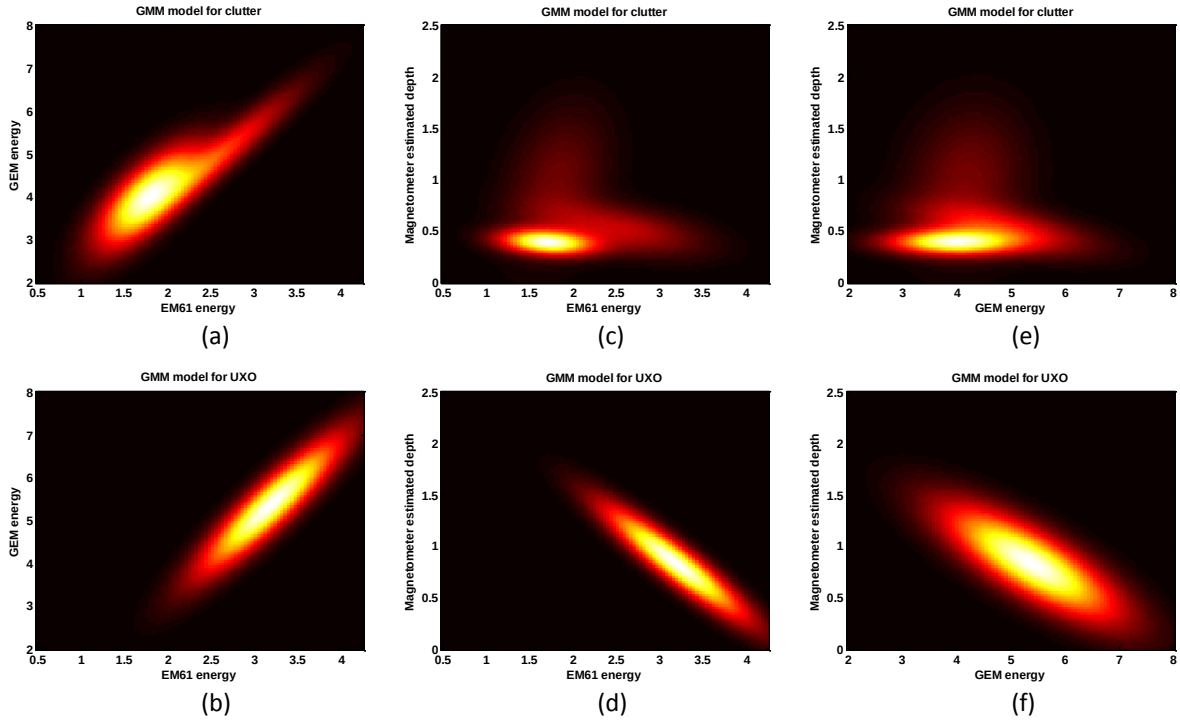

Figure 21. Marginal densities from GMM model fits to the Camp Sibert UXO data. (a)-(f) show the two-dimensional marginal distributions for each of the different feature pairings for both the UXO and clutter models. (a)-(f) may be compared to (b)-(d) from Figure 20 to visually assess the quality of fit that the models provide.

an unmanaged, direct search strategy, which would make observations using all three sensing modalities for each object.

A model is trained for both the UXO and the clutter data using a GMM with one mixture component for the UXO and with three mixture components for the clutter. The models that are fit to the data may be seen in Figure 21 as a series of two-dimensional marginal densities. Figure 21 may be compared to two-dimensional plots in Figure 20 to visually assess how well the models fit the data, and they do in fact provide reasonable fits.

Now that feature models $p(F_M \mid S = 0)$ and $p(F_M \mid S = 1)$ have been determined, the expected information gain for an observation with each of the three sensors may be computed. The marginal densities for individual features with both the UXO model and the clutter model are presented in Figure 22. Examining the marginal densities shows that the best separation between the UXO and clutter densities appears to occur with the EM61 energy feature, and so it would be expected that the EM61 sensor would have the highest expected information gain. Computing the expected information gains using (4), it is found that $\Delta D = 0.1223$ for the EM61, $\Delta D = 0.0472$ for the GEM, and $\Delta D = 0.0365$ for the magnetometer. Therefore, the first observation of an object should always be made with the EM61 sensor, since that sensor provides the largest expected information gain.

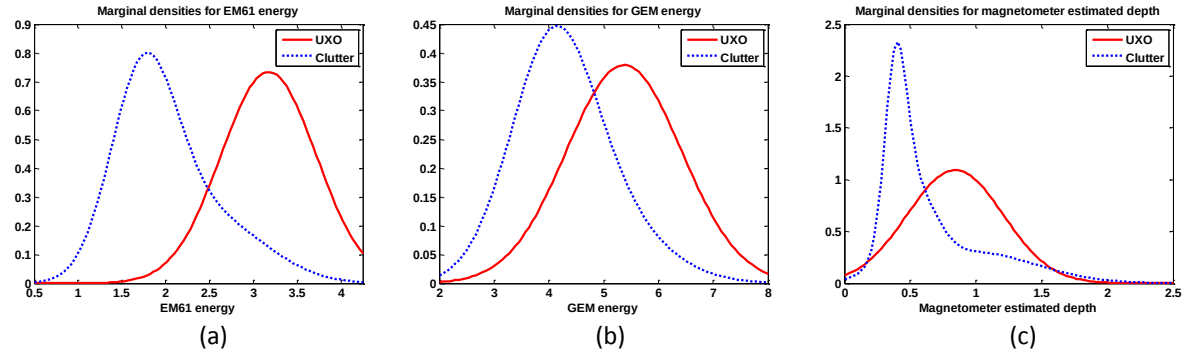

Figure 22. Marginal densities from GMM model fits to the Camp Sibert UXO data for individual features for both the UXO and clutter models. (a) shows the densities for the EM61 energy feature, (b) shows the densities for the GEM energy feature, and (c) shows the densities for the magnetometer estimated depth feature.

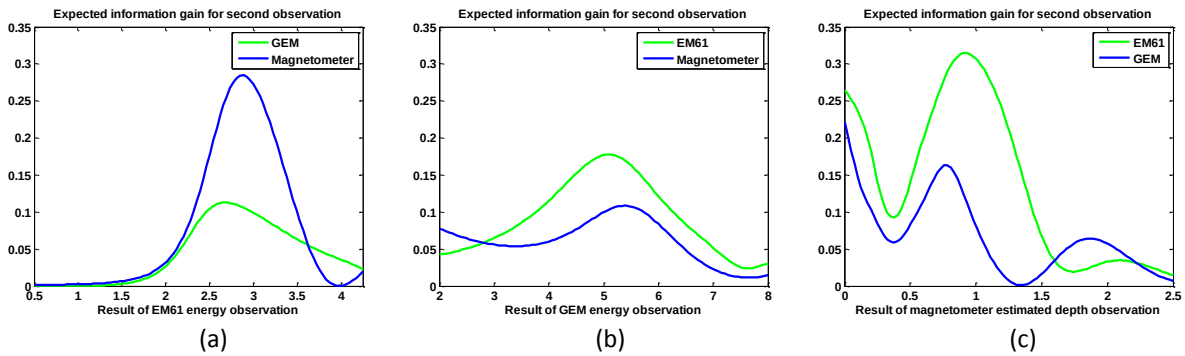

Figure 23. Expected information gain for an observation with one of the two remaining sensing modalities after a first observation has been made with (a) the EM61, (b) the GEM, (c) and the magnetometer.

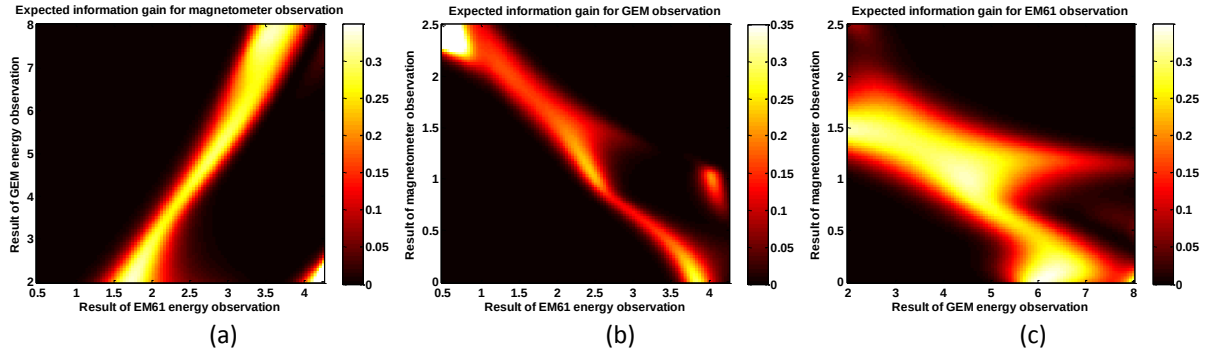

Figure 24. Expected information gain for an observation with the third sensing modality after observations have been made with the other two modalities. The third observation is being made with (a) the magnetometer, (b) the GEM, and (c) the EM61.

After a single observation has been made, the expected information gain that may be obtained from a subsequent observation with a different sensor depends on the result of the first sensor observation. Figure 23 shows the expected information gain from making a second observation with a different sensing modality than the first sensor used. Even though, based on the results computed above, the EM61 sensor is always used to make the first observation, results are still presented for the three different cases when each of the sensors has made the first observation. It may be seen that if either

the GEM or the magnetometer were used for the first sensor observation, the EM61 is nearly always the most informative sensor to use for the second observation. When the EM61 is used as the first sensor, the magnetometer provides a more informative observation for most cases except for high values of the EM61 energy feature, in which case the GEM should be used.

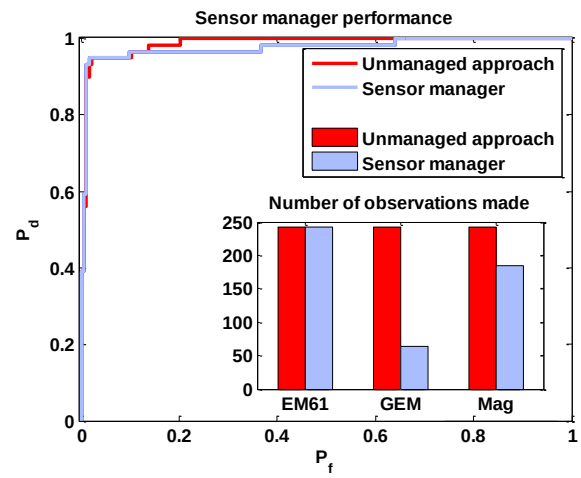

Figure 25. Performance of unmanaged search and the sensor manager with threshold λ = 0.01 on UXO data. The bar plot shows the number of observations made with each modality.

Similarly, after a second sensor observation has been made, the expected information gain for using the third and final sensing modality may be computed as a function of the values of the first two sensor observations. Figure 24 plots the expected information gain for the third sensor observation, again considering each of the three different combinations of the sensor pairs that have made the first two observations. The results in Figure 24 may be compared with subplots (b)-(d) in Figure 20, and it may be observed that the expected information gain for a third sensor observation is highest in the regions where there is difficulty separating UXO from clutter in the two-dimensional scatter plot of the data. The expected information gain may also be high in regions where there is little or no training data with which to train the models, but as long as the training data is representative of the objects that will be encountered, it is unlikely that sensor observations will lie in these regions. Considering the results seen in Figure 22 through Figure 24, the expected information gain behaviors provide highly intuitive results and also demonstrate the mechanisms through which the intelligent tasking of the sensor suite is performed by the sensor manager.

The performance of the sensor manager is now analyzed and compared to the performance of an unmanaged approach in which all three sensors are used to observe each cell. Performance is presented as a receiver operating characteristic (ROC) generated by using the posterior state probability of a target in each cell as a decision statistic. Results are presented in Figure 25, and it may be seen that in terms of probability of correct decision, the sensor manager performs nearly as well as the unmanaged approach while using substantially fewer sensor observations to achieve that performance. In fact, the performance of the sensor manager and the unmanaged approach are nearly identical up to $P_d \approx 0.95$. The sensor manager makes only 51.2% of the GEM and magnetometer observations that are made by the unmanaged approach and 67.4% of the total

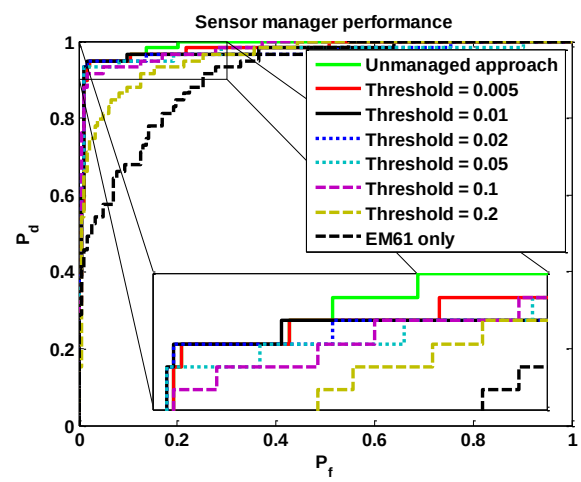

Figure 26. Performance on UXO data of unmanaged search, the sensor manager with a variety of information gain thresholds, and baseline performance using only the EM61 sensor.

observations. In terms of number of unnecessary digs saved (P_{FA} @ P_D=1), however, the sensor manager performance lags the unmanaged approach considerably.

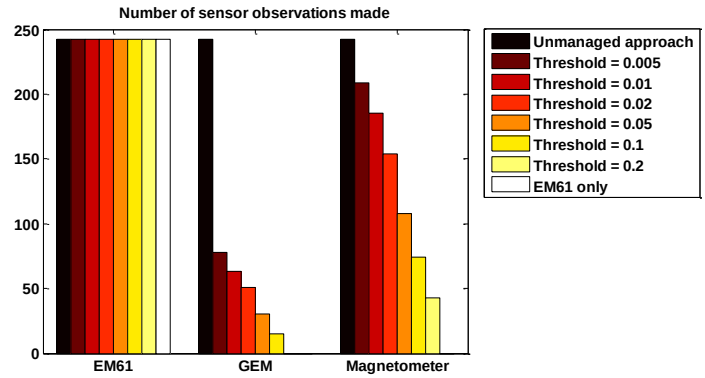

Figure 27: Bar plot of the number of observations made by each sensing modality as a function of the sensing approach used or of the information gain threshold used (for approaches using the sensor manager).

Figure 26 and Figure 27 examine the performance of the sensor manager as the information gain threshold, λ, is varied. A performance baseline using only the EM61 sensor is also included. As the threshold gets lower, the ROC for the sensor manager generally improves, although there are only slight differences in the curves once the threshold is lower than 0.1. Figure 27, which plots the number of sensor observations made, shows how higher thresholds result in progressively fewer sensor observations being made for the GEM and magnetometer sensing modalities. Thus, there is a tradeoff between number of observations and performance in which a smaller information gain threshold results in generally improved performance but at a cost of requiring more observations than are made with a higher threshold. It should be noted, however, that even in the case where λ = 0.005, the sensor manager still makes only 59.3% of the GEM and magnetometer observations that are made by unmanaged, direct search and only 72.9% of the total observations, which is still a substantial savings in the number of observations made.

Now that the sensor manager has been demonstrated to outperform an unmanaged technique, one final test is also considered in which it is examined how effectively the sensor manager can reject a feature that is uninformative. For this test case, the EM61 and magnetometer features from the UXO data are retained, but the GEM feature is replaced with random noise, giving the data that is shown in Figure 28. The EM61 is again the most informative sensor to use for the first observation, and Figure 29

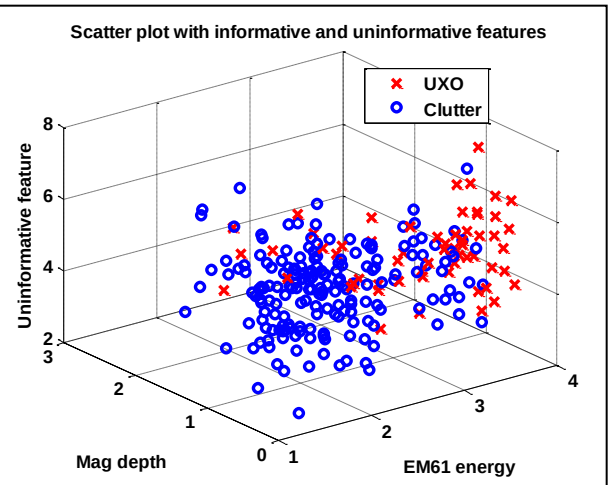

Figure 28. Scatter plot of data that includes random noise as an uninformative feature in addition to the EM61 and magnetometer features.

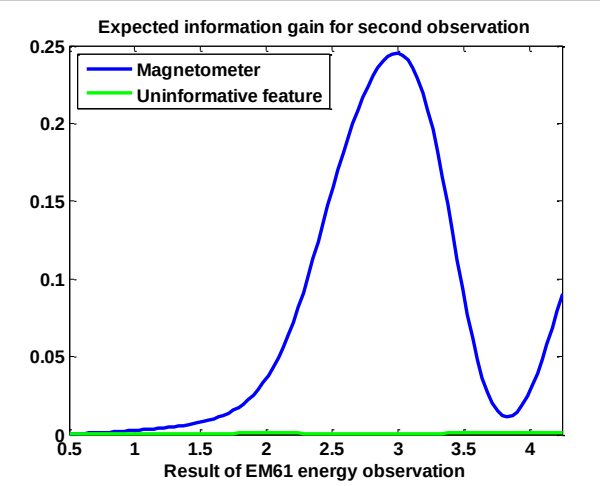

Figure 29. Expected information gain for an observation with one of the two remaining sensing modalities after a first observation has been made with the EM61 sensor.

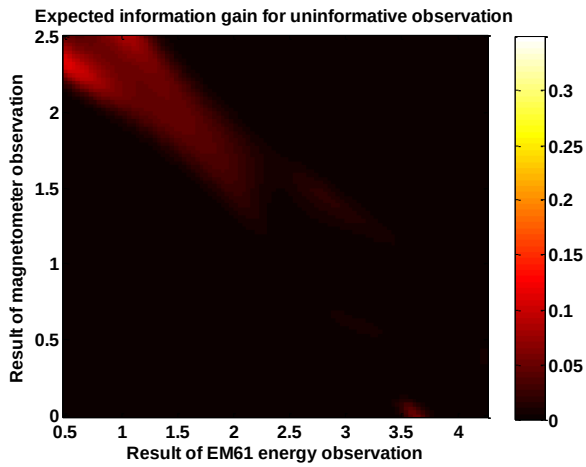

Figure 31. Expected information gain for an observation with the uninformative sensor after observations have been made with the EM61 and the magnetometer.

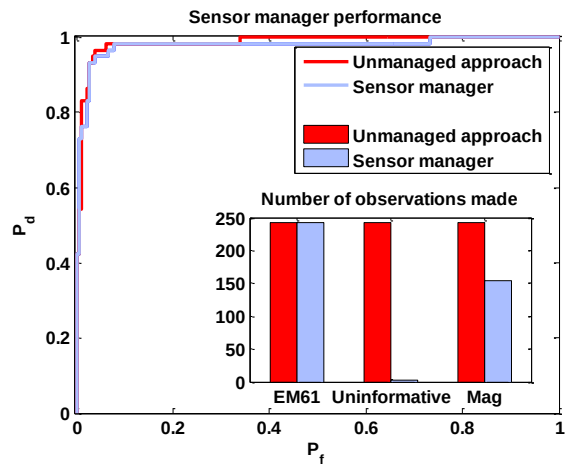

Figure 30. Performance of unmanaged search and the sensor manager with threshold λ = 0.02 on modified UXO data. The bar plot shows the number of observations made with each modality.

shows the expected information gain for using both the magnetometer and the uninformative sensor as a function of the observation result from the EM61. The magnetometer is vastly more informative than the uninformative sensor for all EM61 observation values, indicating that the sensor management framework is correctly able to identify the uninformative feature as containing an extremely low amount of information.

Similarly, Figure 31 shows the expected information gain for the uninformative sensor after observations have been made using both the EM61 and the magnetometer. The information gain is again very low except in the top left-hand corner of the plot. Referring back to Figure 20(c), it may be seen that this region of the EM61-magnetometer feature space has almost no training data, meaning that model behaviors in this region are not based on any actual data and may be unpredictable. Finally, Figure 30 shows the performance of the sensor manager compared to unmanaged search using this modified UXO data. The sensor manager performs almost exactly as well as the unmanaged approach in terms of probability of correct decision, while using only slightly over half as many magnetometer observations and virtually no observations from the uninformative sensor. Considering the number of unnecessary digs saved (P_{FA} @ P_D=1), however, again indicates that the performance of the sensor manager lags performance of the unmanaged approach considerably. The sensor management framework is indeed able to correctly determine that the uninformative sensor is unhelpful in discriminating between UXO and clutter, and that sensor is almost never used.

Sensor Management Summary

We have developed a framework for modeling observations within an information-theoretic sensor management framework. The new modeling framework naturally incorporates the correlations between sensor observations and also provides a capability to model uncertainty by modeling the feature values that are observed by the sensors as being corrupted by noise. The new sensor management framework has been tested using an example with real data from the UXO discrimination application. The sensor manager was found to outperform an unmanaged procedure in terms of probability of correct decision by achieving a very similar ROC using substantially fewer sensor observations. In terms of number of unnecessary digs saved (P_{FA} @ P_D=1), however, the sensor manager

performance lagged performance of the unmanaged approach. The sensor manager presented here was not developed with the goal of minimizing the number of unnecessary digs. It is possible that a sensor manager developed with the goal of minimizing P_{FA} at P_D=1 could perform as well as the unmanaged approach. The sensor manager was also found to be able to successfully ignore an uninformative sensor by almost never choosing to observe with that sensor. The sensor manager presented in this work is well-suited to the grid-based static target detection problem or to a detection problem in which a pre-screener has identified a set of objects or locations which require further interrogation. Future work will continue to develop and analyze the presented sensor management framework; for example, in future work the variance on the observed features, which was assumed to be known in this work, will be allowed to be unknown. The sensor manager will also be tested through additional simulations and with additional real data, particularly multi-axis data.

Feature Selection

In general, the statistical algorithms being considered for UXO discrimination can process either raw data, or features extracted from raw data. Generally, the community has opted to focus on processing features, particularly since these can often be tied to a specific phenomenology. Generally, raw data incorporates enough complexity that feature-based processing is an efficient and effective discrimination methodology. In general, the specific model that is used to invert the data provide different features, and it has been hypothesized that processing the features from multiple models may also be a useful approach for UXO discrimination. The combination of employing approaches that suggest more complicated models, thus producing larger feature sets, utilizing multiple models, and introducing combinations of transmitters/receivers may result in a 'feature explosion'. Generally, statistical algorithms cannot easily be trained in high-dimensional feature spaces, and are often not robust when they operate in high dimensional feature spaces.

There are many techniques in the literature for feature selection – essentially down selecting from a large feature set to a smaller feature set containing the most relevant, discriminative, features. The tradeoff among feature selection techniques is obtaining the optimal feature set versus computational efficiency. Ad-hoc approaches that rely on knowledge of the physics are often good, but degrade if the physics becomes more complex, or if the underlying models are not phenomenological in nature. Exhaustive feature selection provides the optimal feature set, but computational expense is prohibitive as feature size grows. Generally, sequential approaches, which are among the most computationally efficient, involve growing the selected feature set by the 'next best' feature at a time but are not guaranteed to find the optimal feature set.

An example of how sequential feature selection (SFS) works is shown in Figure 32. Consider a set of 7 features, 6 of which are to be selected. SFS picks the single feature which maximizes its performance criteria (e.g. FAR at a given Pd, area under the ROC curve). In the first step of the toy example, it picks feature 5. In the next step, it then considers all possible pairs of two features, one of which is feature 5, maximizes its performance metric and picks {2,5}. It continues adding the best feature to the set in each step. In this way it picks the feature set {2,3,4,5,6,7} when in fact the correct answer is {1,2,3,4,5,6}. This results because feature 7 is

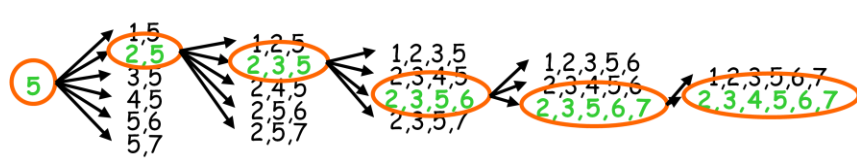

Figure 32. An example of a sequential feature selection of six features from a set of 7. Each step lists the sets being considered, with the circled set selected as best and propagated to the next stage.

deemed better than feature 1 in step 5, and once feature 7 is selected there is no mechanism to reconsider that selection in a future step. SFS is not exhaustive in that it does not consider all possible sets of 7 choose 6 features. However, it is computationally efficient.

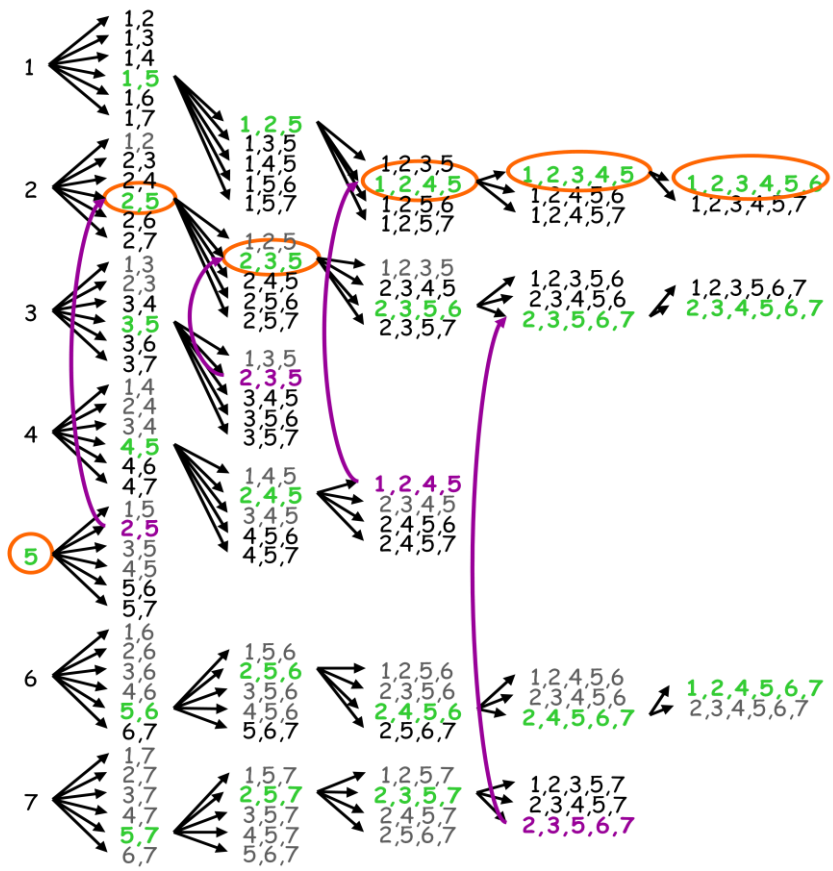

Figure 33. An example of the parallel-sequential (ParSe) feature selection method, again searching for the best set of 6 features out of 7.

The basic idea behind the approach that we are considering is to evaluate N sequential feature selections in parallel. This is inspired by the Viterbi algorithm, which provides a computationally efficient algorithm to find the maximum likelihood estimate of a hidden Markov process. We have introduced modifications to the Viterbi algorithm, however, since a state sequence 2-5 and 5-2 are considered to be different in the Viterbi algorithm, but should not be differentiated in a feature selection algorithm. A pictorial example of how the proposed 'parallel-sequential' feature selection algorithm is provided in Figure 33.

As with the SFS, the first step considers all possible sets of 1 feature and selects the best (5). Also as with SFS, the second step considers all possible sets of 2 features that also contain feature 5. Different from SFS, however, is that potential feature sets continue to be propagated forward and evaluated, as denoted by the arrows in Figure 33. Greyed feature sets are not considered for computational efficiency since they are considered in a branch above (e.g. {1,2} is propagated into the second step in the feature 1 branch, so does not need to be considered in feature 2's branch). A branch is terminated only once it is being considered in a branch above it after feature down select, e.g. in step 3, {2,3,5} is selected in 2's branch so 'children' of that branch will be considered in branch 2 so need not be considered in branch 3. Following the chain of red circles indicating subsequent feature set selection, this algorithm finds the correct feature set {1,2,3,4,5,6} with only a modest increase in the number of feature sets that must be considered over SFS. Notice that in step 4 the feature selection jumps from one sequence (originating from feature 2) to a different sequence (originating from feature 1). This illustrates the capability of ParSe to compensate for errors in feature selection earlier in the process.

Figure 34 considers computational performance for a test case from a feature selection test database with 15 features. It plots number of feature set evaluations as a function of the number of features selected. This plot shows there is a significant decrease in the number of feature sets evaluated for our

Viterbi-like algorithm (note the log scale on the y axis) when compared to an exhaustive search, and a modest increase over SFS.

Figure 35 considers the performance of the resulting algorithm using the selected features. A simple GLRT is used for ease of comparison. The plot shows 1 – Pfa at a Pd of 99% as function of the number of features selected. Sequential search has generally poorer performance, and our proposed algorithm tracks the performance of the exhaustive search quite well.

In summary, it appears that the proposed Viterbi-like parallel-sequential (termed ParSe) feature selection algorithm provides improved performance over simple search algorithms with lower computational expense compared to exhaustive approaches. It has also been applied to the LBL BUD AEM data as described below, and was also applied to the USGS ALLTEM processing. The approach continues to show good performance for the new data being considered from both the USGS ALLTEM and LBL BUD system.

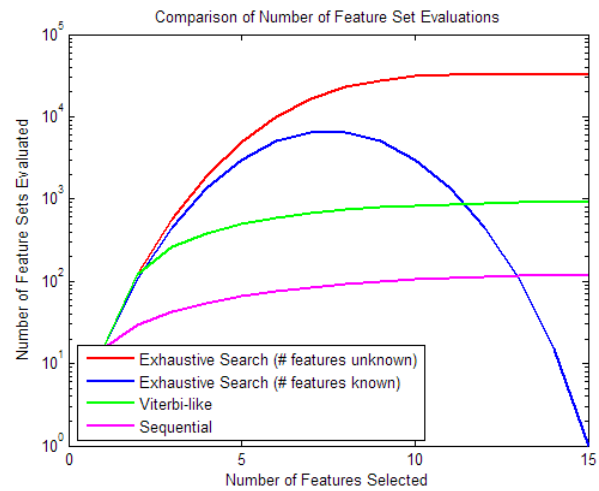

Figure 34. Number of feature sets evaluated for 4 different feature selection algorithms as a function of the number of features selected.

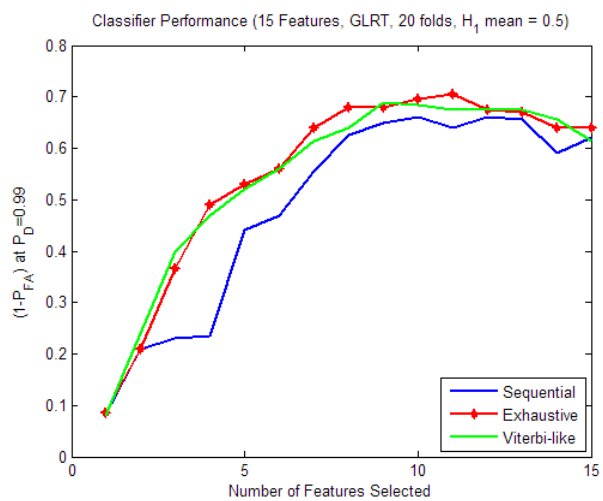

Figure 35. Performance of the classification algorithm for the selected feature sets.

Feature selection was applied to the simple sum of decaying exponentials signal model and dipole model BUD data feature sets individually as well as all the features in aggregate. The performance metric that was optimized was P_{FA} at P_D=1 (P_{FA} was minimized at P_D=1). The decaying exponential model produced as total of 11 features: signal energy, decay rate for N=1, decay rates for N=2 and their ratio, and decay rates for N=3 and their ratios. The dipole model with the BOR assumption produced 7 features: signal energy, M_1, M_2, M_2/M_1, ω_1, ω_2, and ω_2/ω_1. The dipole model without the BOR assumption produced 13 features: the features for the dipole model with the BOR assumption plus M_3, M_3/M_1, M_3/M_2, ω_3, ω_3/ω_1 and ω_3/ω_2. Performance as a function of the number of features selected is shown in the left panel of Figure 36. The simple decaying exponential model features and the aggregated features consistently provided the lowest P_{FA} at P_D=1, the dipole model with the body-of-revolution (BOR) assumption provided the worst performance, and the dipole model without the BOR assumption fell in between. These trends in performance were also seen in the ROC curves associated with the selected feature sets, which are shown in the right panel in Figure 36.

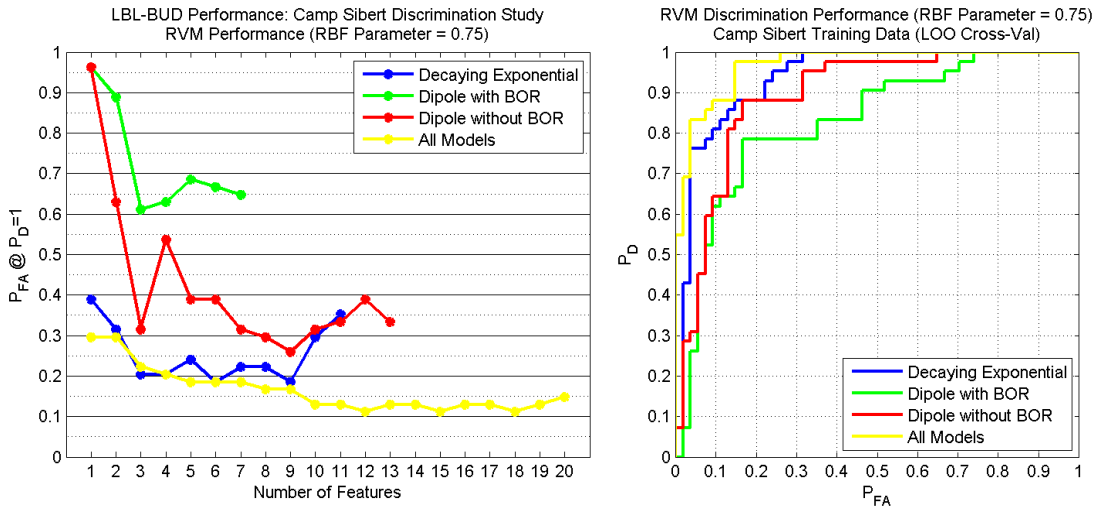

Figure 36. Feature selection classifier performance (minimum P_{FA} at P_D=1) for features extracted from the LBL BUD Camp Sibert data. The left panel shows performance as a function of the number of features selected, and the right panel shows ROC performance for the feature sets containing 8 features (7 features for the dipole model with the BOR assumption).

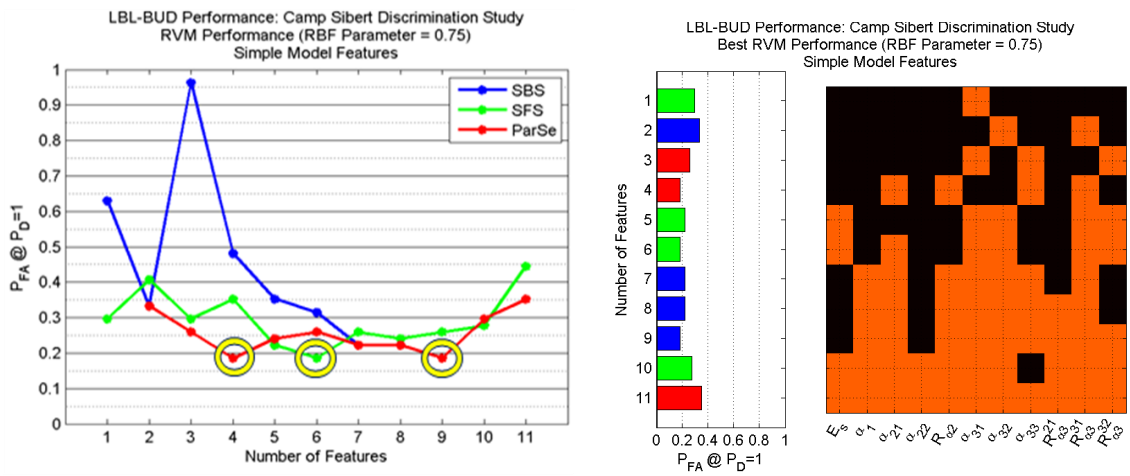

Figure 37. Feature selection results using the entire LBL BUD Camp Sibert data set. The left panel shows performance for each of 3 feature selection algorithms as a function of the number of features selected. The right panel shows performance (PFA at PD=1) for the best feature set for a given number of features along with the selected features (orange pixels).

There is a concern, however, that the minimum P_{FA} at P_D=1 performance metric may not be a robust performance metric, meaning that it may vary greatly with small changes in the data set. A sensitivity analysis was performed in which the data set was minimally perturbed by removing data associated with a single target. The results of this analysis are shown in Figure 37 and Figure 38.

Feature selection results for the decaying exponential signal model features extracted for the entire LBL BUD AEM Camp Sibert data set are shown in Figure 37. The left panel shows performance (P_{FA} at P_D=1) for each of 3 feature selection algorithms: sequential backward search (SBS), sequential forward search (SFS), and parallel-sequential search (ParSe). Overall, ParSe feature selection appears to provide better

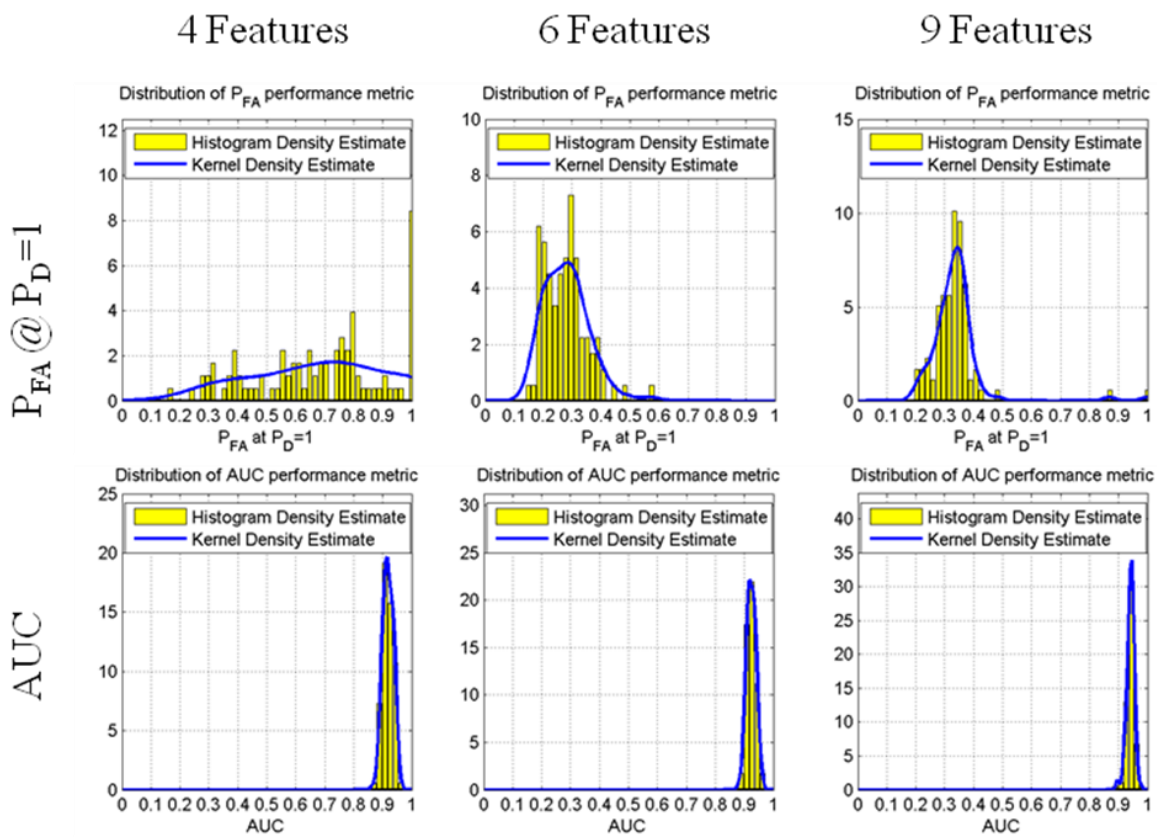

Figure 38. Histograms and corresponding kernel density estimates for the distributions for P_{FA} at P_D=1 and the area under the ROC curve (AUC) when the data set is minimally perturbed by removing data associated with a single target. The 3 feature sets considered were the feature sets with the best performance determined by feature selection on the full data set, denoted by yellow circles in Figure 37.

feature selection performance than either SBS or SFS. The right panel shows the best performance (minimum of the 3 curves in the left panel) along with the selected features (orange pixels).

Sensitivity analysis was performed for the 3 feature sets with the lowest P_{FA} at P_D=1 (4 features, 6 features, and 9 features), indicated by the yellow circles in the left panel of Figure 37. The features included in those feature sets are denoted by the orange pixels in rows 4, 6, and 9, respectively, in the right panel in Figure 37. The original, full, data set was minimally perturbed by removing data associated with a single target. This process resulted in a total of 96 data sets, each of which contained data for 95 targets. Feature selection was performed on each of the 96 data sets, and the distribution of the resulting performance metrics were examined and compared to the feature selection results obtained using the entire data set (96 targets). The distributions for P_{FA} at P_D=1 and the area under the ROC curve (AUC) were examined for the 3 feature sets considered. Histograms and corresponding kernel density estimates for these distributions are shown in Figure 38. The top row of plots shows the distributions for P_{FA} at P_D =1, and the bottom row shows the distributions for AUC. The left column shows the distributions for the feature set with 4 features, the center column for the feature set with 6 features, and the right column for the feature set with 9 features. The distributions show that the P_{FA} at P_D=1 performance metric exhibits much greater variability than the AUC performance metric when the data set is minimally perturbed. For the smallest feature set, consisting of 4 features, the distribution of

the P_{FA} at P_D=1 performance metric is approaching a uniform distribution. However, the distribution for the AUC performance metric for that feature set resembles a Gaussian distribution with an extent of about 0.85 to 0.95. The variability in the performance metrics decreases as the size of the feature set increases for both P_{FA} at P_D=1 and AUC.

This analysis revealed that P_{FA} at P_D=1 varies significantly with small changes in the data set. Area under the ROC curve (AUC), on the other hand, does not vary nearly as much, suggesting that AUC may be a more robust performance metric than P_{FA} at P_D=1. This trend in performance metric sensitivity is consistent over a wide range of feature sets.

Feature Selection Summary

A feature selection algorithm, such as ParSe, which maintains some diversity in the candidate feature sets appears to have the potential to enable improved performance by providing more discriminative feature sets with lower computational expense than an exhaustive feature selection which examines every possible combination of features. Feature selection performance, however, is dependent on the performance metric it optimizes. A sensitivity analysis revealed that when the data set is minimally perturbed by removing the data associated with a single target, the AUC performance metric was much more robust than the P_{FA} at P_D=1 performance metric, indicating that the choice of performance metric to be optimized may have a significant impact on the feature selection results.

Fisher Information for Model Inversion

Previously, Fisher Information was investigated as a potential metric for selecting the more informative data to use in model inversion. Fisher Information can also be used to assess the quality of a model inversion. Since the Fisher Information measures the information conveyed by the measured data about the model parameters, a model which is likely will have high Fisher Information while a model that is not likely will have low Fisher Information. In this way, two candidate target models which produce very similar fit errors can be compared in terms of their Fisher Information. This use of Fisher Information differs from that previously presented in that here it is being used to select the most informative model for the given data, and previously it was used to select the most informative data to invert the model.

The notion of using Fisher Information as the objective function for model inversion was evaluated using EM61 array data collected at Camp Sibert. The target model employed, proposed in [**25**], is a nonparametric model in which

$$M = \begin{bmatrix} k_2(t) & 0 & 0 \\ 0 & k_2(t) & 0 \\ 0 & 0 & k_1(t) \end{bmatrix}.$$

The parameters in this model correspond directly to the EMI61 measured data; a term $k_n(t)$ corresponds

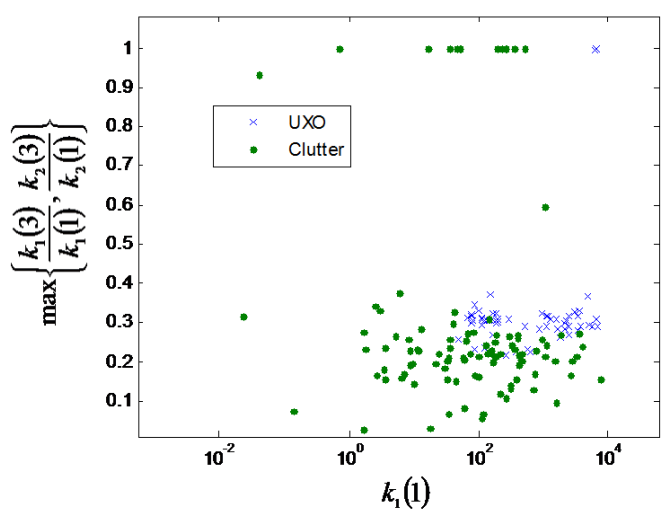

Figure 39: Scatter plot of features derived from the target model proposed in [**25**] using EM61 array data collected at Camp Sibert.

to the tth time gate for the nth polarizability axis, where $n = 1$ corresponds to the largest polarizability axis. The features derived from this model, $k_1(1)$ and $\max\left\{\frac{k_1(3)}{k_1(1)}, \frac{k_2(3)}{k_2(1)}\right\}$, are shown in a scatter plot in Figure 39. While the majority of the UXO features are within two clusters in the lower right portion of the plot, there are a few UXO targets for which the features do not fall near one of the two UXO clusters. Fisher Information for model inversion was evaluated to assess if it could produce features for those UXO targets which would fall near one of the two UXO clusters

As an example of how using Fisher Information inversion may provide improved target features, consider Target 159 from Camp Sibert. The measured data (solids lines) and the model fits (dashed lines) for two possible models are shown in Figure 40 as a function of spatial sample number. The blue lines correspond to the first time gate, while the red lines correspond to the third time gate. Both models provide very similar normalized residual ; 1,042,900 for the model on the left and 1,069,000 for the model on the right. Although the model on the right has a slightly larger normalized residual, and therefore would be considered a slightly inferior model using least squares criteria, it has a much larger

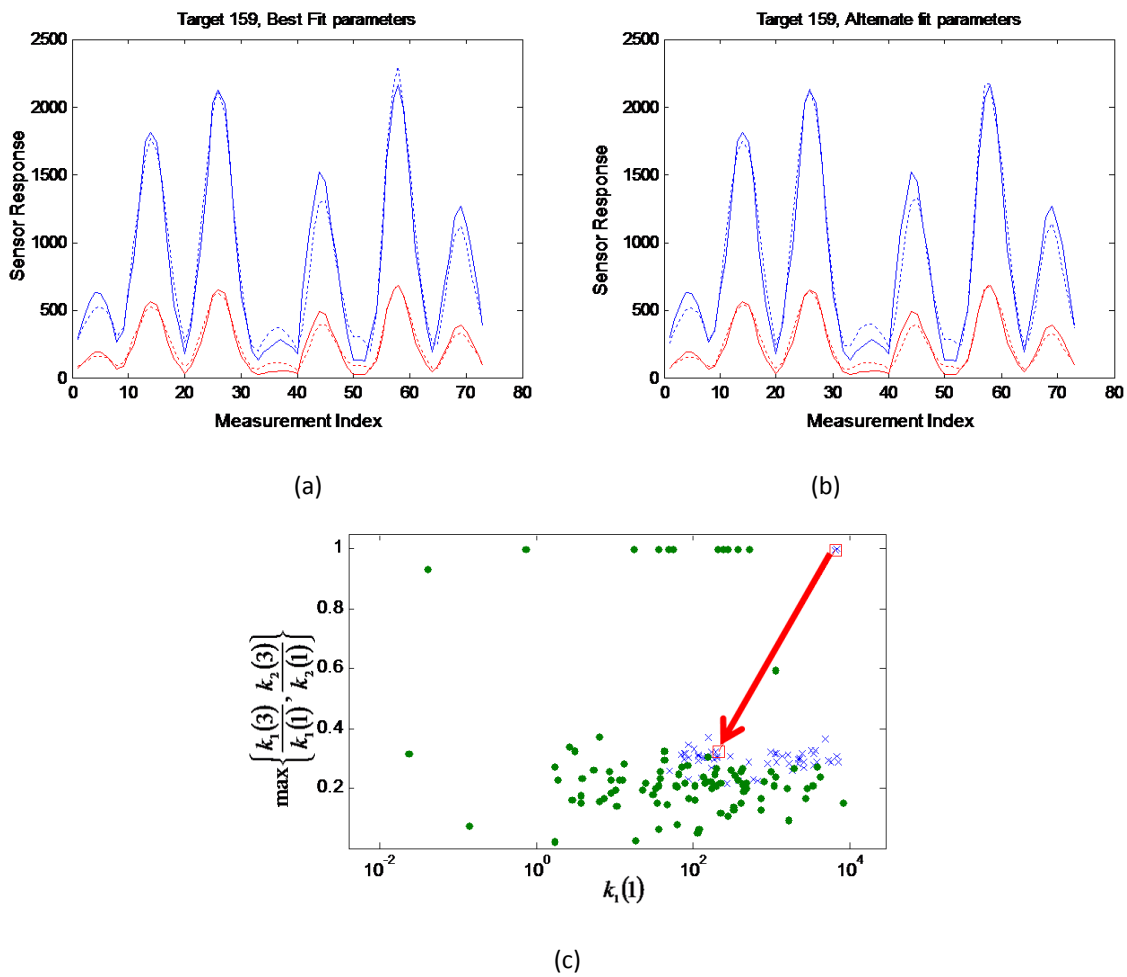

Figure 40: Comparison of two model fittings for Target 159 at Camp Sibert. (a) Model fit with normalized residual 1,042,900 and trace(FIM) 4.15. (b) Model fit with normalized residual 1,069,000 and trace(FIM) 1138.7. (c) Target features for the model fit with larger normalized residual and large FIM (right plot) move toward a UXO feature cluster when compared to target features for the model with smaller normalized residual and small FIM (left plot).

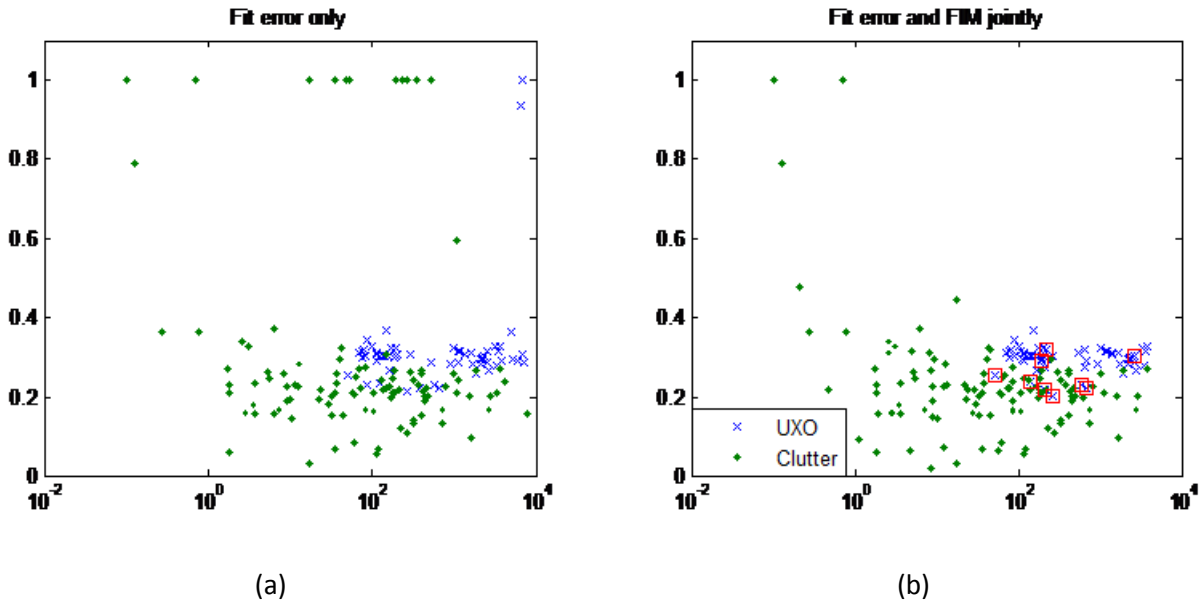

Figure 41: Improvement in UXO target feature clustering when fit error and FIM are jointly optimized. (a) Optimization of normalized residual only. (b) Joint optimization of normalized residual and Fisher Information measure.

Fisher Information measure (1138.7 versus 4.15), and would be considered a vastly superior model using Fisher Information measure criteria. Examination of the corresponding target features shows that the features for the model with the smaller normalized residual and small Fisher Information measure (left plot) are far from the two features clusters for UXO targets, but the features for the model with the larger normalized residual and large Fisher Information measure (right plot) are within one of the two clusters for UXO targets. This example demonstrates the potential utility of the Fisher Information measured to provide valuable guidance in model inversion.

Jointly optimizing the normalized residual and the Fisher Information measure provides substantial UXO discrimination performance improvement using the EM61 array data from Camp Sibert. When only normalized residual is optimized, there are a number of UXO target whose features are some distance from the two UXO feature clusters, as shown in Figure 41(a). When normalized residual and Fisher Information are jointly optimized, however, the features for the UXO targets that were previously far from the UXO feature clusters move toward the two UXO feature clusters, as shown by the red squares in Figure 41(b).

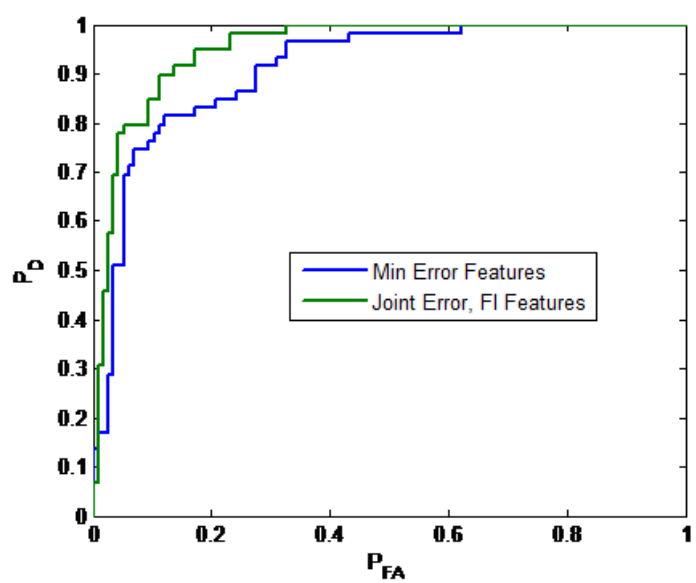

Figure 42: Detection performance for fit error alone and joint optimization of fit error and FIM.

The improved target features provided by the joint normalized residual and Fisher Information measure results in improved UXO discrimination performance, as shown in Figure 42. When normalized residual is optimized alone, P_{FA} is approximately 0.62 at P_D=1, but when normalized residual and Fisher Information measured are jointly optimized, P_{FA} at P_D=1 decreases to approximately 0.32– a 47% improvement in P_{FA} at P_D=1.

Fisher Information for Model Inversion Summary

Fisher Information may be a useful tool for optimizing model inversions. By selecting a model inversion which provides both a small normalized residual and a large Fisher Information, the target features appear to be more likely to be consistent with UXO target features, as demonstrated with the EM61 array data collected at Camp Sibert. The improved target features obtained when Fisher Information is considered within the inversion process enable improved UXO discrimination performance, as also demonstrated with the EM61 array data collected at Camp Sibert.

USGS ALLTEM Test Stand Data Processing

The USGS ALLTEM system provides multi-dimensional data by utilizing three orthogonal transmitting coils and multiple receivers, resulting in a total of 19 polarizations. The goal is to assess whether access to multi-dimensional data provides an advantage over the standard single-direction transmitter/receiver pair. Since we initially did not have access to a phenomenological model, we again initiated our processing with a simple sum of decaying exponentials signal model. It also required a bit of time to sort out how the data was represented and making sure we used consistent notation as the USGS researchers. Results on processing the YPG standardized test site calibration site testing are quite promising.

USGS later provided to Duke additional data from their ALLTEM system. Previous experience with this system at the YPG calibration grid suggested that while this system performs as well as its best single polarity for discriminating targets from blanks, it noticeably outperformed the best single polarity when discriminating targets from clutter. Given this indication for potential performance improvement, USGS has provided finely sampled test-bed data for 12 UXO and 10 clutter items, seven of which are canonical. The objects are measured multiple times at different orientations and depths. Initial assessment of the data has involved the extraction of the same features previously used successfully on the YPG calibration lane: magnitude change, energy of the decay curve, single decay rate estimates, and two-decay rate estimates. While the small data set prevents algorithm-based feature selection, some naïve selection is possible.

First, energy detection performance was considered for each individual polarization. Then the best energy detector performance was compared to multi-axis performance using the energy from each polarization as a feature. While the performance of the energy detector for each individual polarization was close to chance, the performance of using the energies of all polarizations as features in a KNN classifier variant resulted in 100% probability of detection at a probability of false alarm of 6%. The strong performance may be due in part to the selection of clutter items; however, it does demonstrate the potential performance improvement from using multi-axis data rather than single-axis data.

Rather than attempting global feature selection, which is not feasible with the small number of objects available, Duke considered polarization selection. The performance of each polarization, using all the features derived from that polarization, was determined in terms of area under the ROC curve (AUC). For the multi-axis classification, the features from the top performing three polarizations were used in an RVM. The scores in general were poorer than for using energy alone which is likely a function of not

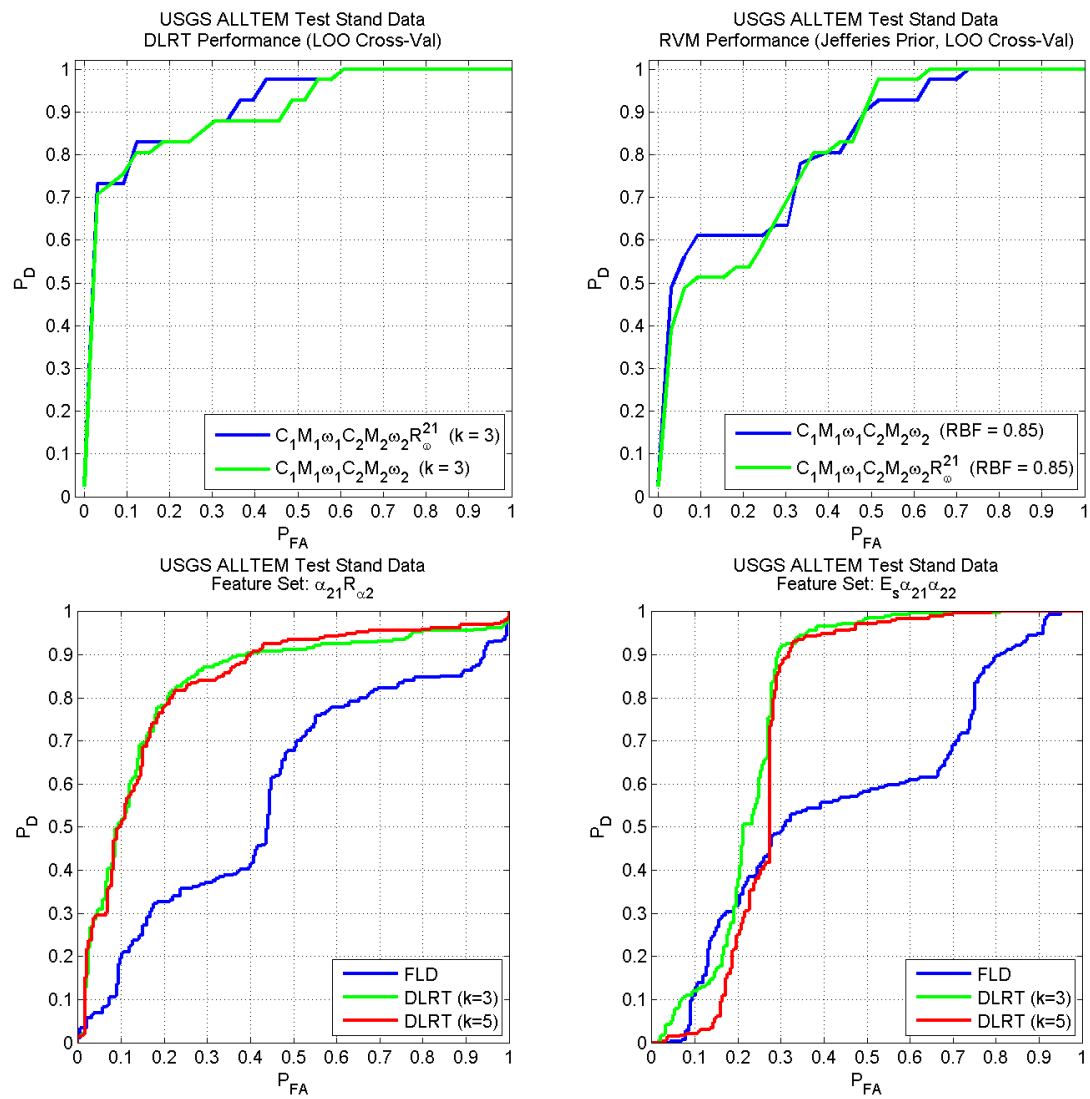

Figure 43. ROCs for USGS ALLTEM Test Stand data. Distance likelihood ratio test (DLRT) operating on features from the dipole model (top left), relevance vector machine (RVM) operating on features from the dipole model (top right), DLRT and RVM operating on the best simple model features excluding signal energy (bottom left), and DLRT and RVM operating on the best simple model features including signal energy.

discarding redundant or counterproductive features; however, multi-axis performance (0.74 AUC) was still higher than the best single-axis performance (0.68 AUC). This indicates that, as suggested by the previous YPG calibration results, multi-axis sensors have the potential to outperform the best performing single-axis sensor for clutter discrimination. Further, since the best performing single-axis polarization for a given scenario is unlikely to be known beforehand, the multi-axis sensor has the advantage of providing all the polarization information in all scenarios.

Several classifiers were considered for features derived from the ALLTEM test stand data using the simple model and the dipole model with the BOR assumption. Generally, classifiers developed for the dipole model outperformed classifiers developed for the simple model, particularly at high P_D. ROCs for the dipole model and simple model are shown in Figure 43. The top row shows ROCs for dipole model features and the bottom row shows ROCs for simple model features. The results of applying the

distance likelihood ratio test (DLRT) are shown in the top left plot and the results of applying relevance vector machines (RVM) are shown in the top right. Both classifiers have relatively stable performance for a wide variety of feature sets and classifier parameters. The DLRT tends to outperform the RVM. The results for the best simple model feature set excluding signal energy are shown in the bottom left and the results for the best simple model feature set including signal energy are shown in the bottom right. Again, there is a fair amount of stability in the results for a variety of classifier parameters. Including energy in the simple model feature set tends to improve performance at high P_D. In both cases, a Fisher linear discriminant (FLD) performs rather poorly, indicating that the features are not linearly separable.

One important distinction between the dipole model results and simple model results shown here is that the dipole features were extracted using multiple measurements, whereas the simple model features were extracted using a single measurement. As noted earlier for the LBL BUD sensor, utilizing multiple measurements to estimate the simple model parameters tends to improve the inversion results and result in better clustering in the estimated parameters. It remains to be seen if increasing the number of measurements considered in the inversion process improves performance using the simple model for the USGS ALLTEM sensor as it did for the LBL BUD AEM sensor.

Extracting features using the simple model highlighted the potential need to consider model order selection as part of the model inversion process. A single decay rate model applied to the 19 channels of data at each spatial location tended to produce a target "image," as the single estimated decay rate varied with the target/sensor orientation. The spatial variance of the decay rate tended to follow the geometry of the target. This suggests that choosing a model order that is too low will result in spatial variability of the decay rate estimates. In contrast, a three decay rate model tended to produce spatially stable decay rate estimates for complex targets, for which a higher order model would be expected to be appropriate, but spatially chaotic decay rate estimates for simple targets, such as the spherical BLU-26, for which a lower order model would be expected to be appropriate. In addition to considering model order selection, using data from multiple spatial measurements for decay rate estimation will be investigated as a way to mitigate the potential spatial variability of the decay rate estimates.

USGS ALLTEM Test Stand Data Processing Summary

USGS ALLTEM data test stand data was processed and analyzed. The results indicate that reasonable UXO discrimination performance can be attained with model-based features and statistical classifiers. Results also revealed that model order selection may be an important consideration in model inversion, as the chosen model order may have a significant impact on the model inversion results.

LBL BUD YPG Data Processing

BUD data measured at the YPG calibration lanes has been analyzed rather extensively as part of the ongoing algorithm development process. While attempting to understand the root causes of persistent false alarms and missed detections in the YPG calibration lanes data set, it was discovered that the time-domain decay curves for all targets exhibited remarkable similarity. The decay constants did not vary with target size as much as we expected them to given our experience with other time-domain EMI sensors and the electromagnetic models that have been developed to predict the EMI signatures for UXO. For example, the time-domain decay curves for the 37mm and 105mm targets were virtually indistinguishable after normalizing the magnitude of the initial sample to 1, despite the targets being very different sizes. In an effort to understand these unexpected characteristics, the data pre-processing and system effects were investigated. Specifically, the variable width half-sine wave filters utilized to convert the raw linearly sampled data to logarithmically sampled responses for detection and

discrimination has been examined, and this pre-processing step does not appear to be causing the similarity in the time-domain decay curves.

LBL BUD Camp Sibert Discrimination Study Data Processing

We received new BUD data in April 2008, and commenced working with it. This initially involved significant back and forth communication with Erika Gasperikova, as there were some issues in interpreting the data and the ground truth.

Features have been extracted from the Camp Sibert data set using a rigorous model which includes all the system geometry and a simplier model which captures the essence of the rigorous model. The rigorous model represents the measured data on each channel as a weighted sum of magnetization tensors, with the weights determined by the target/sensor orientation parameters (target inclination and azimuth relative to the sensor, target depth, and target location relative to the sensor). The initial rigorous model assumed the target can be well-characterized by two magnetization tensors, which is equivalent to assuming it is a cylindrical shape or body of revolution (BOR). This assumption has since been relaxed and the model now assumes the target can be well-characterized by three magnetization tensors. The simpler model captures the essence of the rigorous model without the constraints imposed by including the target/sensor orientation parameters in the model. In this case, each channel of measured data is also assumed to consist of the weighted sum of two or three magnetization tensors (decaying exponentials), but the linear weights on the magnetization tensors are found by a least squares fit to the data rather than through modeling the system geometry. Cluster analysis of the decay rates which characterize the magnetization tensors indicate that the simpler model results in features (decay rates) which are more tightly clustered by target type than the features estimated using the rigorous model. We believe this result occurs because the rigorous model is much more constrained than the simpler model due to the calculation of the magetization tensor weights using the target/sensor geometry and the data are insufficient to reliably estimate the decay rates within the constraints imposed by the geometrical model. The simpler model, on the other hand, does not impose constraints on the magnetization tensor weights; it simply requires that the magnetization tensors be the same across all channels of data utilized in the inversion.

An alternate approach to feature extraction, diffusion analysis, has also been considered. Diffusion analysis is a semi-supervised approach that evaluates the pair-wise distances among all available data (either raw data or extracted features) and through an eigenanalysis rotates the data into a new coordinate system where the data are well-represented with a small number of dimensions. The diffusion analysis classification results seem to indicate that diffusion analysis may be a promising approach.

Among the statistical classifiers considered are linear discriminants, relevence vector machines (RVM), and the distance likelihood ratio test (DLRT), a LRT where the required probabilities are estimated from the available data using K-nearest neighbors density estimation. Leave-one-out cross-validated classification results with the labeled data from Camp Sibert indicate that classification performance is quite good; the area under the ROC curve approaches 0.98 and PFA at PD=1 is as low as 0.2. However, it was observed that the simple sum of decaying exponentials signal model provided better classification performance than either dipole model (with and without the BOR assumption). This result is counter-intuitive because the dipole model was developed with fewer simplifying assumptions than the simple sum of decaying exponentials model. In an effort to understand why the features from the simple model provide better classification than features from the dipole model, the characteristics of the features themselves were examined in finer detail.

The purpose of this study was to further investigate the differences in classification performance for the simple phenomenological model and the physics-based dipole model. The M3 dipole model, using three unique terms in the magnetization tensor matrix, was chosen for comparison in this study since it had the largest degree of underperformance when compared to the simple model and also does not impose a symmetry constraint (which has seen decreasing use by other researchers). Both models use eight features, so the complexity of the features spaces are approximately equivalent. Classification scores $p(\mathrm{H}_1 | x)$ were generated using a logistic discriminant classifier, which should offer similar performance to the RVM used originally, but with the benefits of lower complexity and classification scores that can be interpreted using visualizations of the feature space. These characteristics will be beneficial in the analysis of how individual anomalies were classified. Classification results are shown below, as well as an analysis to identify problematic test samples and outliers, and to determine the factors resulting in outliers in feature space.

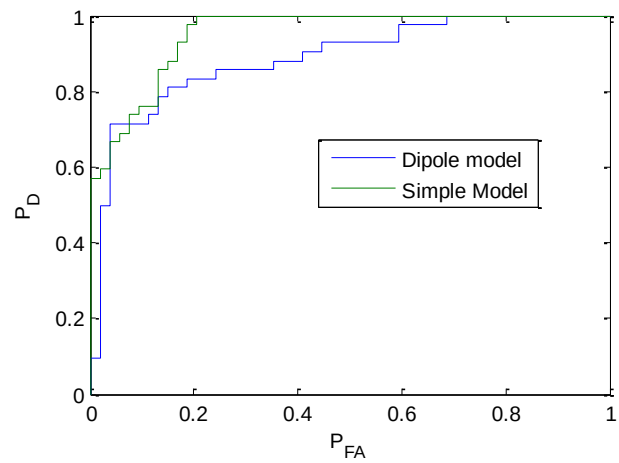

Figure 44. ROC curves for the two feature sets. Leave-one-out cross validation, Distance Logistic Discriminant,

In Figure 44, the ROC curves for the two feature sets are shown. These results are similar to those achieved using the RVM; performance using the dipole model lags behind performance with the simple model, particularly at the operating point $P_D = 1$. The false alarm rate at $P_D = 1$ is similar to the values seen using the RVM classifier.

Figure 45 illustrates the classifier outputs that produced the ROC curves shown in Figure 44. Ideally, the objects would be sorted such that all of the clutter items would be on the left side of the plot with low $p(H_1 | x)$ and all of the UXO would be on the right side with high $p(H_1 | x)$. For the simple model, the classifier outputs are relatively well-behaved. There are no UXO with low $p(H_1 | x)$, which allows a large number of false alarms to be rejected even at high UXO detection rates. There are six clutter objects with probability of being UXO greater than 0.9; these will be considered "Problematic Test Samples" and will be investigated more closely in the following analysis. The term "Problematic Test Samples" is used since Figure 45 does not reveal any information about how these points behave as training samples. However, from

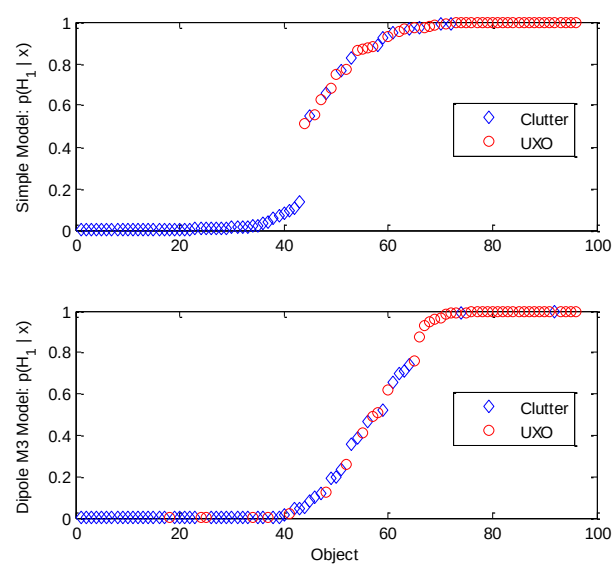

Figure 45. Sorted classifier outputs for the simple model (top) and dipole model (bottom).

Figure 45 we can determine that they produce anomalous classifier outputs when used as test samples; therefore, they are at a minimum problematic samples when testing. The classifier outputs using the dipole model features show a larger degree of overlap between the decision metrics for the UXO and clutter; with six UXO problem test samples having a probability of being UXO less than 0.1, and two clutter problematic test samples identified using the threshold of $p(H_1 \mid x)$ >0.9. In order to diagnose why the classifier using the dipole model features generates more false alarms before identifying all of the UXO, it is necessary to determine why some UXO have such low classifier outputs. In particular, we will consider the six problematic UXO that have $p(H_1 \mid x)$ < 0.1.

Before taking a closer look at the most problematic test samples for the two models, it is worth checking whether there are common anomalies in the identified problem samples for the two models. A scatter plot of the classifier outputs using each model, shown in Figure 47, reveals that the sets of problematic test samples for the two models are unique, with no common anomalies. Additionally, the classifier outputs are less correlated that anticipated. As shown in Figure 46, the relatively small number of points along the diagonal between (0,0) and (1,1) suggests that except for anomalies for which categorization is very clear (both targets can say $p(H_1 \mid x)$ is very near zero or unity), the classifiers based on the two sets of model features are unlikely to agree. Since the classifier parameters were equal in both cases, this result suggests that the representations in feature space provided by the two models are dissimilar for many of the anomalies considered in this data set.

Table V lists the characteristics of the problematic test samples for each model. For the classifier using the simple model, there are six clutter objects with high $p(H_1 \mid x)$. For the classifier using the dipole model, there are six UXO with low $p(H_1 \mid x)$ and two clutter objects with high $p(H_1 \mid x)$. Objects with ID numbers in the 5000 range come from the GPO section of Camp Sibert. The object characteristics listed in Table V do not suggest any obvious feature that explains why certain anomalies were problematic.

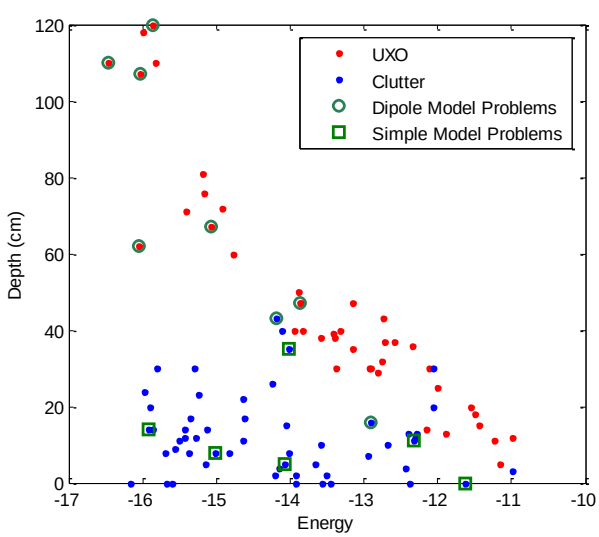

Figure 47. Location of problematic test samples in energy/depth space.

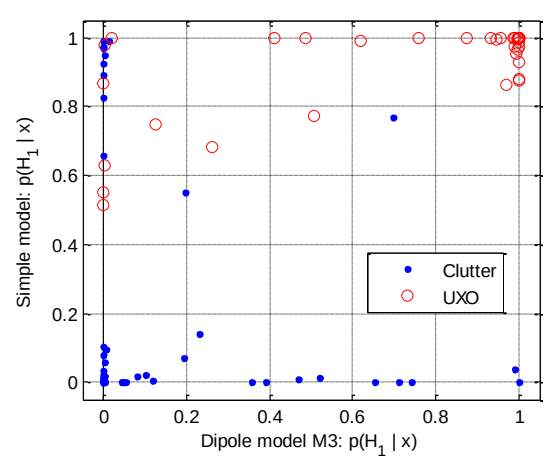

Figure 46. Scatter plot of classifier outputs using dipole model features (x-axis) and simple model features (y-axis).

Simple Model p(H1|X)	Label	Name	ID	Depth (cm)	Length (cm)
0.926	0	HR-7	5037	35	40
0.951	0	HR-3	5033	11	40
0.968	0	Scrap Metal	852	5	15
0.972	0	Frag	74	8	6
0.989	0	Scrap Metal	53	14	12
0.991	0	Halfshell	617	0	20

Dipole Model M3 p(H1|X)	Label	Name	ID	Depth (cm)	Length (cm)
0.000	1	UXO	45	62	45
0.000	1	UXO	643	110	45
0.000	1	42-145	5021	67	40
0.002	1	UXO	57	120	45
0.004	1	42-100	5017	107	40
0.019	1	42-179	5030	47	40
0.992	0	HR-8	5038	43	40
1.000	0	HR-6	5036	16	40

Table V. Characteristics of problematic test samples.

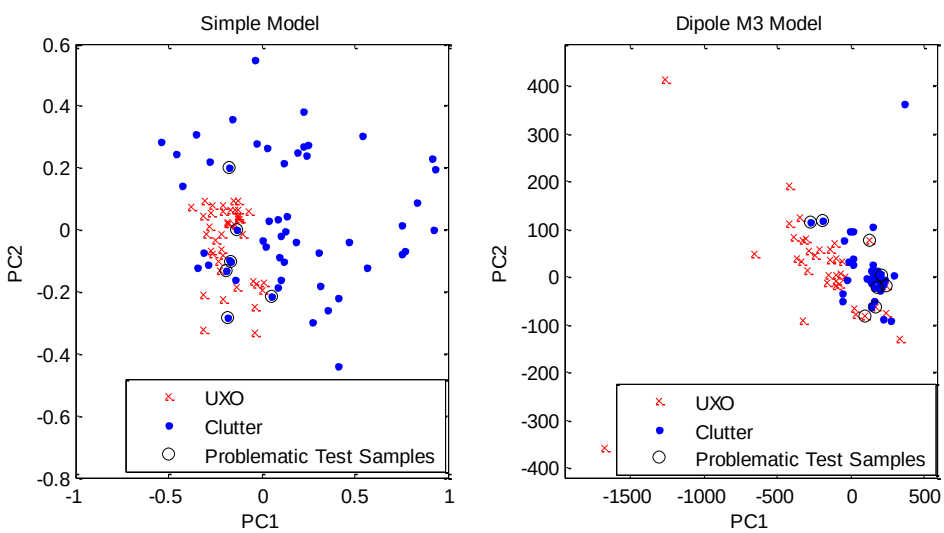

Figure 48. Two-dimensional visualization of the (left) simple model feature space and (right) dipole model feature space.

Additional consideration of anomaly characteristics looked at energy and ground truth depth. The eight dipole model problematic samples are identified by circle markers; the six simple model problematic samples are identified by square markers. The dipole model does have problems with 3 of the 5 deepest UXO targets, which might be indicative of some restrictions in the inversion process (e.g. a too-limited range of initialization values for some parameters). However, few other trends exist.

The next step in the analysis was to visualize the representation of all anomalies in feature space using the principal components. Scatter plots of the first two principal components are shown in Figure 48. In the simple model feature space, the UXO anomalies are relatively clustered and the clutter objects have more diverse representations, which reflects the object population at Camp Sibert (a single size of UXO and various types of clutter anomalies). For the dipole model, the UXO representations display a larger degree of variability than the clutter objects, while a subset of clutter objects form a relatively tight cluster in feature space. This situation does not reflect the characteristics of the anomalies at Camp Sibert; additionally, it suggests that classification performance using the dipole model features was driven by learning the clutter rather than by consistent representation of the UXO. In general, it is desirable to avoid training on the clutter items since clutter can be highly variable in the field and difficult to characterize in advance.

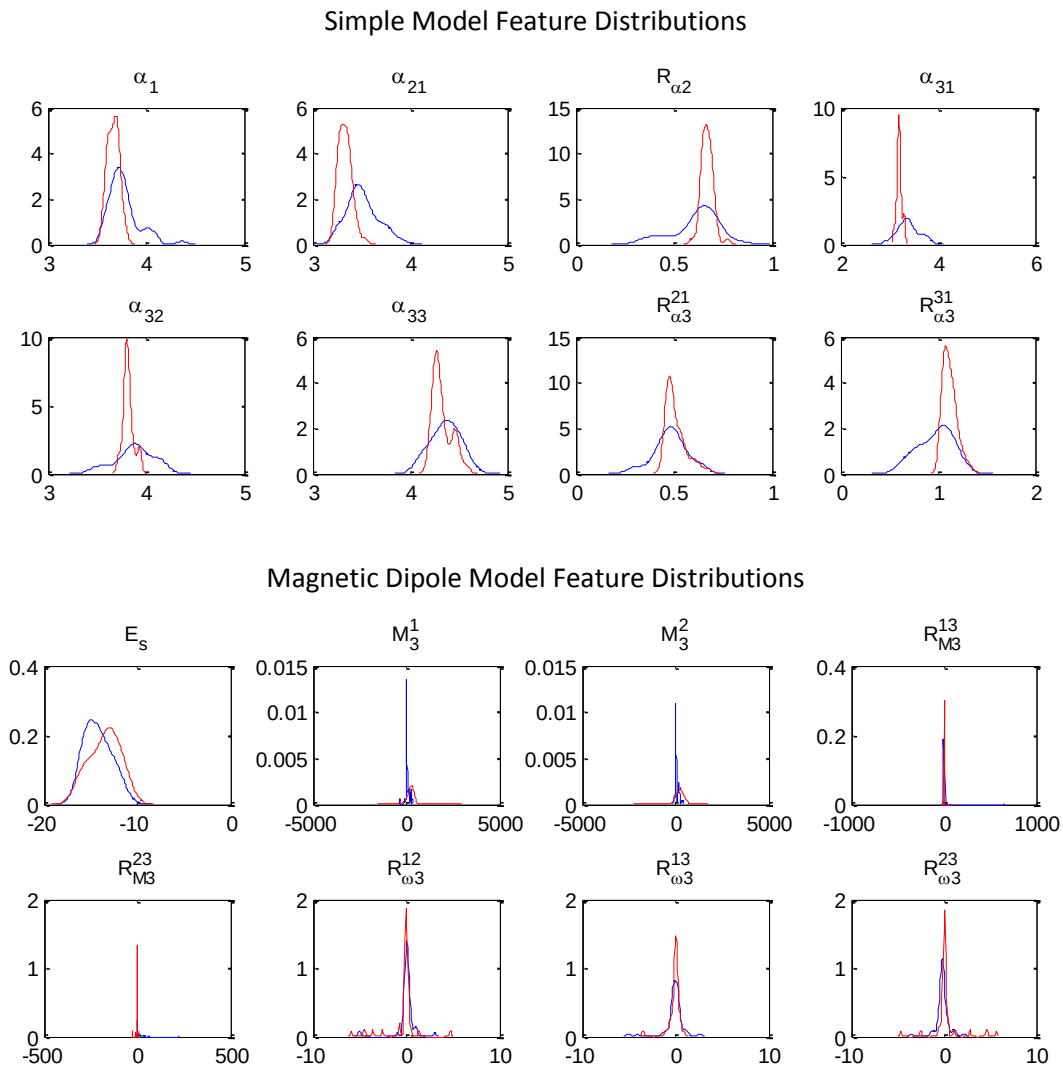

Figure 49. Feature distributions of UXO (red) and clutter (blue) for the simple model (top) and dipole model (bottom).

To ensure that the behavior observed above was not simply an artifact of reducing dimensionality using principal components, a matrix of distances between each pair of anomalies was constructed for each model. Plots of the distance matrices (not shown) verify that the UXO representations are more closely clustered than the clutter objects in the simple model feature space, whereas the opposite behavior is observed in the dipole model feature space.

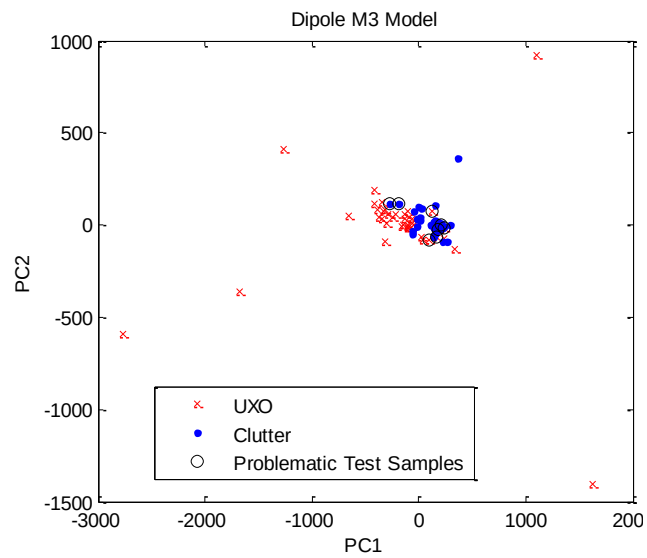

Figure 50. Two-dimensional visualization of the (left) simple model feature space and (right) dipole model feature space.

The distributions of the individual model features, shown in Figure 49, indicate that the simple model representations (top eight subplots) have fewer outliers in the UXO representations and also show greater consistency among the UXO than the clutter, as expected. Two disturbing observations in the distributions of the dipole model features are 1) the narrow distribution of clutter feature values in the M1 and M2 features and 2) the UXO outliers, which are most obvious in the features constructed from ratios of resonance frequencies. Based on these observations, it can be concluded that the simple model features provide better representation of the UXO anomalies that are more consistent with the population of objects at Camp Sibert.

There are five outlier UXO in the dipole model feature space that can be seen if the axis limits in Figure 48 are expanded to show the entire range of values, as depicted in Figure 50. These UXO are not problematic test samples since they are not necessarily difficult to identify as UXO. However, since it is known that the UXO population at Camp Sibert is quite consistent, we are interested in looking further into what causes these UXO to have feature representations outside the range of the other UXO feature

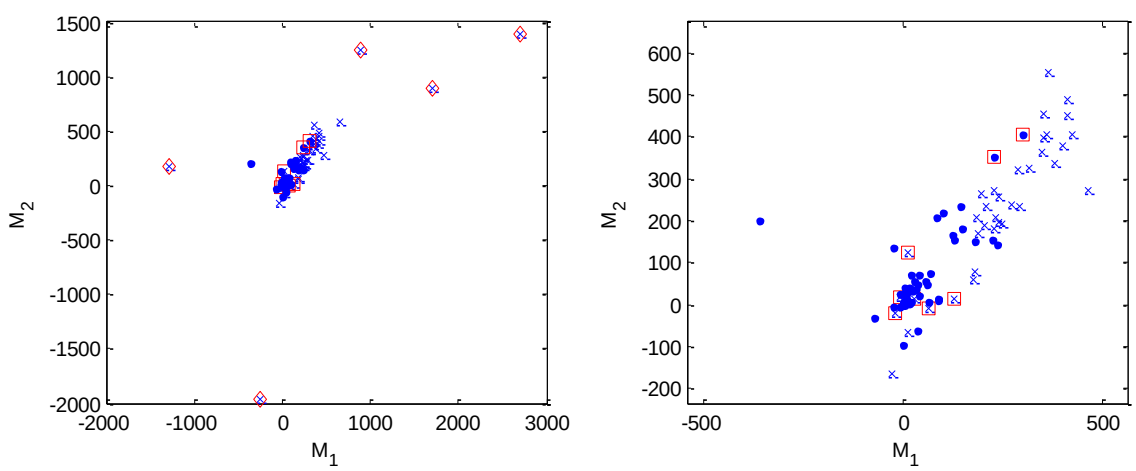

Figure 51. Scatter plot of two dipole model features: M_1 and M_2.

values.

The existence of the outlier UXO is driven by two features, M1 and M2. The five outlier UXO from Figure 50 are identified with diamond markers in Figure 51, and are located at relatively extreme values of the dipole model parameters M1 and M2. Also, a close-up view of the main cluster shows the locations of the problematic test samples (identified with square markers) situated within or near representations of anomalies from the wrong class. It appears that these two dimensions are driving performance. They result in the observed consistent representation of the clutter (rather than the preferable scenario of consistent representation of the UXO) and determine the outlier UXO and problematic test samples. However, since these dimensions produce such consistent representations of the clutter, and that seems to be driving performance, excluding them from the classification has a significant, negative impact on performance.

This study performed a further investigation of the reasons underlying the differences in performance seen for the simple model features and the dipole model features. It was observed that the dipole model classification performance was driven by consistent representation of the clutter objects, whereas the simple model was consistently representing UXO. The distribution of UXO and clutter in the dipole model feature space was driven by two features, M1 and M2. The other features exhibited only minimal discriminability between UXO and clutter. Thus, these are issues with the feature representations generated from the model parameters. They cannot be addressed using training methods, feature scaling and normalizing, or application of different classifiers. The dipole model representations need to be modified either through a different inversion technique or data selection. The simple model parameters were estimated using trimmed time series (the early time measurements were removed). It is not clear whether this was performed on the data for the dipole model inversions, but it could have a significant effect. Also, since the dipole model is much more complicated than the simple model, it may require a more sophisticated inversion technique to achieve more consistent representation of the anomalies in the other model parameters.

Ongoing efforts will continue to focus on developing classifiers for the BUD system. Both data-based and feature-based classifiers will be investigated, and improvements to the feature extraction and model parameter estimation algorithms will be considered. The blind data from the Camp Sibert discrimination study have been processed, and decision statistics for three of the most promising classifiers have been sent to the program office. The results of scoring by an independent third party, shown in the next section, indicate promising performance.

As a preliminary step, the magnetization tensors resulting from the generalized magnetization tensor inversion for the LBL BUD Camp Sibert data have been classified using a simple correlation classifier. The preliminary results, shown below in Figure 52, approach the results reported by LBL for the BUD system,

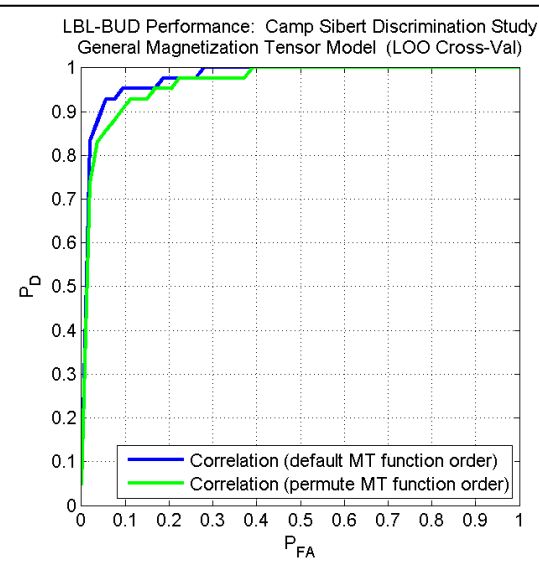

Figure 52. Correlation classifier performance for generalized magnetization tensor inversion for the LBL BUD Camp Sibert data.

and show promise. The magnetization tensors were estimated based on the modified Berkeley model described earlier. Two scenarios were considered. In the first, denoted "default" the order of the magnetization tensors produced by the inversion is assumed to be correct. In the second, denoted "permute" the order of magnetization tensors is not assumed to be correct, and all permutations of the order are considered in the correlation. The magnetization tensor order that produces the highest correlation is taken as the correct order. Classifiers for these data will continue to be developed.

LBL BUD Camp Sibert Discrimination Study Data Processing Summary

Model-based features for the LBL UBD sensor data collected at Camp Sibert was analyzed extensively to better understand why classification using features from the less rigorous sum of decaying exponentials signal model would outperform classification using features from the more rigorous dipole model. The analysis revealed that for the simple model, the UXO features were more clustered than the clutter features, while for the dipole model, the opposite was true – the clutter features clustered more tightly than the UXO features. Having a better understanding of the mechanisms underlying the surprising classification performance results provided insight into how to approach rectifying the relatively poor performance obtained using features from the dipole model. It appeared that the features from the dipole model were not as descriptive or informative of the UXO targets as the features from the simple sum of decaying exponentials signal model. With this in mind, the model inversion and feature extraction were re-examined to investigate the potential for shortcomings within those stages of the data processing. It was discovered, relatively late into the program, that the multi-axis inversions using the dipole model were not as accurate as they could have been. The model inversion process has since been improved, and a subset of the analyses have been repeated with the modified, improved, inversion process. The inversion process modifications and resulting changes in the performance results are described in detail in the following section.

EMI Dipole Model Inversion Modifications

Part of this SERDP effort was to develop phenomenological models for multi-axis sensor systems developed for UXO detection and discrimination, specifically the LBL BUD AEM sensor and the USGS ALLTEM sensor, and to then use these models within the construct of physics-based statistical signal processing strategies to develop robust algorithms for UXO detection and discrimination. Typically, the statistical signal processing algorithms operate on phenomenological model parameters estimated from the measured data. Since the estimated model parameters are the features utilized within the signal processing algorithms, a key step bridging the models and the algorithms is the data inversion process from which the model parameter estimates are obtained.

Model inversion can be viewed as a fairly straight-forward task within the broader physics-based statistical signal processing process which typically consists of forward modeling, model inversion, feature extraction and selection, and a decision or classification algorithm. It often reduces to optimizing an objective function, such as the sum of squared error between the measured data and the modeled data (L_2 norm), for which there are many standard solutions and approaches. It will be shown here, however, that taking care to ensure the inversion process is consistent with the underlying phenomenological model can improve both the inversion accuracy and convergence speed, while failing to do so can result in inaccurate inversions which are time-consuming to obtain. Ensuring consistency of the inversion with the underlying model involves modifying standard inversion approaches so that 1) the objective function is optimized in a parameter space in which it is well-behaved, and 2) initializations are selected so that the regions of the objective function most likely to yield the globally optimal parameters are preferentially searched. The positive impact of accurate model inversions is most directly seen

through improvements in the model fits to the measured data, but it also manifests itself in improvements in discrimination performance. The resulting improvements in the end performance (discrimination) are of more interest in the UXO discrimination application.

Motivations for Considering EMI Dipole Model Inversion Modifications

Over the course of several months, some seemingly inconsistent performance results were obtained. The common thread shared by all these results is that a phenomenological model which neglected to incorporate the spatial variations in the measured signal amplitudes provided better performance than a phenomenological model which fully incorporated these spatial variations. In other words, features provided by a simple and incomplete signal model gave better classification performance than a more sophisticated and more complete signal model.

Briefly, there were three instances of seemingly inconsistent performance results. The first instance involved the clustering, or lack thereof, of the estimated poles, the second instance involved the discrimination performance results, and the third instance involved the model fits themselves.

Example scatter plots of the poles estimated from BUD data measured in the YPG calibration lanes are shown in Figure 53. The left panel shows the poles estimated using the original dipole model inversion process with the body-of-revolution (BOR) assumption. The BUD sensor geometry and effects of propagation between the sensor and target are included in this model. The right panel shows the poles estimated using a weighted sum of *M* decaying exponentials signal model with *M*=2. Neither the BUD geometry nor the effects of propagation between the sensor and target are included in this model. The estimated poles for each target type are plotted in a unique color, as shown in the key to the right of each scatter plot. Each target is represented several times in the calibration lanes, at various orientations and depths, resulting in several sets of estimated poles for each target type. Theoretically, we would expect the estimated poles for each target to be intrinsic to the target and independent of any extrinsic factors such as the target orientation relative to the sensor. Thus, we would expect the estimated poles for a given target to be fairly consistent and to cluster fairly well. This behavior is

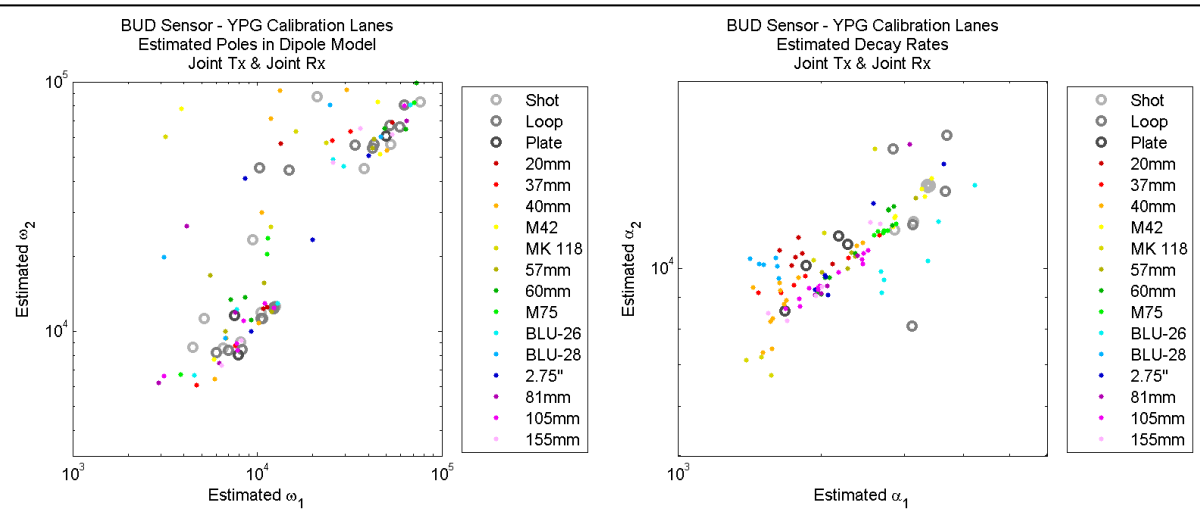

Figure 53: Scatter plots of poles estimated from BUD data measured in the YPG calibration lanes. Left panel: Poles estimated using the dipole model with the BOR assumption. Right panel: Poles estimated using a weighted sum of *M* decaying exponentials signal model with *M*=2.

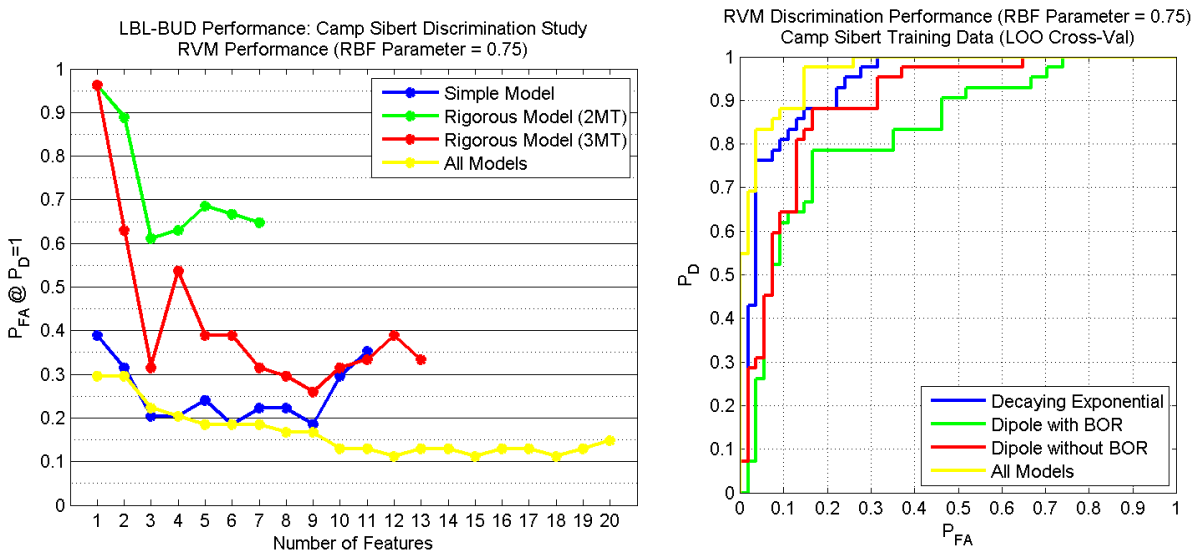

Figure 54: Discrimination performance using BUD data measured as part of the Camp Sibert discrimination study for feature sets generated from three different signal models and the aggregation of the three feature sets. Left Panel: P_{FA} @ P_D=1 as a function of the number of features selected. Right panel: Discrimination performance using the best 8 features (7 features for Rigorous Model (2MT)).

exhibited in the results obtained using the weighted sum of decaying exponentials signal model, but it is not seen in the results obtained using the dipole model.

Discrimination performance results obtained with features estimated from the BUD data measured as part of the Camp Sibert discrimination study are shown in Figure 54. The left panel shows probability of false alarm (P_{FA}) at probability of detection (P_D) of 1 as a function of the number of features selected for three different signal models, and the right panel shows discrimination performance using the best 8 features (7 features for the Dipole with BOR signal model). Simple Model (Decaying Exponential) features are the poles estimated using the weighted sum of *M* decaying exponentials with *M*={1,2,3} and their ratios. Rigorous Model (2MT) (Dipole with BOR) features are the magnetization constants and poles estimated using the dipole model with the BOR assumption and their ratios. Rigorous Model (3MT) (Dipole without BOR) features are the magnetization constants and poles estimated using the dipole model without the BOR assumption and their ratios. All Models features are the aggregation of the features from all three signal models under consideration. Both rigorous models utilize features derived from the original dipole model inversion process, and include the BUD sensor geometry as well as the effects of propagation between the sensor and target. The simple model, on the other hand, does not include the effects of propagation between the sensor and target. Theoretically, we would expect to see performance improve as the accuracy of the signal model increases. Thus, we would expect performance improvements from the Decaying Exponential model to the Dipole with BOR model to the Dipole without BOR model. However, this performance trend is not evident in either performance evaluation. In both cases, the Decaying Exponential (Simple Model) model outperforms the Dipole (Rigorous) models; P_{FA} at P_D=1 is lower and the ROC curve is higher.

Example model fits using the original dipole model inversion process (top row) and the weighted sum of decaying exponential signal model (bottom row) are shown in Figure 55. Model complexity increases from left to right in each row. In each plot, there are 12 curves, each representing a different transmitter-receiver channel. The blue asterisks represent the measured data, and the red lines represent the model fits to the measured data. The residuals are shown below the data. Theoretically,

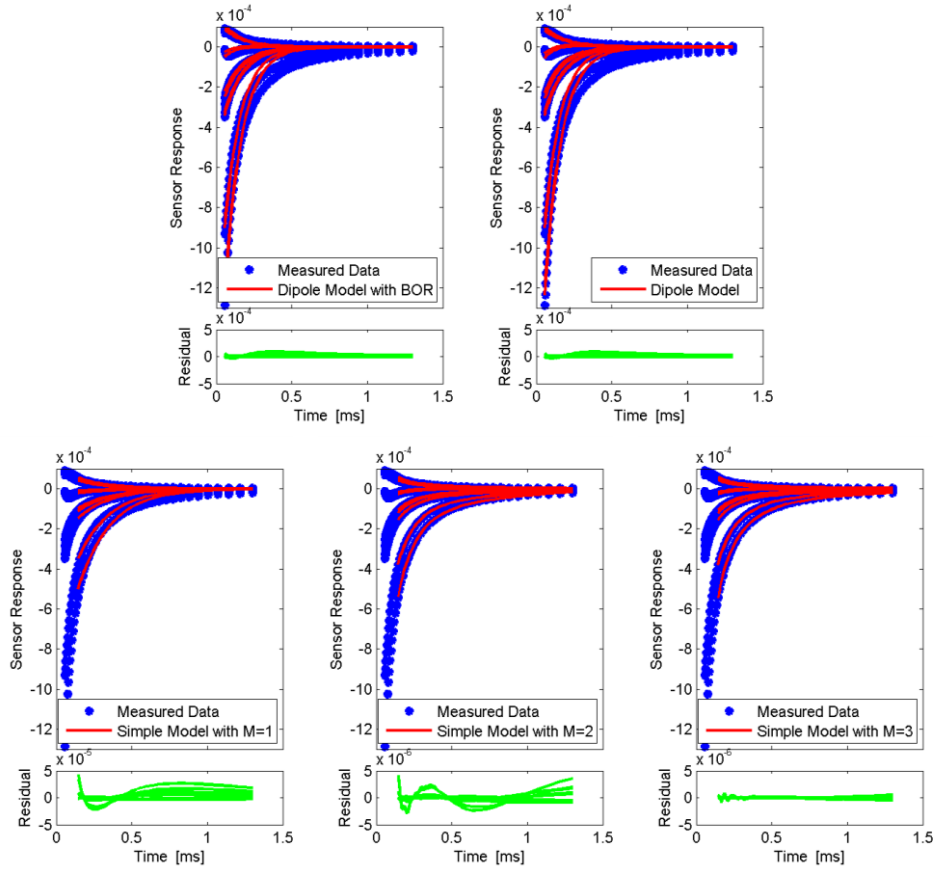

Figure 55: Example model fits and residual for the dipole (top row) and weighted sum of decaying exponential (bottom row) signal models. Model complexity increases from left to right in each row.

we would expect to see the dipole model provide better fits to the data than the weighted sum of decaying exponentials signal model, and for the model fits to improve with increasing model complexity for both models. While we do see the model fits improve with increasing model complexity, it is clear that the weighted sum of decaying exponentials signal model inversion provides better model fits (lower residual) than the original dipole model inversion. Once again, the weighted sum of decaying exponentials signal model inversion performs better than the original dipole model inversion.

In each of these three instances, the results are both counter-intuitive and contrary to what theory would suggest. While the principle of parsimony, the notion that the least complicated model which adequately represents the data is preferable, could possibly explain these seemingly inconsistent results, this explanation seemed unlikely because the more sophisticated of the two models, the dipole model, approximates the interrogated target as an idealized (dipole) target and therefore was not likely to be overfitting the data or experiencing difficulties in the inversion due to its high precision. We believed the more likely explanation was that something unexpected was happening in the model inversion process for the more sophisticated model which was leading to inaccuracies in the model inversion, and so modifications to the EMI dipole model inversion process were investigated.

EMI Dipole Model Inversion Modifications

Three modifications have been considered: a modification in the geometric aspects of the EMI propagation model, a re-parameterization of the dipole model for inversion, and the introduction of a sequential inversion process. The effects of each modification individually have not yet been investigated, so it cannot yet be conclusively stated which modification, or set of modifications, is driving the changes in inversion performance. In addition to modifying the inversion process, the time samples selected for inversion have been chosen to be consistent with those selected by LBL. The time samples originally utilized for inversion were not always consistent with LBL's recommendation.

EMI Propagation Model Modifications

The target models for both the USGS ALLTEM sensor and the LBL BUD AEM sensor are based on a magnetization tensor model which takes the form

$$M(t) = \begin{bmatrix} f_1(t) & 0 & 0 \\ 0 & f_2(t) & 0 \\ 0 & 0 & f_3(t) \end{bmatrix}.$$

When the target is assumed to be rotationally symmetric (a body-of-revolution, or BOR), then $f_1(t) = f_2(t)$. When the transmitter and receiver are (nearly) co-located, the received magnetic field is proportional to $U^T M U$, where M is the magnetization tensor given above and U is a unitary rotation matrix that rotates the fields from the sensor coordinate system to the target (dipole) coordinate system. The rotation matrix U is a function of the target orientation angles: azimuth (θ), inclination (ϕ), and rotation (ψ), as illustrated in Figure 56, and is given by

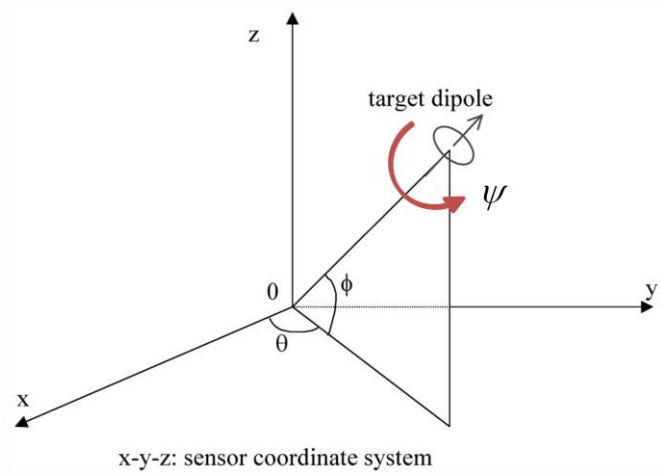

Figure 56: xyz sensor coordinate system and the target (dipole) orientation angles azimuth (θ), inclination (ϕ), and rotation (ψ).

$$U = U_3 U_2 U_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\psi & \sin\psi \\ 0 & -\sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} \cos\phi & 0 & \sin\phi \\ 0 & 1 & 0 \\ -\sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

When the target is rotationally symmetric, the rotation ψ need not be considered since rotating the target about its longitudinal axis has no effect on the sensor response to the target. Thus, the rotation ψ can be neglected for BOR targets, in which case $U = U_2 U_1$. Neglecting ψ is equivalent to setting $\psi = 0$ since U_3 is identity when $\psi = 0$. When the target is not a BOR, however, the rotation ψ is an important component of the propagation model, and should be retained in the rotation matrix U.

The EMI propagation model originally did not include the rotation angle ψ for the models with the BOR assumption ($f_1(t) = f_2(t)$) or without that assumption. The model has been modified to include this orientation angle. In addition the BOR assumption has been relaxed slightly in that it is still assumed that $f_1(t) = f_2(t)$, but the rotation angle ψ is include which allows for cross-sections, such as squares, for which the dipoles are equal and their orientation relative to the sensor need be considered. For a

	Previous model		Modified Model	
M=2 (BOR-like)	$f_1(t) = f_2(t),\ f_3(t)$	$U = U_2 U_1$	$f_1(t) = f_2(t),\ f_3(t)$	$U = U_3 U_2 U_1$
M=3 (non-BOR)	$f_1(t),\ f_2(t),\ f_3(t)$	$U = U_2 U_1$	$f_1(t),\ f_2(t),\ f_3(t)$	$U = U_3 U_2 U_1$

Table VI: Comparison of magnetization tensor functions $f_n(t)$ and rotation matrix U for previous and modified EMI dipole propagation models.

BOR, such as a target with a circular cross-section, the dipoles are equal and their orientation relative to the sensor need not be considered. The magnetization tensor functions $f_n(t)$ and the rotation matrix U for the previous and modified EMI dipole propagation models are summarized below in Table VI. By neglecting the rotation angle ψ, the previous models implicitly assume $\psi = 0$. While this is a moot point for targets that are truly rotationally symmetric, it may impact model inversion and subsequent target classification for targets that are not.

Model Inversion Process Modifications

Model inversion generally involves minimizing an objective function that measures how well the model fits the measured data. A typical objective function is the sum of the squared errors between the model and the measured data, or the L_2 norm, and there are several standard solutions and approaches to solving this least squares optimization problem. While the objective function can often be minimized analytically when the signal model is linear, a nonlinear signal model precludes an analytical solution to this problem. In these cases, the objective function must be solved numerically. This approach typically involves gradient descent, where the gradients of the objective function are estimated numerically. Two aspects of nonlinear least squares minimization problems which may affect optimization performance are the model parameterization and the selection of initial conditions. The model parameterization affects the shape of the objective function, which can render it better-suited (or less well-suited) for numerical minimization. In addition, the results may be highly dependent on the initial condition if there are multiple local minima in the objective function. In this case, the choice of initial conditions can affect which regions of the objective function are searched and if the end result is the global minimum or just one of several local minima.

Two aspects of the numerical optimization have been modified to address the potential issues posed by an objective function ill-suited for numerical optimization and the possible presence of multiple local minima. First, the dipole model has been re-parameterized to make the objective function more concave in the vicinity of the global minimum, and therefore better suited for numerical minimization. Second, a sequential inversion process has been implemented to guide the selection of initial conditions so that regions of the objective function more likely to yield the global minimum are preferentially searched. Both of these strategies have been shown to improve decay rate estimation, and since, as will be shown below, the dipole model is essentially a decay rate model which incorporates a model of the spatial variation in the exponential signal amplitudes, it was hypothesized that these modifications to the model inversion process could yield similar improvements for dipole model inversion.

Model Parameterization

The model for the sensor response, which includes propagation between the sensor and target and interaction with a dipole target, can be expressed as

$$H(r,\theta,\phi,\psi,t) \propto \frac{\mathbf{r}_{st}}{R^6}\mathbf{r}_{st} \cdot U(\theta,\phi,\psi)^T \cdot M(t) \cdot U(\theta,\phi,\psi) \cdot \mathbf{r}_{st}$$

$$= \frac{\mathbf{r}_{st}}{R^6}\mathbf{r}_{st} \cdot U(\theta,\phi,\psi)^T \cdot \begin{bmatrix} f_1(t) & 0 & 0 \\ 0 & f_2(t) & 0 \\ 0 & 0 & f_3(t) \end{bmatrix} \cdot U(\theta,\phi,\psi) \cdot \mathbf{r}_{st} .$$

In this model, a target is modeled as consisting of three orthogonal dipoles where each dipole corresponds to a term on the diagonal of the magnetization tensor. For a BOR, it is assumed that the two dipoles oriented perpendicular to the longitudinal axis of the target are equal ($f_1(t) = f_2(t)$), while the dipole oriented with the longitudinal axis of the target, $f_3(t)$, is unique. With $f_n(t) = M_n \omega_n e^{-\omega_n t}$, this model can also be expressed as

$$H(r,\theta,\phi,\psi,t) = \sum_{n=1}^{3} H_n(r,\theta,\phi,\psi) M_n \omega_n e^{-\omega_n t} ,$$

which shows that an equivalent expression for the sensor response due to the target is modeled as a weighted sum of decaying exponential signals, with the weights determined by the target/sensor geometry (r,θ,ϕ,ψ), the target magnetization (M_n), and the signal decay rate, or pole (ω_n). Each decaying exponential signal corresponds to a single dipole. This expression more clearly shows the relationship between the dipole model and the weighted sum of decaying exponentials signal model,

$$H(t) = \sum_{n=1}^{\mathrm{N}} A_n e^{-\alpha_n t} .$$

Inspection reveals that the two models are equivalent with $A_n = H_n(r,\theta,\phi,\psi) M_n \omega_n$ and $\alpha_n = \omega_n$. Thus, the only difference between the two signal models is the incorporation of a model for spatial variations of the weights on the decaying exponential signals in the dipole model, whereas there is no model for the spatial variation of the weights in the weighted sum of decaying exponentials signal model.

The above signal models are expressed in "exponential" form. Expressing the weighted sum of decaying exponentials signal model in "pole" form has been shown to improve decay rate estimation performance by altering the shape of the objective function, rendering it better-suited for numerical minimization. For the weighted sum of decaying exponentials signal model, the reparameterization is achieved by substituting $p_n = e^{-\alpha_n T_0}$, where the constant T_0 serves to scale the resulting pole so the range of possible p_n is more or less centered between 0 and 1. Thus, the "pole" form of the weighted sum of decaying exponentials signal model is

$$H(t) = \sum_{n=1}^{\mathrm{N}} A_n p_n^{t/T_0} .$$

This model is utilized within the nonlinear least squares framework to determine the p_n that minimize the residual between the model and the measured data. The estimated decay rates α_n are then found by transforming the pole back to a decay rate, $\alpha_n = -\ln(p_n)/T_0$.

The dipole model can similarly be expressed in "pole" form

$$H(r,\theta,\phi,\psi,t)=\sum_{n=1}^{3}H_n(r,\theta,\phi,\psi)B_0B_nq_n^{t/T_0}\ .$$

In this parameterization, $B_0B_n = M_n\omega_n$ and $q_n = e^{-\omega_n T_0}$. The constant B_0 serves a purpose similar to T_0; it is a scaling constant so that B_n is on the same order of magnitude as the pole q_n. As for the weighted sum of decaying exponentials signal model, the parameters B_n and q_n are optimized within the nonlinear least squares framework to minimize the residual between the model and the measured data. The desired parameters M_n and ω_n are then calculated from the estimated parameters by $\omega_n = -\ln(q_n)/T_0$ and $M_n = B_0B_n/\omega_n$.

Sequential Inversion

The notion of sequential inversion is based on two observations. First, the decay rate estimated when N=1 typically lies between the two decay rates estimated when N=2, and similarly, the two decay rates estimated when N=2 typically lie between the smallest and largest decay rates estimated when N=3. Second, model inversions with lower model orders typically encounter fewer difficulties with convergence and exhibit less sensitivity to the initial conditions. From these observations, it was hypothesized that a sequential process in which the parameter estimates from model order N were utilized to guide the selection of initial conditions for the inversions with model order N+1 could place the initial conditions for model order N+1 closer to the solution and thus improve model inversion.

The sequential inversion process is graphically depicted in Figure 57. First, a single decay rate is estimated using the decaying exponential signal model. That estimated decay rates is then used to select the initial conditions for the decaying exponential signal model with model order N=2. Those two

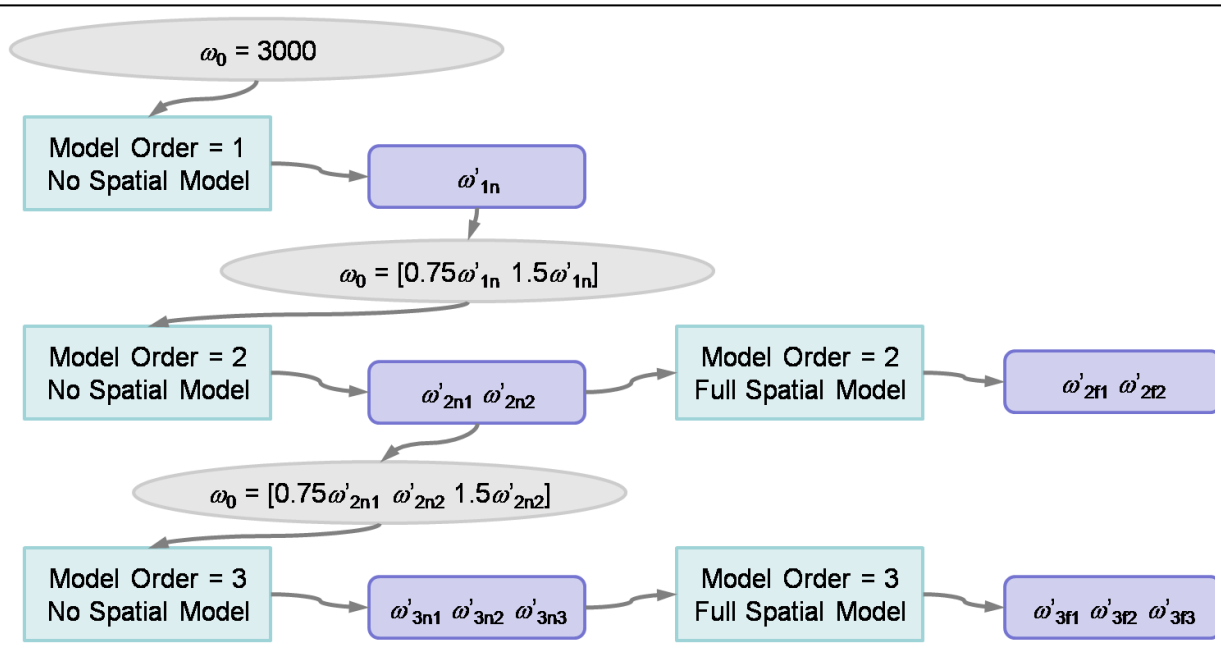

Figure 57: Sequential inversion flow chart. No Spatial Model refers to the weighted sum of decaying exponentials signal model for which there is no model of the spatial variations of the coefficients on the decaying exponentials. Full Spatial Model refers to the dipole model for which there is a model of the spatial variations of the coefficients on the decaying exponentials.

decay rate estimates are then used to select the initial conditions for the dipole model with model order N=2 as well as the decaying exponential signal model with model order M=3. Finally, the decay rates estimated from the decaying exponential signal model with model order M=3 are used to select the initial conditions for the dipole model with model order N=3. At this point, there is not a clear understanding of the mechanisms that allow this approach to provide improved inversion performance. That remains the subject of future investigation.

Modified Model Inversion Parameter Estimates

The BUD data collected as part of the Camp Sibert discrimination study were re-inverted using the dipole model with the modified inversion process and using the time samples recommended by LBL. Inversions were completed both assuming the target was a BOR, so the first two terms in the magnetization tensor are identical , and without the BOR assumption, so the three terms in the magnetization tensor may all be unique. In addition, all three orientation angles were included in the

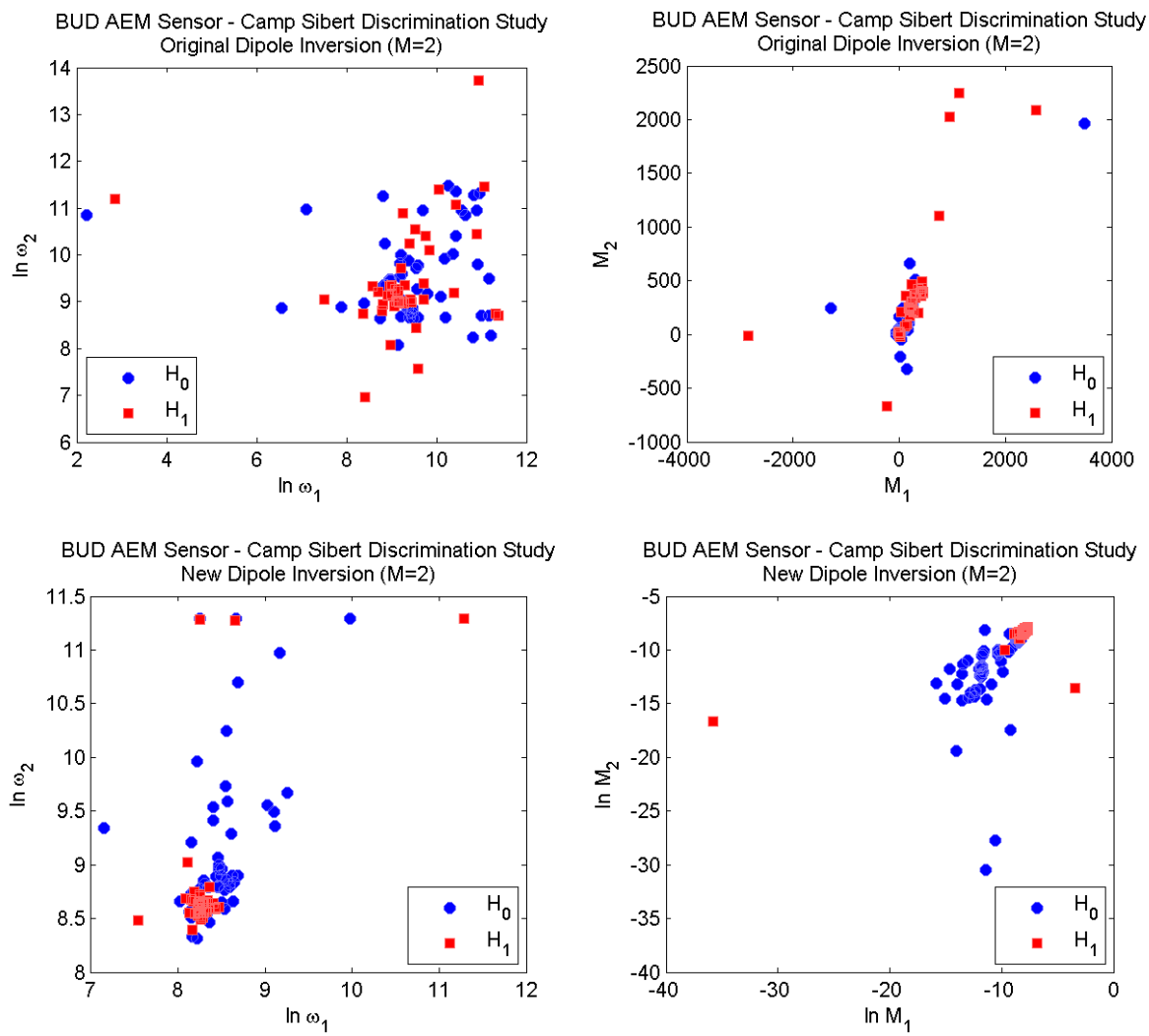

Figure 58: Parameter estimated obtained with the original (top row) and modified (bottom row) model inversions for the dipole model with the BOR assumption. The target poles are shown in the left column and the target magnetizations are shown in the right column. The non-UXO targets (H_0) are plotted with blue circles and the UXO targets (H_1) are plotted with red squares.

model inversion. The parameters estimated obtained with the modified inversion strategy are compared to those obtained with the original model inversion process

Scatter plots of the estimated dipole parameters when the BOR assumption is employed are shown in Figure 58. The top row shows the parameters obtained using the original inversion, and the bottom row shows the parameters obtained using the modified inversion process. The left column shows the poles (ω), and the right column shows the target magnetizations (M). In all four plots, the parameter estimates corresponding to the non-UXO targets (H_0) are plotted using blue circles, while those corresponding to the UXO targets (H_1) are plotted with red squares. This figure shows that the modified inversion process provides parameter estimates for the UXO targets that are more tightly clustered, and therefore better suited for non-UXO vs. UXO classification. This is true for both the target poles and the magnetizations.

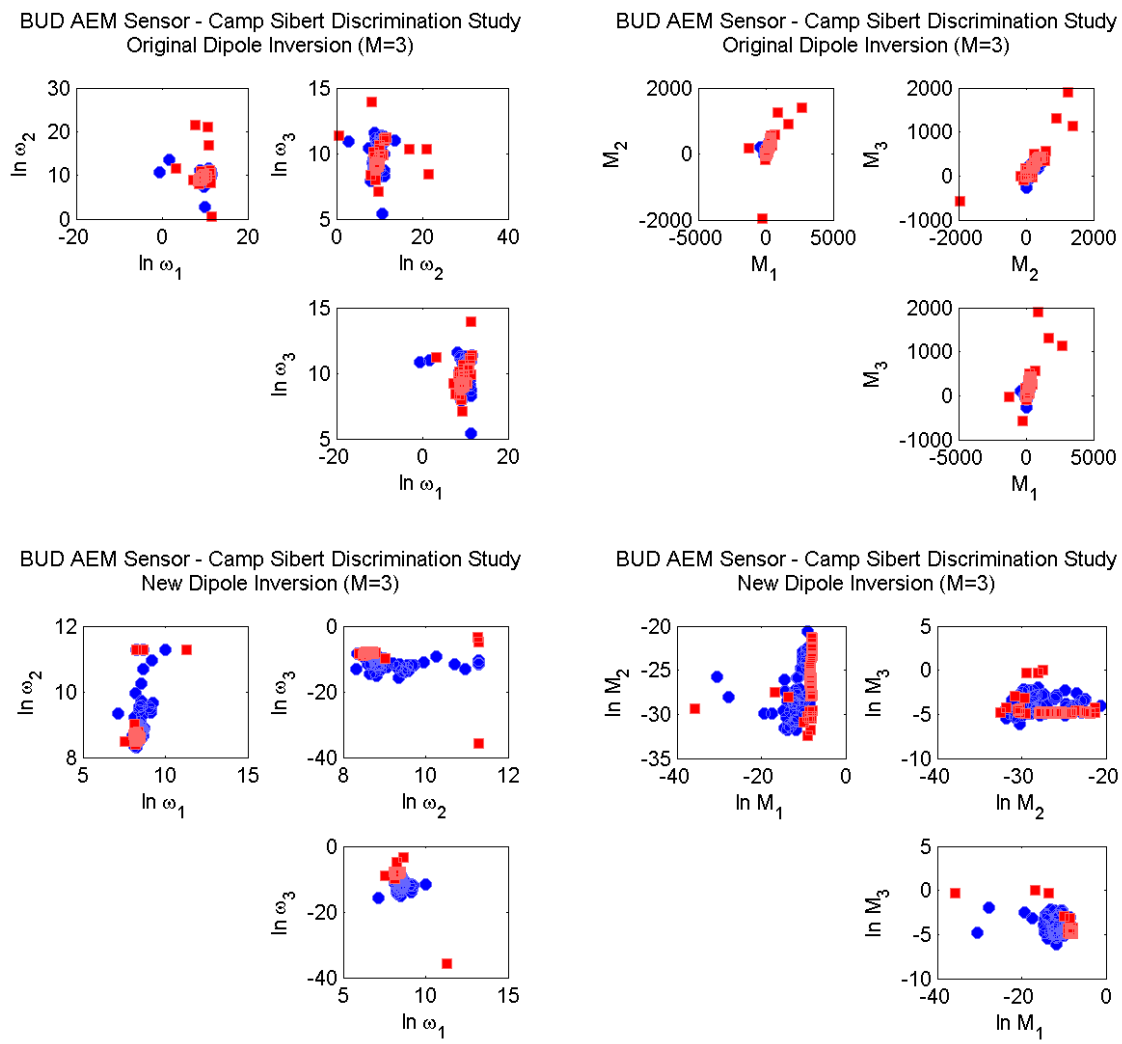

Figure 59: Parameter estimated obtained with the original (top row) and modified (bottom row) model inversions for the dipole model without the BOR assumption. The target poles are shown in the left column and the target magnetizations are shown in the right column. The non-UXO targets (H_0) are plotted with blue circles and the UXO targets (H_1) are plotted with red squares.

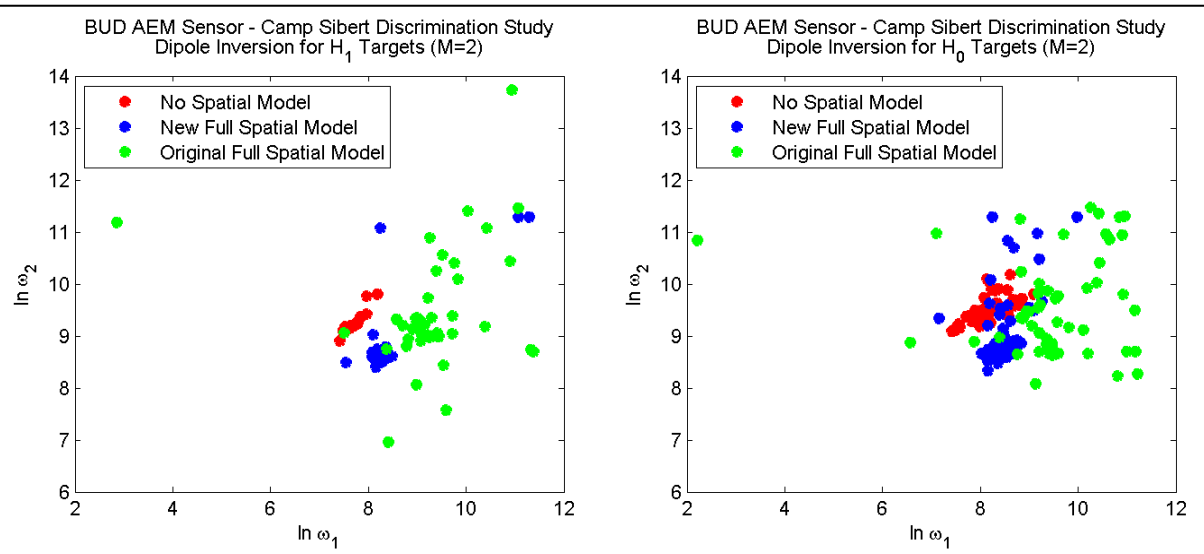

Figure 60: Target pole estimates obtained with the weighted sum of decaying exponentials signal model (No Spatial Model, red), the modified dipole model inversion (New Full Spatial Model, blue), and the original dipole model inversions (Original Full Spatial Model) for the dipole model with the BOR assumption. The target poles are shown in the left column and the target magnetizations are shown in the right column. The non-UXO targets (H_0) are plotted with blue circles and the UXO targets (H_1) are plotted with red squares.

A similar set of plots of estimated dipole parameters when the BOR assumption is not employed are shown in Figure 59. Again, the top row shows parameter estimated obtained via the original inversions while the bottom row show parameter estimates obtained from the modified inversions, and the left column shows the target poles while the right column shows the target magnetizations. In all four plots, the parameter estimates corresponding to the non-UXO targets (H_0) are plotted using blue circles, while those corresponding to the UXO targets (H_1) are plotted with red squares. These results also indicate that the UXO parameter estimates (H_1, red squares) are more tightly clustered for the modified inversions than they are for the original inversion.

The estimated target poles obtained from the original and modified dipole model inversions are compared to the target pole estimates obtained using the simple weighted sum of decaying exponential signal model in Figure 60. These parameter estimates were obtained with the BOR assumption employed for the dipole model, which corresponds to a model order of 2 for the weighted sum of decaying exponentials signal model. No Spatial Model refers to the weighted sum of decaying exponentials signal model, New Full Spatial Model refers to the dipole model inverted with the modified inversion process, and Original Full Spatial Model refers to the dipole model inverted with the original inversion process. Recall that both the dipole model and the weighted sum of decaying exponentials signal model represent the measured signal as a weighted sum of decaying exponential signals; the fundamental difference between the two is the dipole model incorporates a model for the spatial variations in the weights on the decaying exponential signals, whereas the weighted sum of decaying exponentials signal model does not. The results for UXO targets (H_1) are shown on the left, and the results for non-UXO targets (H_0) are shown on the right. In both cases, the target poles estimated with the modified inversion process for the dipole model are more similar to those obtained with the simple weighted sum of decaying exponential signal model than are the estimates obtained with the original dipole inversion process. Since the poles (decay rates) in both models have been shown to be equivalent, they would be expected to be similar, though not identical due to the constraints imposed

by the model for the spatial variation in the amplitudes imposed by the dipole model. Thus, the target pole estimates obtained from the dipole model with the modified inversion are preferable to those obtained with the original inversion not only because they are more tightly clustered, as shown previously, but also because they are more similar to the poles obtained with the weighted sum of decaying exponentials signal model and, therefore, are more consistent with the theoretical analysis of the two signal models. In addition, in all but a few cases, the modified inversion process provided robust, reliable inversions without the need for human intervention or oversight.

Feature Selection and Classifier Design

The BUD data collected as part of the Camp Sibert discrimination study were re-inverted using the dipole model with the modified inversion process, as described above. Feature selection and classification results obtained utilizing features obtained using the modified inversion strategy are compared to those obtained with the original model inversion process.

Feature Selection

The features utilized for classification are the model parameters intrinsic to the target; M_n and ω_n for the dipole model and α_n for the weighted sum of decaying exponentials model. In addition, the ratios of the estimated parameters, the residual normalized by signal length (ε), and the residual normalized by the signal energy (ε_{Es}) are also included in the feature set. The residual is the sum of the squared differences between the measured data and the data predicted by the forward model using the inverted parameters. Since the number of samples in the measured data is not consistent and the signal energy, defined as the integral of the squared signal, varies widely, normalized quantities based on the residual were considered as features. The residual normalized by signal length is the residual divided by the number of samples in the measured data. Similarly, the residual normalized by the signal energy is the residual divided by the energy in the measured signal. Thus, for the dipole model, the feature set consists of $M_1, M_2, M_3, M_2/M_1, M_3/M_1, M_3/M_2, \omega_1, \omega_2, \omega_3, \omega_2/\omega_1, \omega_3/\omega_1, \omega_3/\omega_2, \varepsilon, \varepsilon_{Es}$. The full feature set for the features derived from the weighted sum of decaying exponentials signal model was constructed in a similar fashion. Features to be utilized for classification were selected using a parallel-sequential (ParSe) feature selection algorithm, with the area under the ROC curve as the performance metric to be optimized.

Classifier

The classifier utilized for target classification as UXO or non-UXO/clutter was a relevance vector machine (RVM) with a radial basis function (RBF) kernel (kernel parameter = 0.75). The classifier parameters have not yet been optimized for the model parameters generated by the modified inversion process. All performance results have been generated using leave-one-out (LOO) cross-validation.

LBL BUD AEM Sensor Camp Sibert Discrimination Study Results

Measurement Selection

The data collected with the BUD sensor as part of the Camp Sibert discrimination study were inverted and the parameter estimates obtained via the modified dipole model inversion process analyzed to assess their utility for target classification. Many anomalies in this data collection had multiple measurements taken at different locations relative to the anomaly location. LBL analyzed the multiple measurements and shared with us the measurements that they believed were the best suited for processing. In order to maintain consistency and to enable comparison to the LBL results, the

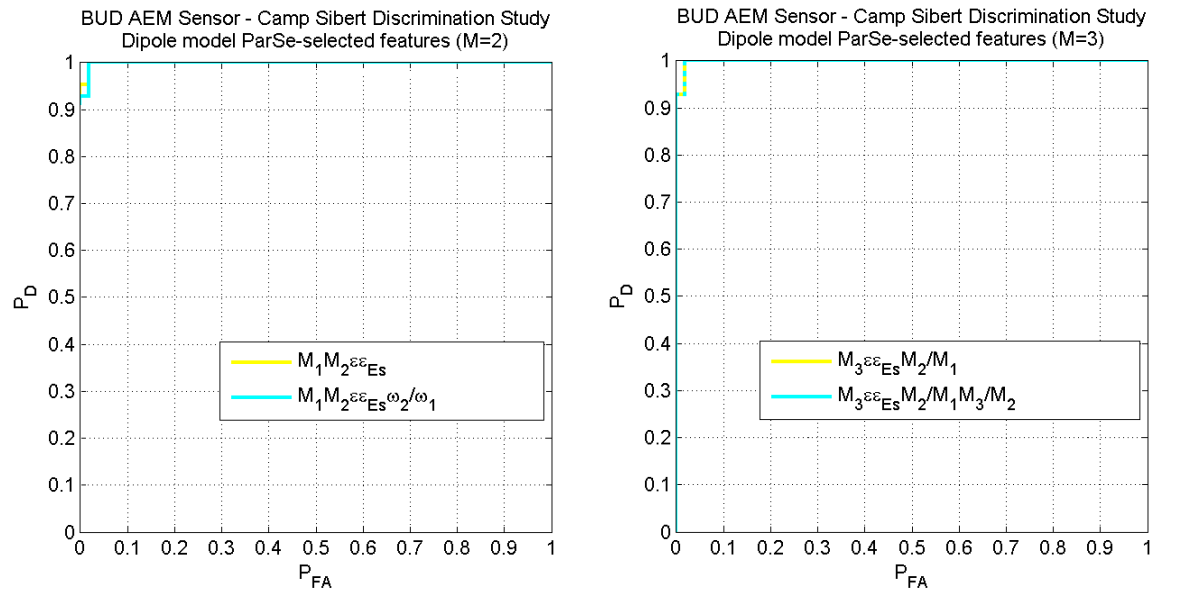

Figure 61: Classification performance for features obtained using the modified dipole model inversion with the BOR assumption (left) and without the BOR assumption (right) on the Camp Sibert discrimination study training data collected with the BUD sensor.

measurements selected by LBL were analyzed first. The other measurements are currently under consideration as part of a subsequent analysis.

Training Data Results

The training data collected was utilized to select features for classification and train the classification algorithm, an RVM. The feature selection process resulted in four feature sets with identical AUC performance; all four feature sets had just a single false alarm at 100% detection. Two of the feature sets were obtained with the BOR assumption employed, and the remaining two feature sets were obtained when the BOR assumption was relaxed. The ROCs corresponding to the feature sets identified through parallel-sequential (ParSe) feature selection are shown in Figure 61. The two feature sets which result in only a single false alarm at P_D=1 when the BOR assumption is included in the dipole model (left subplot) are $\{M_1, M_2\varepsilon, \varepsilon_{Es}\}$ (yellow curve) and $\{M_1, M_2\,\omega_2/\omega_1, \varepsilon, \varepsilon_{Es}\}$ (cyan curve). The two feature sets which result in only a single false alarm at P_D=1 when the BOR assumption is not included in the dipole model (right subplot) are

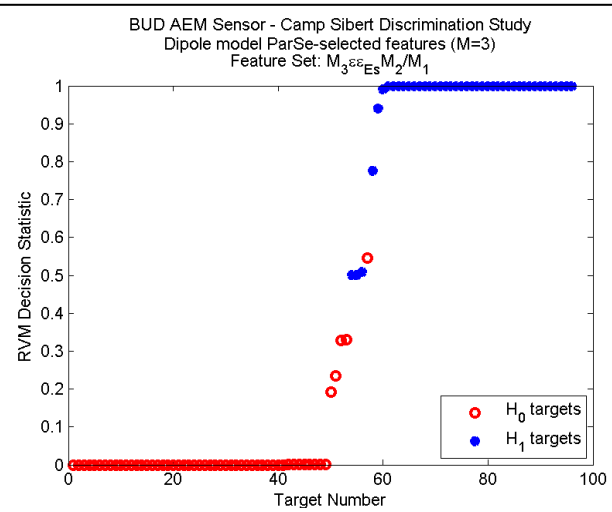

Figure 62: Decision statistic distribution for feature set $\{M_3, M_2/M_1, \varepsilon, \varepsilon_{Es}\}$ obtained using the dipole model without the BOR assumption on the Camp Sibert discrimination study training data collected with the BUD sensor.

$\{M_3, M_2/M_1, \varepsilon, \varepsilon_{Es}\}$ (yellow curve) and $\{M_3, M_2/M_1, M_3/M_2, \varepsilon, \varepsilon_{Es}\}$ (cyan curve).

In addition to the excellent discrimination performance shown in the ROCs, the decision statistics obtained with the RVM operating on these features are quite separable; there is a large gap between the non-UXO are UXO targets, with only a few targets (typically about 4 to 6 targets) in a small region where the decision statistics overlap. One example of this is shown in Figure 62. In summary, results obtained using the modified inversion process on Camp Sibert training data are quite promising.

Blind Test Data Results

Further analysis of the decision statistics revealed that of the four features sets that resulted in only a single false alarm at P_D=1, the feature set consisting of 4 features provided by inversion of the dipole model without the BOR assumption $\{M_3, M_2/M_1, \varepsilon, \varepsilon_{Es}\}$ yielded the best class separation (non-UXO vs. UXO), and so that feature set was selected for evaluation on the blind test set This feature set provided very good separation between non-UXO (H_0) and UXO (H_1) targets, with only 4 targets in a small overlapping region in the middle.

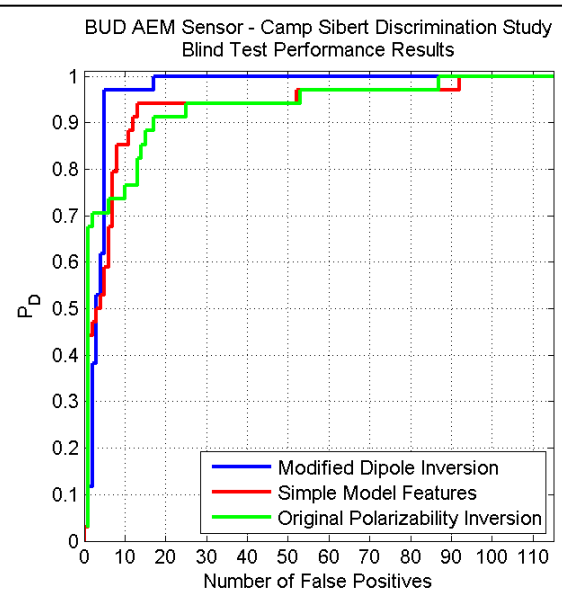

Figure 63: Camp Sibert discrimination study blind performance results for data collected with the BUD sensor, as scored and reported by IDA.

Discrimination performance results, as scored and reported by the Institute for Defense Analyses (IDA), for this feature set found via the modified inversion process are shown below in Figure 63. All but one UXO target are detected with 5 false alarms, and all UXO targets are detected with 17 false alarms. These results show that performance using features obtained with the modified dipole inversion process is quite good, and compares favorably both to performance obtained using features found using the simple weighted sum of decaying exponentials signal model and performance using features found by the original polarizability inversion, also scored and reported by IDA. Both of these performance curves are also shown in Figure 63. The decision statistics for features derived from the blind data using the original dipole model inversion were not submitted for blind scoring due to the relatively poor performance on the training data compared to the weighted sum of decaying exponentials model features. Reevaluation of the inversion process for the polarizability model has not yet been considered. These blind performance results are consistent with the discrimination performance on the training data; discrimination performance using features found by the modified inversion of the dipole model is better than performance using features found by the simple weighted sum of decaying exponentials model, and furthermore, they are quite promising.

EMI Dipole Model Inversion Modifications Summary and Discussion

The modifications made to the inversion process have resulted in substantial improvements in both the inversion performance (parameter estimates) and UXO classification performance, as demonstrated by the performance on both the Camp Sibert training data and the Camp Sibert blind test data collected with the BUD sensor and scored and reported by IDA. The three modifications considered are a

modification to the geometric aspects of the propagation model, a re-parameterization of the dipole model for inversion, and the introduction of a sequential inversions process. In all but a few cases, the modified inversion processes provided robust, reliable inversions without the need for human intervention or oversight. The effects of each modification individually have not yet been investigated, however, so it cannot yet be stated conclusively which modification, or set of modifications, is driving the changes in inversion performance.

The characteristics of the four training targets whose decision statistics are in the ambiguous region have been investigated. One of the four targets is a non-UXO target that happens to possess feature values that are similar to the UXO feature values. The other three targets are UXO targets for which the inversion diverged. One of these three UXO targets has been evaluated in detail. The analysis revealed that the measurement selected by LBL produced a divergent inversion, but one of the other measurements which was rejected by LBL provided a good inversion with parameter estimates consistent with other UXO targets. This analysis suggests that the measurements selected by LBL as the best measurements for processing may not uniformly be the best measurements for the processing approach taken here. The subject of how to best select measurements for this approach to UXO classification remains the subject of investigation.

Overlapping Signatures

Significant improvements in unexploded ordnance (UXO) discrimination performance have been realized through the development of statistically-based discrimination algorithms. However, many of these gains have been achieved in tests at sites seeded with UXO and clutter that ignored some of the challenges faced in actual clean up scenarios, such as the presence of native clutter and natural UXO placement, which will deleteriously impact performance if not addressed. One particular problem that is usually not considered is posed by data that consists of overlapping signatures that result from adjacent anomalies. Many UXO remediation strategies use phenomenological models to generate features for discrimination; these models are typically representative of a single anomaly and often fail to produce consistent features in highly cluttered scenarios. Previous approaches to address overlapping signatures have utilized independent component analysis to first separate the measured data into signals corresponding to individual anomalies which were then processed independently. However, the separated signals recovered using independent component analysis do not necessarily resemble the signals produced by the anomalies individually, which prevents subsequent analysis using phenomenological models to generate

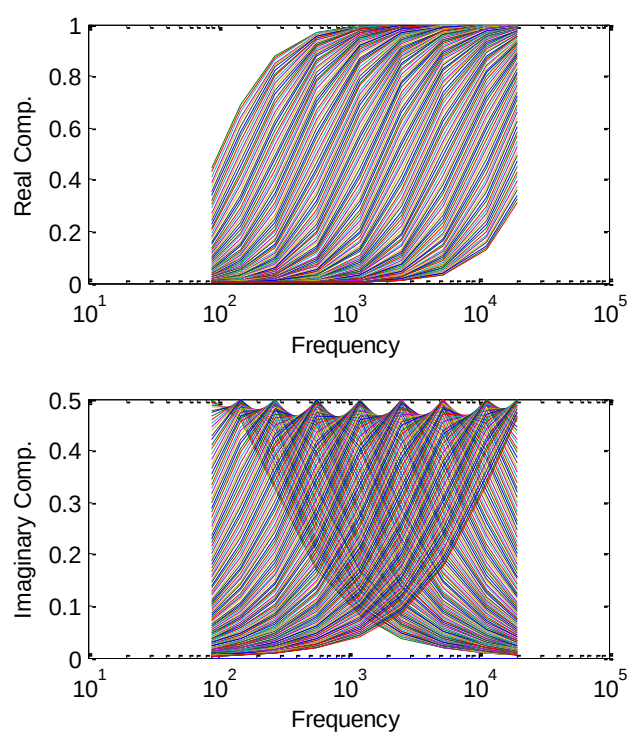

Figure 64. Matching pursuits dictionary elements.

physically-relevant features for discrimination. We are investigating matching pursuits as an alternative approach to decomposing the measured data, motivated by its previous use to generate relevant decompositions and physics-based features in other scattering problems.

The approach adopted is based on standard matching pursuits decomposition with back projection. Matching pursuits is a method for decomposing a waveform into a linear combination of dictionary elements *D** selected from the full dictionary *D* through a greedy, iterative process of finding the best fitting dictionary term and calculating the residual waveform. The back-projection routine updates the weights on the terms in *D** after each iteration to further minimize the residual waveform; thus, while the selection of dictionary terms is greedy, the weights on the selected dictionary terms are jointly optimized. In this work, an application-specific dictionary, shown in Figure 64, was constructed for the matching pursuits decomposition of data measured with a frequency-domain electromagnetic induction (EMI) sensor. Each dictionary term, calculated using Equation 1, corresponded to a single frequency-domain dipole mode in the Carin et al. dipole model [**26**] with a single parameter Ω, equivalent to the resonant frequency of the dipole mode. The dictionary terms were evaluated at the nine measurement frequencies of the EMI sensor, ranging from 90 Hz to 20010 Hz, and separated into real and imaginary components for a total length of eighteen elements. The dictionary contained 150 entries: the first entry had a resonant frequency $\Omega = 0$, resulting in a constant real response at all frequencies, and the Ω values for the remaining dictionary elements were logarithmically spaced over the frequency range of the EMI sensor, 90 Hz to 20010 Hz.

$$d = \frac{\omega}{\omega - j\Omega} \tag{1}$$

The data available for identifying each anomaly as UXO or non-UXO is measured spatially, producing an 18 by *N* matrix with dimensions corresponding to measurement frequency and location, respectively. The measured data at each spatial location is a function of the target/sensor orientation. The dictionary terms, however, are not parameterized to account for the position and orientation of the subsurface objects relative to the sensor. This dictionary construction allows for a simplified dictionary element and significantly reduces the size of the dictionary but requires some modification in order to utilize the set of spatial measurements. The distance between the subsurface anomaly and the sensor location will affect only the necessary weighting coefficient on the dictionary term. Thus, if a selected dictionary term is allowed to have a different weighting for each spatial measurement it should be capable of appropriately decomposing the data. This modification to the matching pursuits procedure resulted in a set of weights, A_n, for each selected dictionary term with n = 1 to N. In each iteration of the decomposition, the dictionary term in *D* with the maximum corresponding value of α is added to *D** where α is equal to the sum of $|A|$; the residual waveform was then calculated using the individual weights A_n for each of the terms in *D**. The weights A_n for each term in *D** were updated using the back-projection routine. Thus, all of the spatial data could be utilized in the selection of the elements of *D** without significant expansion of the dictionary or dictionary terms. The outputs of the matching pursuits decomposition were the Ω values for the terms in *D** and the corresponding values of the separated signature.

The primary interest in this study was the applicability of the matching pursuits approach to the scenario of overlapping signatures. Measurements from isolated anomalies were considered as well to determine a baseline level of performance for features generated from this approach. The matching pursuits algorithm was evaluated using field data collected from Camp Sibert with the GEM-3 frequency-domain EMI sensor. A set of overlapping signatures were generated from the set of isolated signatures

by simulating the placement of two clutter objects in close proximity to the original, isolated object (at distances of 0.3 m and 0.6 m) and interpolating data at the measurement locations of the original object. Scatter plots of the first two principle components of the matrices of distances between sets of matching pursuit parameters for the isolated anomalies and the overlapping anomalies are shown in Figure 65. These results demonstrate consistency in the feature space for the UXO targets despite the presence of other anomalies in the measured response.

The isolated and overlapping signatures were fit using six dictionary terms in the matching pursuits decomposition, an equivalent number of intrinsic model parameters as in the previously mentioned dipole model based on the singularity expansion method (SEM). Matching pursuits representations of

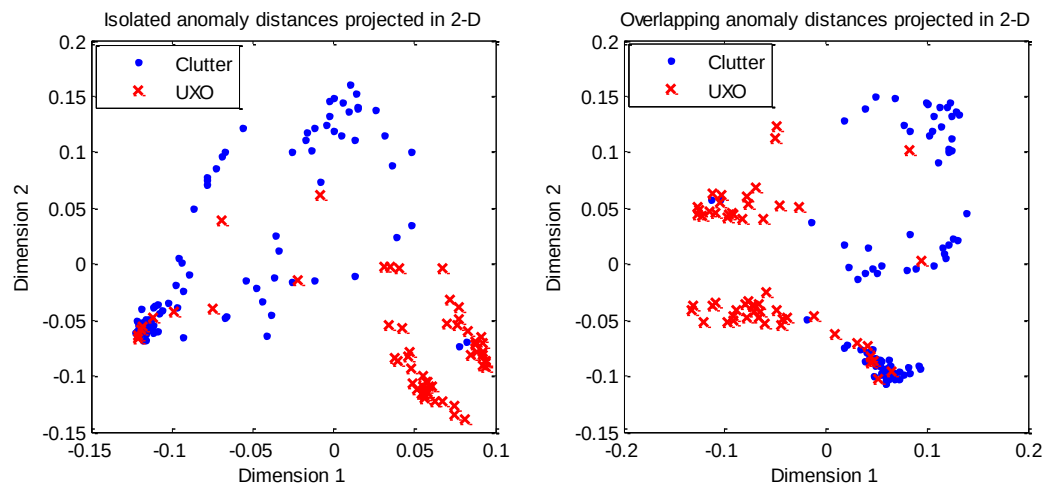

Figure 65. First two principal components of the matrices of distances between sets of matching pursuit parameters. Note: the reference set to which the distances for the interpolated overlapping signatures were calculated contained only isolated anomalies, to indicate consistency in feature space despite the presence of other anomalies in the measured response

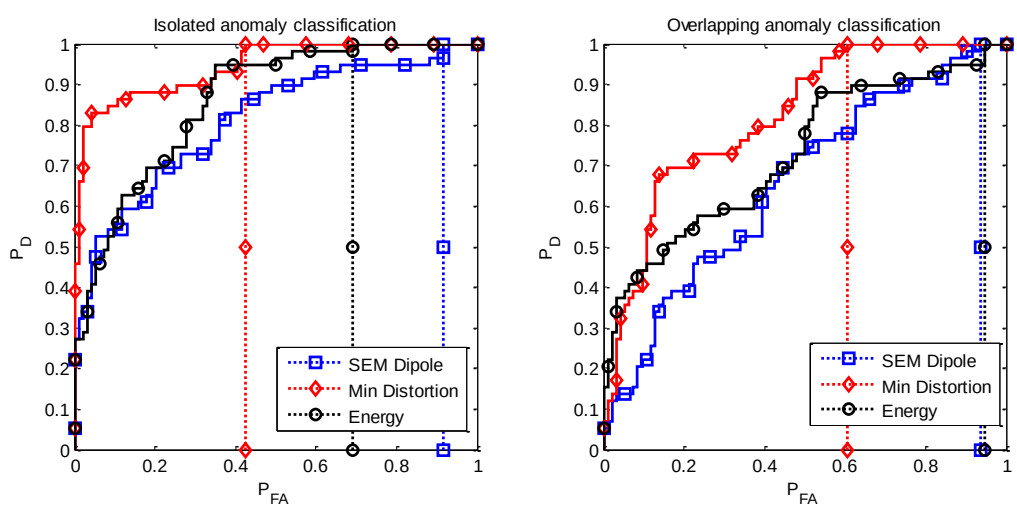

Figure 66. Classification performance for (left) isolated anomalies and (right) interpolated overlapping signatures from multiple isolated anomalies. In both scenarios, the set of training data contained only isolated anomaly signatures. The vertical dashed lines indicate the probability of false alarm (P_{FA}) at which all UXO are detected.

data generated using the forward SEM dipole model had an average signal-to-noise ratio of +80 dB. The matching pursuits approach and the SEM dipole model are being compared in terms of model fit and consistency of parameters across similar types of objects in addition to discrimination performance on the isolated and overlapping signatures target sets, the overarching goal of the proposed approach.

Preliminary results, shown in Figure 66, suggest that the matching pursuits method may yield a promising approach to UXO discrimination and robust performance in the case of overlapping signatures. Most notably, P_{FA} at P_D=1 is considerably improved for the case of overlapping anomaly classification. The matching pursuits approach provides a stable and robust method for estimating a small set of physically-relevant parameters that are capable of producing the same data fits as the standard dipole model. Other mechanisms for fully utilizing the capabilities of the matching pursuits approach to provide discriminative features for UXO classification will also be considered.

We continued our examination of matching pursuits decomposition to generate physically-relevant features, specifically to address overlapping signatures. A test stand data set of overlapping signatures from 2004, containing a variety of combinations of UXO and clutter objects has been processed using matching pursuits decomposition. For comparison, we considered two other modeling schemes for feature generation using a dipole model based on the singularity expansion method: one scheme used a single dipole model; the second used a pair of dipole models inverted simultaneously. For this data set, the matching pursuits analysis was also extended to the time-domain EM63 EMI sensor.

The matching pursuits representations of anomalies consist of an unordered set of paired features, with each pair containing a dictionary term parameter and a dictionary term weight. The matching pursuit representations for different anomalies were compared by calculating the distance between the features using a measure appropriate for unordered sets of points. A two-dimensional representation of the distances between all of the anomalies in the test stand data set was constructed using multidimensional-scaling analysis. In two-dimensions, the matching pursuits representations did not separate into clusters that indicated the presence or absence of a UXO as desired. The EM63 representations were marginally more consistent when a UXO was present than the GEM-3 measurements decompositions.

Overlapping Signatures Summary

Matching pursuits has been proposed to address the problem of discriminating UXO from clutter when the measured data consist of multiple overlapping anomaly signatures. This approach decomposes the data onto a set of pre-defined dictionary terms, subject to the constraint that the selected dictionary terms must be the same for all spatial sensor measurements. Preliminary discrimination performance results indicate that matching pursuits may hold promise for UXO discrimination in the case of overlapping signatures. In particular, P_{FA} was significantly reduced for very high P_D.

Active Learning and Kernel-Based Algorithms

We have also developed new semi-supervised classification tools, with these applicable to next-generation EMI systems. Specifically, we utilize a general kernel, to quantify the similarity between the features associated with any two targets, for labeled and unlabeled examples. This kernel is used to constitute an NxN matrix, for N items, with most of our work focused on the radial basis function (RBF). The RBF is a general kernel, applicable to any feature vector, including those associated with the aforementioned BUDS systems. The matrix is appropriately normalized, such that the rows constitute probability mass functions. The jth column of row i corresponds to the probability of "walking" from the ith item to the jth in one step of a random walk, where the random walk is constituted on the data

manifold. This random walk framework captures the properties of the data manifold, providing important information about the inter-relationships of all data, labeled and unlabeled. The random-walk probabilities are then utilized within the context of a logistic-regression model, to constitute a semi-supervised classifier. The algorithm has been successfully tested using EMI and magnetometer data collected at the Badlands. Performance has been compared against other state-of-the-art semi-supervised algorithms, wherein the new algorithm has been demonstrated to yield superior performance. This algorithm will next be tested with the data from the BUDS system, after the mentioned feature extraction is completed.

An important challenge in EMI sensing involves the need for training data. Duke has previously developed an active-learning algorithm, which performs in situ learning, by adaptively determining which items should be excavated for the purpose of learning. This procedure provides adaptive learning, but it does not exploit information from previous experience at other sites. This latter challenge has been addressed by using an algorithm called concept drift, which learns the relationship between data at a previous site and the current site under test. While this algorithm is very useful, and has been demonstrated to add value for UXO sensing, it does not allow one to exploit information from multiple previous sites. The challenge in this latter problem is determination of which previous sites considered are relevant to the current site under test, and which are not. This has motivated development of a new algorithm, which we term "life-long learning" (L3). The idea is that over time, after more sites have been interrogated, the algorithm should get smarter, and should require less training data, if it can learn relationships between the current site and all previous sites investigated previously. The L3 algorithm is based upon a fully Bayesian formalism, with the relationships between different sites learned through use of a Dirichlet process (DP) statistical prior. The Bayesian inference is solved efficiently via a variational Bayes (VB) formalism, which yields computational efficiency comparable to traditional EM-based maximum-likelihood solutions, while yielding a full posterior on the model parameters. The algorithm has been tested and validated on several data sets, including those associated with EMI systems applied to UXO.

Duke also developed a semi-supervised classification algorithm based on a random walk on a graph. This algorithm extends previous Duke semi-supervised studies in two ways. Previously Duke used the graph to define a statistical prior on the parameters of a classifier. In the new work, rather than incorporating the graph within the prior, it is employed directly in the parametric classifier itself. This stronger imposition of classifier smoothness on the data manifold has yielded significantly improved classification performance, relative to state-of-the-art classifiers in the literature. Secondly, the form of the new semi-supervised model is well suited to multi-task learning and concept drift. This is because the new algorithm doesn't utilize the prior to represent the graph, and therefore the form of the prior may be utilized for other purposes, specifically in the context of concept drift, via a Dirichlet process prior. The algorithm has been successfully tested on Badlands data, and will be tested on next-generation sensors as more-extensive databases come on line.

Semi-Supervised Active Learning

Classification using EMI and magnetometer sensors is typically not performed directly on the measured data, but on features extracted therefrom. Specifically, parametric models have been developed for the response of targets as viewed from such sensors, with most of these models based on a dipole approximation[**27**]. The parameters extracted from the models, when fitting is performed to the measured data, are typically employed to constitute feature vectors within the subsequent classification algorithm. Most of these algorithms are supervised, in the following sense. A set of labeled feature vectors are assumed given (the identity, UXO/non-UXO, of each feature vector is known), and these data

are used to design a classifier. Numerous such classifiers have been considered for UXO detection, such as kernel matching pursuits [**28**], support vector machines [**27**], and likelihood-ratio tests [**27**]. There are two limitations of such approaches: (i) the assumption of the presence of an *appropriate* labeled data set is tenuous in many cases, and (ii) even when such labeled data are available, a purely supervised algorithm doesn't exploit the contextual information provided by the unlabeled data.

General interest in these latter two issues has motivated recent research in the machine learning community. Specifically, active learning [**29**,**30**] is a framework whereby the acquisition of labeled data is integrated within classifier design. Using appropriate information-theoretic measures, an active-learning algorithm asks which of the unlabeled feature vectors would be most informative for classifier design if the associated labels could be made available. This idea has been applied previously in the context of UXO detection [**28**]. The new aspect of the work considered here is that this active-learning framework is placed within the context of a semi-supervised learning setting. Specifically, in addition to actively acquiring the labeled data (performing item excavations selectively for the purpose of algorithm learning), a semi-supervised algorithm exploits contextual information provided by all of the unlabeled data (the classification of any one unlabeled feature vector is placed within the context of all unlabeled feature vectors). In the UXO problem the EMI/magnetometer data are often all collected at once, typically using a cart-based system [**31**]; recall that the UXO of interest are all buried, and therefore they are not dangerous until excavation begins. Therefore, one may perform feature extraction on all of the signatures at once, and the contextual information provided by these data may be of utility in improving classification performance.

Semi-supervised learning has been an area of significant recent interest in the machine-learning community [**32**,**33**,**34**,**35**,**36**,**37**,**38**,**39**], where exploitation of the information available in the unlabeled data has been demonstrated to often add value. To our knowledge this work represents the first use of such an approach as applied to the UXO problem. The semi-supervised approach presented here was first proposed in a general setting, within the machine-learning community [**40**], wherein it was demonstrated to yield superior performance relative to other semi-supervised algorithms, based on canonical machine-learning data sets. In this work this algorithm is applied to actual UXO data, compared to other supervised and semi-supervised approaches, and extended to an active-learning setting. An important goal of this work is to introduce semi-supervised learning to the geophysical community, since it is likely to have general utility for associated sensing problems, and typically semi-supervised approaches yield performance that is significantly better than the widely used supervised approaches.

To date, there have been several semi-supervised methods developed. The generative-model method, an early semi-supervised method, estimates the joint probability of data and labels via expectation-maximization (EM), treating the missing labels of unlabeled data as hidden variables; this method was studied in statistics for mixture estimation[**41**] and has been reformulated for semi-supervised classification[**39**]. Co-training [**37**], another early method, exploits two independent subvectors of features, using one to provide the label estimates for the other; co-training has received renewed interest recently, particularly theoretically. The semi-supervised support vector machine (SVM) [**36**] represents a more recent method, which maximizes the margin between classes, taking into account both labeled and unlabeled data. Graph-based methods [**33**,**34**,**35**,**38**], the main focus of current research in semi-supervised learning, exploits the assumption that strongly connected data points (in feature space) should share the same label, and utilizes spectral graph theory to quantify the between-data connectivity. For a more complete review of the literature, see [**32**].

Most graph-based algorithms operate in a transductive fashion, *i.e.*, they directly learn the labels of the unlabeled data, instead of learning a classifier first and then using the classifier to infer the unseen labels (inductive learning). While transductive algorithms avoid the problem of model selection for a classifier, they lack a principled way of predicting the labels of data out of the training set. The work in [**33**] addresses this problem by constructing a graph-based prior distribution on the parameters of a classifier and learns the classifier by maximizing the posterior (MAP estimation); the prior utilizes both labeled and unlabeled data, thus enforcing semi-supervised learning. Several drawbacks are inherent in the algorithm in [**33**]. For example, the hyper-parameter balancing the importance of the prior relative to the data likelihood needs to be learned.

In this work, with a focus on the UXO-sensing application, we apply an algorithm for learning parametric classifiers on a partially labeled data manifold [**40**], by representing the manifold as a graph; each vertex on the graph represents a data point and the weighted edge between two vertices manifests the immediate connectivity between the corresponding data points. This algorithm is motivated by the work in [**35**] and builds the *t*-step connectivity between data points via a Markov random walk on a manifold. To account for heterogeneities in the data manifold, the random walk takes different step-sizes at different data locations; each step-size dictates a Markov transition matrix and we select the step-size to assemble the transition matrix for the entire manifold.

In the discussion below we present the semi-supervised learning algorithm summarized above as applied to actual UXO data. The algorithm is also considered in an active-learning setting. While it is shown that the different semi-supervised algorithms considered here yield comparable performance on the UXO data considered, and therefore the relative utility of one semi-supervised approach over another is not pronounced, we do observe a marked improvement of all of the semi-supervised approaches relative to conventional supervised algorithms.

To evaluate the proposed algorithm, we applied it to a UXO data set from an actual former bombing range. This data set was collected by the Multi-sensor Towed Array Detection System (MTADS) [**31**]. This system is composed of arrays of full-field cesium vapor magnetometers and time-domain electromagnetic pulsed induction sensors. The magnetometers were Geometrics Model 822ROV, while the EMI sensors were highly-modified Geonics EM-61 sensors. The data were collected at a bombing target on the Badlands Bombing Range on the Ogala Sioux Reservation in Pine Ridge, South Dakota. The UXO items present at the site included M 38 (100 lb.) sand-filled practice bombs, M 57 (250 lb.) practice bombs, 2.25 in. and 2.75 in. rocket bodies and rocket warheads, and ordnance scrap (such as tail fins and casing parts). The details of how these measured data are analyzed with magnetometer and EMI dipole models has been described in detail elsewhere [**27**], and these same techniques were applied to extract feature vectors from the data considered here.

The data associated with a given item under test was manifested in one of three variations: (i) only magnetometer data were available, (ii) only EMI data were available, or (iii) both magnetometer and EMI data were available. These different variations were tied to the details of the data collection, and to the quality of the data acquired for each of the two modalities. For EMI sensor data alone, there are 230 clutter cases (non-UXOs) and 44 UXOs. For the magnetometer sensor data alone, there are 719 non-UXOs and 79 UXOs. Concerning the case for which data from both the magnetometer and EMI sensors are available, there are 228 Non-UXOs and 44 UXOs. For the EMI data the feature vector is of dimension 10, for the magnetometer the feature vector is of length 9, and when both are used the two types of features are concatenated. Before processing each feature is centered and normalized. Specifically, we compute the mean and variance for each dimension of the features; each feature is

shifted by subtracting its mean and then divided by its variance. The feature vectors from this data set are available to other researchers, upon request to the authors.

In semi-supervised learning, as discussed above, there are two frequently applied settings. In a transductive algorithm [**35**] it is assumed that all of the labeled and unlabeled data are available simultaneously, and the algorithm is designed to classify the unlabeled data, employing the data-manifold information provided by both the labeled and unlabeled data. Importantly, if a new unlabeled example was added, then the whole transductive learning process would have to begin anew. In an inductive semi-supervised learning algorithm [**33**] one again has both labeled and unlabeled data with which an algorithm is designed, exploiting the data manifold. Once this algorithm is designed, it may be applied to the existing unlabeled data, as well as to new unlabeled data, without having to redo the learning process.

In many UXO-sensing settings all of the data are collected at once, and therefore a transductive semi-supervised learning algorithm may be sufficient. However, if data is collected incrementally on a large UXO-cleanup site, the inductive framework may be attractive. The semi-supervised algorithm developed here is inductive, but clearly it may be applied in a transductive setting as a special case. However, there are existing semi-supervised algorithms of interest that are only transductive, the algorithm of Szummer and Jaakkola [**35**] being an important example.

We compare our results with performance achieved using [**35**]. We also make a comparison to results computed using a logistic-regression classifier, where the graph considered here was as a prior to regularize the learning process (imposing smoothness of the classifier along the data manifold [**33**]). Like our algorithm, the approach in [**33**] may operate in an inductive setting. However, such that the comparisons are fair, for all examples considered in this section, the unlabeled data on which classification is performed is the same unlabeled data used for semi-supervised algorithm learning (consistent with the requirements of a transductive algorithm). The performance is evaluated in terms of classification accuracy, defined as the ratio of the number of correctly classified UXOs and non-UXOs over the total number of data being used. For this, a threshold 0.5 is used to the classification probability. In the discussion that follows the algorithms considered are referred to as follows: (i) the method in [**35**] is denoted RW-Transductive; (ii) the method developed in this work is termed RW-Inductive; (iii) the method in [**33**] is termed Logistic-GRF (for Gaussian random field prior); and (iv) the supervised solution is termed Logistic-Regression, with this equivalent to the algorithm in [**33**] without

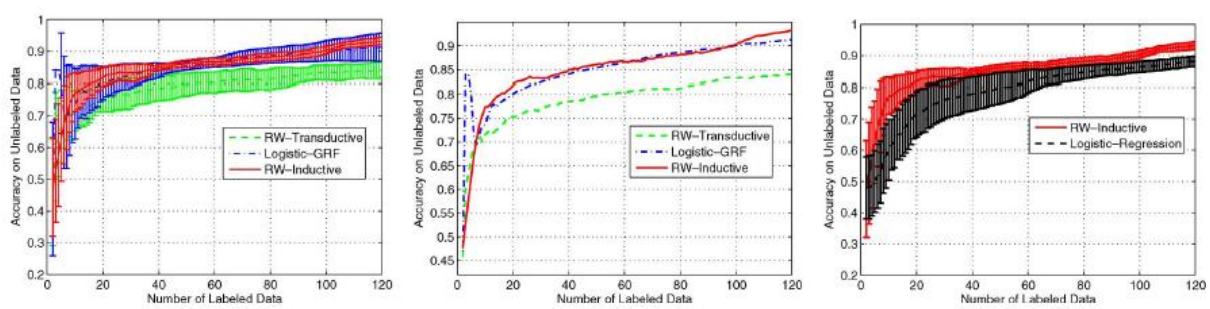

Figure 67: Active-learning results for EMI data. The first two labeled data include one UXO and one non-UXO (both randomly sampled). Each curve is an average from 50 independent trials. The horizontal axis is the size of the labeled data set (actively selected). The algorithms are tested on the unlabeled data set. Where shown, error bars are standard deviations. (Left) Average results of three semisupervised algorithms, with error bars. (Center) Average results of three semisupervised algorithms, without error bars. (Right) Comparison of the semisupervised algorithm proposed here, i.e., RW-Inductive, with a supervised logistic-regression method [5].

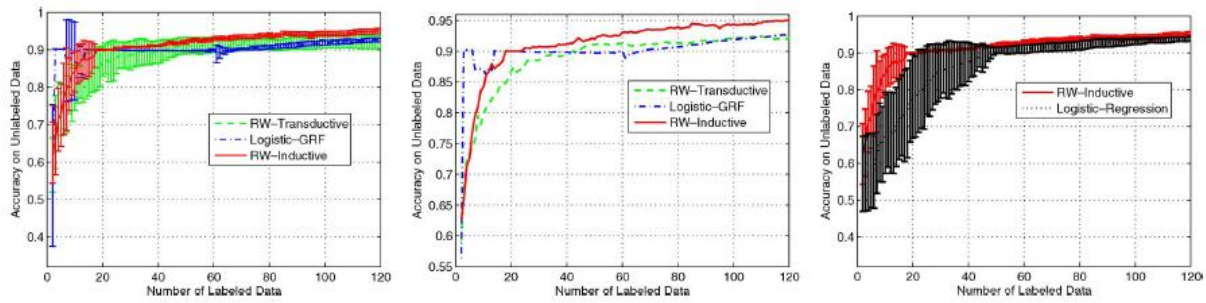

Figure 68: Active-learning results for magnetometer data. The first two labeled data include one UXO and one non-UXO (both randomly sampled). Each curve is an average from 50 independent trials. The horizontal axis is the size of the labeled data set (actively selected). The algorithms are tested on the unlabeled data set. Where shown, error bars are standard deviations. (Left) Average results of three semisupervised algorithms, with error bars. (Center) Average results of three semisupervised algorithms, without error bars. (Right) Comparison of the semisupervised algorithm proposed here, i.e., RW-Inductive, with a supervised logistic-regression method [5].

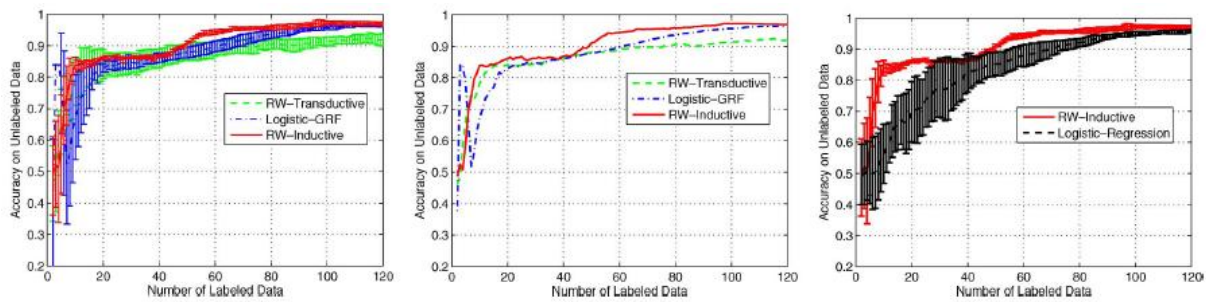

Figure 69: Active-learning results for integrated magnetometer and EMI data. The first two labeled data include one UXO and one non-UXO (both randomly sampled). Each curve is an average from 50 independent trials. The horizontal axis is the size of the labeled data set (actively selected). The algorithms are tested on the unlabeled data set. Where shown, error bars are standard deviations. (Left) Average results of three semisupervised algorithms, with error bars. (Center) Average results of three semisupervised algorithms, without error bars. (Right) Comparison of the semisupervised algorithm proposed here, i.e., RW-Inductive, with a supervised logistic-regression method [5].

the graphical prior. To ensure a fair comparison, we apply the same Markov random walk graph for both RW-Transductive and RW-Inductive.

In Figure 67 through Figure 69 we consider active learning for the three data sets; we first randomly select one UXO and one non-UXO feature vector, and the other labeled data are selected by the active learning algorithm; to design the classifier we require at least one feature vector from each class, but after active learning proceeds sufficiently the large number of labeled examples determined adaptively typically dominate the two labeled examples with which we commence. In these examples we also consider active learning using a supervised classifier, as in [**28**], and it is demonstrated that the semi-supervised algorithm proposed here provides improved performance, particularly for the fusion of the EMI and magnetometer data. In addition to the improved average performance, the results of the semi-supervised RW-Inductive algorithm, when performing active learning, generally provides tighter variation than the supervised approach in [**28**]. This phenomenon is attributed to the fact that the semi-supervised algorithms exploit the relatively large quantity of unlabeled data, which tends to stabilize the solution (make it less sensitive to the particular samples that are labeled).

Although all three semi-supervised active-learning algorithms perform well, the RW-Inductive results appear to be best on average, particularly after a relatively large number of labeled examples are acquired. Although active learning applied to Logistic-GRF [**33**] does perform better on average for a small quantity of actively acquired labeled data, the error bars are very large in this regime. Note that for a relatively small number of excavations for label acquisition, the magnetometer results are superior but, encouragingly, as the number of labels acquired extends to 120, the fusion of the magnetometer and EMI data yields slightly improved performance relative to either sensor alone.

Semi-Supervised Active Learning Summary

In this work we have considered the use of semi-supervised learning in the context of UXO detection, based on electromagnetic induction (EMI) and magnetometer data. The algorithms were applied to features extracted from these data, with the features linked to EMI and magnetometer dipole-based parametric models. Semi-supervised learning is particularly well suited to the UXO-sensing problem, because one typically deploys a cart-based system to collect all EMI and magnetometer data at once, for an entire site. Hence, one may perform feature extraction simultaneously on all buried items of interest, and the classification of any one feature vector may be placed within the context of all feature vectors. This contextual information yields information on the characteristics of the data manifold, which has proven useful to improve classification performance in many settings. By contrast, in traditional supervised learning the labeled data alone are employed to learn a classifier, and this classifier is employed one-by-one to each unlabeled example, in isolation, and consequently contextual information is not employed.

Semi-supervised algorithms typically impose the following condition: two feature vectors that are ``close" in feature space should be classified similarly. This implies that the classifier outputs should vary smoothly over the high-density portion of the data manifold, and consequently that the decision boundary in feature space should reside in areas of low data density. These concepts may only be implemented if knowledge of the distribution of all unlabeled data is exploited when performing algorithm learning. The most advanced semi-supervised algorithms developed to date are based on graphical techniques. Specifically, the nodes on the graph correspond to the feature vectors (labeled and unlabeled), and the edge between any two feature vectors is defined by a distance between the two in feature space, where here this is defined by a radial basis function. There are many different ways in which the graph may be employed within a semi-supervised algorithm. For example, one m ay perform inference directly on the nodes of the graph, thereby inferring labels on the unlabeled nodes. This approach does not generalize to the classification of a general (new) feature vector that is not on the original graph, and therefore if new unlabeled data are acquired, the graph must be reconstituted and learning performed anew. This has been referred to as transductive semi-supervised learning. By contrast, one may also use the graph to learn an ``inductive" semi-supervised algorithm, which may be applied to new unlabeled data without having to reconstitute the graph or relearn. In the work presented here we have developed a new inductive semi-supervised algorithm, which extends the transductive algorithm developed in [**35**]. We have also performed comparisons to another (distinct) inductive semi-supervised algorithm [**33**], as well as to supervised learning. We have demonstrated that for the measured UXO-sensing data considered here, from an actual UXO cleanup site, that the semi-supervised algorithms perform better than purely supervised learning, implying that there is value in the manifold information associated with UXO sensing, at least for the UXO site considered.

In the UXO-excavation problem, clearly there will be many items manually removed, and the cost of unnecessary excavation of non-UXO items often constitutes the principal cleanup cost. One may therefore ask whether initial excavation may be performed with the purpose of learning. This is termed

active learning, and is characterized by asking in an information-theoretic sense which unlabeled feature vectors would be most informative for improving the classifier if the associated label could be acquired (here implemented via targeted excavation). In this sense the algorithm learns adaptively, directly on the site under test. In the examples considered here active learning yielded substantial improvement in UXO-classification performance, relative to selecting the labeled data randomly. One limitation of the active-learning framework, as implemented, is that to commence one needs at least one UXO and one non-UXO labeled example. In practice this is often not a significant limitation, because one typically knows the type of UXO that may be encountered at a given site (from historical information, and also from the items observed on the surface), and an archive of existing labeled UXO data may be used for this target class. Further, since at a typical site the quantity of non-UXO is much larger than the number of UXO, almost any initial excavations will yield at least one non-UXO signature. In the results presented here we examined the sensitivity of the algorithm to the initial UXO and non-UXO labeled exemplars, and found the algorithm to be robust in practice.

For the UXO-sensing data considered, we observed a substantial gain in the performance of the semi-supervised algorithm developed here relative to a corresponding supervised-learning algorithm. However, for the semi-supervised algorithm, the performance of learning using active-learning-determined labeled data was only slightly better to learning with randomly selected labeled data (the latter still using semi-supervised learning). This is a phenomenon we have observed on several different data sets: Since the semi-supervised algorithm exploits the information in the entire data manifold, using labeled and unlabeled data, we have found in practice that it is less sensitive to exactly which labeled data are considered; by contrast, when employing supervised learning the particular labeled data considered is often of significant importance [**28**].

The most significant direction for future research involves appropriate design of the graph for UXO applications. The weights on the graph edges are adapted to the characteristics of the manifold, via the data-dependent variance in the measure of the similarity between data points. However, in the analysis that followed a t-step walk on the graph was employed, where in the examples considered here t=50. The size of t plays an important role in defining what it means for two feature vectors to be ``close" in feature space. It is of interest to develop a principled means of defining an appropriate t for a given data set. We note that the use of t=50 was not carefully tuned for the data considered here, and many similar values ($20 < t < 80$) yielded similar results on the Badlands UXO data. We also note that the need to develop a technique for selecting t is not unique to the RW-Inductive algorithm introduced here, but is of interest for any of the graph-based semi-supervised algorithms.

Improved Concept Drift

In supervised classification problems, the goal is to design a classifier using the training examples (labeled data) such that the classifier predicts the labels correctly for unlabeled test data. The accuracy of the predictions is significantly affected by the quality of the training examples, which are assumed to contain essential information about the test instances for which predictions are desired. A common assumption utilized by learning algorithms is that the training examples and the test instances are drawn from the *same* source distribution.

As a practical example, consider the detection of a concealed entity based on sensor data collected in a non-invasive manner. This problem is of relevance in several practical problems, for example in the medical imaging of potential tumors or other hidden anomalies. In the context of remote sensing, one is often challenged with the problem of detecting and characterizing a concealed (e.g., underground) target based on remotely collected sensor data. The application that will be considered explicitly here is

the detection of buried unexploded ordnance (UXO) [**27**]. Sensors used for detecting and characterizing UXO include magnetometers and electromagnetic induction [**27**]. In designing an algorithm for characterization of anomalies detected by such sensors, to determine if a given buried item is UXO or clutter, one typically requires training data. Such training data typically comes from previously considered sites, and there is a significant issue as to whether such extant labeled sensor data are relevant for a new site under test. The challenge addressed in this work involves learning the relevance and relationship of existing labeled (training) data for analysis of a new unlabeled or partially labeled data set of interest. This type of problem has significant practical relevance for UXO sensing, for which results are presented on measured data, as well as for the aforementioned classes of problems, for which there is uncertainty concerning the appropriateness of existing labeled data for a new set of unlabeled data of interest.

To place this problem in a mathematical setting, let $\mathcal{T}(\mathbf{x}, y)$ be the probability distribution (or concept, borrowing a term from psychology[1]) from which test instances (each including a feature vector $\mathbf{x}$ and the associated class label y) are drawn. The goal in classifier design is to minimize the expected loss [**42**,**43**], i.e.,

$$\min_{\zeta} \sum_{y} \int_{\mathbf{x}} L\big(y_i, \zeta(\mathbf{x})\big) \mathcal{T}(\mathbf{x}, y)$$

where $L\big(y_i, \zeta(\mathbf{x})\big)$ is a loss function, which provides a quantitative measure for the loss incurred by the classifier when it predicts for $\boldsymbol{x}$ whose true label is y. In practice, one only has access to N independent training examples $(\mathbf{x}, y)$ drawn from $\mathcal{T}(\mathbf{x}, y)$, therefore the minimization is performed to minimize the empirical loss [**42**,**43**], i.e.,

$$\min_{\zeta} \frac{1}{N} \sum_{i=1}^{N} L\big(y_i, \zeta(\mathbf{x}_i)\big), \qquad \text{with } (\mathbf{x}_i, y_i) \sim \mathcal{T}(\mathbf{x}_i, y_i).$$

The empirical loss is an Monte Carlo approximation of the expected loss and therefore will approach the expected loss when $N \to \infty$.

A learning algorithm based on empirical loss minimization implicitly assumes that the future test instances are also drawn from $\mathcal{T}(\mathbf{x}, y)$. It is this assumption that assures that the classifier generalizes to test instances when it is trained to minimize empirical loss on training examples. This assumption, however, is often violated in practice, since training examples and test instances may correspond to different collections of measurements (likely performed at different times under different experimental conditions) and the class memberships of the measurements may also change. These issues can introduce statistical *differences* between the training examples and the test instances; the UXO-sensing problem discussed above constitutes an important example for which the aforementioned statistical issues hold, concerning the utility of existing labeled (training) data.

Assume that one has training examples from a distribution $\mathcal{A}(\mathbf{x}, y)$ which is different from $\mathcal{T}(\mathbf{x}, y)$. For convenience of exposition, we call $\mathcal{T}(\mathbf{x}, y)$ the primary or target distribution and call $\mathcal{A}(\mathbf{x}, y)$ the auxiliary distribution. Accordingly, the examples drawn from $\mathcal{T}(\mathbf{x}, y)$ are called primary data and the examples drawn from $\mathcal{A}(\mathbf{x}, y)$ are called auxiliary data.

[1] Traditionally, the (probabilistic) mapping $\Pr(y|\mathbf{x})$ is called a concept, and $\Pr(\mathbf{x})$ is called a virtual concept (language describing the concept) [**69**]. For simplicity, usually they are collectively called a concept.

In order to write the empirical loss for $\mathcal{T}$ in terms of examples drawn from $\mathcal{A}$, one may employ the technique of importance sampling [**44**]. By doing so, one makes the following modifications to the expression of empirical loss

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathcal{T}(\mathbf{x}_i, y_i)}{\mathcal{A}(\mathbf{x}_i, y_i)}L\big(y_i, \zeta(\mathbf{x}_i)\big), \qquad \text{with } (\mathbf{x}_i, y_i) \sim \mathcal{A}(\mathbf{x}_i, y_i).$$

where $\mathcal{T}(\mathbf{x}, y)/\mathcal{A}(\mathbf{x}, y)$ is the importance weight. It is known that the modification does not change the asymptotic behavior provided that $\mathcal{A}(\mathbf{x}, y)$ has the same nonzero support as $\mathcal{T}(\mathbf{x}, y)$.

Unfortunately, both $\mathcal{A}(\mathbf{x}, y)$ and $\mathcal{T}(\mathbf{x}, y)$ are unknown to the algorithm; all that is available are samples from $\mathcal{A}(\mathbf{x}, y)$. The challenge, therefore, is to learn a classifier on training examples drawn from $\mathcal{A}(\mathbf{x}, y)$ such that the resulting classifier still generalizes to test instances drawn from $\mathcal{T}(\mathbf{x}, y)$. Clearly, without assuming any knowledge about the relationship between $\mathcal{A}(\mathbf{x}, y)$ and $\mathcal{T}(\mathbf{x}, y)$, there is little one can do but to treat the training examples as if they are from the target distribution. This will, of course, introduce errors, which may be intolerable when the difference between $\mathcal{A}(\mathbf{x}, y)$ and $\mathcal{T}(\mathbf{x}, y)$ is large.

In this work we propose an efficient algorithm for solving the general problem of learning on examples from $\mathcal{A}$ with the goal of generalizing to instances from $\mathcal{T}$, when $\mathcal{A}$ is different from $\mathcal{T}$. We consider the case in which we have a fully labeled auxiliary data set $\mathcal{D}^a$ and a partially labeled primary data set $\mathcal{D}^p = \mathcal{D}_l^p \cup \mathcal{D}_u^p$, where $\mathcal{D}_l^p$ are labeled and $\mathcal{D}_u^p$ unlabeled. We assume that $\mathcal{D}^p$ are examples of the primary concept $\mathcal{T}$ (the concept we are interested in) and $\mathcal{D}^a$ are examples of the auxiliary concept $\mathcal{A}$ (the one providing indirect and low-quality information about $\mathcal{T}$). Our objective is to use a mixed training set $\mathcal{D}^{tr} = \mathcal{D}_l^p \cup \mathcal{D}^a$ to train a classifier that predicts the labels of $\mathcal{D}_u^p$ accurately, with the hope that $\mathcal{D}_l^p$ is required to have a small number of examples.

Assume $\mathcal{D}^p \sim \Pr(\mathbf{x}, y)$. We can write $\mathcal{D}^a \sim Pr(\mathbf{x}, y | s = 1)$, where the variable s is an indicator variable indicating that $(\mathbf{x}, y)$ is in the training set ($s = 1$), or not in the training set ($s = 0$), as long as the source distributions of $\mathcal{D}^p$ and $\mathcal{D}^a$ have the same support of nonzero probability. It is difficult to correct the mismatch by directly estimating $\Pr(s = 1)/\Pr(s = 1|\mathbf{x}, y)$. Therefore we take an alternative approach. We introduce an auxiliary variable μ_i for each $(\mathbf{x}_i^a, y_i^a) \in \mathcal{D}^a$ to reflect its mismatch with $\mathcal{D}^p$ and to control its participation in the learning process. The auxiliary variables μ are estimated along with the classifier in the learning. We employ logistic regression as a specific classifier and develop our method in this context.

The performance of *migratory logistic regression* (MigLogit) is demonstrated and compared to the standard logistic regression. The MigLogit is trained using $\mathcal{D}^a \cup \mathcal{D}_l^p$, where $\mathcal{D}_l^p$ are either randomly selected from $\mathcal{D}^p$, or actively selected from $\mathcal{D}^p$. When $\mathcal{D}_l^p$ are randomly selected, 50 independent trials are performed and the results are obtained as an average over the trials. Three logistic regression classifiers are trained using different combinations of $\mathcal{D}^a$ and $\mathcal{D}_l^p$: $\mathcal{D}^a \cup \mathcal{D}_l^p$, $\mathcal{D}_l^p$ alone, and $\mathcal{D}^a$ alone, where $\mathcal{D}_l^p$ are identical to the $\mathcal{D}_l^p$ used by MigLogit. The four classifiers are tested on $\mathcal{D}_p^u = \mathcal{D}^p \setminus \mathcal{D}_l^p$ to produce the test-error rate or the area under the ROC curve.

The performance of MigLogit is demonstrated on detection of unexploded ordnance (UXO) where the UXO signatures are site-sensitive. The sensor signature of a given UXO item is dependent on the soil properties as well as the history of the site in which it is located, the latter having a particular strong influence on the signature. The site history is dictated by complex factors such as co-located ordnance, the way the ordnance impacted the soil, and the surrounding man-made conducting clutter and UXO fragments. Therefore UXO detection is a typical site-sensitive problem.

The site-sensitivity makes standard supervised classification techniques an inappropriate choice for UXO detection, due to the difficulty in constituting a universal training set for classifier design. The training examples collected at previous sites are often not appropriate for use for analysis of the current site since the current site is often different from the previous ones (in the sense described above). Despite these disparities, the examples from previous sites are not totally useless; indeed, they can provide quite useful information about the examples for the current site (particularly for the UXO, since the ordnance types at different sites are often the same or similar; the clutter signatures are most often site specific). The usefulness of existing labeled data for a new site of interest is dictated by the characteristics of the new site, as well as on the characteristics of the sites from which the labeled data were acquired; these inter-relationships are complex and often difficult to characterize *a priori* (often accurate records are not available about the history of a former bombing site).

Let the examples at the current UXO cite be distributed according to $\mathcal{T}(\mathbf{x}, y)$, and the examples at a previous UXO cite be distributed according to $\mathcal{A}(\mathbf{x}, y)$. To demonstrate the utility of MigLogit in UXO detection, we here consider two UXO sites and design the classifier for the primary site (the one we are interested in) by using examples from another site (the auxiliary site). The auxiliary site is called *Jefferson Proving Ground (JPG)*, for which one is provided with the EMI and magnetometer measurements as well the associated labels (which are binary: UXO or non-UXO). The examples from the auxiliary site constitutes the auxiliary data $\mathcal{D}^a$. The primary site we are interested in is called *Badlands*, for which we have unlabeled EMI and magnetometer measurement for constituting the primary data $\mathcal{D}^p$. The labeled JPG data consists of 104 total items, of which 16 are UXO and 88 are non-UXO. The Badlands site consists of a total of 492 items, 57 of which are UXO and the remaining 435 are non-UXO. These two former bombing ranges exist at two very different geographical locations within the United States.

The UXO sensor measurements are mapped to four dimensional feature vectors $[\log(M_p), \log(M_z), z, \log(M_p/M_z)]$, where M_p and M_z are the dipole moments perpendicular and parallel to the target axis, respectively, and z is the approximate target depth [**27**]. These parameters are estimated by fitting the EMI and magnetometer measurements to a physical model [**27**]; the features from this study are

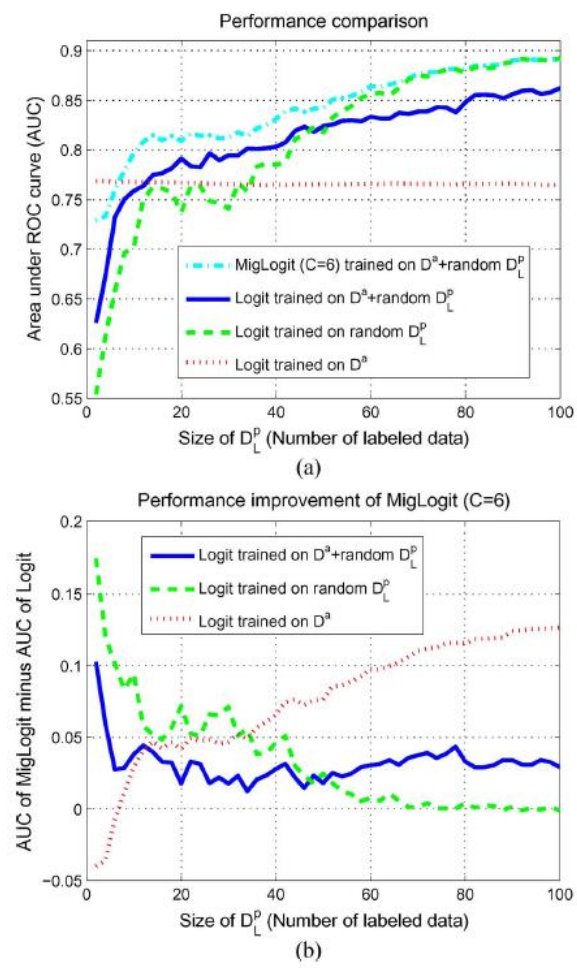

Figure 70: AUC of MigLogit minus AUC of logistic regression (a) The area under ROC curve (AUC) of MigLogit and logistic regression on the UXO data, as a function of size of $\mathcal{D}_l^p$ (b) The AUC of MigLogit minus the AUC's of logistic regression. The auxiliary data are collected at *Jefferson Proving Ground (JPG)* and the primary data are collected at *Badlands*. The primary labeled data $\mathcal{D}_l^p$ are randomly selected from $\mathcal{D}^p$. Each curve is an average over 50 independent trials of random selection of $\mathcal{D}_l^p$.

available upon request. Each feature is normalized to have zero mean and unitary variance. In UXO detection, one is interested in the receiver's operating characteristic (ROC) curve, particularly the area under ROC curve (AUC) [**45**].

The results are presented in Figure 70(a) and Figure 71(a), where each curve is the area under ROC curve as a function of the size of $\mathcal{D}_l^p$. The results in Figure 70(a) are obtained by randomly labeling primary data and by averaging the AUC's over 50 independent trials. The results in Figure 71(a) are obtained by actively labeling primary data using Fisher Information. For a better view of the improvement achieved by MigLogit, we plot in Figure 70(b) and Figure 71(b) the AUC of MigLogit with the AUC of each logistic regression classifier subtracted. A positive difference indicates performance improvement while a negative difference indicates performance degradation. We have the following observations:

1) With $\mathcal{D}_l^p$ determined randomly, MigLogit outperforms all logistic regression classifiers except at the early part of the curves, where there are very few examples in $\mathcal{D}_l^p$. The primary labeled examples are critical to the performance of MigLogit. With a few randomly selected examples one may not be able to find the appropriate auxiliary variables, leading to a poor compensation of the mismatch between $\mathcal{D}^p$ and $\mathcal{D}^a$ and therefore performance degradation.

2) With $\mathcal{D}_l^p$ actively determined, MigLogit outperforms all logistic regression classifiers, regardless of the number of primary labeled examples. This verifies that a good choice of $\mathcal{D}_l^p$ is important to the performance of MigLogit.

3) Active learning is not only beneficial to MigLogit, but to other classifiers as well, again demonstrating the advantage of active learning.

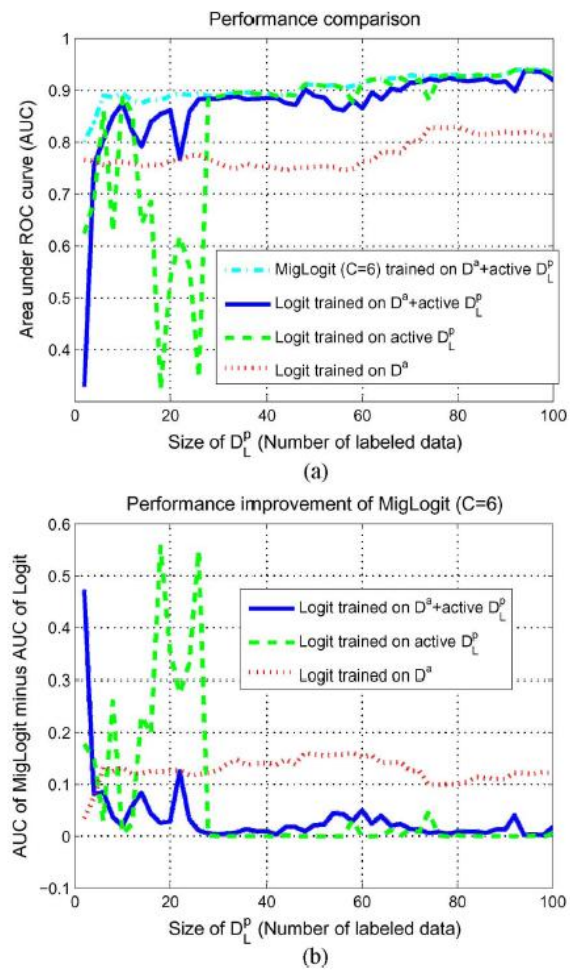

Figure 71: AUC of MigLogit minus AUC of logistic regression. (a) The area under ROC curve (AUC) of MigLogit and logistic regression on the UXO data, as a function of size of $\mathcal{D}_l^p$ (b) The AUC of MigLogit minus the AUC's of logistic regression. The auxiliary data are collected at *Jefferson Proving Ground (JPG)* and the primary data are collected at *Badlands*. The primary labeled data $\mathcal{D}_l^p$ are actively selected from $\mathcal{D}^p$.

Improved Concept Drift Summary

We have proposed an algorithm, called *migratory logistic regression* (MigLogit), for learning in the presence of concept change between the (auxiliary) training data $\mathcal{D}^a$ and the (primary) testing data $\mathcal{D}^p$. The basic idea of our method is to introduce an auxiliary variable μ_i for each example $(\mathbf{x}_i^a, y_i^a) \in \mathcal{D}^a$, which allows $\mathbf{x}_i^a$ to migrate to the class y_i^a when it cannot be correctly classified along with $\mathbf{x}_p$ by the classifier.

The migrations of $\mathcal{D}^a$ are controlled by the inequality constraint $(1/N^a)\sum_{i=1}^{N^a} y_i^a \mu_i \leq C$, where $C \geq 0$ is an appropriate bound limiting the average migration. The primary labeled data $\mathcal{D}_l^p$ play a pivotal role in correctly learning the classifier, and we have presented a method to actively selecting $\mathcal{D}_l^p$, which enhances the adaptivity of the entire learning process. We have developed a fast learning algorithm to enhance the ability of MigLogit to handle large auxiliary data sets.

The results from data collected at actual unexploded ordnance (UXO) sites show that MigLogit yields significant improvements over the standard logistic regression, demonstrating that if the classifier trained on $\mathcal{D}^a$ is to generalize well to $\mathcal{D}^p$, the mismatch between $\mathcal{D}^a$ and $\mathcal{D}^p$ must be compensated.

Although the core MigLogit algorithm has been developed in the setting of two data sets, we have also presented the methods for applying the MigLogit algorithm to multiple data sets, with no modification or some small modifications to the core algorithm. MigLogit utilizes the notion of information borrowing as do multi-task learning (MTL) and example selection/weighting. Yet, the way in which information borrowing is accomplished in MogLogit is unique and offers many advantages over other methods. These advantages, along with the convexity of the associated optimization problem and the computational efficiency, makes MigLogit particularly suitable for practical applications such as UXO detection.

There are several interesting directions for future research. The first one involves theoretical analysis of MigLogit, particularly examination of its performance bounds. The second direction is to extend the work to other models like support vector machines (SVM). One possible extension to SVMs is to use the μ variables as a set of data-dependent slack variables. Since the SVM is a convex problem, one may still get analytic solutions for the μ variables. Finally, it is also interesting to investigate the practical meaning that the μ variables may possess in certain applications (this may be possible because the μ variables have strong geometric interpretations).

Multi-Task Learning

When developing classification algorithms, one typically assumes access to a set of labeled data with which one may train a classifier. The set $\{\mathbf{x}_i, y_i\}_{i=1:n_L}$ constitutes the labeled data, where $\mathbf{x}_i$ represents the ith feature vector and y_i the associated (integer) label; n_L defines the number of available labeled data. The goal is to train a classifier to estimate a label y for a new unlabeled sample $\mathbf{x}$ under test. In many classification scenarios one measures all of the data of interest at once, where $\{\mathbf{x}_i\}_{i=1:n_L+1:n_L+n_U}$ denotes the n_U unlabeled data for which labels are to be estimated. For example, in airborne radar and electro-optic sensors, one may measure a large swath of terrain, and the unlabeled data $\{\mathbf{x}_i\}_{i=1:n_L+1:n_L+n_U}$ may be defined simultaneously for this entire terrain. Consequently, it is possible to place the classification of any one sample in $\{\mathbf{x}_i\}_{i=1:n_L+1:n_L+n_U}$ within the context of all such data.

In many previous classification studies it has been assumed that one has a sufficient set of labeled data $\{\mathbf{x}_i, y_i\}_{i=1:n_L}$ with which to build a classifier. However, in practice n_L may often be small and there is danger that one may over-train the classifier, and hence not generalize well to the unlabeled data $\{\mathbf{x}_i\}_{i=1:n_L+1:n_L+n_U}$ (i.e., $n_L \ll n_U$). To mitigate this challenge, one may note that typically one may have performed many previous classification "tasks" in the past, or multiple classification tasks may be performed simultaneously (e.g., in the context of airborne sensing, data may be collected simultaneously at different specific regions within an overall region of interest). If we have M such data sets, denoted $\{\mathcal{D}_m\}_{m=1:M}$, there is the potential to share data (labeled and unlabeled) between the M classification tasks, and therefore improve the classification performance relative to building a classifier only based on task-specific data. However, not all of the M tasks may be related to one another, and

therefore the technical challenge involves inferring the inter-relationships between the multiple data sets, such that the sharing of data across multiple tasks is performed appropriately. If this can be achieved, we may realize a second form of context: placing a given sensing task within the context of all previous or concurrent sensing tasks.

Algorithm training based only on labeled data is referred to as supervised learning, while learning based only on unlabeled data is termed unsupervised learning. The concept of integrating all available data, labeled and unlabeled, when training a classifier is typically referred to as semi-supervised learning. Semi-supervised learning will be applied here to realize the first class of context discussed above, provided by all of the unlabeled data. The second form of context, manifested by the appropriate sharing of data from M data sets $\{\mathcal{D}_m\}_{m=1:M}$, will be implemented via multi-task learning. As summarized below, semi-supervised and multi-task learning have been investigated separately, and this work represents their integration, with example results presented for realistic sensing challenges. The algorithm presented here is similar to that discussed in [**46**] with the difference being that here we consider non-linear classifiers where in [**46**] only linear classifiers were considered.

Semi-supervised learning has been an area of significant recent interest in the machine learning community [**32**,**33**,**34**,**35**,**36**,**37**,**38**,**39**]. To date, there have been several semi-supervised methods developed. The generative-model method, an early semisupervised method, estimates the joint probability of data and labels via expectation-maximization (EM), treating the missing labels of unlabeled data as hidden variables; this method was studied in statistics for mixture estimation [**41**] and has been reformulated for semi-supervised classification [**39**]. The semi-supervised support vector machine (SVM) [**36**] represents a more recent method, which maximizes the margin between classes, taking into account both labeled and unlabeled data. Graph-based methods [**33**,**34**,**35**,**38**], the main focus of current research in semi-supervised learning, exploit the assumption that strongly connected data points (in feature space) should share the same label, and utilizes spectral graph theory to quantify the between-data connectivity. For a more complete review of the literature, see [**32**].

Most graph-based algorithms operate in a transductive fashion, i.e., they directly learn the labels of the unlabeled data, instead of learning a classifier first and then using the classifier to infer the unseen labels (inductive learning). Transductive algorithms lack a principled means of predicting the labels of data out of the training set. The work in [**33**] addresses this problem by constructing a graph-based prior distribution on the parameters of a classifier and learns the classifier by maximizing the posterior (MAP estimation); the prior utilizes both labeled and unlabeled data, thus enforcing semi-supervised learning. Several drawbacks are inherent in the algorithm in [**33**]. For example, the hyper-parameter balancing the importance of the prior relative to the data likelihood needs to be learned. In this work we develop a non-transductive semisupervised algorithm, exploiting graph theory, with the final algorithm appropriate for integration within a multi-task-learning construct. Importantly, our semi-supervised algorithm is implemented without an explicit prior; this is important, because it allows a prior on the model parameters to be reserved for the multi-task component of the model.

Multi-task learning has been the focus of much recent interest in the machine learning community. Typical approaches to information transfer among tasks include: sharing hidden nodes in neural networks [**47**,**48**,**49**]; placing a common prior in hierarchical Bayesian models [**50**,**51**,**52**,**53**]; sharing parameters of Gaussian processes [**54**]; learning the optimal distance metric for K-Nearest Neighbors [**55**]; sharing a common structure on the predictor space [**56**]; and structured regularization in kernel methods [**57**], among others.

In statistics, the problem of combining information from similar but independent experiments has been studied in meta-analysis [**58**]. The objective of multi-task learning is different from that of meta analysis, since the objective of the latter is to combine data from different sources within a single model. Despite the difference in objectives, many of the techniques employed in the statistical literature on meta-analysis can be applied to multi-task learning as well.

Hierarchical Bayesian modeling is one of the most important methods for meta analysis [**59**,**60**,**61**,**62**,**63**]. Hierarchical Bayesian models provide the flexibility to model both the individuality of tasks (experiments), and the correlations between tasks. Usually the bottom layer of the hierarchy is individual models with task-specific parameters. On the layer above, tasks are connected together via a *common* prior placed on those parameters. In a nonparametric hierarchical Bayesian model, the common prior is often drawn from the Dirichlet process (DP). The advantage of applying the DP prior to hierarchical models has been addressed in the statistics literature; see for example [**62**,**63**,**64**]. Research on the Dirichlet process model goes back to 1973 and Ferguson [**65**], who proved that there is positive (non-zero) probability that some sample function of the DP will be as close as desired to any probability function defined on the same support set. Therefore, the DP is rich enough to model the parameters of individual tasks with arbitrarily high complexity, and flexible enough to fit them well without any assumption about the functional form of the prior distribution. The DP has been employed successfully in a recent MTL application [**66**].

In this work we seek to integrate semi-supervised learning and multi-task learning. As discussed in detail below, rather than directly employing a DP prior within the MTL component, which often yields relatively complex inference (learning), we alternatively employ a simpler prior that is based on and motivated by the DP; this yields a relatively simple expectation-maximization (EM) learning procedure. The proposed MTL framework is particularly well matched to the semi-supervised setting of interest, and it also yields fast inference.

Appropriate UXO data was not available, so these algorithms were tested using acoustic backscattering data for underwater mine detection. We consider the underwater-mine classification problem studied in [**67**], where the acoustic imagery data were collected with four different imaging sonars from two different environments (see [**67**] for details)[3]. This is a binary classification problem aiming to separate mines from non-mines based on the synthetic-aperture sonar (SAS) imagery. For each sonar image, a detector finds the objects of interest, and a 13-dimensional feature vector is extracted for each target. In this problem, there are a total of 8 data sets, constituting 8 tasks. The 8 data sets are collected with four sonar sensors from two different environmental conditions. The total number of data points in each task as well as the distribution of sensors and environments is listed in Table VII. The number of mines in each of the

Task	Number of Data	Sonar	Environment
1	1813	1	B
2	3562	1	A
3	1312	2	B
4	1499	3	B
5	2853	4	B
6	1162	2	A
7	1134	3	A
8	756	4	A

Table VII: Number of data points and distribution of sensors and environments for each of the eight tasks in the underwater-mine data set under consideration. Environment A is relatively challenging while environment B is relatively benign, with these characteristics manifested by the details of the sea bottom.

eight tasks varies from 9 to 65, and each task contains from 10 to 100 times more non-mines (clutter) than mines.

Following [**67**], we perform 100 independent trials, in each of which we randomly select a subset of data

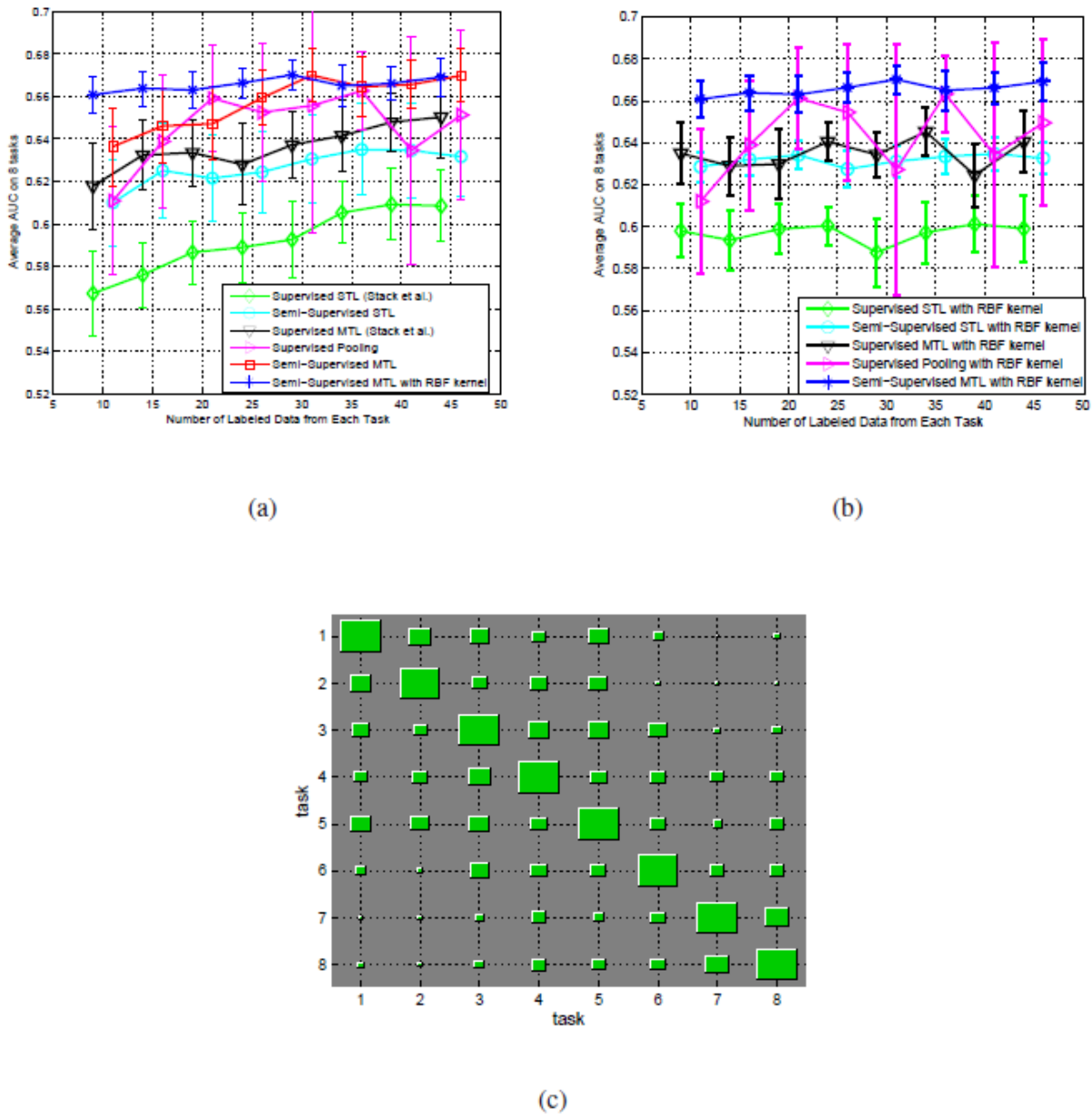

Figure 72: Comparison of classification results for the underwater mine data, considering supervised, semi-supervised, linear, and non-linear classifiers; results are also shown for single-task learning (STL) and multi-task learning (MTL). Results are presented in terms of the area under the receiver operating characteristic (ROC) curve, denoted on the vertical axis as AUC. The number of labeled data considered are 10, 20, 30, etc., with the points offset to enhance readability of the error bars (manifested by considering 100 different randomly selected labeled data). The number of data samples across the 8 tasks is given in Table VII. (a) Linear classifiers, with comparison to the proposed semisupervised MTL with a nonlinear (RBF) classifier; (b) non-linear classifiers; (c) the inferred sharing mechanisms between the 19 tasks, for the proposed semi-supervised MTL algorithm based on the nonlinear classifier (based on 20 labeled data).

for which labels are assumed known (labeled data). The AUC averaged over 8 tasks is presented in Figure 72(a) for the linear classifiers and in Figure 72(b), for the non-linear classifiers; AUC results are again presented as a function of the number of labeled data in each task, where each curve represents the mean calculated from the 100 independent trials. The results of supervised STL and supervised MTL are cited from [**67**]. In this case the linear and non-linear semi-supervised MTL classifiers yield comparable results when the number of labeled data is relatively large, while for smaller quantities of labeled data the non-linear classifier shows advantages. While the pooling results are sometimes good, they do not appear to be as stable as the MTL methods.

The results on underwater target classification reveal relative performance among the twelve algorithms that is consistent with those shown above for the landmine sensing example. One difference, for this example, is that the pooling results were similar to those of the associated MTL results (*e:g:*, the semi-supervised MTL results were comparable to those of the semi-supervised pooling results). This is attributed to the observation that, for this data, the inter-task sharing properties inferred from the MTL analysis are not as stark as they were for the landmine example above. The MTL-inferred inter-task sharing properties are depicted in Figure 72(c), as computed for the semi-supervised MTL algorithm with an RBF kernel. This demonstrates that in some cases pooling may be appropriate, but this information is generally unknown *a priori*, and the MTL results are comparable to pooling under such circumstances. This suggests that in general, when the sharing mechanisms are unknown, it may be prudent to perform MTL rather than pooling.

Multi-Task Learning Summary

We have developed a new classification algorithm that merges ideas from semisupervised and multi-task learning. The algorithm exploits context from two perspectives: (i) the classification of any one unlabeled sample is placed within the context of all unlabeled data, and (ii) the classification of all data within a particular data set (task) is placed within the context of all other data sets (tasks) that may be available. Concerning (ii), these multiple data sets may be collected simultaneously, or they may be collected sequentially over the "lifetime" of the system or sensor. The main challenge in this latter form of context involves determining which data sets are relevant for the data set of interest, since not all data sets are anticipated to be relevant for one another. This problem has been addressed using a variant of the Dirichlet process [**65**]. An important aspect of the proposed method is that it is implemented efficiently using an expectation-maximization (EM) framework, and therefore all computations presented here have been performed on a PC, using un-optimized Matlab software. The software is very fast computationally, with run times much faster than other related approaches based on a variational Bayesian analysis using the Dirichlet process for multi-task learning [**66**].

Concerning future research, one may observe $M-1$ data sets previously, with these data partially labeled, as considered here. When observing the Mth data set, one may not have any labeled data, but will often have access to a large quantity of unlabeled data from this task. It is of interest to develop transfer-learning algorithms that may infer a classifier for task M, even though one has no associated labeled data. It is believed that the manifold information alone may be used to infer which of the $M-1$ tasks (if any) are relevant for the Mth task, and if there is at least one relevant task the associated labels may be "transferred" to the Mth task. It therefore seems feasible to perform inference on a new task based only on unlabeled data, particularly if the $M-1$ tasks observed previously are sufficiently rich. This concept may also be integrated with active learning [**68**], to infer which data from task M would be most informative to algorithm training if the associated labels could be acquired. Ideally, if the manifold information from the unlabeled data from task M is sufficient, the required number of actively-determined labels required for task M would be small or even zero.

V. Project Summary and Future Work

The work completed during this program focused on developing algorithms that provide UXO classification capability for multi-axis sensors that are superior to single-axis solutions. Previous SERDP-sponsored research was leveraged and extended to 1) develop exact and approximate phenomenological models for multi-axis sensors, 2), develop, and apply to field-collected data, statistical signal processing algorithms which maximize UXO classification performance using data measured by multi-axis sensors, and 3) develop active learning and kernel-based algorithms to enable algorithm training and application across test sites.

The modeling efforts focused on developing phenomenological models for the LBL BUD AEM sensor and the USGS ALLTEM sensor. Both of these sensors employ multiple transmit and receive coils at different orientations. Although the sensors were not fully modeled under this effort (i.e., some of the transmit and/or receive coils were not yet incorporated into the model), there is generally good agreement between the measured data and the modeled sensor response. In addition, the sensor models are currently being completed so that all transmit and receive coils will be fully incorporated into the models. As sensors continue to be introduced and refined, it will be important to continue modeling efforts for the new sensors so the benefits of model-based signal processing will continue to be realized.

The statistical signal processing efforts focused on developing physics-based statistical signal processing approaches appropriate for data measured by multi-axis sensors, including model inversion, feature selection, classifier design, and sensor management, and evaluating the proposed approaches on field-measured LBL BUD AEM and USGS ALLTEM sensor data. Results indicate that applying statistical signal processing algorithms to multi-axis sensor data has the capability to provide improved UXO classification performance. Results also showed that adapting standard algorithms, such as numerical least squares for model inversion, to the specific sensor and task at hand can improve their robustness. To this end, further work focused on continuing to customize these algorithms to ensure robust model inversion would be worthwhile. Similarly, other approaches which may improve robustness should be explored. Results obtained under the present effort have indicated that inversion performance may be improved by judiciously selecting the measurements to include in the model inversion, or by selecting in a principled manner one inversion from a group of several which are equivalent in terms of goodness-of-fit, and continuing to explore these ideas to improve model inversion performance and robustness could be beneficial. Continuing to improve robustness throughout the processing chain by extending the work on feature selection and classifier design would also be advantageous. For example, results obtained here indicate that feature selection performance may be sensitive to the performance metric that is optimized, so there may be opportunities to improve robustness by optimizing the performance metric utilized in feature selection.

The efforts related to active learning and kernel-based algorithms focused on developing approaches that would enable data collected across multiple sites to be integrated within the classifier. Results obtained here show that active and multi-task learning may provide frameworks in which the classifier parameters can be adapted within a data collection, or across data collections at multiple sites, to improve classifier training and UXO classification performance. Continuing to explore improvements to classifiers that would make them more robust to uncertainties inherent in the field would be valuable. For example, matching pursuits appears to hold promise for UXO classification in situations in which there are multiple spatially overlapping target signatures within a set of measurements.

The models and algorithms developed under this program provide improved UXO classification for multi-axis sensor data. Further improvements may be possible by focusing on improving the robustness

of model inversion, of the classifiers, and in the presence of uncertainties inherent in the field, such as overlapping signatures or varying clutter characteristics. An additional benefit of improving algorithm and processing robustness is that more of the processing may be automated, reducing the need to rely on human experts to analyze the data.

VI. Technology Transfer

The results of this work are available to the research community through reports, and journal and conference publications. In addition, a software suite of pattern recognition tools has been given to Skip Snyder of Snyder Geoscience.

VII. Manuscripts Published/Submitted and Presentations Given

Manuscripts

[1] Ya Xue, Xuejun Liao, Balaji Krishnapuram and Lawrence Carin, "Multi-task learning for classification with Dirichlet Process priors." *Journal of Machine Learning Research*, **8**:35–63. (January 2007).

[2] Stacy L. Tantum, Yongli Yu, and Leslie M. Collins, "Bayesian mitigation of sensor position errors to improve unexploded ordnance detection." *IEEE Geoscience and Remote Sensing Letters*, **5**(1): 103–107 (January 2008).

[3] Qiuhua. Liu, Xuejun Liao and Lawrence Carin, "Detection of unexploded ordnance via efficient semisupervised and active learning." *IEEE Transactions on Geoscience and Remote Sensing*, **46**(9):2558–2567 (September 2008).

[4] Stacy L. Tantum, Quan Zhu, Peter A. Torrione, and Leslie M. Collins, "Modeling position error probability density functions for statistical inversions using a Goff-Jordan rough surface model." *Stochastic Environmental Research and Risk Assessment*, **23**(2):153–261 (February 2009).

[5] Xuejun Liao and Lawrence Carin, "Migratory logistic regression for learning concept drift between two data sets with application to UXO sensing." *IEEE Transactions on Geoscience and Remote Sensing*, **47**(5):1454–1466 (May 2009).

[6] Qiuhua Liu, Xuejun Liao, Hui Li, Jason R. Stack, and Lawrence Carin, "Semisupervised multitask learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **31**(6):1074–1086 (June 2009).

Presentations

[1] Leslie Collins, "The Current State of Signal Processing and Sensor Fusion for Shallow Geophysical Remote Sensing," *Workshop on Shallow Tunnel Detection*, University of Mississippi, February, 2006 (INVITED).

[2] Stacy L. Tantum, Yuequan Wang, and Leslie M. Collins, "Statistical and Adaptive Signal Processing For UXO Discrimination for Next-Generation Sensor Data," *PIERS*, Boston, MA, July, 2006 (INVITED).

[3] Stacy L. Tantum, Yuequan Wang, Yongli Yu, and Leslie M. Collins, "Mitigating Measurement Position Uncertainty to Improve UXO Detection," *UXO/Countermine Forum*, Las Vegas, NV, July, 2006.

[4] Stacy L. Tantum, Yuequan Wang, Yongli Yu, and Leslie M. Collins, "Statistical And Adaptive Signal Processing For UXO Discrimination For Next-Generation Sensor Data," *SERDP/ESTCP Symposium*, Washington, D.C., November, 2006.

[5] Jeremiah J. Remus and Leslie M. Collins, “Matching pursuits decomposition for discrimination of unexploded ordnance: Isolated and overlapping signatures.” *Proceedings of the 2008 IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2008*. Boston, Massachusetts. (July 6-11, 2008).

[6] Jeremiah J. Remus, Kenneth D. Morton, Peter A. Torrione, Stacy L. Tantum, and Leslie M. Collins, “Comparison of a distance-based likelihood ratio test and k-nearest neighbor classification methods.” *Proceedings of the 2008 IEEE International Workshop on Machine Learning for Signal Processing*. Cancun, Mexico. (October 16-19, 2008).

VIII. References

[1] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley-Interscience, 1991.

[2] X. Liao and L. Carin, "Application of the Theory of Optimal Experiments to Adaptive Electromagnetic-Induction Sensing of Buried Targets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 961-972, 2004.

[3] R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*. New York: Springer-Verlag, 1987.

[4] J. C. Alverez, R. C. Gonzalez, D. A. Prieto, A. Shkel, and V. Lumelsky, "Sensor Management for Local Obstacle Detection in Mobile Robots," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002, pp. 2401-2406.

[5] S. Dasika, S. Vrudhula, K. Chopra, and R. Srinivasan, "A Framework for Battery-Aware Sensor Management," in *Proceedings of the IEEE Design, Automation, and Test in Europe Conference and Exhibit*, vol. 2, 2004, pp. 962-967.

[6] M. A. Perillo and W. B. Heinzelman, "Optimal Sensor Management Under Energy and Reliability Constraints," in *Proceedings of the IEEE Conference on Wireless Communications and Networking*, vol. 3, 2003, pp. 1621-1626.

[7] A. El-Fallah et al., "Multisensor-Multitarget Sensor Management with Target Preference," in *Proceedings of SPIE*, vol. 5429, 2004, pp. 222-232.

[8] M. L. Hernandez and P. R. Horridge, "Dynamic Multisensor Resource Management with Increased Sensor Utility," , vol. 5204, 2003, pp. 374-389.

[9] M. Kalandros and L. Y. Pao, "Multisensor Convariance Control Strategies for Reducing Bias Effects in Interacting Target Scenarios," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 1, pp. 153-173, 2005.

[10] M. Kalandros, L. Trailovic, L. Y. Pao, and Y. Bar-Shalom, "Tutorial on Multisensor Management and Fusion Algorithms for Target Tracking," in *Proceedings og the 2004 American Control Conference*, vol. 5, 2004, pp. 4734-4748.

[11] C. Kreucher and A. O. Hero, "Non-myopic Approaches to Scheduling Agile Sensors for Multistage Detection, Tracking, and Identification," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. V, 2005, pp. 885-888.

[12] C. Kreucher, K. Kastella, and A. O. Hero, "Multitarget Tracking Using the Joint Multitarget Probability Density," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 41, no. 4, pp. 1396-1414,

2005.

[13] R. Mahler, "Objective Functions for Bayesian Control-Theoretic Sensor Management, I: Multitarget First-Moment Approximation," in *Proceedings of the IEEE Aerospace Conference*, vol. 4, 2002, pp. 1905-1923.

[14] R. Malhotra, "Temporal Considerations in Sensor Management," in *Proceedings of the IEEE National Aerospace and Electronics Conference*, 1995, pp. 86-93.

[15] G. A. McIntyre and K. J. Hintz, "Sensor Management Scheduling: An Enhanced Dynamic, Preemptive Algorithm," *Optical Engineering*, vol. 37, pp. 517-523, 1998.

[16] M. R. Moreland, C. Kreucher, and K. Kastella, "A Bayesian Approach to Multiple Target Detection and Tracking," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1589-1604, 2007.

[17] S. Musick and R. Malhotra, "Chasing the Elusive Sensor Manager," in *Proceedings of the IEEE National Aerospace and Electronics Conference*, 1994, pp. 606-613.

[18] D. Sinno, D. Cochran, and D. R. Morell, "Multi-mode Detection with Markov Target Motion," in *Proceedings of the IEEE 3rd International Conference on Information Fusion*, vol. WeD1, 2000, pp. 25-31.

[19] M. P. Kolba and L. M. Collins, "Information-Based Sensor Management in the Presence of Uncertainty," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2731-2735, 2007.

[20] M. P. Kolba and L. M. Collins, "Sensor Managment for Static Target Detection with Non-Binary Sensor Observations and Observation Uncertainty," in *IEEE/SP 14th Workshop on Statistical Signal Processing, SSP '07*, 2007, pp. 74-78.

[21] M. P. Kolba and L. M. Collins, "Sensor Management for Landmine Detetion Using Correlated Sensor Observations," in *Proceedings of SPIE*, vol. 6953, 2008.

[22] M. P. Kolba, W. R. Scott Jr., and L. M. Collins, "A Framework for Information-Based Sensor Management for the Detection of Static Targets," *IEEE Transactions on Systems, Man, and Cybernetics -- Part A: Systems and Humans*.

[23] K. Kastella, "Discrimination Gain to Optimize Detection and Classification," *IEEE Transactions on Systems, Man, and Cybernetics -- Part A: Systems and Humans*, vol. 27, no. 1, pp. 112-116, 1997.

[24] Y. Yu and L. M. Collins, "Multi-Modal Iterative Adaptive Processing (MIAP) Performance in the Discrimination Model for Landmine Detection," in *Detection and Remediation Technologies for Mines and Minelike Targets X*, 2005, pp. 1108-1117.

[25] N. Lhomme, D. W. Oldenburg, L. R. Pasion, D. B. Sinex, and S. D. Billings, "Assessing the Quality of Electromagnetic Data for the Discrimination of UXO Using Figures of Merit," *Journal of Environmental and Engineering Geophysics*, vol. 13, no. 3-SI, pp. 165-176, September 2008.

[26] Lawrence Carin et al., "On the Wideband EMI Response of a Rotationally Symetric Permeable and Conducting Target," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 6, pp. 1206-1213, June 2001.

[27] Y. Zhang, L. M. Collins, H. Yu, C. E. Baum, and L. Carin, "Sensing of Unexploded Ordnance with Magnetometer and Induction Data: Theory and Signal Processing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 41, no. 5, pp. 1005-1015, May 2003.

[28] Y. Zhang, X. Liao, and L. Carin, "Detection of Buried Targets Via Active Selection of Labeled Data: Application to Sensing Subsurface UXO," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 11, pp. 2535-2543, November 2004.

[29] D. Cohn, Ghahramani Z., and M. Jordan, "Active Learning with Statistical Models," in *Advances in Neural Information Processing systems (NIPS)*, Cambridge, MA, 1996.

[30] V. V. Fedorov, *Theory of Optimal Experimentsd*. New York: Academic Press, 1972.

[31] H. H. Nelson and J. R. McDonald, "Multisensor Towed Array Detection System for UXO Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 39, no. 6, pp. 1139-1145, June 2001.

[32] O. Chapelle, B. Scholkopf, and A. Zien, Eds., *Semi-Supervised Learning*. Cambridge, MA: MIT Press, 2006.

[33] B. Krishnapuram et al., "On Semi-Supervised Classification," in *Advances in Neural Information Processing Systems (NIPS)*, Cambridge, MA, 2005.

[34] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised Learning Using Gaussian Fields and Harmonic Functions," in *Proceedings of the 20th ICML*, 2003, pp. 912-919.

[35] M. Szummer and T. Jaakkola, "Partially Labeled Classification with Markov Random Walks," in *Advances in Neural Information Processing Systems (NIPS)*, Cambridge, MA, 2002.

[36] T. Joachims, "Transductive Inference for Text Classification Using Support Vector Machines," in *Proceedings of the 16th ICML*, 1999, pp. 200-209.

[37] A. Blum and T. Mitchell, "Combining Labeled and Unlabeled Data with Co-Training," in *Proceedings Annu. COLT*, 1998, pp. 92-100.

[38] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and Semisupervised Learning on Large

Graphs," in *Proceedings Annu. COLT*, 2004, pp. 624-638.

[39] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, "Text Classification from Labeled and Unlabeled Documents Using EM," *Machine Learning*, vol. 39, no. 2/3, pp. 103-134, May/June 39.

[40] Q. Liu, X. Liao, and Carin L., "Semi-supervised Multitask Learning," in *Advances in Neural Information Processing Systems 20*, Cambridge, MA, 2008, pp. 937-944.

[41] S. Ganesalingam, "Classification and Mixture Approaches to Clustering Via Maximum Likelihood," *Applied Statistics*, vol. 38, no. 3, pp. 455-466, 1989.

[42] V. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.

[43] V. N. Vapnik, "An Overview of Statistical Learning Theory," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 988-999, September 1999.

[44] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. New York: Springer-Verlag, 1999.

[45] J. Hanley and B. McNeil, "The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve," *Radiology*, vol. 143, no. 1, pp. 29-36, April 1982.

[46] Q. Li, X. Liao, and L. Cari, "Semi-Supervised Multi-Task Learning," in *Advanced in Neural Information Processing Systems*, Cambridge, MA, 2007.

[47] J. Baxter, "Learning Internal Representations," in *COLT: Proceedings of the Workshop on Computational Learning Theory*, 1995.

[48] J. Baxter, "A Model of Inductive Bias Learning," *Journal of Artificial Intelligence Research*, 2000.

[49] R. Caruana, "Multitask Learning," *Machine Learning*, vol. 28, pp. 41-75, 1997.

[50] K. Yu, A. Schwaighofer, V. Tresp, W.-Y. Ma, and H. Zhang, "Collaborative Ensemble Learning: Combining Collaborative and Content-Based Information Filtering Via Hierarchical Bayes," in *Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence*, 2003.

[51] K. Yu, V. Tresp, and S. Yu, "A Nonparametric Hierarchical Bayesian Framework for Information Filtering," in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2004.

[52] K. Yu, A. Schwaighofer, and V. Tresp, "Learning Gaussian Processes From Multiple Tasks," in *Proceedings of the 22nd International Conference on Machine Learning*, 2005.

[53] J. Zhang, A. Ghahramani, and Y. Yang, "Learning Multiple Related Tasks Using Latent Independent Component Analysis," in *Advances in Neural Information Processing Systems 18*, Cambridge, MA,

2006.

[54] N. D. Lawrence and J. C. Platt, "Learning to Learn with the Informative Vector Machine," in *Proceedings of the 21st International Conference on Machine Learning*, 2004.

[55] S. Thrun and J. O'Sullivan, "Discovering Structure in Multiple Learning Tasks: The TC Algorithm," , 1996, p. Proceedings of the 13th International Conference on Machine Learning.

[56] R. K. Ando and T. Zhang, "A Framework for Learning Predictive Structures from Multiple Tasks and Unlabeled Data," *Journal of Machine Learning Research*, vol. 6, pp. 1817-1853, 2005.

[57] T. Evgeniou, C. A. Micchelli, and M. Pontil, "Learning Multiple Tasks with Kernel Methods," *Journal of Machine Learning Research*, vol. 6, pp. 615-637, 2005.

[58] G. V. Glass, "Primary, Secondary and Meta-Analysis of Research," *Educational Researcher*, vol. 5, 1976.

[59] D. Burr and H. Doss, "A Bayesian Semiparametric Model for Random-Effects Meta-Analysis," *Journal of the American Statistical Association*, vol. 100, no. 469, pp. 242-251, March 2005.

[60] F. Dominici, G. Parmigiani, R. Wolpert, and K. Reckhow, "Combining Information from Related Regressions," *Journal of Agricultural, Biological, and Environmental Statistics*, vol. 2, no. 3, pp. 294-312, 1997.

[61] P. D. Hoff, "Nonparametric Modeling of Hierarchically Exchangeable Data," University of Washington Statistics Department, Technical Report 421 2003.

[62] P. Muller, F. Quintana, and G. Rosner, "A Method for Combining Inference Across Related Nonparametric Bayesian Models," *Journal of the Royal Statistical Society Series B*, vol. 66, no. 3, pp. 735-749, 2004.

[63] B. K. Mallick and S. G. Walker, "Combining Information from Several Experiments with Nonparametric Priors," *Biometrika*, vol. 84, no. 3, pp. 697-706, 1997.

[64] S. Mukhopadhyay and A. E. Gelfand, "Dirichlet Process Mixed Generalized Linear Models," *journal of the American Statistical Association*, vol. 92, no. 438, pp. 633-639, 1997.

[65] T. Ferguson, "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, vol. 1, pp. 209-230, 1973.

[66] Y. Xue, X. Liao, L. Carin, and B. Krishnapuram, "Multi-Task Learning for Classification with Dirichlet Process Priors," *Journal of Machine Learning Research*, vol. 8, pp. 35-63, 2007.

[67] J. R. Stack, F. Crosby, R. J. McDonald, Y. Xue, and L. Carin, "Multi-Task Learning for Underwater Object Classification," in *Proceedings of the SPIE Defense and Security Symposium*, 2007, pp. 1-10.

[68] M. I. Jordan and D. A. Ghahramani, Z. Cohn, "Active Learning with Statistical Models," in *Advances in Neural Information Processing Systems 21*, 1996.

[69] G. Widmer and M. Kubat, "Effective Learning in Dynamic Environments by Explicitly Concept Tracking," in *Proceedings of the European Conference on Machine Learning*, 1993, pp. 224-227.

[70] S. Ganesalingam, "Classification and Mixtre Approaches to Clustering Via Maximum Likelihood," *Applied Statistics*, vol. 38, no. 3, pp. 455-466, 1989.