Incorporating Uncertainty in Unexploded Ordnance Discrimination

Laurens Beran, Stephen Billings, and Doug Oldenburg

***Abstract*—We examine representations of feature vector uncertainty in the context of unexploded ordnance (UXO) discrimination with electromagnetic data. We compare a local uncertainty estimate derived from the curvature of the misfit function with global estimates of the model posterior probability density (PPD) obtained with Markov chain sampling. For well-posed experiments (i.e., with high SNR and adequate spatial coverage), the two methods of uncertainty appraisal agree. However, when the inverse problem is ill posed, we find out that the PPD can be multimodal. To incorporate these uncertainties in discrimination, we first develop an extension of discriminant analysis which integrates over the posterior distribution of the model. When dealing with multimodal PPDs, we show that an effective solution is to input all modes of the PPD—corresponding to all models at local minima of the misfit—into discrimination and, then, to classify on the basis of the model which is most likely a UXO.**

***Index Terms*—Discrimination, electromagnetics, uncertainty, unexploded ordnance (UXO).**

I. Introduction

UNEXPLODED ordnance (UXO) discrimination with digital geophysical data is a challenging problem which requires careful application of inversion and decision theory. In the inversion stage, we try to estimate the parameters of a physical model such that the data predicted by a model adequately reproduce the observed data acquired over a buried target. Magnetic and electromagnetic (EM) sensors are most commonly deployed for ordnance detection, and for these data types, simple dipole models are usually fit to the data. A magnetostatic dipole, parameterized in terms of a Cartesian location and dipole moment vector, can be used to predict the anomalous magnetic field observed over a susceptible target.

For EM data, the induced secondary field radiated by a conductive target is often modeled as a superposition of orthogonal magnetic dipoles with characteristic magnetic polarization decays (or spectra in the frequency domain). The model vector $\mathbf{m}$ is then comprised of extrinsic (target location and orientation) and intrinsic (target polarizations) parameters and is related to the predicted data ($\mathbf{d}^{\text{pred}}$) via the forward modeling functional $\mathbf{d}^{\text{pred}} = F[\mathbf{m}]$. To obtain estimates of these model parameters in an inversion, we minimize a function quantifying the misfit between the observed ($\mathbf{d}^{\text{obs}}$) and predicted data. A common choice of misfit function is the least squares data misfit

$$\phi = \frac{1}{2}\left\|\mathbf{W}_d(\mathbf{d}^{\text{obs}} - \mathbf{d}^{\text{pred}})\right\|^2. \tag{1}$$

The data weighting matrix $\mathbf{W}_d$ quantifies the uncertainty in the data. A typical choice when dealing with geophysical data is to assign a weighting to each datum as a percentage of the observed datum plus a floor value. The noise floor represents a baseline background noise level, and it can be estimated from regions of a survey where there are no detected targets.

Parameters of empirical models used for UXO discrimination are proxies for physical properties such as target size and shape. For example, the strength and decay rate of the induced dipole moment are diagnostic of target size and wall thickness [1]. As such, the parameters estimated in an inversion can be used to decide whether a target is likely to be an ordnance item. Rule-based approaches can perform well when the discrimination task is relatively simple. For example, when there is a single ordnance type which is much larger than the metallic clutter, shrapnel, etc., that we wish to leave in the ground, then the EM-induced moment amplitude can be an effective metric for distinguishing ordnance from clutter.

Statistical classification algorithms have also been employed for UXO discrimination [2]–[6]. In this approach, a subset of model parameters is selected to span a feature space. We also have a training data set comprised of feature vectors for which ground truth is known. Using this training data set, we try to formulate a decision rule in the feature space, which we then apply to the remaining unlabeled feature vectors (the test data). A common approach is to fit probability distributions to the training data and then use these distributions to predict the probability that a test feature vector is an ordnance item.

In this paper, we focus upon the role of parameter uncertainty in inversion and discrimination. Fig. 1 shows a motivating example for the problem of discrimination, with features extracted from time-domain EM (TEM) data (more details of the forward modeling, inversion, and discrimination algorithms used in this example are provided in Section III). The best-fitting model for one target (indicated by an arrow in Fig. 1) is far removed from the typical feature vectors that we obtain for this type of ordnance. This outlier highlights that the TEM parameter estimation problem is fundamentally ill posed: a small perturbation in the data caused by noise can produce a large change in the estimated model (ill conditioning), and multiple models can fit

Manuscript received July 9, 2010; revised December 31, 2010; accepted January 30, 2011. Date of publication March 24, 2011; date of current version July 22, 2011. This work was supported by SERDP project MR 1629: Robust statistics and regularization for feature extraction and UXO discrimination.

L. Beran and S. Billings are with Sky Research Inc., Vancouver, BC VCT 1Z4, Canada, and also with the University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: lberan@eos.ubc.ca; stephen.billings@skyresearch.com).

D. Oldenburg is with the University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: doug@eos.ubc.ca).

Digital Object Identifier 10.1109/TGRS.2011.2112772

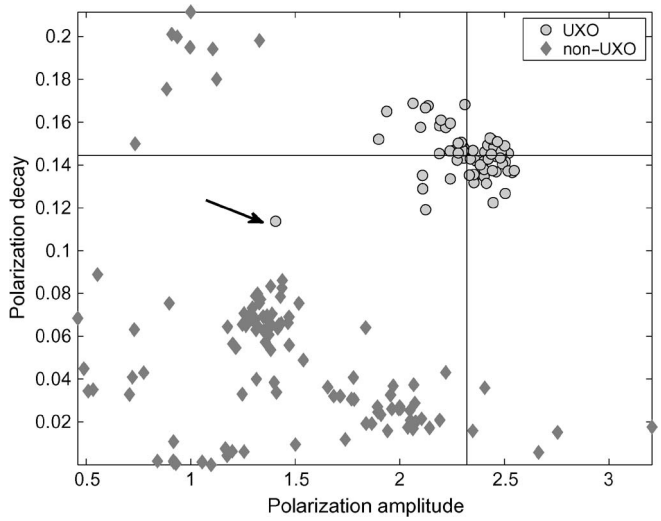


Fig. 1. Estimated model parameters for the ordnance and nonordnance targets for the Camp Sibert EM63 data. The feature vector indicated by an arrow corresponds to the fit shown in Fig. 3(a). The crosshairs show the location of the mean of the UXO class.

the data equally well (nonuniqueness). In this example, noise related to errors in measurement of sensor position confounds the inversion and causes the estimated feature vector to deviate significantly from its class (see [7] for a detailed discussion of this target).

We can address ill-posed inversions in a number of ways. TEM sensors developed specifically for UXO detection and discrimination can better constrain the inverse problem by providing measurements of orthogonal components of the received secondary magnetic field. Nonuniqueness can be somewhat alleviated by careful parameterization of the forward model (especially with respect to decay parameters, as in [8]), by incorporation of prior information (e.g., bound constraints) into the inversion [9], or by utilizing alternate criteria to assess the quality of the numerical solution [10]. However, there will always be uncertainty in our parameter estimates arising from the approximate nature of the forward model and the noise on the data. In this paper, we therefore investigate how model uncertainty can be propagated through inversion and into a discrimination algorithm.

In Section II, we explore two methods of uncertainty appraisal: a local linearized analysis which approximates the model parameters as Gaussian distributions and a global nonlinear appraisal which makes no assumptions regarding the parametric form of the model probability density function (pdf). We find in Section III that the two techniques agree for well-constructed experiments where the data quality is high. However, if the inversion is poorly constrained (i.e., when SNR is low), the model pdf can be multimodal and is not accurately represented by a Gaussian distribution. In Section IV, we develop a discrimination algorithm which incorporates uncertainties in the discrimination procedure by integrating over the model posterior probability density (PPD). Finally, in Section V, we demonstrate that this method can improve discrimination performance by improving the detection of the outlying feature vectors in real data sets.

II. Model Uncertainty

For a nonlinear forward modeling $\mathbf{d}^{\text{pred}} = F(\mathbf{m})$, the misfit can be minimized iteratively by solving for the model perturbation $\delta\mathbf{m}$

$$\mathbf{H}\delta\mathbf{m} = \mathbf{J}^T\mathbf{W}_d^T\mathbf{W}_d\delta\mathbf{d} \tag{2}$$

with $\mathbf{H} = \nabla^2\phi$ the Hessian of the misfit, $\mathbf{J}$ the Jacobian matrix of sensitivities, and $\delta\mathbf{d} = (\mathbf{d}^{\text{obs}} - F(\mathbf{m}))$. At the minimizer $\hat{\mathbf{m}}$, we can approximate the model covariance as

$$\begin{aligned}\text{cov}(\hat{\mathbf{m}}) &= E(\delta\hat{\mathbf{m}}\delta\hat{\mathbf{m}}^T)\\ &= E\left(\mathbf{H}^{-1}\mathbf{J}^T\mathbf{W}_d^T\mathbf{W}_d\delta\mathbf{d}\delta\mathbf{d}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{J}^T\mathbf{H}^{-1}\right)\\ &= \mathbf{H}^{-1}\mathbf{J}^T\mathbf{W}_d^T E\left((\mathbf{W}_d\delta\mathbf{d})(\mathbf{W}_d\delta\mathbf{d})^T\right)\mathbf{W}_d\mathbf{J}^T\mathbf{H}^{-1}.\end{aligned} \tag{3}$$

If the errors on the data are independent and Gaussian and the weightings in $\mathbf{W}_d$ are the inverse standard deviations of the noise, then the expectation of the terms in the last expression is the identity, so that

$$\text{cov}(\hat{\mathbf{m}}) = \mathbf{H}^{-1}\mathbf{J}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{J}^T\mathbf{H}^{-1}. \tag{4}$$

The expression can be further simplified by noting that $\mathbf{H} \approx \mathbf{J}^T\mathbf{W}_d^T\mathbf{W}_d\mathbf{J}^T$, giving us an approximate expression for the covariance of the model parameters [11]

$$\text{cov}(\hat{\mathbf{m}}) \approx \mathbf{H}^{-1}. \tag{5}$$

From this result we can see that if there is a large curvature to the misfit function at the model estimate $\hat{\mathbf{m}}$, then the model is well constrained and the variance of the model parameters is small. The probability distribution of the model parameters may then be approximated as a normal pdf with mean $\hat{\mathbf{m}}$ and covariance computed with (5) [11]. The above expression for the model covariance is the Cramer–Rao lower bound for the model uncertainty: the covariance of any unbiased estimator is *at least* that of (5) [12], [13].

The feature vector $\mathbf{x}$ used in a discrimination algorithm is typically a function of the model $\mathbf{m}$. For a transformation $\mathbf{x} = g(\mathbf{m})$ operating on the model vector, the first-order approximation to the transformed model covariance S is

$$\text{cov}(\mathbf{x}) = \mathbf{G}^T\text{cov}(\mathbf{m})\mathbf{G} \tag{6}$$

with the elements of the sensitivity matrix (or Jacobian) of the transformation $\mathbf{G}$ computed as

$$G_{ij} = \frac{\partial g_i(m)}{\partial m_j}. \tag{7}$$

A linearized uncertainty analysis may not be valid if the objective function is nonconvex. In this case, the local quadratic approximation provides a poor approximation to the actual objective function, and uncertainties estimated with (5) may not be reflective of the actual uncertainties in the model. An alternative approach to estimating uncertainties is to use a

Bayesian framework [14], [15] to estimate the model posterior probability distribution

$$p(\mathbf{m}|\mathbf{d}^{\text{obs}}) \propto p(\mathbf{d}^{\text{obs}}|\mathbf{m})p(\mathbf{m}). \tag{8}$$

The posterior $p(\mathbf{m}|\mathbf{d}^{\text{obs}})$ is the product of a likelihood function $p(\mathbf{d}^{\text{obs}}|\mathbf{m})$ and a prior probability $p(\mathbf{m})$. For continuous models, the likelihood function is often assumed to have the form

$$p(\mathbf{d}^{\text{obs}}|\mathbf{m}) \propto \exp[-\phi_d] \tag{9}$$

with ϕ_d the least squares misfit function. The posterior probability is then computed as

$$p(\mathbf{m}|\mathbf{d}^{\text{obs}}) = \frac{\exp[-\phi_d]p(\mathbf{m})}{\int \exp[-\phi_d]p(\mathbf{m})d\mathbf{m}}. \tag{10}$$

Here, the normalizing integral is over all of model space. For a nonlinear forward problem, the normalizing integral is often difficult to evaluate analytically, especially in high-dimensional model spaces. However, the PPD for a nonlinear problem can be estimated numerically using the Metropolis–Hastings algorithm. This algorithm works by randomly perturbing the current model $\mathbf{m}^{\text{current}}$ to a proposed model $\mathbf{m}^{\text{proposed}}$ and by accepting the proposed model according to the Metropolis criterion [16]

$$\begin{aligned}\eta &\leq \exp[-\Delta\phi] \\ &= \exp\left[-\left(\phi(\mathbf{m}^{\text{proposed}}) - \phi(\mathbf{m}^{\text{current}})\right)\right].\end{aligned} \tag{11}$$

If $\exp[-\Delta\phi]$ is less than or equal to η, then the perturbation is accepted and the proposed model becomes the current model. Perturbations that decrease the misfit function are always accepted, while perturbations that increase ϕ are accepted randomly according to the aforementioned criterion. At each model perturbation, η is drawn from a uniform random distribution on the interval [0 1].

This scheme is a Markov chain. Acceptance of the proposed model depends only on the current model. After a sufficient number of samples, the chain of accepted models will converge to a stationary distribution, which is the posterior distribution. We adopt the fast sampler algorithm developed in [17] to estimate the posterior distribution of the model parameters. A key feature of this algorithm is the use of two independent samplers. Convergence of these samplers to the same distribution, as measured by the maximum difference in their cumulative distributions, ensures that the sample provides a reasonable estimate of the PPD.

The PPD is a function in an N-dimensional model space, with N the number of model parameters. It is therefore useful to consider the 1-D marginal distribution of each parameter

$$p(m_i|\mathbf{d}^{\text{obs}}) = \int p(\mathbf{m}|\mathbf{d}^{\text{obs}})dm_1dm_2\ldots dm_{i-1}dm_{i+1}\ldots dm_N. \tag{12}$$

III. Comparison of Linearized and Nonlinear Uncertainty Appraisals for TEM Dipole Model Parameters

As an example of feature vector estimation and uncertainty propagation in the context of UXO discrimination, we first consider TEM data acquired with Geonics EM61 and EM63 sensors at Camp Sibert, AL [7]. The discrimination task for this demonstration project was to distinguish between emplaced 4.2-in mortars and nonhazardous clutter (munitions related scrap, cultural items, etc.).

A. Forward Modeling and Inversion

In this paper, we model the observed data using a dipole forward model [9]. The secondary magnetic field is computed as

$$\mathbf{B}^s(\mathbf{r},t) = \frac{m(t)}{r^3}\left(3\left(\hat{\mathbf{m}}(t)\cdot\hat{\mathbf{r}}\right)\hat{\mathbf{r}} - \hat{\mathbf{m}}(t)\right) \tag{13}$$

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

$$\mathbf{m}(t) = \frac{1}{\mu_o}\mathbf{M}(t)\cdot\mathbf{B_o}. \tag{14}$$

The induced dipole is the projection of the primary field $\mathbf{B_o}$ onto the target's polarization tensor $\mathbf{M}(t)$. The polarization tensor can be decomposed as

$$\mathbf{M}(t) = \mathbf{A}^T\mathbf{L}(t)\mathbf{A} \tag{15}$$

with $\mathbf{A}$ as an orthogonal matrix which rotates the coordinate system from the geographic coordinates to a local body-centered coordinate system. In body-centered coordinates, the tensor is diagonal

$$\mathbf{L}(t) = \begin{bmatrix} L_1(t) & 0 & 0 \\ 0 & L_2(t) & 0 \\ 0 & 0 & L_3(t) \end{bmatrix} \tag{16}$$

with the eigenvalues ordered by convention so that $L_1(t) \geq L_2(t) \geq L_3(t)$. When inverting EM61 data, we estimate a 15-element model vector comprised of target location (x, y, and z), orientation (ϕ, θ, and ψ), and the instantaneous amplitudes of the polarization tensor at three time channels ($L_i(t_j)$, $i = 1,\ 2,\ 3$, and $j = 1,\ 2,\ 3$). In this example, we have computed 2-D feature vectors ($\mathbf{x}$) spanned by the polarization amplitude and the polarization decay

$$x_1 = \text{ Polarization amplitude} = \left(\sum_{i=1}^{3} L_i^2(t_1)\right)^{1/2} \tag{17}$$

$$x_2 = \text{Polarization decay} = \left(\sum_{i=1}^{3} L_i^2(t_m)\right)^{1/2} \Big/ \left(\sum_{i=1}^{3} L_i^2(t_1)\right)^{1/2}. \tag{18}$$

The polarization amplitude is then proportional to the magnitude of the induced dipole moment at the first time channel,

and the polarization decay is the ratio of the magnitude of the induced dipole at the first time channel and some later time t_m. For the EM61, $t_1 = 0.22$ ms, and we use $t_m = t_3 = 0.66$ ms. The EM63 measures the decay of the secondary field at 26 time channels spanning a wider range of off-times, and for these data, Pasion [9] shows that a suitable parameterization of the polarization decays is

$$L_i(t_j) = k_i t_j^{-b_i} \exp(-t_j/g_i), \qquad i = 1, 2, 3. \tag{19}$$

The model vector for inverting EM63 data is then the target location, orientation, and polarization parameters (k_i, b_i, and g_i, $i = 1,\ 2,\ 3$). The recovered parameters can then be used to compute polarization amplitude and decay using (17) and (18), with $t_1 = 0.18$ ms and $t_m = t_{15} = 2.17$ ms for the EM63. An intermediate time channel is used for this sensor because early times do not yet allow discrimination between slow-decaying UXO and fast-decaying clutter, while late times are contaminated by noise [7].

B. Uncertainty Appraisal

Fig. 2(a) shows the distributions of ordnance and nonordnance feature vectors for the EM61 data set. In Fig. 2(b), we zoom in on the region of the feature space populated with ordnance feature vectors and show the linearized uncertainties in the model parameters. We note that the predicted uncertainty of the individual ordnance target parameters is generally much smaller than the overall variance of the ordnance class, with a little correlation between parameter uncertainty and the deviation of a feature vector from the class mean.

To validate the computation of uncertainties shown in Fig. 2, Fig. 3 compares the variances of model parameters estimated by linearized and nonlinear uncertainty appraisals for a single ordnance item in the Camp Sibert EM61 data. This target is indicated by the leftmost arrow in Fig. 2(b). The nonlinear appraisal is run with uniform priors on the model so that agreement between linearized and nonlinear appraisals is expected. Consistent with results in [9] obtained with the neighborhood sampling algorithm [18], the parameter uncertainties propagated to the feature vectors using (6) agree well with those estimated by nonlinear appraisal. The nonlinear marginal pdf of the polarization amplitude has a slight positive skew, resulting in a shift of the mean of this distribution to a slightly larger value than what is obtained with the linearized appraisal.

Fig. 4 shows a second example of model uncertainty for an ordnance item, indicated by the rightmost arrow in Fig. 2(b). In this case, the distribution of model parameters is a multimodal distribution. Linearized uncertainty analysis is a good approximation to one mode of this distribution centered about the minimum misfit model. However, the second mode of the nonlinear model pdf is in better agreement with the expected features for targets belonging to this ordnance class. To further understand this result, in Fig. 5 we show the misfit versus depth (MVD) curve for this inversion. This curve is generated by carrying out multiple inversions of the same data set, with each inversion constrained to lie within a narrow interval of

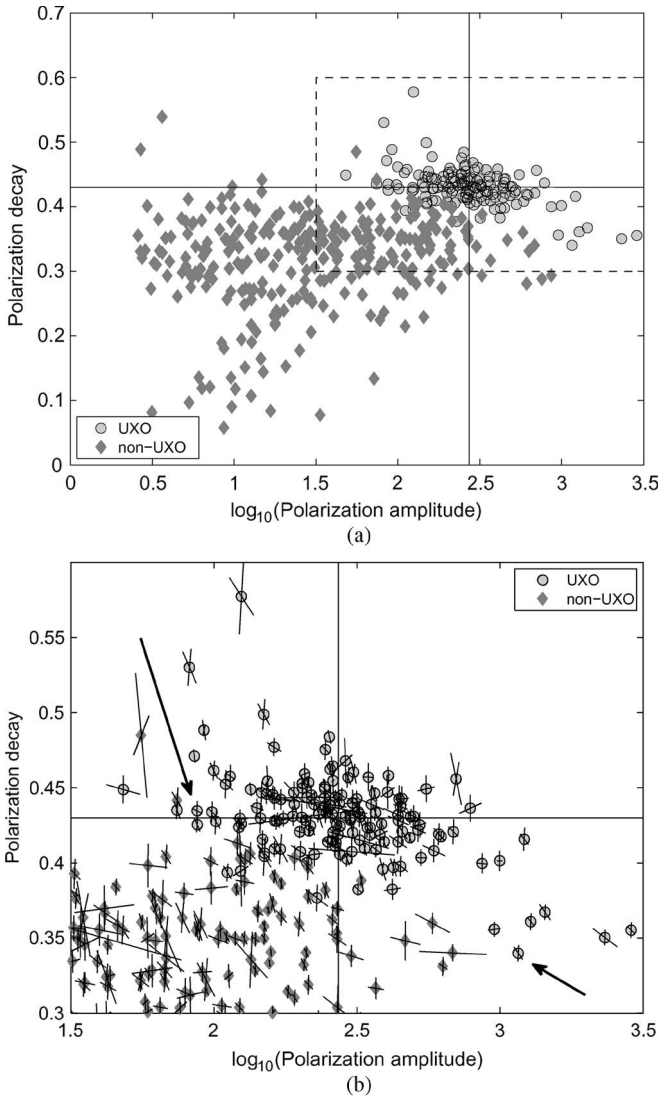


Fig. 2. Feature vectors estimated from the Camp Sibert EM61 data. (a) Estimated feature vectors. The dashed area indicates the plotted region in (b), and the crosshairs show the location of the mean of the UXO class. (b) Estimated feature vectors with local standard deviations along the principal axes. The principal axes do not appear orthogonal because of the unequal scalings of the plot axes. The arrows indicate the targets appraised in Figs. 3 and 4.

target depth [19]. In Fig. 5(a), we display the relative misfit, i.e., the misfit relative to the global minimum value on the MVD curve. Two minima are identified on the MVD curve, and they correspond to the modes of the model pdf in Fig. 4. The shallower local minimum of the MVD curve is a better estimate of true target depth for this case. We can see in Fig. 5(b) that polarization amplitude increases monotonically with depth so that the optimal model at depth overestimates polarization amplitude. This is a consequence of the inverse relationship between target–sensor separation and dipole moment amplitude in (13). Fig. 5(c) shows that for models on the MVD curve, the polarization decay varies nonlinearly with polarization amplitude.

Fig. 6 shows estimated feature vectors and associated linearized uncertainties for the Camp Sibert EM63 data. Parameter uncertainties are generally smaller for the EM63 than for the

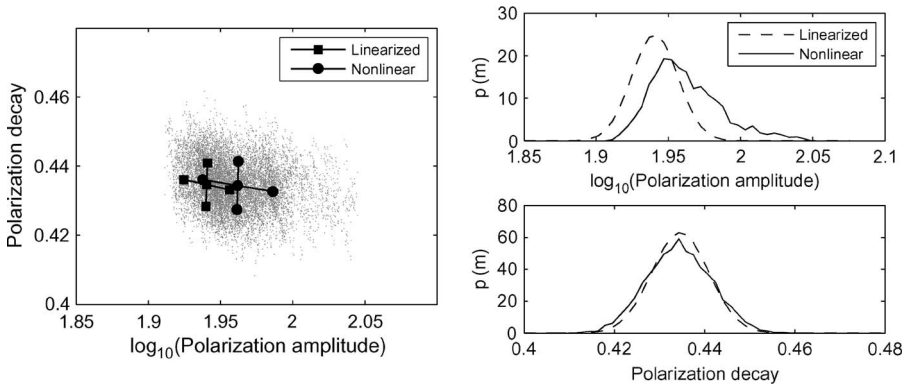


Fig. 3. (Left) Principal axes of the covariance matrix estimated by the linearized and nonlinear uncertainty appraisals of the EM61 data for the target indicated by the left arrow in Fig. 2(b). Also shown are all accepted models from nonlinear sampling. (Right) Marginal parameter distributions from the linearized and nonlinear appraisals.

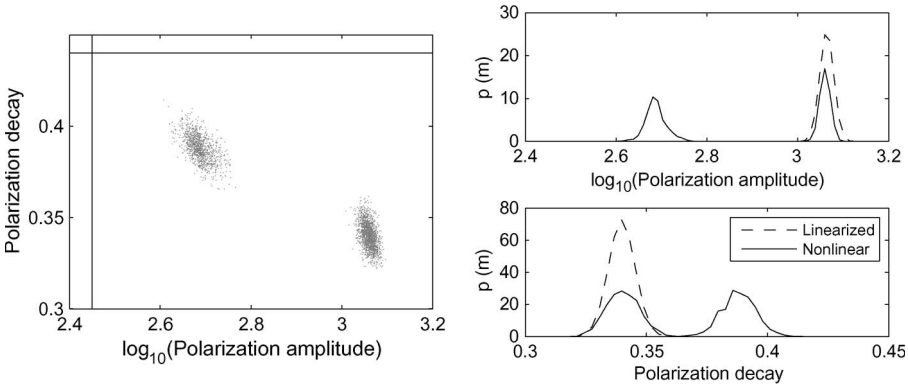


Fig. 4. (Left) Accepted models for the nonlinear appraisal of a single target in the EM61 data, indicated by the right arrow in Fig. 2(b). The crosshairs intersect at the mean of the UXO class. (Right) Marginal parameter distributions from the linearized and nonlinear appraisals.

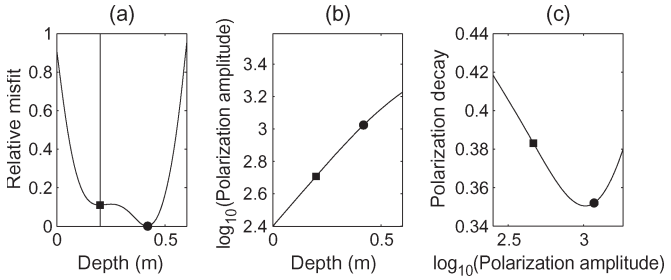


Fig. 5. (a) MVD curve for the EM61 target considered in Fig. 4. The vertical line indicates the minimum of the misfit, which best agrees with the actual target depth. (b) Dependence of the MVD curve model polarization amplitude on the estimated target depth. (c) Trajectory of the MVD curve models in the feature space. In all plots, the markers correspond to models at the minima of the MVD curve.

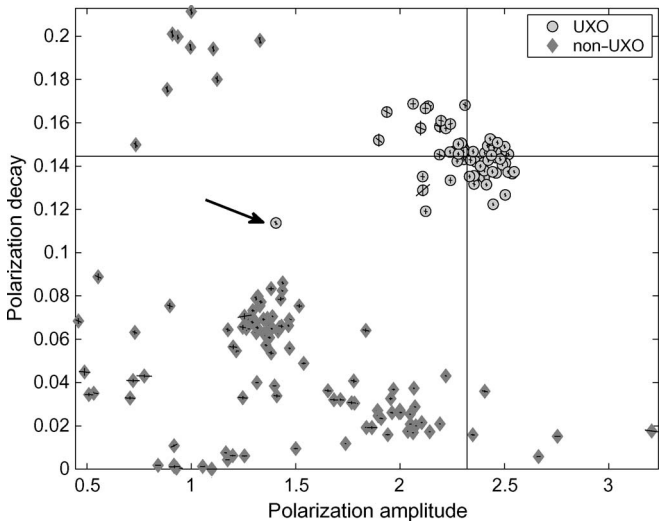


Fig. 6. Feature vectors and linearized uncertainties estimated from the Camp Sibert EM63 data. The arrow indicates the target appraised in Fig. 7.

EM61. The EM63 data were acquired in high-quality cued interrogation surveys over previously identified targets, and so, we expect smaller errors on both the data and the model than on the EM61 detection mode data set.

A similar result to that in Fig. 4 is obtained in Fig. 7 for appraisal of the EM63 target highlighted in Fig. 6: local minima of the misfit function produce a multimodal pdf. In both of these cases, however, the individual modes are approximately Gaussian and the linearized model pdf is a good approximation to its respective mode. The MVD curve for this target has a false global minimum at a shallow depth, resulting in an underestimate of the polarization amplitude expected for this class (Fig. 8).

When data are acquired in a controlled experiment (e.g., test stand measurements), then we expect that the global minimum of the misfit function will correspond to the true model. However, when inverting field measurements, we generally have only the naive estimates of the noise on the observed data, and

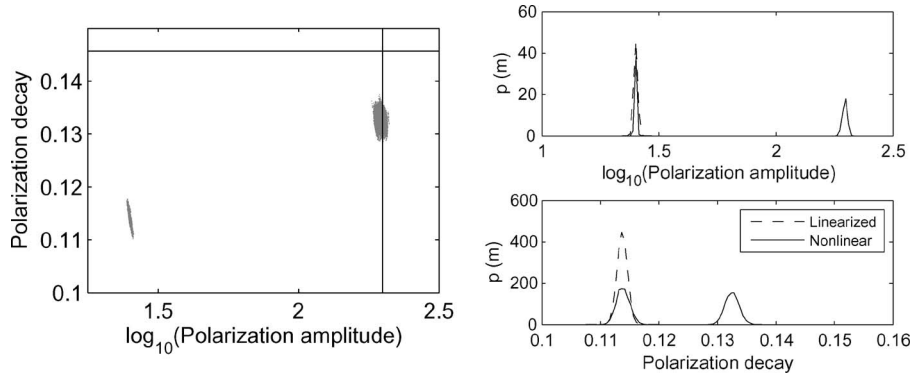


Fig. 7. Accepted models for the nonlinear appraisal of a single target in the EM63 data, highlighted in Fig. 6. The crosshairs intersect at the mean of the UXO class. (Right) Marginal parameter distributions from the nonlinear appraisal.

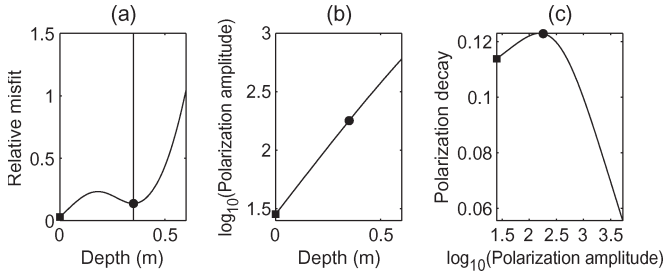


Fig. 8. (a) MVD curve for the EM63 target considered in Fig. 7. The vertical line indicates the minimum of the misfit which best agrees with the actual target depth. (b) Dependence of the MVD curve model polarization amplitude on the estimated target depth. (c) Trajectory of the MVD curve models in the feature space. In all plots, the markers correspond to models at the minima of the MVD curve.

the weightings $(\mathbf{W}_d)$ that we choose have a direct effect on determining the minimum misfit model through (1). For this paper, we have assigned errors as 10% of each observed datum plus a floor value estimated at each time channel. While this choice generally allows us to recover useful estimates of model parameters, in these last two examples, the data are weighted so that a "false" minimum at an incorrect depth becomes the global minimum. From these examples, we therefore conclude the following.

1) The minimum misfit model is not always the best model to use for discrimination.
2) The linearized uncertainty appraisal cannot always account for the deviation of an estimated feature vector from the expected value of that target class.

We can address this problem in a number of ways. First, we might investigate different choices of data weightings or data norm. For example, in [20], robust inversion techniques were applied to the same TEM data sets presented here. We found that robust norms can sometimes reweight minima of the misfit so that the desired minimum (i.e., closest to the true target depth) becomes the global minimum. Robustness, however, does not imply a foolproof inversion procedure, and it may not always unambiguously eliminate local minima as possible solutions to the inverse problem. Similarly, different choices of noise standard deviations can correctly reweight the minima of the misfit, but again, no noise estimation procedure is likely to be universally successful. In light of this ambiguity, here we instead investigate how the estimates of model uncertainty can be incorporated into the discrimination process.

IV. Incorporating Uncertainties in Discrimination

As mentioned previously, feature vectors are normally treated as point estimates when they are used as an input into a discrimination algorithm. Shivaswamy *et al.* [21] developed a support vector machine algorithm which accounts for uncertainty in training feature vectors. Uncertainty is characterized by an ellipsoid about each training feature vector, and the algorithm tries to find a separating hyperplane such that all points in an ellipsoid lie on the correct side of the hyperplane. However, this method does not account for the uncertainty in test vectors in the prediction stage.

Here, we account for uncertainties in both test and training data. We adopt a generative approach to classification where we model the probability distributions of classes and then use these probability distributions to make predictions for the uncertain test data. In this development, we assume for simplicity that the posterior distributions of individual training and test feature vectors are Gaussian. Multimodal distributions, as encountered in the previous section, can be approximated as a mixture of Gaussians [22], and so, this algorithm can readily be extended to classification with nonparametric pdfs.

In a conventional generative classifier, we typically fit a parametric distribution, such as a Gaussian, to the empirical distribution of feature vectors in each class. Consider a sample of N training vectors (x^{train}) in one dimension, all belonging to the same class ω (e.g., $\omega \in \{T, F\}$, with T denoting a UXO and F denoting a clutter). In the absence of uncertainty, the empirical distribution of these data can be represented as a superposition of delta functions

$$p(x|\omega) = \frac{1}{N}\sum_{i=1}^{N} \delta\left(x - x_i^{\text{train}}\right). \tag{20}$$

The mean and variance of this distribution are the sample mean and variance

$$\mu = E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i^{\text{train}}$$

$$\sigma^2 = E\left((x-\mu)^2\right) = \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i^{\text{train}} - \mu\right)^2. \tag{21}$$

For the sample variance, the normalization must be changed to $N-1$ to obtain an unbiased estimator. When feature vectors are uncertain, the empirical distribution becomes

$$p(x|\omega) = \frac{1}{N}\sum_{i=1}^{N} p_i^{\text{train}}(x) \tag{22}$$

where p_i^{train} is the posterior distribution $p_i(x|\mathbf{d}^{\text{obs}})$ obtained from uncertainty appraisal of the ith target. Assuming the p_i^{train} are normally distributed with means x_i^{train} and variances $(\sigma_i^{\text{train}})^2$, the mean and standard deviation of $p(x|\omega)$ are

$$\begin{aligned}
\mu &= E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i^{\text{train}} \\
\sigma^2 &= E\left((x-\mu)^2\right) \\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\int \left(x^2 p_i^{\text{train}}(x) - 2x\mu p_i^{\text{train}}(x) \right.\right. \\
&\qquad \left.\left. + \mu^2 p_i^{\text{train}}(x)\right)dx\right) \\
&= \frac{1}{N-1}\sum_{i=1}^{N}\left(x_i^{\text{train}}-\mu\right)^2 + \frac{1}{N}\sum\left(\sigma_i^{\text{train}}\right)^2 \\
&= \sigma_B^2 + \sigma_L^2
\end{aligned} \tag{23}$$

with the expression for the variance following from the identity $E(x^2) = \sigma^2 + \mu^2$. Again, the normalization on the first term of the variance is modified to obtain an unbiased estimator of the variance of the training vectors. We can see that the total variance of the class distribution can be decomposed into contributions arising from the variability between feature vectors (σ_B) and the local uncertainties about feature vectors (σ_L). When there is no uncertainty in the feature vectors, we again obtain the expressions in (21). This is analogous to the variance decomposition which is employed in analysis of variance or canonical analysis [23]. The estimates of the mean and standard deviation from the aforementioned expressions can then be used as the moments of the class distribution $p(x|\omega)$. In the generalization of discriminant analysis developed here, we approximate the mixture of Gaussians in (22) by a single Gaussian, with mean and variance given by (23). When dealing with multivariate data, we can replace the variances in (23) with covariances.

We now turn to classification of uncertain test feature vectors, generalized to multidimensional feature spaces. A binary discrimination algorithm computes the probability of membership of an estimated vector $\hat{\mathbf{x}}$ in class ω using Bayes' rule

$$\begin{aligned}
p(T|\hat{\mathbf{x}}) &= \frac{p(\hat{\mathbf{x}}|T)p(T)}{p(\hat{\mathbf{x}}|T)p(T) + p(\hat{\mathbf{x}}|F)p(F)} \\
p(F|\hat{\mathbf{x}}) &= 1 - p(T|\hat{\mathbf{x}})
\end{aligned} \tag{24}$$

where $p(\hat{\mathbf{x}}|T)$ is the T class distribution estimated from the training data and $p(T)$ is the prior probability of the T class (and similarly for class F).

When $\mathbf{x}$ is uncertain with posterior probability distribution $p(\mathbf{x}|\mathbf{d}^{\text{obs}})$, then the joint probability distribution can be factored

$$p(\mathbf{d}^{\text{obs}}, \mathbf{x}, \omega) = p(\mathbf{d}^{\text{obs}}|\mathbf{x})p(\mathbf{x}|\omega)p(\omega). \tag{25}$$

This factorization implies that given the feature vector $\mathbf{x}$ (via the model $\mathbf{m}$), the observed data are independent from the class ω. We assume this because the mapping from the model to observed data

$$\mathbf{d}^{\text{obs}} = F[\mathbf{m}] + \epsilon \tag{26}$$

is independent of class ω (here, ϵ is the additive random noise). Applying Bayes rule to (25), we have

$$p(\mathbf{x}, \omega|\mathbf{d}^{\text{obs}}) \propto p(\mathbf{x}|\mathbf{d}^{\text{obs}})p(\mathbf{x}|\omega)p(\omega). \tag{27}$$

Marginalizing over the uncertain feature vector $\mathbf{x}$ gives

$$p(\omega|\mathbf{d}^{\text{obs}}) \propto \int p(\mathbf{x}|\mathbf{d}^{\text{obs}})p(\mathbf{x}|\omega)p(\omega)d\mathbf{x}. \tag{28}$$

Normalizing $p(\omega|\mathbf{d}^{\text{obs}})$ over all values of ω yields the posterior probabilities of the class membership, shown at the bottom of the page. Intuitively, expression (29) evaluates the Bayes rule over all possible values of $\mathbf{x}$, weighted by the probability of each respective value. When there is no uncertainty $(p(\mathbf{x}|\mathbf{d}^{\text{obs}}) = \delta(\mathbf{x}-\hat{\mathbf{x}}))$, the expressions in (29) reduce to (24). The required integrals must be evaluated over the entire feature space. For multivariate normal distributions with means μ_i and covariances S_i, the integral can be solved analytically, with the result that the integral of two Gaussians is itself a Gaussian distribution with moments [24]

$$\begin{aligned}
\mu &= S_1^{-1}\mu_1 + S_2^{-1}\mu_2 \\
S &= \left(S_1^{-1} + S_2^{-1}\right)^{-1}.
\end{aligned} \tag{30}$$

We therefore term a classifier employing (29) with Gaussian distributions for both classes and feature vectors a "Gaussian product" (GP) classifier.

The GP algorithm incorporates the covariance of a test feature vector in classification. As illustrated in Fig. 9, the relative

$$\begin{aligned}
p(T|\mathbf{d}^{\text{obs}}) &= \frac{\int p(\mathbf{x}|\mathbf{d}^{\text{obs}})p(\mathbf{x}|T)p(T)d\mathbf{x}}{\int p(\mathbf{x}|\mathbf{d}^{\text{obs}})p(\mathbf{x}|T)p(T)d\mathbf{x} + \int p(\mathbf{x}|\mathbf{d}^{\text{obs}})p(\mathbf{x}|F)p(F)d\mathbf{x}} \\
p(F|\mathbf{d}^{\text{obs}}) &= 1 - p(T|\mathbf{d}^{\text{obs}})
\end{aligned} \tag{29}$$

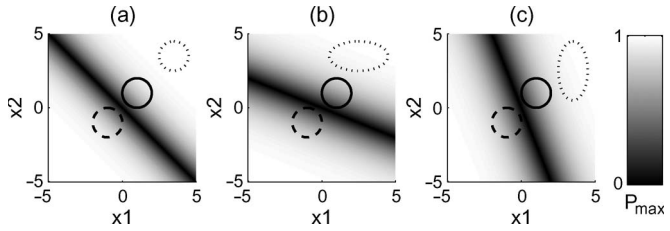


Fig. 9. Classification with the GP in two dimensions. (a) Two class discrimination problem with one standard deviation ellipses of classes shown as solid and dashed lines. Grayscale image is the decision surface of the Gaussian product classifier for a test feature vector with equal standard deviations along principal axes (dotted line). (b) As in (a) but with the decision surface computed for a test feature vector with $\sigma_{x1} = 2\sigma_{x2}$, as indicated by dotted line. (c) Decision surface for $\sigma_{x2} = 2\sigma_{x1}$.

sizes of the variances along principal directions of test vector covariance affect the orientation of the decision boundary. The decision surface for a classifier is generated by evaluating the maximum probability output by the classifier

$$p_{\max}(\mathbf{x}) = \max\left(p(T|\mathbf{x}), p(F|\mathbf{x})\right) \tag{31}$$

over a grid of points. This generates an image of the classifier output in the feature space, with $P(T|\mathbf{x}) = P(F|\mathbf{x}) = 0.5$ corresponding to the decision boundary. For the GP classifier, a separate decision surface is generated for each possible test vector covariance. In Fig. 9, we show three decision surfaces corresponding to three different specifications of test vector covariances. In Fig. 9(a), the covariance is isotropic, and so, a linear decision boundary, identical to that generated when there is no uncertainty, is obtained. In Fig. 9(b) and (c), we successively increase the variance of the test vector along each orthogonal direction in the feature space. This rotates the decision boundary such that its normal is closer to orthogonal to the direction of the maximal test vector uncertainty. This implies that when we make classification decisions for an individual test vector, we rely most upon the element in that vector which is least uncertain. The rotation of the decision boundary shown here is similar to the result in [21]: their support vector machine formulation rotates the decision boundary so that the covariance ellipses of the training data lie on the correct side of the boundary. In this case, however, the orientation of the boundary is determined by uncertainties in both test and training data.

V. Application to the Camp Sibert Data Sets

Application of the GP classifier requires estimation of the uncertainty for all inverted targets. As demonstrated above, linearized uncertainty evaluated about the minimum misfit model does not adequately characterize the multimodal pdfs encountered for TEM model parameters. While nonlinear appraisal can be carried out for all targets, here we find that an efficient solution is to approximate the multimodal distributions with an ensemble of models obtained by repeatedly minimizing the misfit with an iterative algorithm. We initialize these inversions with a number of models selected over a range of possible target depths. To define the feature vectors for test items, we consider the set of converged models from our repeated iterative inversions. If a subset of these models is within a range of 5 cm

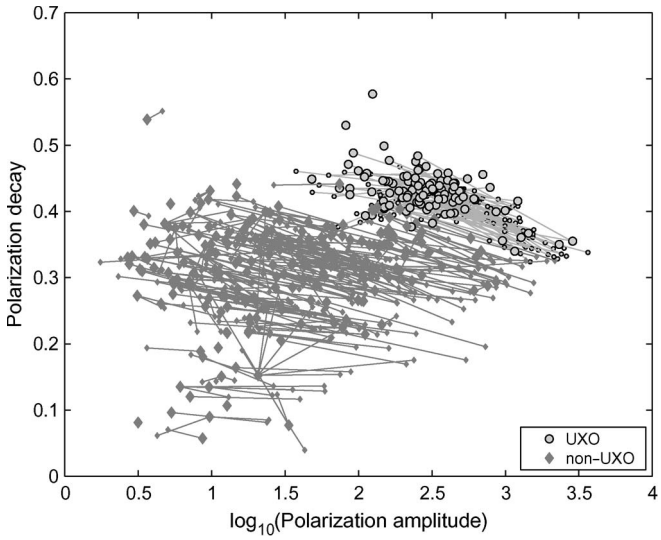


Fig. 10. EM61 test data showing multiple feature vectors per target. The feature vectors for a given target are connected by a line, with the largest marker indicating the minimum misfit model.

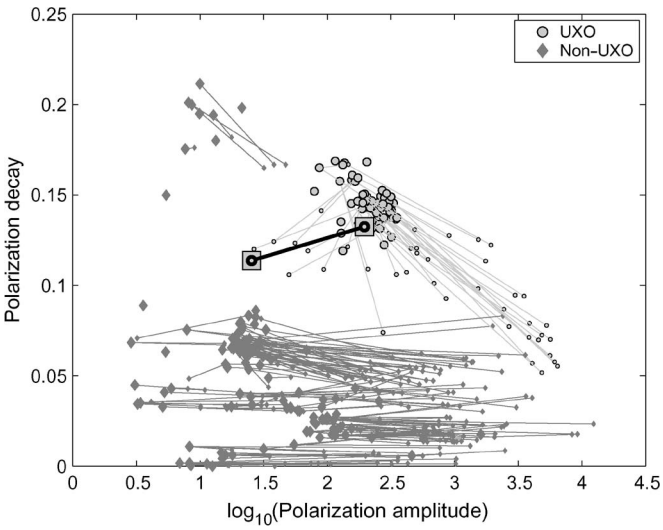


Fig. 11. Feature vectors estimated from the Camp Sibert EM63 data. The feature vectors for a given target are connected by a line, with the largest marker indicating the minimum misfit model. The highlighted feature vector is an outlier to the ordnance class.

in depth, then we select the minimum misfit model from this subset to use for discrimination. Most targets have at least two unique feature vectors identified in this manner (from a total of ten converged inversions per target). Figs. 10 and 11 show EM61 and EM63 test data generated with this approach.

We then evaluate the linearized uncertainty about each of these feature vectors, using (5). This provides us with N kernels of a Gaussian mixture model which approximates the PPD as

$$p(\mathbf{x}|\mathbf{d}^{\text{obs}}) \approx \sum_{i=1}^{N} w_i g(\mathbf{x}|\mu_i, S_i) \tag{32}$$

with w_i the kernel weights and $g(\mathbf{x}|\mu_i, S_i)$ a multivariate Gaussian with mean μ_i and covariance S_i. Expectation maximization is typically used to obtain the maximum likelihood

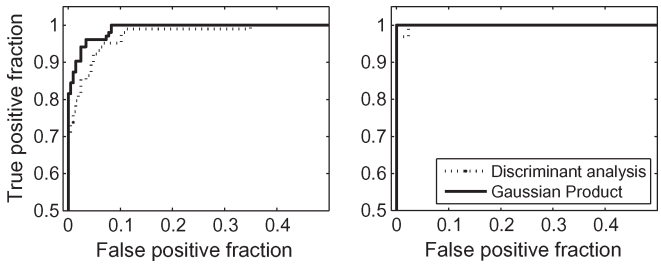


Fig. 12. ROCs for the classifiers applied to the (left) EM61 and (right) EM63 test data.

estimates of weights, means, and covariances of Gaussian mixtures [22]. However, in the absence of a sample from the full nonlinear pdf, we cannot estimate the weights w_i. While various approximations for the weights might be pursued, we find that a simple and effective solution is to classify each target on the basis of the feature vector which is most likely a UXO. That is, we evaluate the Gaussian product (29) using the linearized uncertainty about each model in our ensemble (i.e., each kernel in the mixture model). The kernel with maximal probability of membership in the UXO class is then used to classify the respective target.

Fig. 12 compares receiver operating characteristics (ROCs) obtained with this approach with conventional quadratic discriminant analysis using only the minimum misfit feature vector to classify each target. For both EM61 and EM63 data sets, the area under the ROC and the false alarm rate at $\mathrm{Pd} = 1$ (i.e., the proportion of false positives required to identify all true positives) are improved by the GP classifier. While the improvement for the EM63 data appears negligible, the identification of one outlying 4.2-in mortar is a significant result from the perspective of a regulator charged with site remediation. Similarly, the significant reduction in false alarm rate at $\mathrm{Pd} = 1$ for the EM61 data improves the likelihood that all ordnance will be identified with this sensor.

VI. Conclusion

In this paper, we have investigated uncertainty in the dipole model parameters estimated from TEM data. We found that linearized uncertainty estimates are a good approximation to the distribution of the model parameters about local minima. However, in many cases, the data misfit has multiple minima, and a linearized uncertainty appraisal about the global minimum alone cannot fully capture the variability of the model. A nonlinear appraisal with Markov sampling can be used to explore model space and to estimate a multimodal density. The multimodal form of the model probability density may have implications for sensor and survey design. For example, [25] minimized local uncertainty of dipole model parameters as a function of receiver placement and orientation. This analysis could be repeated with an aim to eliminating local minima of the MVD curve, thereby improving parameter estimation and subsequent discrimination. Initial results with multistatic systems which can measure multiple components of the secondary magnetic field already suggest that these systems overcome many of the complications encountered here with monostatic sensors.

To account for model uncertainty in discrimination, we have developed the GP classifier. This algorithm is a generalization of discriminant analysis, which incorporates uncertainty in both test and training feature vectors. Rather than carrying out a full nonlinear appraisal for all targets in the Camp Sibert test data, we characterized model uncertainties via repeated iterative inversions with different initializations. This approach identified modes of the model pdf, and linearized appraisal was used to characterize the uncertainty about each mode. Targets were then classified with the GP algorithm using the mode that was most probably a UXO. This technique improved detection of the outlying feature vectors in both EM61 and EM63 data sets.

Aliamiri *et al.* [26] investigated dipole parameter uncertainty by forward modeling data for a range of target locations and orientations and then inverted these synthetic data to obtain expected parameter distributions for a given target class in the feature space. These class distributions were non-Gaussian, and so, a nonparametric model of class distributions was used for discrimination. In this paper, we have used discriminant analysis (or its GP generalization). This algorithm assumes that class distributions are normally distributed. While this assumption is certainly not true for the distribution of clutter in the feature space, discriminant analysis can often be successfully employed on non-Gaussian data [22], and it gave good results for the data sets considered here. There is no difficulty, however, in generalizing the GP classifier to represent arbitrary class distributions as mixtures of Gaussian kernels.

The concept of using multiple feature vectors when classifying test data can be regarded as an extension of the technique in [26] to the test data. While synthetics can augment training data and help to gain a sense of the variability of features, it is impossible to fully anticipate the complications (i.e., noise) which will be encountered in field data. By allowing a test target to be represented by a number of possible models, we can classify the target conditional upon the observed data and possibly detect outliers which are not represented in the training data.

References

[1] F. Shubitidze, K. O'Neill, K. Sun, and K. D. Paulsen, "Investigation of broadband electromagnetic induction scattering by highly conductive, permeable, arbitrarily shaped 3-D objects," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 540–556, Mar. 2004.

[2] Y. Zhang, X. Liao, and L. Carin, "Detection of buried targets via active selection of labeled data: Application to sensing subsurface UXO," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 11, pp. 2535–2543, Nov. 2004.

[3] L. S. Beran and D. W. Oldenburg, "Selecting a discrimination algorithm for unexploded ordnance remediation," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 9, pp. 2547–2557, Sep. 2008.

[4] B. Zhang, K. ONeill, J. Kong, and T. Grzegorczyk, "Support vector machine and neural network classification of metallic objects using coefficients of the spheroidal MQS response modes," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 1, pp. 159–171, Jan. 2008.

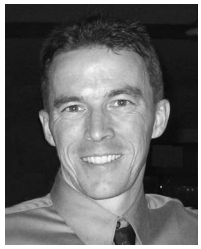
[5] S. L. Tantum, Y. Li, and L. M. Collins, "Bayesian mitigation of sensor position errors to improve unexploded ordnance detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 103–107, Jan. 2008.

[6] X. Liao and L. Carin, "Migratory logistic regression for learning concept drift between two data sets with application to UXO sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1454–1466, May 2009.

[7] S. D. Billings, L. R. Pasion, L. Beran, N. Lhomme, L. Song, D. W. Oldenburg, K. Kingdon, D. Sinex, and J. Jacobson, "Unexploded ordnance discrimination using magnetic and electromagnetic sensors: Case study from a former military site," *Geophysics*, vol. 75, no. 3, pp. B103–B114, 2010.
[8] S. L. Tantum and L. M. Collins, "Performance bounds and a parameter transformation for decay rate estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, p. 2241-2231, Oct. 2003.
[9] L. R. Pasion, "Inversion of time-domain electromagnetic data for the detection of unexploded ordnance," Ph.D. dissertation, Univ. British Columbia, Vancouver, BC, Canada, 2007.
[10] J. J. Remus and L. M. Collins, "Phenomenological model inversion with Fisher information metrics for unexploded ordnance detection," in *Proc. Int. Geosci. Remote Sens. Symp.*, 2010, pp. 691–694.
[11] W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory*. New York: Academic Press, 1989.
[12] H. L. van Trees, *Detection, Estimation, and Modulation Theory, Volume I*. Hoboken, NJ: Wiley, 2001.
[13] X. Liao and L. Carin, "Application of the theory of optimal experiments to adaptive electromagnetic-induction sensing of buried targets," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 8, pp. 961–972, Aug. 2004.
[14] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM, 2005.
[15] M. Sen and P. Stoffa, *Global Optimization Methods in Geophysical Inversion*. Amsterdam, The Netherlands: Elsevier, 1995.
[16] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.
[17] S. E. Dosso, "Environmental uncertainty in ocean acoustic source localization," *Inverse Problems*, vol. 19, no. 2, pp. 419–431, Apr. 2003.
[18] M. Sambridge, "Geophysical inversion with a neighbourhood algorithm I. Searching a parameter space," *Geophys. J. Int.*, vol. 138, no. 2, pp. 479–494, Aug. 1999.
[19] N. Lhomme, D. W. Oldenburg, L. R. Pasion, D. Sinex, and S. D. Billings, "Assessing the quality of electromagnetic data for the discrimination of UXO using figures of merit," *J. Eng. Environ. Geophys.*, vol. 13, pp. 165–176, Sep. 2008.
[20] L. Beran, "Discrimination algorithms for the remediation of unexploded ordnance," Ph.D. dissertation, Univ. British Columbia, Vancouver, BC, Canada, 2010.
[21] P. K. Shivaswamy, C. Battacharyya, and A. J. Smola, "Second order cone programming approaches for handling missing and uncertain data," *J. Mach. Learn. Res.*, vol. 7, pp. 1283–1314, Dec. 2006.
[22] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. New York: Spring-Verlag, 2001.
[23] R. E. Walpole, R. H. Myers, S. L. Myers, and K. Ye, *Probability and Statistics for Engineers and Scientists*. Englewood Cliffs, NJ: Prentice Hall, 2007.
[24] M. Brookes, *The Matrix Reference Manual*. London, U.K.: Imperial College, 2005.
[25] J. T. Smith and H. F. Morrison, "Optimizing receiver configurations for resolution of equivalent dipole polarizabilities in situ," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 7, pp. 1590–1498, Jul. 2005.
[26] A. Aliamiri, J. Stalnaker, and E. L. Miller, "Statistical classification of buried unexploded ordnance using nonparametric prior models," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 9, pp. 2794–2806, Sep. 2007.

Laurens Beran received the M.Sc. and Ph.D. degrees in geophysics from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2005 and 2010, respectively.

He is currently a Geophysicist with Sky Research Inc., Vancouver, and a Postdoctoral Fellow with UBC. His research focuses on processing of geophysical data for the detection and discrimination of unexploded ordnance.



Stephen Billings received the Ph.D. degree from the University of Sydney, Sydney, Australia, in 1998.

He is currently the Vice President of Research and Development with Sky Research Inc., Vancouver, BC, Canada, and an Adjunct Professor with the Department of Earth and Ocean Sciences, University of British Columbia (UBC), Vancouver, BC, Canada. He has spent the past ten years researching methods for the detection and discrimination of unexploded ordnances. Between 2001 and 2003, he was a Postdoctoral Fellow with the Geophysical Inversion Facility, UBC.

Doug Oldenburg received the B.Sc. (Honors) degree in physics and the M.Sc. degree in geophysics from the University of Alberta, Calgary, AB, Canada, and the Ph.D. degree in earth sciences from the University of California, San Diego, in 1974.

After a three-year Postdoc, he joined the University of British Columbia, Vancouver, BC, Canada, where he is currently a Professor and the Director of the Geophysical Inversion Facility.