

Robust Inversion of Time-domain Electromagnetic Data: Application to Unexploded Ordnance Discrimination

Laurens Beran^{1,2}, Stephen Billings² and Doug Oldenburg¹

¹Department of Earth and Ocean Sciences, University of British Columbia, 6339 Stores Road, Vancouver, British Columbia, Canada, V6T 1Z4

²Sky Research Inc., Suite 112A, 2386 East Mall, Vancouver, British Columbia, Canada, V6T 1Z3

ABSTRACT

We invert time-domain electromagnetic data for the purpose of discriminating between buried unexploded ordnance (UXO) and non-hazardous metallic clutter. The observed secondary magnetic field radiated by a conductor is forward modeled as a linear combination of decaying, orthogonal dipoles. We show via a perturbation analysis that errors in the measurement of sensor position propagate to non-normal errors on the observed data. A least squares (L2) inversion assumes normal errors on the data, so non-normal errors have the potential to bias dipole parameter estimates. In contrast, robust norms are designed to downweight the effect of outlying (noisy) data and so can provide useful parameter estimates when there is a non-normal component to the noise.

When positional errors are modeled as independent Gaussian perturbations, we find that weighted least squares and robust inversions have comparable performance. Both inversion techniques estimate data uncertainties from observed data, and this has the effect of making the least squares inversion robust to outliers. However, when simulated errors are correlated, robust inversion with a bisquare norm provides a marked improvement over L2 inversion. Application of robust inversion to real data sets from Camp Sibert, Alabama produced an incremental improvement to the initial L2 inversion, identifying outlying ordnance items and improving discrimination performance.

Introduction

Estimating model parameters from observed geophysical data is a crucial step in discriminating between hazardous unexploded ordnance (UXO) and benign clutter items. Inversion involves minimizing some norm of the discrepancies between observed (\mathbf{d}^{obs}) and predicted (\mathbf{d}^{pred}) data. The parameters of the best-fitting model can then be used to infer a target's intrinsic properties (size, shape, material composition, etc.). The least squares (L2) norm is most commonly chosen as a misfit function

$$\phi = |\mathbf{W}_d (\mathbf{d}^{\text{obs}} - \mathbf{d}^{\text{pred}})|^2 \quad (1)$$

with \mathbf{W}_d a weighting matrix. When the weightings are selected to be the inverse standard deviations of the data ($W_{ii} = 1/\sigma_i$), then minimization of the L2 norm is equivalent to maximizing the likelihood function of the data (Menke, 1989). This assumes that

$$d_i^{\text{obs}} = d_i^{\text{pred}} + \varepsilon_i, \quad (2)$$

the noise on the data is independent and Gaussian distributed ($\varepsilon_i \sim \mathcal{N}(0, \sigma_i)$). While the central limit theorem can be employed to justify the assumption of Gaussian noise, it is often difficult in practice to characterize the uncertainties on the data.

Of particular concern when inverting time-domain electromagnetic (TEM) data for UXO discrimination is the effect of sensor positional error on the uncertainty in the data. Bell (2005) showed that reliable parameter estimation and discrimination from TEM data requires a relative positional accuracy on the order of 2 cm. This accuracy can be nominally achieved with laser positioning systems, or by deploying sensors in a cued-interrogation mode where measurements are made on a template. For data that do not meet these requirements, algorithms have been developed to account for positional error in the inversion process. For example, Tarokh and Miller (2007) impose box constraints on positional errors and then estimate model parameters with a min-max approach: they maximize the misfit function with respect to positional errors, and then minimize this same function

with respect to the model parameters. They show that this former maximization can be determined by evaluating the misfit function at the extrema (corners) of the positional box constraints. Tatum *et al.* (2008) adopt a Bayesian approach, treating positional errors as nuisance parameters and marginalizing these parameters from the posterior distribution of the model parameters. Both these approaches provide improvements over conventional least squares minimization, though some additional computational cost is incurred, especially by the Bayesian approach, which requires Monte Carlo integration of the posterior distribution.

In this study we further examine the role of positional error on inversion of TEM data. We first show that the distributions of data residuals in the presence of positional error do not follow a Gaussian distribution. This motivates consideration of robust norms which are designed to provide reliable parameter estimates when a proportion of the data is contaminated with non-Gaussian noise. We carry out simulations to understand improvements in parameter estimation afforded by robust estimation in the presence of both uncorrelated and correlated noise. Finally we apply robust inversion to real datasets acquired in detection and cued-interrogation surveys.

Forward Modeling

We use the dipole model (Bell, 2005; Pasion, 2007) to predict observed TEM data and to investigate the effect of positional error on the data and the model estimated from those data. The secondary magnetic field is computed as:

$$\mathbf{B}_s(\mathbf{r}, t) = \frac{\mathbf{m}(t)}{r^3} (3(\hat{\mathbf{m}}(t) \cdot \hat{\mathbf{r}})\hat{\mathbf{r}} - \hat{\mathbf{m}}(t)), \quad (3)$$

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

$$\mathbf{m}(t) = \frac{1}{\mu_0} \mathcal{M}(t) \mathbf{B}_o. \quad (4)$$

The induced dipole is the projection of the primary field \mathbf{B}_o onto the target's polarizability tensor $\mathcal{M}(t)$. The primary field radiated by a finite transmitter loop can be computed with the Biot-Savart law and is approximately dipolar in the far-field. In the analysis of positional errors presented in the next section we assume, for simplicity, a dipolar primary field, in which case the above expressions indicate that the observed data will have a $1/r^6$ dependence. However, for forward modeling and inversion of sensor data we always use the Biot-Savart law to accurately compute the primary field.

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

$$\mathcal{M}(t) = \mathcal{A}^T \mathbf{P}(t) \mathcal{A} \quad (5)$$

with \mathcal{A} an orthogonal matrix which rotates the coordinate system from geographic coordinates to a local, body centered coordinate system and

$$\mathbf{P}(t) = \begin{pmatrix} P_1(t) & 0 & 0 \\ 0 & P_2(t) & 0 \\ 0 & 0 & P_3(t) \end{pmatrix} \quad (6)$$

with the eigenvalues ordered by convention so that $P_1(t) \geq P_2(t) \geq P_3(t)$. The eigenvalues decay independently in time and can be parameterized according to:

$$P_i(t) = k_i t^{-\beta} \exp\left(-\frac{t}{\gamma_i}\right), \quad i = 1, 2, 3, \dots \quad (7)$$

The parameters k_i, β_i , and γ_i are then related to the intrinsic properties of the target (Pasion, 2007). Another option is to parameterize the model at each time channel directly in terms of the instantaneous polarizability amplitudes $P_i(t)$, with no functional dependence between polarizabilities at different times.

Decomposing the polarizability tensor with Eq. 5 parameterizes the model in an orthogonal coordinate system which is assumed to correspond to axial and transverse coordinates of a target. However, this parameterization introduces additional nonlinearity into the forward model: the rotation matrix \mathcal{A} is a nonlinear function of target orientation. Parameterizing the polarizability decay with an expression such as Eq. 7 also produces a nonlinear dependence between data and model. A final source of nonlinearity in the forward model is the $1/r^6$ dependence arising from the dipolar primary and secondary fields. All of these nonlinearities complicate the corresponding inverse problem. Iterative algorithms may converge to local minima of the chosen objective function and thereby produce model estimates that are far from the true parameters.

One way to address these complications is to repeatedly solve a related linear problem. If the target location is known, then the forward model for the data at a single time channel is linear in terms of the elements of $\mathcal{M}(t)$, the non-diagonalized polarizability tensor. Solving this linear problem over a range of proposed target locations provides a preliminary search of model space and can help identify local minima of the misfit and starting models for subsequent nonlinear inversion.

Positional Errors

In this section we provide an analysis of the distribution of the data when the received signal follows

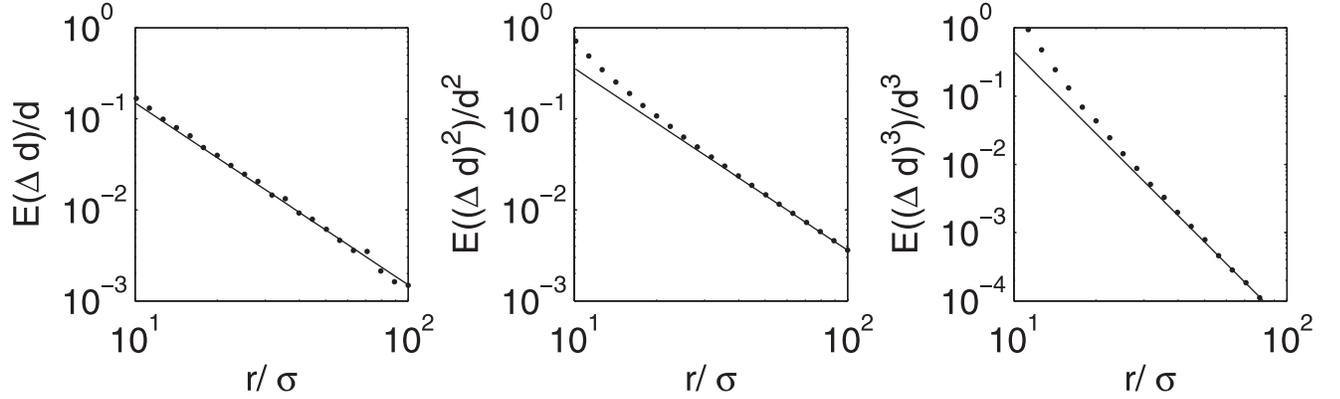


Figure 1. Dependence of the first three normalized moments of the data residuals on the ratio of target-sensor separation (r) and positional standard deviation (σ). Moments are normalized by the true datum (d). Points indicate Monte Carlo simulations, solid lines are predictions for moments given by Eqs. 10 to 12.

a $1/r^6$ decay. We show that the distribution of data residuals about the true datum has positive skew and so we might expect to see large positive outliers and non-normal distributions of residuals.

Consider a datum

$$d = f(\mathbf{r}) = \frac{1}{r^6} \quad (8)$$

with \mathbf{r} the position vector between sensor and target, and r its norm. Assume that positional error is a normally distributed random variable $\Delta \mathbf{r} = N(0, S)$, with S the covariance matrix. For simplicity, assume independent positional errors with identical variance (σ^2), so that $S = \sigma^2 \mathbf{I}$. Addition of positional error is approximated by a Taylor series:

$$f(\mathbf{r} + \Delta \mathbf{r}) = f(\mathbf{r}) + \mathcal{J} \Delta \mathbf{r} + \frac{1}{2} (\Delta \mathbf{r})^T \mathcal{H} \Delta \mathbf{r} \quad (9)$$

The distribution of a datum about its true value is then given by $p(\Delta d) = p(f(\mathbf{r} + \Delta \mathbf{r}) - f(\mathbf{r}))$. We can characterize this distribution by its moments. The first three moments are, to leading order,

$$E(\Delta d) = \frac{1}{2} \sum_{i=1}^3 \frac{\partial^2 f}{\partial x_i^2} \sigma_i^2 \propto \frac{\sigma^2}{r^8} \quad (10)$$

$$E((\Delta d)^2) = \sum_{i=1}^3 \left(\frac{\partial f}{\partial x_i} \right)^2 \sigma_i^2 \propto \frac{\sigma^2}{r^{14}} \quad (11)$$

$$E((\Delta d)^3) = 3 \sum_{i=1}^3 \left(\frac{\partial f}{\partial x_i} \right)^2 \frac{\partial^2 f}{\partial x_i^2} \sigma_i^4 + 3 \sum_{i=1}^3 \sum_{j \neq i} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \frac{\partial^2 f}{\partial x_i \partial x_j} \sigma_i^4 \sigma_j^4 \propto \frac{\sigma^4}{r^{22}}. \quad (12)$$

Here x_i is the i^{th} basis vector in the chosen coordinate system. Now for $f(\mathbf{r})$ given by Eq. 8, the third moment is strictly positive. This implies that the distribution of residuals has positive skew, *i.e.*, there are larger magnitude positive outliers than negative outliers. This is in contrast to the generating distribution of positional errors, which has zero skew.

Figure 1 shows the agreement of the moments of the distribution of residuals computed by simulation and using the above expressions. It is important to note that the distribution of residuals at a given location is determined by the dimensionless ratio of r/σ . There is some disagreement between simulations and the predicted second and third moments for small r/σ , likely because of contributions of higher order terms. We see that the data residuals have nonzero mean, with the mean proportional to the variance of the positional error (σ^2) and inversely proportional to target sensor separation. Similarly, the variance and skew of the residuals decreases nonlinearly as we increase r or decrease σ . This result suggests that the closer the sensor is to a target, or the larger the positional error, then the larger the variance and the more likely we are to encounter outlying data.

In Fig. 2 we show the distributions of residuals generated by simulation for a few cases, together with maximum likelihood estimates of generalized extreme value (GEV) distributions (Gumbel, 1958). There is a good correspondence between the empirical and fitted distributions, however any distribution that can have nonzero skew will likely provide an adequate fit to the data residuals. We see that there is significant skew in the data residual distribution for $r/\sigma = 30$, but the distribution is effectively normal for $r/\sigma = 100$. The former case can be interpreted as the distribution of residuals we might expect for a target 30 cm directly

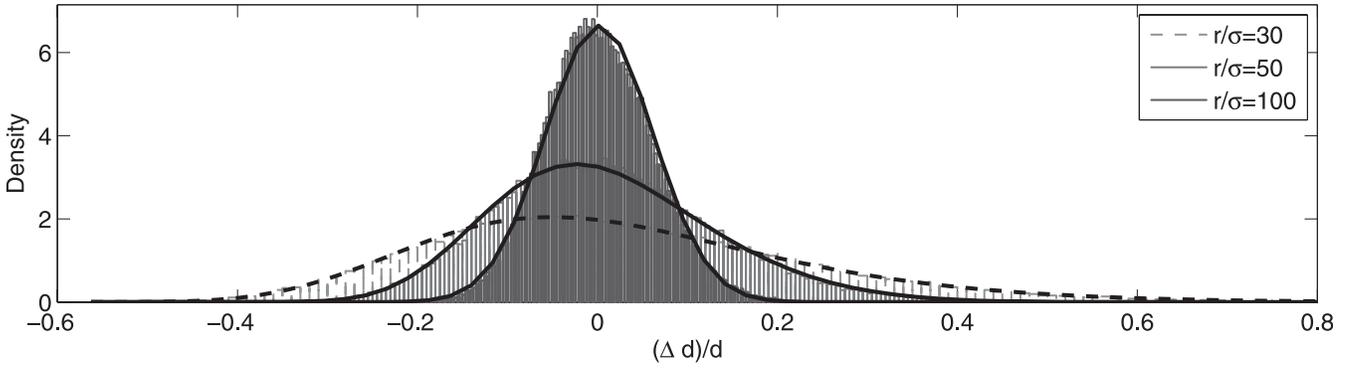


Figure 2. Distributions of relative residuals $(\Delta d)/d$ for increasing r/σ .

below the sensor, with a 1-cm standard deviation on positional error.

Noise in observed data is not limited to positional error. Other sources such as orientation error, cultural noise, and instrument noise will also contribute additively to the observed data. While many of these noise sources are likely non-normally distributed as well, their superposition will tend towards a normal distribution. However, individual noise sources may dominate a particular datum and thus cause large deviations. For example, if the sensor quickly traverses uneven topography that cannot be adequately sampled by a positioning system, then these data will have relatively large deviations from their true values and may have undue influence on inversion with a least squares norm. Furthermore, if errors are correlated, then the central limit theorem does not apply and a Gaussian error distribution will not be a valid assumption. Finally, accounting for non-dipolar effects in the near field of a transmitter loop yields a secondary field response that decays slower than the $1/r^6$ dependence assumed in the preceding analysis (see Nabighian (1979) for a comparison of fields radiated by a dipole and by an arbitrary loop source). The skew of the distribution of residuals in Eq. 12 is therefore increased relative to a dipolar primary when the target is in the near field (*i.e.*, a distance less than the characteristic dimension of the transmitter). In the following section we therefore investigate inversion with robust norms.

Robust Norms

Robust inversion algorithms in the geophysical and statistical literature modify the misfit function so that outlying data have down-weighted contributions to the total misfit. In the context of geophysical inversion, Farquharson and Oldenburg (1998) have demonstrated the use of various robust norms for measuring both data misfit and model norm. Here we follow their description

of an iterative approach to minimizing these norms for linear and nonlinear inverse problems. Consider the norm

$$\phi = \rho(\mathbf{x}) \quad (13)$$

with

$$\mathbf{x} = \mathcal{W}_d(\mathbf{d}^{\text{obs}} - \mathbf{d}^{\text{pred}}) \quad (14)$$

the weighted discrepancy between observed and predicted data, and $\rho(\mathbf{x})$ defining some norm on this vector (*e.g.*, $\rho(\mathbf{x}) = \mathbf{x}^T \mathbf{x}$ for the L2 norm). For a linear problem, we can write $\mathbf{d}^{\text{pred}} = \mathcal{G}\mathbf{m}$, with \mathcal{G} the forward modeling matrix mapping from model to data. Now to minimize ϕ with respect to the model vector \mathbf{m} we have:

$$\begin{aligned} \frac{\partial \rho}{\partial \mathbf{m}} &= \frac{\partial \rho}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{m}} \\ &= \mathcal{B}^T \mathcal{R} \mathbf{x} \end{aligned} \quad (15)$$

with

$$\mathcal{R} = \text{diag}\left(\frac{d\rho}{dx} / \mathbf{x}\right) \quad (16)$$

and

$$B_{ij} = \frac{\partial x_i}{\partial m_j}. \quad (17)$$

For a linear forward problem then:

$$\frac{d\phi}{d\mathbf{m}} = \mathcal{B}^T \mathcal{R} \mathbf{x} = \mathcal{G}^T \mathcal{W}_d \mathcal{R} \mathcal{W}_d (\mathbf{d}^{\text{obs}} - \mathcal{G}\mathbf{m}). \quad (18)$$

Setting the above expression equal to zero and solving for \mathbf{m} yields an expression identical to that in a standard least squares problem, except for the presence of \mathcal{R} . This matrix depends upon \mathbf{x} the weighted residual, and so is a function of the model \mathbf{m} . Hence, minimization of an arbitrary norm becomes a nonlinear problem even when the forward problem is linear. This nonlinearity

Table 1. Example norms used for robust inversion.

Norm	$\rho(x)$
Huber	$x^2, x \leq k$ $2k x - k^2, x > k$
Eckblom	$(x^2 + k^2)^{p/2}$
Bisquare (Tukey)	$k^2/6[1 - (1 - (x/k)^2)^3], x \leq k$ $k^2/6, x > k$

can be circumvented with the “iteratively reweighted least squares” (IRLS) algorithm, which iteratively updates the model with the following procedure:

1. Set $\mathcal{R}_{ii} = 1$ and solve for \mathbf{m} with Eq. 18.
2. Update the elements of \mathcal{R} (Eq. 16) using the current estimate of \mathbf{m} .
3. Recompute \mathbf{m} and iterate from step 2 until convergence.

For a nonlinear forward problem $\mathbf{d}^{\text{pred}} = \mathbf{F}(\mathbf{m})$, we can use the linearization

$$\mathbf{F}(\mathbf{m} + \delta\mathbf{m}) \approx \mathbf{F}(\mathbf{m}) + \mathcal{J} \delta\mathbf{m} \quad (19)$$

to derive an expression for the model perturbation $\delta\mathbf{m}$ at each iteration, with the sensitivity matrix \mathcal{J} taking the place of the forward modeling operator \mathcal{G} in Eq. 18. IRLS provides a general procedure for minimizing a norm, Table 1 summarizes a number of norms that appear in the literature.

Figure 3 compares these norms as a function of x and also shows the resulting weightings \mathcal{R}_{ii} . The

bisquare norm is unique amongst the norms considered here in that it has the capability to completely disregard outlying data ($\mathcal{R}_{ii} = 0$ for $|x_i| > k$). However, some care must be taken in initializing IRLS for the bisquare norm, since the algorithm is not guaranteed to converge for this norm (Marrona *et al.*, 2006). This is evident in Fig. 3 which shows that if the starting model generates predicted data that is far from the observed data, then the bisquare misfit function has no curvature and cannot converge to the global minimum of the misfit. We find that initializing IRLS with the least squares solution works well in this application.

A practical question when applying robust norms is specification of the parameters of the norm (*e.g.*, parameter k in Table 1). In overdetermined inversion, norm parameters are usually set based upon consideration of the variance of the model estimate when the data are normally distributed. When solving a linear inverse problem, the model covariance at the IRLS solution is

$$\text{cov}(\mathbf{m}) = (\mathbf{G}^T \mathcal{W}_d^T \mathcal{R} \mathcal{W}_d \mathbf{G})^{-1}. \quad (20)$$

Referring to Fig. 3, we see that the weightings \mathcal{R} for robust norms are always less than or equal to the weightings for the L2 norm, for which \mathcal{R} is the identity. This implies that the model variance will always be greater when carrying out robust versus L2 estimation if the data are normally distributed. The ratio of L2 estimator variance to robust estimator variance is called estimator efficiency. Recommended values of norm parameters are then set to achieve 95% efficiency (Marrona *et al.*, 2006).

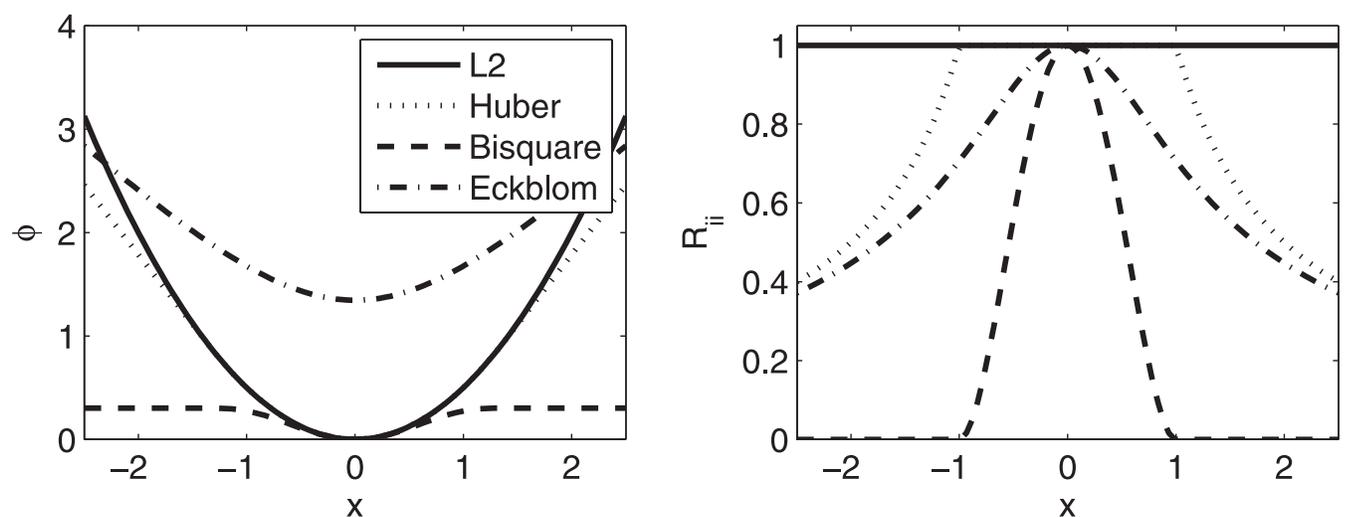


Figure 3. Left: comparison of norms as a function of x , the weighted discrepancy between observed and predicted data. Right: weightings (\mathcal{R}_{ii}) applied to data for IRLS minimization of robust norms. For all norms $k = 1$, and $P = 1$ for the Eckblom p norm.

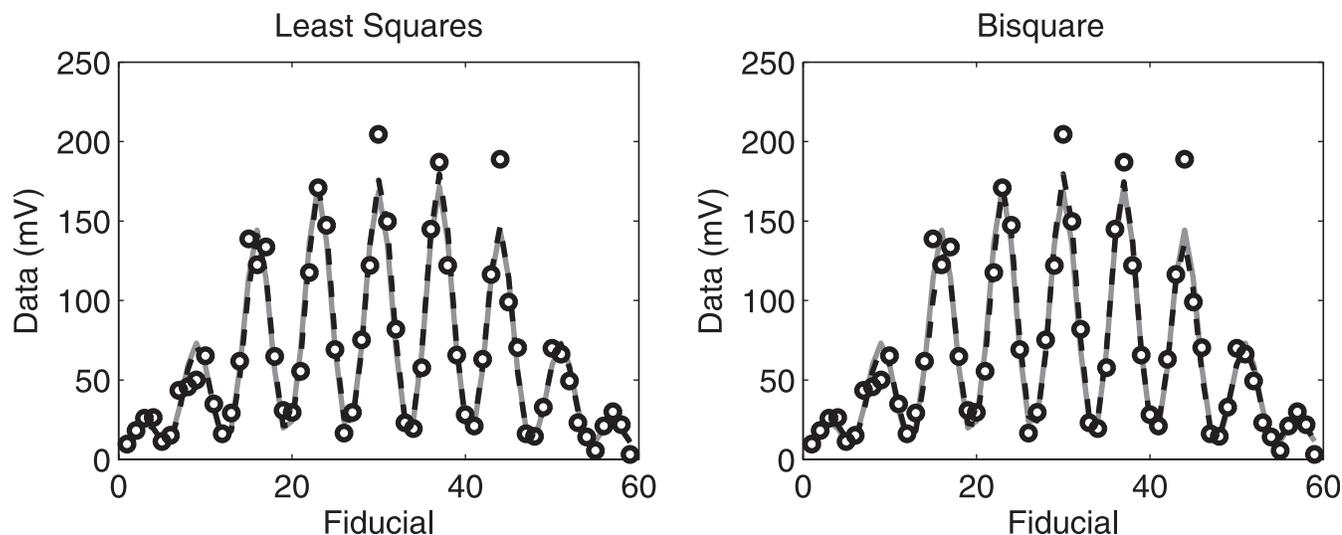


Figure 4. Fits (dashed line) to synthetic data (dots) for a single realization of noise. True data are shown as a gray solid line.

Simulations

To investigate the effect of robust norms on parameter estimation from EM data, we first consider solution of the inverse problem by repeatedly solving the linear problem for the elements of the non-diagonalized polarizability tensor over a range of target locations. We simulate data sets for a monostatic TEM sensor over two targets, in both horizontal and vertical orientations, with polarizabilities of comparable magnitude. One target is axisymmetric and is approximately the size of a mortar baseplate from the Camp Sibert, Alabama ESTCP demonstration study (see Fig. 9). For an axisymmetric target, we expect secondary and tertiary polarizabilities to be equal ($P_2(t) = P_3(t)$). The second target is slightly smaller and non-axisymmetric and is representative of clutter encountered at the same site. In this example, we simulate positional errors as independent Gaussian perturbations to the sensor locations with a standard deviation of 5 cm. We select a relatively large positional error here to investigate the performance of various norms when positional accuracy does not meet the data quality requirements in Bell (2005). Background noise is simulated as 1 percent plus a 5 mV floor Gaussian noise. Stations are spaced uniformly at 20 cm between lines and 20 cm along lines, and only data that are twice the noise floor are used in each inversion. We choose a relatively sparse along line sampling in this case to allow comparison with the results for independent positional error in Tarokh and Miller (2007).

Figure 4 shows an example realization of data over a single target, as well as fits predicted by each norm. As expected from the preceding analysis of positional errors, large positive outliers correspond to

large magnitude data close to the target. Figure 5 shows the results of these simulations for 50 realizations of positional and background noise, with each point representing the minimum misfit model identified in the scan of target positions. We present the least squares and bisquare norms only, other robust norms in Table 2 had comparable performance to the bisquare norm. We see that there is no significant improvement in our ability to discriminate between targets provided by the robust norm relative to conventional least squares when errors are independent.

The difference in the distributions of polarizabilities for horizontal versus vertical targets in Fig. 5 is a result of the positive correlation between polarizability strength and target depth. When a target is oriented horizontally, the data are more sensitive to the transverse (secondary and tertiary) polarizabilities. The difficulty in constraining target depth is therefore manifested as a positive correlation between depth and transverse polarizabilities for horizontal targets. For a vertical target, the data are more sensitive to the primary, axial polarizability, so that the correlation between target depth and transverse polarizabilities is reduced. We also note that the clustering of parameter estimates in Fig. 5 is a function of scanning over a discrete set of target depths; scanning over a finer range of target locations will produce more continuous parameter distributions.

These results seem inconsistent with previous research in Tarokh and Miller (2007), which found that a robust minimax inversion technique provides marked improvement of conventional least squares for independent positional error. We note that in Tarokh and Miller (2007) the authors did not weight their L2 inversion with

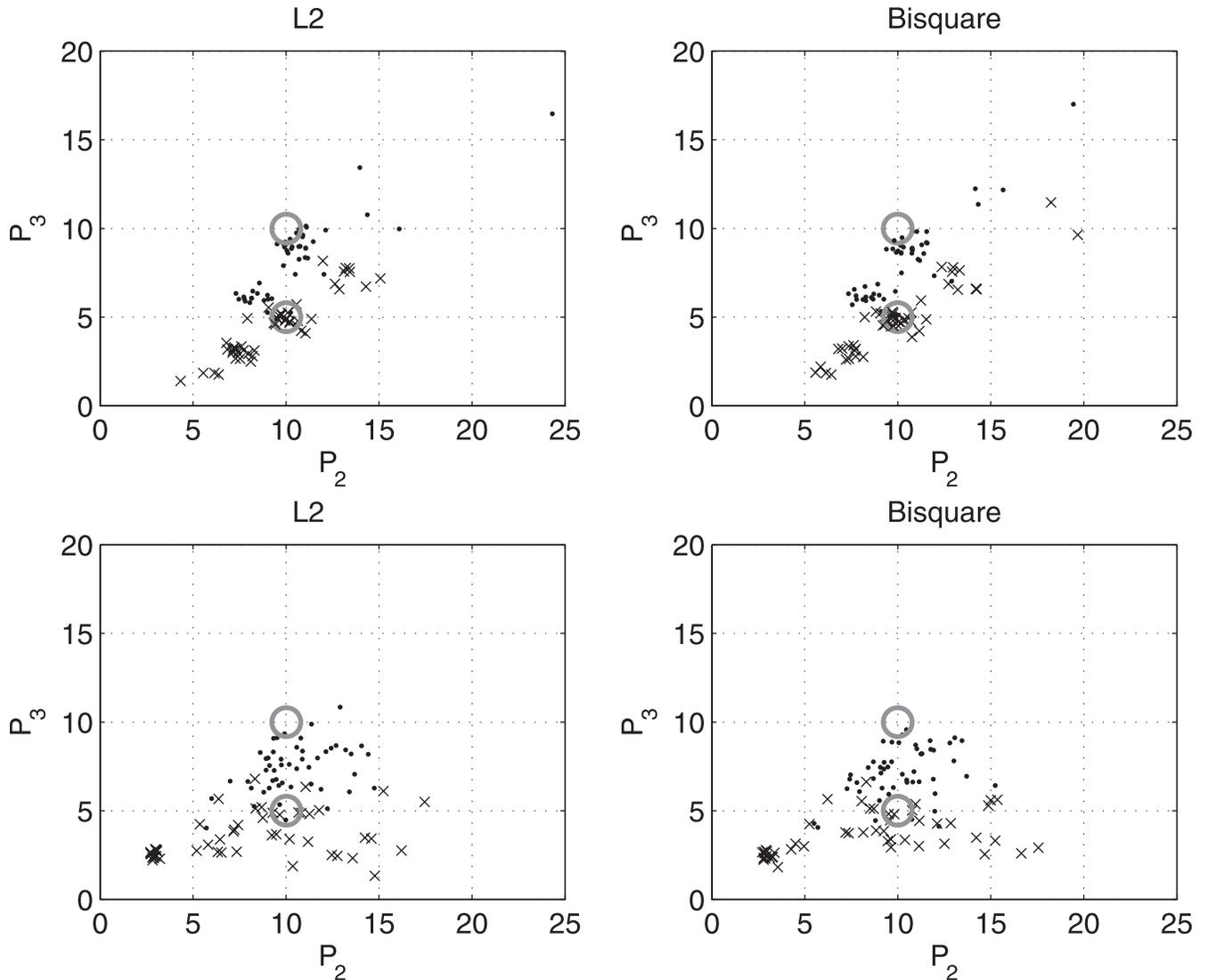


Figure 5. Effect of independent positional error upon L2 and bisquare secondary polarizability estimates (P_2 , P_3) for two targets (dots and crosses). Dots are parameter estimates for an axisymmetric target with true transverse polarizabilities $P_2 = 10$, $P_3 = 10$, and crosses are parameter estimates for non-axisymmetric target with $P_2 = 10$, $P_3 = 5$. True target parameters are shown as open circles. Top row is for targets with primary polarizability oriented horizontally; bottom row is for targets with primary polarizability oriented vertically.

an estimated standard deviation. In contrast, here we estimate a standard deviation for each datum as a percentage of the observed datum (5% in this case) plus a floor. This is especially crucial for inversion of TEM data, where the exponential decay of data requires a weighting, which ensures that large amplitude early time data do not dominate the misfit. Hence, even when minimizing the L2 norm we essentially carry out a single iteration of IRLS, so that we expect the weighted L2 to be robust to outliers, which might have a significant impact if no reweighting is employed. To illustrate this effect, we compute the sensitivity curves of the unweighted L2, weighted L2, and bisquare estimators

for the simple problem of estimating the mean of a random data set (x_1, x_2, \dots, x_N) contaminated by a single outlying datum x_o . The sensitivity curve shows the bias in the estimated parameter

$$\text{bias}(\hat{\mu}) = \hat{\mu}(x_1, x_2, \dots, x_N, x_o) - \hat{\mu}(x_1, x_2, \dots, x_N) \quad (21)$$

as a function of the location of the outlying datum (Marrona *et al.*, 2006). Figure 6 shows sensitivity curves estimated from 100 realizations of Gaussian data, with $N = 100$ data in each realization. We see that while the bias in the L2 estimator increases linearly with outlier location, the weighted L2 estimator is able to control the bias by downweighting outliers. As expected, the

Table 2. Comparison of bootstrapped performance metrics for robust vs. L2 inversion applied to Camp Sibert TEM datasets. *AUC* denotes “area under the receiver operating curve”, with $AUC = 1$ indicating perfect performance. *FAR* denotes “false alarm rate”, with $FAR = 0$ indicating perfect performance.

Sensor	L2		Robust	
	<i>AUC</i>	<i>FAR</i>	<i>AUC</i>	<i>FAR</i>
EM61	0.973	0.297	0.976	0.156
MTADS	0.997	0.111	0.999	0.007
EM63	0.999	0.001	1.000	0.000

bisquare estimator ignores large outliers so that there is zero bias in the estimate when x_o is far from μ .

In the above simulations, we considered positional errors to be independent. In practice, this assumption may not necessarily hold. For example, transient noise sources may cause neighboring data to deviate systematically from their actual position, so that positional error must be treated as correlated. To understand the effect of correlated noise on parameter estimation, we repeat the above simulations with correlated noise added as follows:

1. Background Gaussian noise is added to the true data for each target.
2. Along each line, data exceeding the noise floor are identified.
3. A positive shift is then added to a number, n_ε of adjacent points exceeding the noise threshold. This simulates a correlated error, possibly caused by correlated positional error, though in this case the

shift is generated deterministically as a percentage of the largest perturbed datum, rather than as a random perturbation on position. An additional random Gaussian error is added to each shifted datum, with standard deviation defined as 5% of the shift (we do not expect all points to be shifted by the same amount in the presence of a correlated error).

4. For subsequent realizations, correlated errors are then added to all sets of adjacent n_ε points along all lines (the larger n_ε , the fewer perturbations along each line).

Given a noise realization computed in this manner, we then re-estimate the standard deviation of each datum as a percentage of the noisy datum, as discussed above. These standard deviations then define the elements of \mathbf{W}_d , the data weighting matrix for inversions carried out with L2 and robust norms. These simulations are motivated by the observation that the robustness of an estimator can be measured as a function of the proportion of data ε that is contaminated by non-Gaussian noise (Marrona *et al.*, 2006).

Figure 7 shows the resulting distributions of target parameters when n_ε is 5 and 10. These correspond to correlated error applied to 2% and 4% of the data used in each inversion, respectively. The two cases for n_ε are not shown separately here, but increasing the proportion of data contaminated with correlated noise increases the spread in L2 estimates, while robust norms are able to maintain a relatively constant performance. The bisquare norm provides an appreciable improvement over the L2 norm in terms of overlap between parameters for the targets considered, and so we might expect an improvement in parameter estimation and discrimination when inverting real data contaminated with

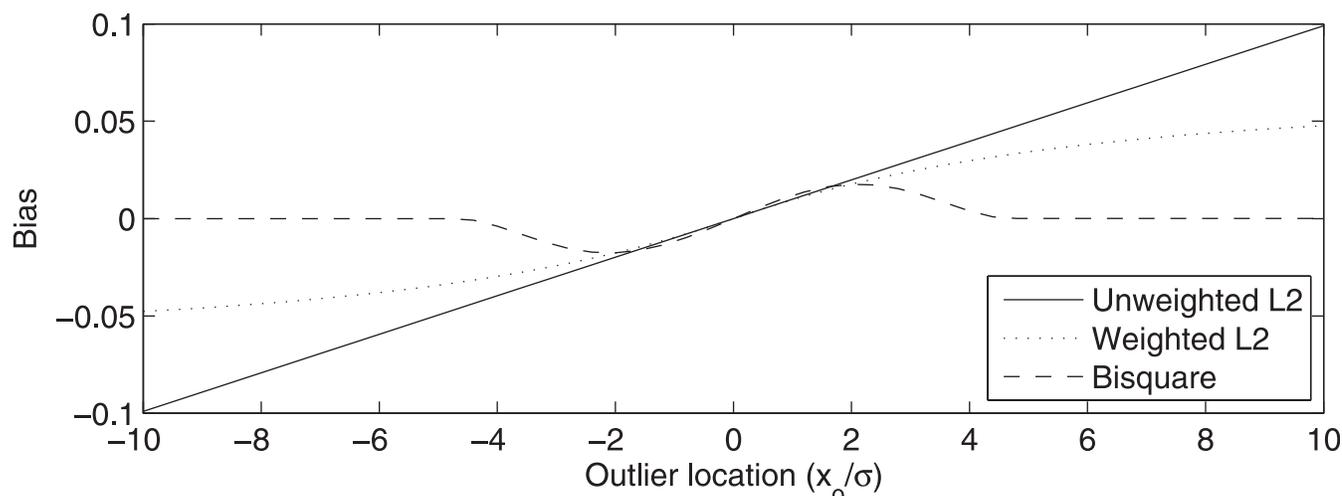


Figure 6. Sensitivity curves of estimates of the sample mean obtained from realizations of $N = 100$ Gaussian data contaminated with a single outlier x_o .

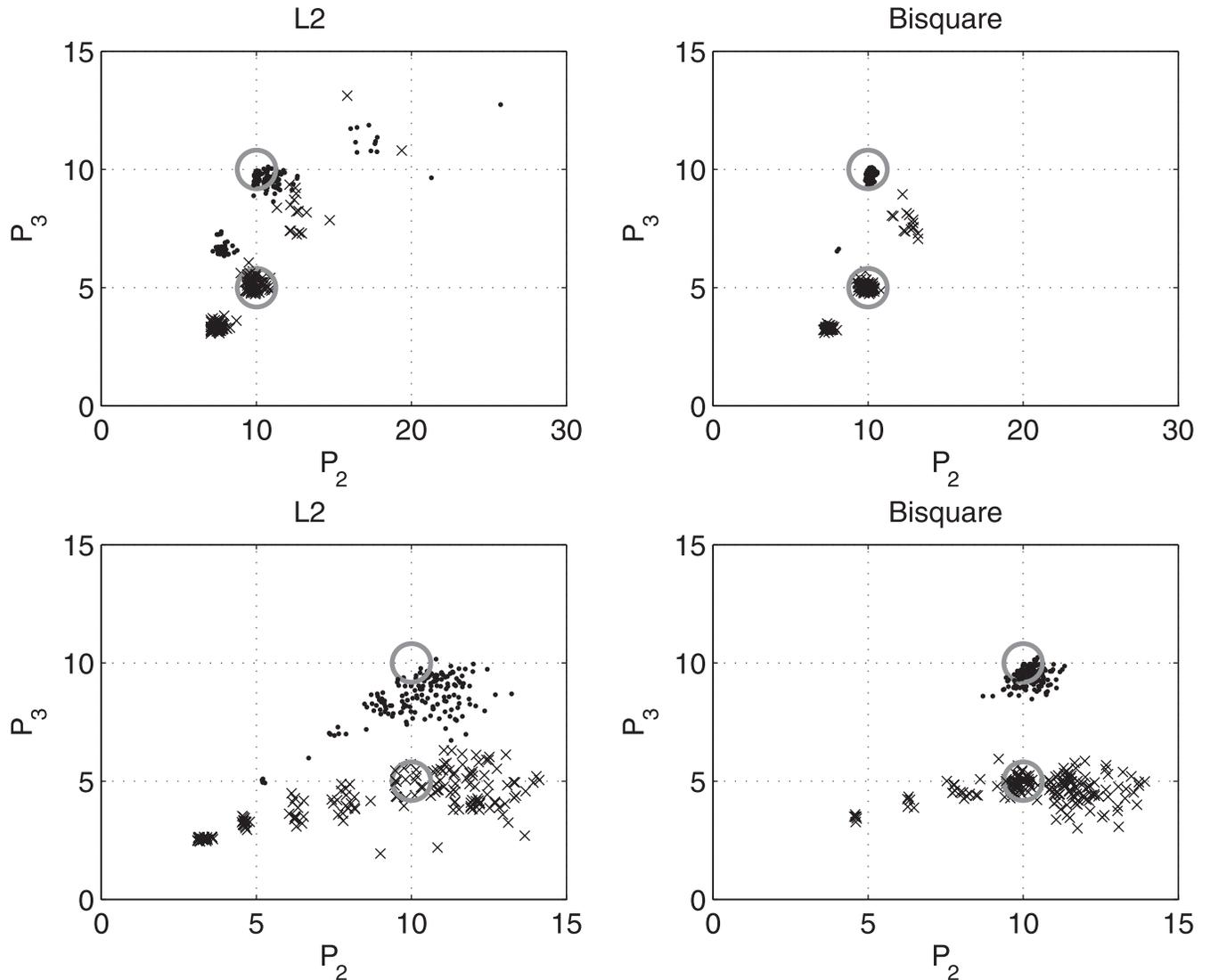


Figure 7. Effect of correlated noise upon L2 and bisquare secondary polarizability estimates (P_2 , P_3) for two targets (dots and crosses). Dots are parameter estimates for an axisymmetric target with true transverse polarizabilities $P_2 = 10$, $P_3 = 10$, and crosses are parameter estimates for non-axisymmetric target with $P_2 = 10$, $P_3 = 5$. True target parameters are shown as open circles. Top row is for targets with primary polarizability oriented horizontally; bottom row is for targets with primary polarizability oriented vertically.

correlated noise. Consistent with the recommendation in Marrona *et al.* (2006), we also find that the bisquare provided tighter parameter distributions than the other robust norms considered in Table 2. Figure 8 shows an analogous result to Fig. 4, illustrating how correlated noise skews the L2 estimate while the robust estimate can effectively ignore these contaminated data.

Application to Real Data

To study the effect of robust parameter estimation upon discrimination, we consider inversion of three TEM data sets acquired at Camp Sibert, Alabama

(Billings, 2008). Data were collected with a Geonics EM61 sensor, the MTADS towed array of three EM61 sensors, and a Geonics EM63 sensor deployed in a cued-interrogation mode. Further description of these sensor platforms can be found in Billings (2008). The primary targets of interest at this site were 4.2" mortars; Fig. 9 shows representative items from target classes encountered at this site.

In this case target location is not known, and so we solve a nonlinear inverse problem for target location, orientation, and polarizability parameters. For the EM63 data, we parameterize each polarizability decay according to Eq. 7, while for EM61 data sets we recover

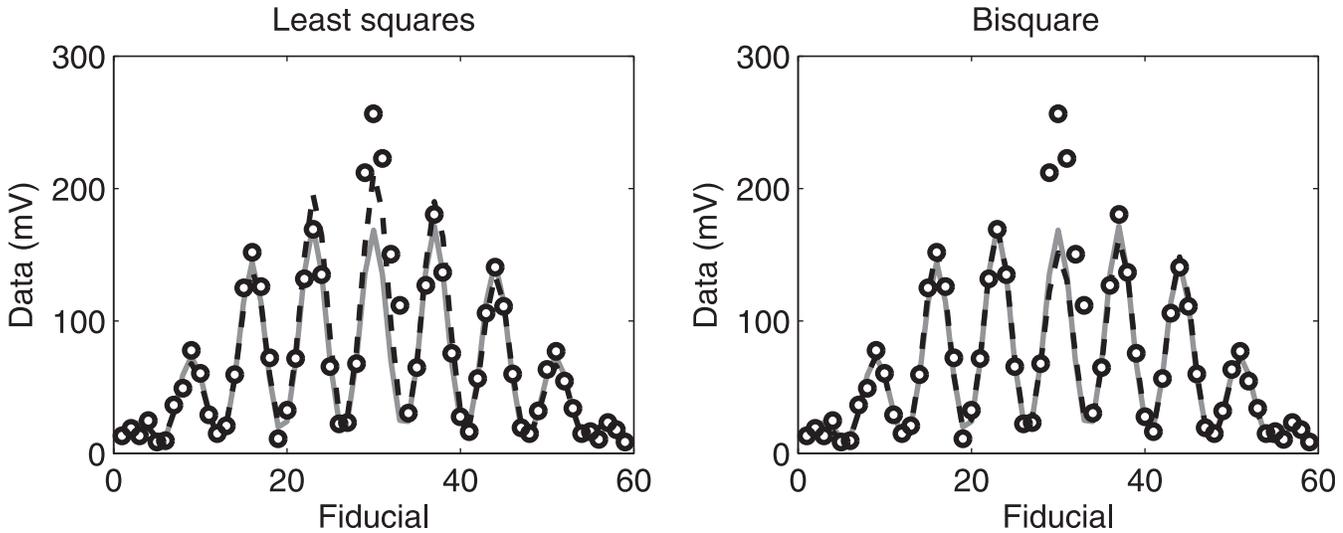


Figure 8. Fits (dashed line) to synthetic data (dots) for a single realization of correlated noise. True data is shown as a gray solid line.

instantaneous polarizability amplitudes at each time channel. We find that the nonlinearity of the forward problem produces local, suboptimal minima of the misfit function, and so we select a set of starting models that adequately explore the misfit surface as follows:

1. We perform a scan over target location and depth, solving the linear problem for the polarizability tensor, as in our analysis of synthetic data.
2. To define N (e.g., 10) starting models for subsequent nonlinear inversion, we identify the minimum misfit model in each of $N - 1$ depth intervals, ranging from $z = 0$ m to $z = z_{\max}$, a maximum target depth (here $z_{\max} = 1.2$ m). An additional N^{th} starting model is specified as the best fitting model at the surface ($z = 0$ m).
3. We then solve each of the N nonlinear inverse problems using the set of starting models, and select the minimum misfit model. The final model is parameterized in terms of target location, orientation, and the eigenvalues of the polarizability tensor at each time channel.

Application of a robust norm in this procedure can occur at several stages. For example, we might apply a robust norm to the linear problem in step 1 as well as for the nonlinear problem at step 2. We have tried several approaches and found that the most computationally efficient is to solve both the linear and nonlinear problems first with the L2 norm, and then subsequently apply the bisquare norm at 95% efficiency, using the final L2 models as starting models for the robust inversion. We do not repeat all N inversions with the robust norm, rather we identify the unique models obtained in the initial L2 inversion, where uniqueness is

determined by the depth of the target. That is, we identify all final models with similar depths (say, within 5 cm of each other), and then use the best fitting model from each subset to initialize a refit with the robust norm.

Based upon recommended practice in Marrona *et al.* (2006) for robust linear regression, for each robust refit we also re-scale the uncertainties at each time channel based upon the residuals predicted by the L2 inversions. The scaling γ_i on the i^{th} time channel is the median absolute deviation (MADN) of the residuals from the L2 inversion

$$\gamma_i = \text{MADN}(\mathbf{x}_i), \quad (22)$$

with \mathbf{x}_i as defined in Eq. 14 and

$$\text{MADN}(\mathbf{x}) = \frac{1}{0.675} \text{Median}(|\mathbf{x} - \text{Median}(\mathbf{x})|). \quad (23)$$

The MADN is a robust estimate of the standard deviation, with the normalization factor ensuring that the expected MADN equals one for $\mathbf{x} \sim \mathcal{N}(0,1)$ (Marrona *et al.*, 2006). In this application, the rescaling ensures that all time channels have approximately equal contributions to the bisquare misfit, at least for the data predicted by the starting model taken from the weighted L2 inversion. Without the scaling, we found that later time channels can sometimes be overemphasized by the bisquare norm, leading to an inadequate fit at early times.

Figure 10 compares features extracted from Camp Sibert EM63 data using L2 and robust inversions. The feature space selected for discrimination is spanned by parameters related to the amplitude and decay of the induced polarizability, both normalized to zero mean



Figure 9. Representative items from Camp Sibert target classes. Clockwise from top left: 4.2" mortar (UXO class), partial mortar (Partial class), mortar fragment (OE class), cultural debris (Culture class), mortar base-plate (Base class).

and unit standard deviation. The feature space for the robust inversion is qualitatively very similar to that of the L2 inversion. This is perhaps unsurprising, since the data are acquired along tightly spaced lines with robotic

total station positioning and therefore can be expected to yield reliable parameter estimates regardless of the misfit function. To understand the benefit of robust inversion, we highlight three targets in Fig. 10. Target 1

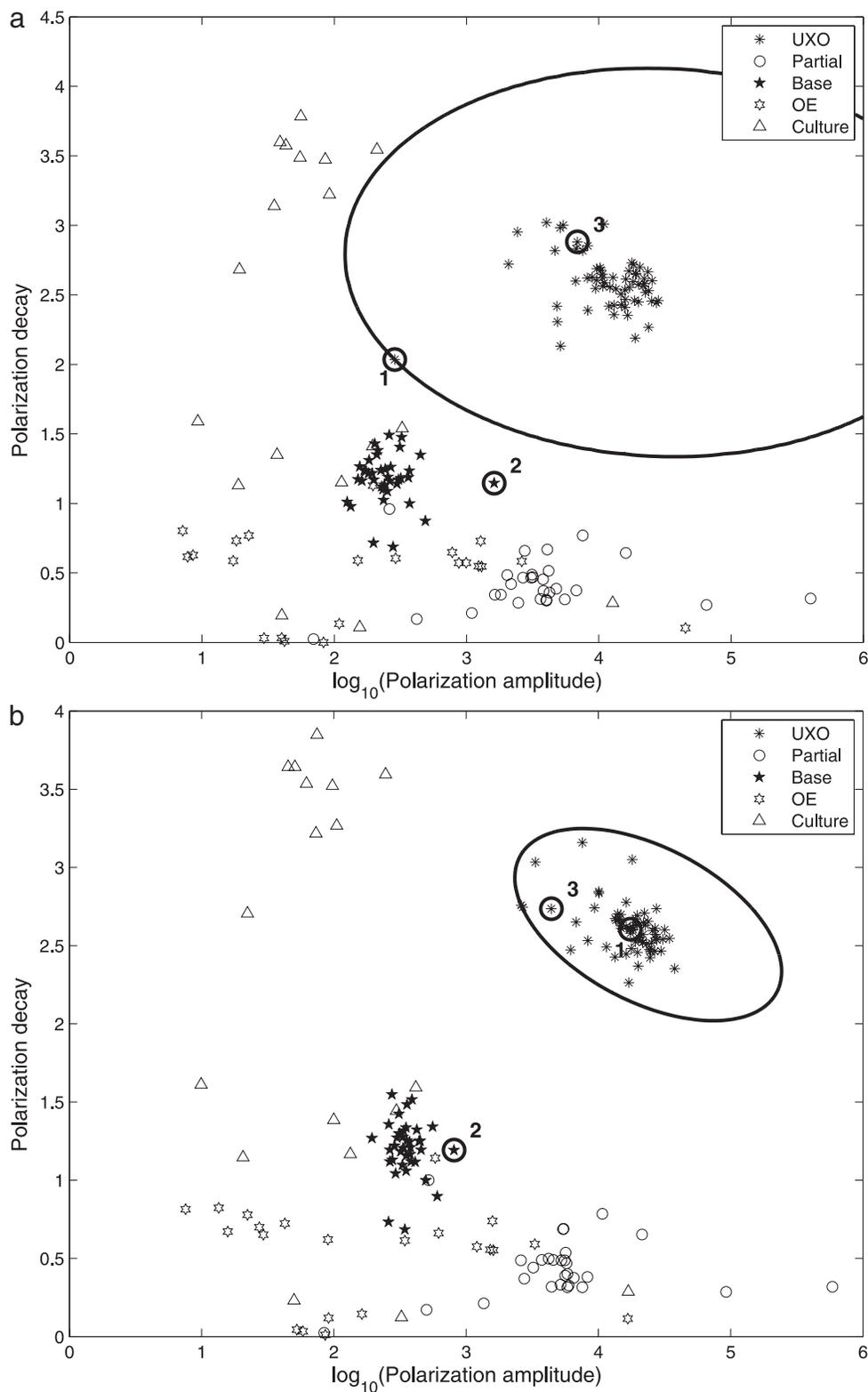
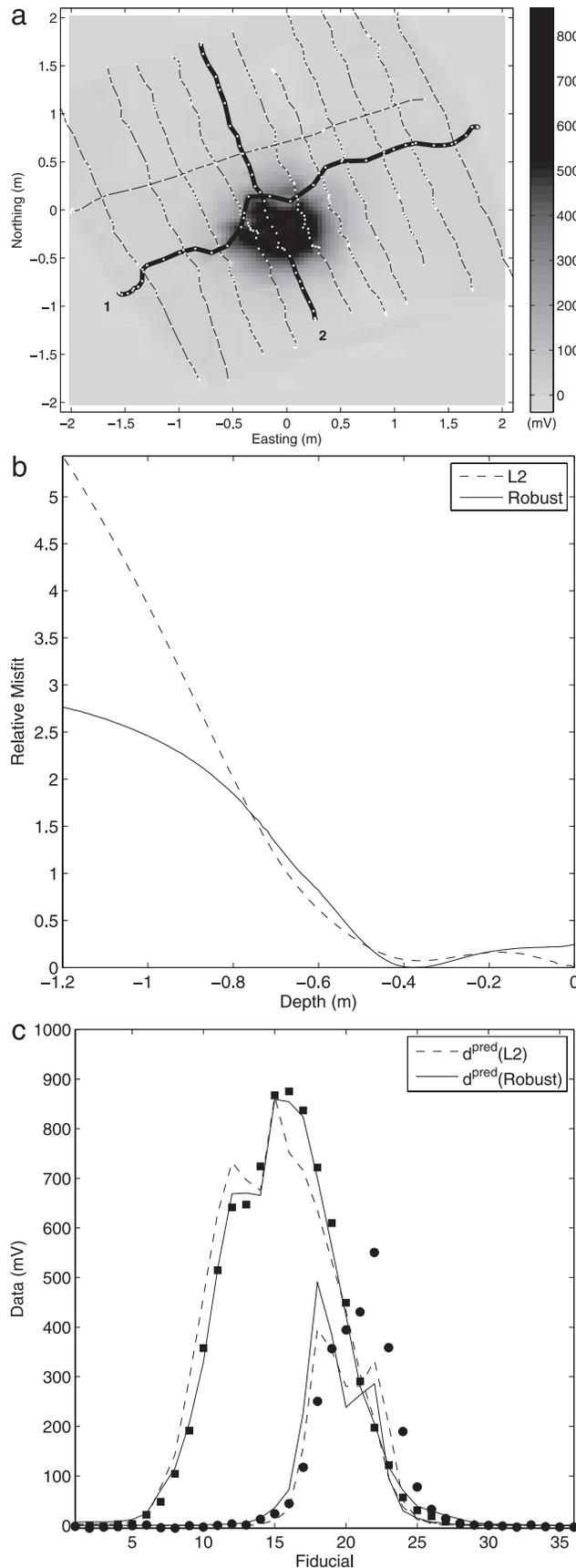


Figure 10. Comparison of features estimated from Camp Sibert EM63 data with L2 (a) and robust (bisquare) (b) norms. Contour indicates output of quadratic discriminant analysis classifier required to find all ordnance items in the feature space.

Beran et al.: Inversion of TDEM Data for UXO Discrimination



is a 4.2" mortar that was identified in subsequent processing to have a single line of data with large positional errors (Fig. 11). When these data are included in the L2 inversion, the best fitting model places the target at $z = 0$ m, and the estimated feature vector is an outlier to its class. Robust inversion determines that a better fit exists at a depth of approximately $z = 0.4$ m and the resulting feature vector lies well within the class distribution. In this example, robust inversion correctly reweights the data so that the minimum at depth becomes the global minimum (Fig. 11(b)), and a better fit to the observed data is obtained (Fig. 11(c)). Targets 2 and 3 are both scenarios where robust inversion remains in the same neighborhood as the optimal L2 model. For target 2 (a base-plate), robust inversion moves the feature vector closer to the class mean, whereas for target 3 (a mortar) the feature vector moves slightly away from the class mean. In the case of ordnance related scrap and culture, we do not expect robust inversion to provide an improvement in clustering in the feature space, since these targets are highly physically variable. For targets with consistent physical properties (UXO and base-plates), we find that robust inversion decreases the variance of parameter distributions. Figure 10 shows a contour of a quadratic discriminant analysis classifier trained in each feature space to discriminate between 4.2" ordnance items and all other target classes (partials, base-plates, etc.). The contour corresponds to the location in the feature space where all ordnance items are identified by the classifier.

A second application of robust inversion is shown in Fig. 12, which considers inversion of MTADS EM61 data acquired in a detection mode survey during the Camp Sibert geophysical prove-out. We again see that robust inversion brings the highlighted outliers closer to their respective class distributions.

Table 2 summarizes the discrimination performance improvements gained by applying robust inversion to Camp Sibert TEM data sets. We use a

←

Figure 11. Comparison of fits for ordnance target from Camp Sibert EM63 data. a) Gridded EM63 data. Data locations are shown as points. Line 1 has probable positional errors. b) Relative misfit as a function of depth for inversion of observed EM63 data with L2 and robust norms. The relative misfit is computed so that the minimum misfit model has relative misfit equal to zero. c) Predicted data from L2 and robust inversions. Circles and squares indicate observed data for lines 1 and 2, respectively.

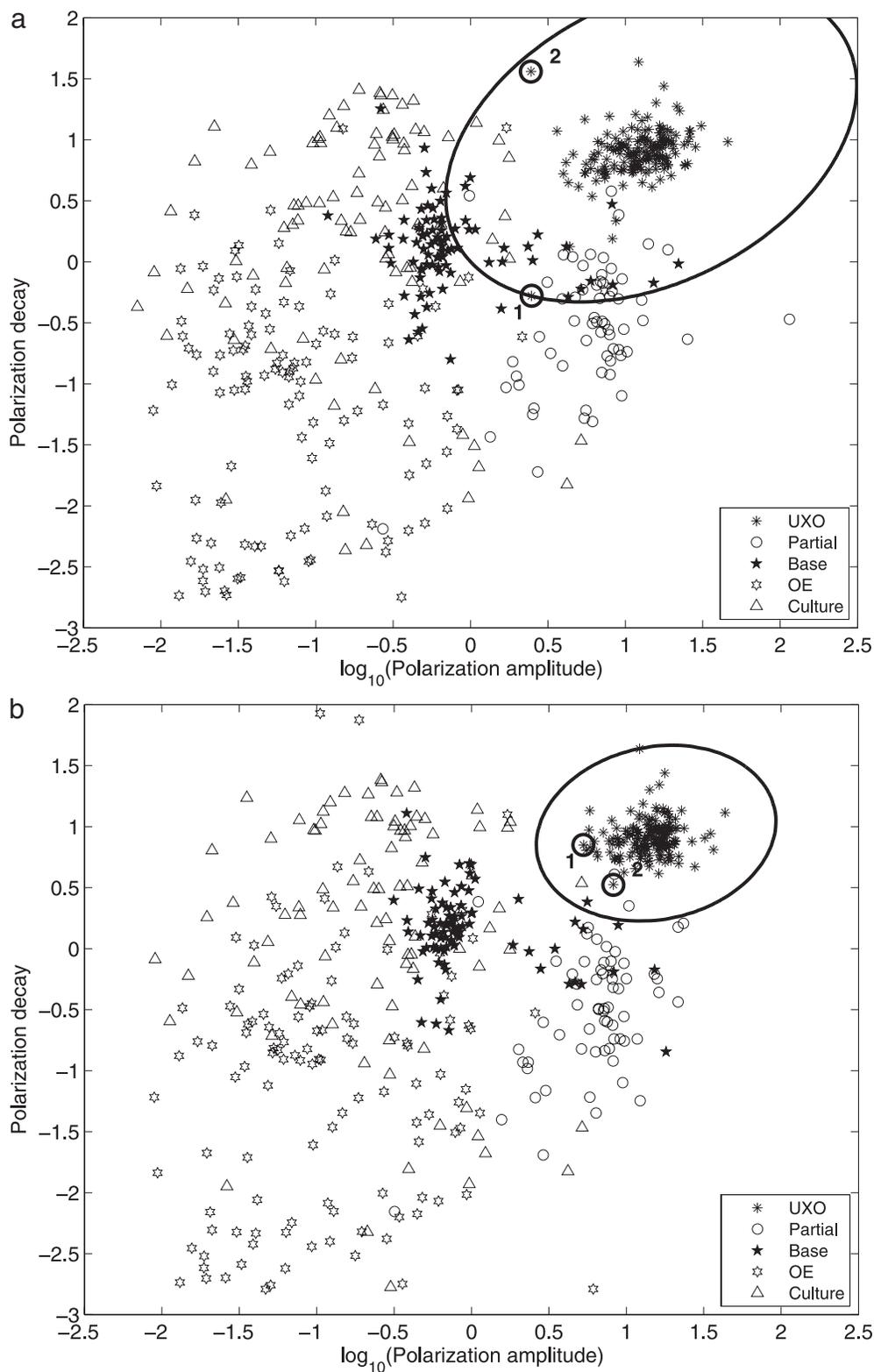


Figure 12. Comparison of features estimated from Camp Sibert MTADS data with L2 (a) and robust (bisquare) (b) norms. Contour indicates output of quadratic discriminant analysis classifier required to find all ordnance items in the feature space.

bootstrapping algorithm to estimate metrics of discrimination performance for quadratic discriminant analysis (Beran and Oldenburg, 2008). We have also included results for the EM61 cart data in this table, though for brevity we have not shown the corresponding feature spaces from L2 and robust inversions of these sensor data. We see an improvement in discrimination performance as we move from a detection mode survey with a single sensor (EM61) to an array of sensors (MTADS) to a cued-interrogation survey (EM63). In all cases, robust inversion with the bisquare norm improves expected discrimination performance versus weighted L2 inversion.

Conclusions

In this work we have applied robust norms to estimation of parameters from time-domain electromagnetic data. When positional errors are modeled as independent Gaussian perturbations, we find that least squares and robust inversion have comparable performance. Both inversion techniques estimate data uncertainties from observed data, and this has the effect of making the least squares inversion robust to outliers. However, when simulated errors are correlated, robust inversion with the bisquare norm provides a marked improvement over L2 inversion. Application of robust inversion to real data sets produced an incremental improvement to the initial L2 inversion, identifying outlying ordnance items and generally tightening parameter distributions in the feature space.

When applying robust norms, we assume that the errors on the data are uncorrelated and non-normally distributed. In simulations, however, we found that robust norms can perform well when errors are correlated. An alternative approach to handling correlated errors is to introduce off-diagonal terms in the data covariance. For example, Dosso *et al.* (2006) iteratively update the data covariance using the data residuals predicted by the current model estimate. This seems very similar to the IRLS procedure employed here, and comparison of these techniques is an avenue for future research.

Acknowledgments

This work was supported by SERDP-ESTCP project MR-1629: Robust statistics and regularization for feature extraction and UXO discrimination.

References

- Bell, T., 2005, Geo-location requirements for UXO discrimination: Technical report, Science Applications International Corporation.
- Beran, L.S., and Oldenburg, D.W., 2008, Selecting a discrimination algorithm for unexploded ordnance remediation: IEEE Transactions on Geoscience and Remote Sensing, **46**, 2547–2557.
- Billings, S.D., 2008, Data modeling, feature extraction, and classification of magnetic and EMI data, ESTCP discrimination study, Camp Sibert, AL: Technical report, Environmental Security Technology Certification Program.
- Dosso, S.E., Nielsen, P.L., and Wilmut, M.J., 2006, Data error covariance in matched-field geoacoustic inversion: Journal of the Acoustical Society of America, **119**, 208–219.
- Farquharson, C., and Oldenburg, D.W., 1998, Nonlinear inversion using general measures of data misfit and model structure: Geophysical Journal International, **134**, 213–227.
- Gumbel, E.J., 1958. Statistics of extremes, Columbia University Press, New York.
- Marrona, R.A., Martin, R.D., and Yohai, V.J., 2006. Robust statistics: Theory and methods, John Wiley and Sons, New Jersey.
- Menke, W., 1989. Geophysical data analysis: Discrete inverse theory: Academic Press, San Diego.
- Nabighian, M.N., 1979, Quasi-static transient response of a conducting half-space—An approximate representation: Geophysics, **44**, 1700–1703.
- Pasion, L.R., 2007. Inversion of time-domain electromagnetic data for the detection of unexploded ordnance, Ph.D. thesis, University of British Columbia, Vancouver, British Columbia.
- Tantum, S.L., Li, Y., and Collins, L.M., 2008, Bayesian mitigation of sensor position errors to improve unexploded ordnance detection: IEEE Geoscience and Remote Sensing Letters, **5**, 103–107.
- Tarokh, A.B., and Miller, E.L., 2007, Subsurface sensing under sensor positional uncertainty: IEEE Transactions on Geoscience and Remote Sensing, **45**, 675–688.