Incorporating geostatistical constraints in nonlinear inversion problems

T. C. Johnson,[1] P. S. Routh,[2] T. Clemo,[2] W. Barrash,[2] and W. P. Clement[2]

Received 17 May 2006; revised 8 June 2007; accepted 13 July 2007; published 19 October 2007.

[1] In this paper we present a method of incorporating semivariogram constraints into nonlinear inversion problems. That is, we describe a method of sampling the space of inverse solutions that honor a specified semivariogram or set of semivariograms and also explain a set of state data. The approach can be considered a method of conditional simulation where model conditioning is based upon state data (as opposed to parameter data). The difference between this approach and other simulation approaches is that the simulation is posed as an optimization problem with the joint objective of matching the semivariograms and honoring the state data. This approach requires computing the sensitivities of the semivariograms with respect to the distributed parameter. We derive these sensitivities and find that they are efficient to compute and store, making the method tenable for large models. We demonstrate the method with one synthetic and one field example using radar velocity tomography, where radar velocity is related through a petrophysical transform to saturated porosity. We address biasing issues and demonstrate ensemble generation and the resulting resolution and uncertainty analysis using ensemble statistics. We also demonstrate how the method can be applied to existing deterministic inversion codes with the field example.

Citation: Johnson, T. C., P. S. Routh, T. Clemo, W. Barrash, and W. P. Clement (2007), Incorporating geostatistical constraints in nonlinear inversion problems, *Water Resour. Res.*, *43*, W10422, doi:10.1029/2006WR005185.

1. Introduction

[2] In order to solve many subsurface environmental and engineering problems, the physical properties that govern subsurface processes must be known. Because geologic properties are often highly spatially variable, it is typically infeasible or prohibitively expensive to obtain adequate information about the property in question through direct sampling. Instead properties must be inferred by the responses they influence at discrete sampling points. For instance, a hydrogeologist might infer the distribution of hydraulic conductivity by pumping from one well, measuring the drawdown response in another, and then finding a hydraulic conductivity distribution that explains the drawdown response. Or a geophysicist may infer the porosity distribution between two wells by measuring the traveltimes of electromagnetic waves propagating from one well to another, and then finding a petrophysical model and a porosity distribution that explain the traveltime data. The data, such as traveltime, are not direct measurements of the property in question, such as porosity, but are influenced by the property in question through some physical process.

[3] The distribution of a physical property is often estimated by mathematically inverting the equations that describe the physical process to find a model (or a solution) that explains the data. However, for most problems many reasonable solutions exist that adequately explain the data. That is, the solutions are nonunique. In addition to complicating the inversion process, solution nonuniqueness naturally adds an element of uncertainty in the estimated physical properties. To be rigorous, the solution uncertainty must be quantified in order to identify what a particular data set can tell us about the subsurface and with what amount of confidence.

[4] Several techniques for estimating subsurface properties and their associated uncertainties have been suggested and most of them can be broadly categorized as either deterministic [*Day-Lewis et al.*, 2002; *Zhou et al.*, 2001; *Anderman et al.*, 1996; *Hyndman et al.*, 1994] or stochastic inversion methods [*Medina and Carrera*, 2003; *Hubbard and Rubin*, 2000; *Capilla et al.*, 1998; *Zimmerman et al.*, 1998; *Copty and Rubin*, 1995]. In general, deterministic methods attempt to estimate a single model that best describes the resolvable subsurface features given a particular data set. Uncertainty in the solution is estimated on the basis of the sensitivity of data to the physical property of interest. Geostatistical inversions attempt to generate the posterior probability density function (pdf) of the property in question given the available data [*Clifton and Neuman*, 1982; *Kitanidis and Vomvoris*, 1983].

[5] The primary limitation of deterministic parameter estimation is directly related to the nonuniqueness of the inverse problem. Data rarely provide enough information to uniquely characterize a discretized model. Thus many models will minimize the difference between the predicted

[1] Energy Resource Recovery and Management, Idaho National Laboratory, Idaho Falls, Idaho, USA.
[2] Center for Geophysical Investigation of the Shallow Subsurface, Department of Geosciences, Boise State University, Boise, Idaho, USA.

0043-1397/07/2006WR005185$09.00

and observed data. In order to remove this nonuniqueness, a regularization term is often explicitly added to the objective function of the inversion. The regularization term of the objective function is minimized at a unique solution, often a homogeneous solution [*Doherty*, 2003; *Tonkin et al.*, 2003; *Tikhonov*, 1963] or a prior reference solution [*Neuman*, 1973], which removes the nonuniqueness in the inverse problem. The inversion is formulated so that the regularization term of the objective function is minimized (e.g., the solution is homogeneous or equal to the prior model) in regions of the model that are not sensitive to the data. The resulting solutions are smoothed versions of the subsurface that generally provide information concerning only the larger-scale subsurface structure.

[6] Smaller-scale structures that may be present in the subsurface are in the null-space of the data. Thus they are not required to fit the data, but they are not excluded by the data either. Instead they are excluded by the regularization operator. These small-scale features are often important when considering the objective of characterizing the heterogeneity of the subsurface to solve environmental or engineering problems. For instance, it is known that contaminant migration is often sensitive to small-scale variations in hydraulic conductivity [*Anderson*, 1997; *Neuman*, 1994; *Sudicky and Huyakorn*, 1991].

[7] In groundwater modeling, regularization is often added implicitly by adhering to the principle of parsimony [*Hill*, 1998], which states that an estimated model should have no more complexity than necessary to explain the data. Parsimony is often implemented by dividing the inversion grid into large zones where physical parameters are assumed to be homogeneous. This reduces the number of parameters estimated and decreases or eliminates solution nonuniqueness at the expense of removing the possibility of having complexity in the model that may be present in the subsurface. In short, the trade-off for using regularization to remove the nonuniqueness in the inverse problem is a solution that favors relatively large-scale structural information and lessens the utility of using indirect data to quantitatively characterize subsurface properties and the corresponding uncertainty via inversion.

[8] In contrast to deterministic inversion, the goal of geostatistical inversion is to estimate the conditional posterior pdf of the estimated property. The original versions of geostatistical inversion required the practitioner to choose the spatial covariance properties of the estimated model, such as the model semivariogram [*Clifton and Neuman*, 1982]. More current methods estimate the spatial covariance properties as part of a two-step inversion process [*Kitanidis and Vomvoris*, 1983]. The first step is to estimate the spatial covariance structure, and the second is to use the spatial covariance estimate as prior information in the generation of the posterior pdf. One of the advantages of geostatistical inversion over deterministic inversion is that the posterior pdf naturally gives an estimate of the parameter uncertainty. Unfortunately, the direct results of geostatistical inversions (the posterior pdf) provide the expected mean of the estimated parameter which tends to filter out actual variability and therefore fails to provide a model which reproduces the spatial covariance [*Carrera et al.*, 2005]. In order to produce a model that reproduces actual variability and explains the data, one must sample the posterior pdf through a method such as Markov Chain Monte Carlo (MCMC) simulation [*Gilks et al.*, 1996], which is generally computationally expensive.

[9] In this work we demonstrate a method which provides realizations of all possible solutions that both fit the data and obey some spatial covariance properties specified by one or more semivariogram(s). This approach can be considered a method of conditional simulation where the two objectives of fitting the state data and honoring the semivariogram(s) are explicit goals of a joint optimization procedure. In order to optimize the semivariogram objective, the sensitivities of the experimental semivariogram with respect to the model parameters must be computed. We derive these sensitivities and find that they are efficient to compute and store, making the approach tenable for large models. In addition, the approach can be implemented into existing deterministic inversion codes by augmenting the data sensitivity matrix with the semivariogram sensitivity matrix and modifying the damping and convergence criteria of the inversion routine.

[10] A critical assumption in this method is that something is known about the semivariogram with some degree of certainty. While geostatistical properties of the subsurface are rarely known precisely, spatial covariance estimates can often be obtained from indirect sources such as analog sites, geophysical data, outcroppings, logs, and expert judgment [*Hubbard et al.*, 1999; *McKenna and Poeter*, 1995; *Jussel et al.*, 1994]. For instance, it may be possible to estimate the vertical semivariogram quite well from well logs which provide dense vertical data. It may be more difficult to estimate the horizontal semivariogram. However, it may be possible to deduce from outcrops or analog sites that the range of the horizontal semivariogram is likely not less than some value, and not greater than some other value. When there is uncertainty in the semivariogram, that uncertainty should be manifest in the uncertainty of the parameter. In this work, we show how semivariogram uncertainty may be included in the parameter estimates using a field example.

[11] Our approach is illustrated conceptually in Figure 1. The model space consists of all possible solutions to the inverse problem. Within the model space, a number of solutions adequately explain the data. These solutions are bounded by the blue line labeled ''Models that fit data without noise'' in Figure 1. Here noise refers to randomly distributed measurement errors. The existence of multiple solutions illustrates the nonuniqueness of the problem and the presence of a null-space. In practice, data are always contaminated by noise which effectively increases the null-space and the number of solutions that explain the data [*Parker*, 1994]. The space of solutions that explain the noisy data is bounded by the dashed line in Figure 1.

[12] In addition to solutions that explain the data, there are also a number of solutions that fit the specified spatial covariance properties. This space is bounded by the red line labeled ''Models that fit known geostatistics'' in Figure 1. Sampling models from this space alone constitutes unconditional stochastic simulation [*Deutsch and Journel*, 1998; *Journel and Huijbregts*, 1978]. In practice the optimal spatial covariance structure is typically not well known, but reasonable bounds may be placed on semivariogram parameters. For instance, we may be able to place upper and lower bounds on the range and/or sill values of the semi-

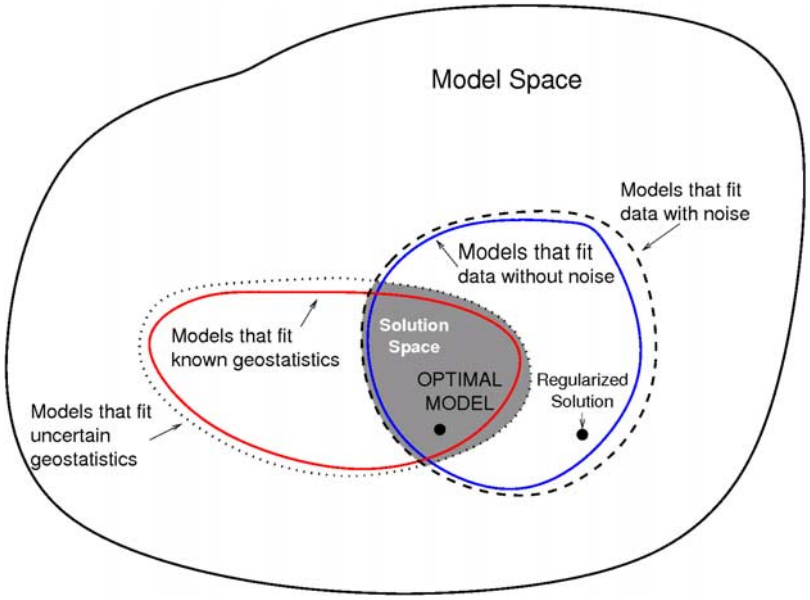

Figure 1. Conceptual representation of model space.

variogram(s). The uncertainty in the spatial covariance structure is analogous to the presence of noise in the data. The "noise" in the spatial covariance structure expands the range of solutions that fit the geostatistics and increases the geostatistical null-space. The space of solutions that fit the uncertain covariance properties is represented by the dotted line in Figure 1.

[13] Assuming data measurement errors are random and the semivariogram space bounds the optimal semivariogram(s) (i.e., the semivariogram(s) that best represent the spatial covariance), there exists a region in the model space that includes solutions which both explain the data and fit the geostatistics. This region, which is represented by the shaded area labeled 'solution space', includes the optimal model. Here we define the optimal model as the solution that most accurately depicts the subsurface at the model grid scale. In this paper, we assume the grid scale is uniform and is much smaller than the semivariogram range so that small-scale spatial correlation represented by the semivariogram can be captured by the model.

[14] Traditional regularization precludes the possibility of obtaining the optimal solution through inversion unless the subsurface is homogeneous (or zonally homogeneous in the implicit regularization case) or equal to the prior model; this is a consequence of obtaining a unique solution through regularization. Because traditional regularization imposes smoothness in unresolved portions of the model, it is difficult to obtain a model that captures the natural complexity of the subsurface as described by the semivariogram(s). Thus it is difficult to use a regularized model to accurately predict processes that are sensitive to that complexity, such as contaminant transport. By generating an individual solution that both explains the data and fits a desired spatial covariance structure, we aim to capture the natural complexity in the subsurface at scales greater than the grid scale (but much smaller than the semivariogram range), and thereby provide the possibility of producing accurate predictions of processes sensitive to variability at scales greater than the grid scale. By generating an ensemble of such solutions we can use statistical analysis to examine the resolving capability of the data (and of the covariance constraints) and estimate the uncertainty in the models and in the predictions they produce.

[15] The ensemble we produce in this work is statistically different than the ensemble produced by MCMC methods primarily because the posterior probability distribution is different for each approach. The traditional posterior pdf used in MCMC methods describes the probability of every model within the model space shown in Figure 1. Each model sampled by the MCMC algorithm has a different probability of being the optimal model. Models closer to the optimal model will be sampled more densely than models much different than the optimal model. However, some models which fit the data and spatial structure poorly will be included in the ensemble in order to represent the tails of the posterior pdf. In other words, with respect to Figure 1, the MCMC algorithm will sample models more densely within and near the solution space and less densely outside of the solution space, but the entire model space will be represented. In the approach we present here, each model has equal probability of being the optimal model, and each is included within the solution space shown in Figure 1. The probability of each model is the maximum possible probability given the objective of the inversion, the noise in the data, and the uncertainty in the semivariogram(s). With respect to other parameter estimation approaches in the literature, our approach is most similar in concept to the self-calibrated algorithm [*Gomez-Hernandez et al.*, 1997]. In the self-calibrated algorithm, an unconditional simulation (or a conditional simulation conditioned on parameter data) is generated as a starting model in the first phase. In the second phase, a perturbation to the starting model is determined which, when added to the started model, causes the resulting model to also honor the state data. However, there is no spatial covariance constraint in the second step, and thus no guarantee that the final model will honor the semivariogram(s). In the approach we present here, state data and structural constraints are both included in the

inversion objective, and both are honored in the final models.

[16] We first describe the twofold objective of the inverse problem which is (1) to minimize the difference between the predicted and observed data and (2) to minimize the difference between the experimental and target semivariogram(s). Although the method we present is generic in scope (i.e., it is applicable to many types of parameter estimation problems), we describe the theory in terms of ray-based radar velocity tomography. In order to accurately generate uncertainty estimates through ensemble solution statistics, models taken from the solution space must be equally probable and must be sampled in an unbiased manner. We use the conditional probability distribution to demonstrate that each solution is equally probable. We discuss bias caused by the spatial covariance constraints which we term 'spatial covariance bias' to facilitate discussion, and we present a practical approach to eliminating spatial covariance bias for solutions that are both unconditional (i.e., solutions constrained to fit the semivariograms only) and conditional (i.e., solutions constrained to fit both the semivariograms and the data). Note here that 'data' do not refer to measurements of the parameter we are estimating (i.e., velocity), but to measurements of state variables that are related to the field through a physical process (i.e., traveltime).

[17] Next we show a field example using radar traveltime data collected at the Boise Hydrogeophysical Research Site (BHRS). The BHRS has been shown to exhibit a zonal geostatistical porosity structure [*Barrash and Clemo*, 2002]. We use an approximation to Topp's equation [*Ferre et al.*, 1996; *Topp et al.*, 1980] to convert zonal porosity semivariograms to velocity semivariograms in order to constrain the inversion. In the field example, we demonstrate how semivariogram uncertainty is incorporated into the ensemble, and how it affects ensemble statistics by decreasing resolution in the ensemble mean and increasing the ensemble variance. Finally we discuss the results and offer concluding remarks.

2. Theory

[18] Although the inversion method described in this section is general in mathematical terms, we develop the discussion in terms of radar traveltime tomography and provide specific examples in the results section. We begin the theoretical development by stating the twofold objective of the inversion which is to minimize the function

$$\Phi(\mathbf{m_{est}}) = \|\mathbf{W_d}(\mathbf{G}(\mathbf{m_{est}}) - \mathbf{d_{obs}})\|^2 + \beta\|\mathbf{W_v}(\mathbf{\Gamma}(\mathbf{m_{est}}) - \mathbf{v_{obs}})\|^2. \tag{1}$$

Here $\mathbf{m_{est}}$ is the estimated model parameter vector (e.g., velocity distribution), $\mathbf{G}$ is the forward traveltime operator (e.g., the eikonal equation solver [*Aldridge and Oldenburg*, 1993]), $\mathbf{d_{obs}}$ is the observed traveltime data vector, $\mathbf{\Gamma}$ is the forward semivariogram operator such that $\mathbf{\Gamma}(\mathbf{m_{est}})$ produces the predicted semivariogram(s), $\mathbf{v_{obs}}$ is a vector containing the desired (or target) model semivariograms, and $\mathbf{W_d}$ and $\mathrm{W_v}$ are the data and semivariogram weighting matrices respectively. The parameter β determines the importance placed on fitting the semivariograms at each iteration as $\Phi(\mathbf{m_{est}})$ is minimized in the inversion procedure. The $\|\ \|$ operator represents the L2 norm.

[19] The second term of the objective function has the same form as a traditional (e.g., Tikhonov) regularization constraint [*Tikhonov*, 1963]. It is also similar in form to what *Neuman* [1973] called a plausibility criteria, which included a reference or prior model. However, with traditional regularization there is a unique solution that minimizes the regularization term. This removes the null-space and the ill posedness of the inverse problem. Conversely, many solutions exactly minimize the second term of equation (1). That is, many parameter distributions produce semivariograms that match those specified in $\mathbf{v_{obs}}$. In fact, minimizing the second term of equation (1) alone constitutes a method of unconditional stochastic simulation. In general, many of these geostatistically valid solutions will also minimize the first term of equation (1). Hence many solutions exist that minimize both terms of equation (1), so the solutions are still nonunique and the inverse problem is still ill posed, (i.e., a null-space exists even though the solutions satisfy the geostatistical constraints). Conceptually, those solutions that minimize both terms of the objective function are contained within the solution space shown in Figure 1.

[20] The inversion objective allows us to sample the space of solutions that fit both the observed (e.g., traveltime) data and the geostatistics, thereby providing a means of estimating solution statistics or uncertainty. By repeatedly solving the inverse problem with different random starting models, we generate an ensemble of geostatistically accurate solutions that explain the data. We take precautions to ensure that each solution is equally probable so that the solution space is represented in an unbiased manner. We use the ensemble to estimate statistical descriptions of the solution space (e.g., mean and variance for each grid cell). To generate the ensemble, we must be able to minimize equation (1) which requires simultaneously minimizing the two nonlinear and nonunique objective function terms. For the remainder of this paper we will refer to this method as inverse stochastic sampling (ISS) in order to simplify forthcoming discussion.

[21] To minimize equation (1) we expand the objective function by perturbing the model by $\delta\mathbf{m_{est}}$ which gives, ignoring the second-order and higher-order terms,

$$\begin{aligned}\Phi(\mathbf{m_{est}} + \delta\mathbf{m_{est}}) = {} & \|\mathbf{W_d}(\mathbf{G}(\mathbf{m_{est}}) + \mathbf{J_G}\delta\mathbf{m_{est}} - \mathbf{d_{obs}})\|^2 \\ & + \beta\|\mathbf{W_v}(\mathbf{\Gamma}(\mathbf{m_{est}}) + \mathbf{J_\Gamma}\delta\mathbf{m_{est}} - \mathbf{v_{obs}})\|^2,\end{aligned} \tag{2}$$

where $\mathbf{J_G}(\mathbf{m_{est}})$ and $\mathbf{J_\Gamma}(\mathbf{m_{est}})$ are, respectively, the traveltime and semivariogram Jacobian operators. The traveltime sensitivities that comprise $\mathbf{J_G}$ are computed by ray tracing [*Aldridge and Oldenburg*, 1993]; the computation of $\mathbf{J_\Gamma}$ is discussed in section 2.2. Taking the gradient of equation (2) with respect to $\delta\mathbf{m}$, setting the results to zero, and gathering terms gives

$$\left(\mathbf{J_G^T}\tilde{\mathbf{W}}_\mathbf{d}\mathbf{J_G} + \beta\mathbf{J_\Gamma^T}\tilde{\mathbf{W}}_\mathbf{v}\mathbf{J_\Gamma}\right)\delta\mathbf{m_{est}} = \mathbf{J_G^T}\tilde{\mathbf{W}}_\mathbf{d}\delta\mathbf{d_{obs}} + \beta\mathbf{J_G^T}\tilde{\mathbf{W}}_\mathbf{v}\delta\mathbf{v_{obs}}, \tag{3}$$

where $\delta\mathbf{d_{obs}} = \mathbf{G}(\mathbf{m_{est}}) - \mathbf{d_{obs}}$, $\delta\mathbf{v_{obs}} = \mathbf{\Gamma}(\mathbf{m_{est}}) - \mathbf{v_{obs}}$, $\tilde{\mathbf{W}}_\mathbf{d} = \mathbf{W_d^T}\mathbf{W_d}$, and $\tilde{\mathbf{W}}_\mathbf{v} = \mathbf{W_v^T}\mathbf{W_v}$. Note that $\delta\mathbf{m_{est}}$ is a perturbation

to the model $\mathbf{m_{est}}$ that decreases the value of the objective function. That is, the updated solution at iteration $\mathbf{k}+\mathbf{1}$ is given by $\mathbf{m_{est}^{k+1}} = \mathbf{m_{est}^{k}} + \delta\mathbf{m_{est}^{k}}$. At each iteration $\mathbf{k}$, we solve equation (3) for $\delta\mathbf{m_{est}^{k}}$ and update $\mathbf{m_{est}^{k}}$ for the next iteration. We determine $\delta\mathbf{m_{est}^{k}}$ using the LSQR [*Paige and Saunders*, 1982] algorithm to solve the equation

$$\begin{bmatrix} \mathbf{W_d J_G} \\ \sqrt{\beta}\mathbf{W_v J_\Gamma} \\ \alpha\mathbf{I} \end{bmatrix} [\delta\mathbf{m_{est}}] = \begin{bmatrix} \mathbf{W_d}\delta\mathbf{d_{obs}} \\ \sqrt{\beta}\mathbf{W_v}\delta\mathbf{v_{obs}} \\ 0 \end{bmatrix} \tag{4}$$

for $\delta\mathbf{m_{est}}$, where α is a constant and $\mathbf{I}$ is the identity matrix. Here α is a damping term that controls the magnitude of the elements of $\delta\mathbf{m_{est}}$ at each iteration. If α is set too large then the solution will converge slowly. If α is set too small, then the solution will oscillate, and may not converge at all. It is important to note that α and β do not affect the properties of the final model, but are tools used to guide the inversion to a solution. Because both terms of the objective function are nonlinear, the algorithms for choosing α and β are nontrivial and merit a separate discussion presented in section 2.3.

[22] Let the data norm $\Phi_d = \|\mathbf{W_d}(\mathbf{G(m)} - \mathbf{d_{obs}})\|^2$ and let the semivariogram norm $\Phi_v = \|\mathbf{W_v}(\mathbf{\Gamma(m)} - \mathbf{v_{obs}})\|^2$. The solution converges when $\Phi_d \approx \epsilon$ and $\Phi_v \approx \tau$. Here ϵ is chosen on the basis of the noise in the data and τ is chosen on the basis of how well the semivariogram is to be matched. We assume the data and semivariogram values are uncorrelated so that $\mathbf{W_d}$ and $\mathbf{W_v}$ are diagonal vectors. Then $\mathbf{W_d}$ contains the reciprocal of the standard deviation of the data noise along the diagonal. Similarly, $\mathbf{W_v}$ contains the reciprocal of the standard deviation of the desired semivariogram misfit, which is user chosen. For example, in the following examples we use a standard deviation of 1% of the target semivariogram value for the corresponding diagonal element of $\mathbf{W_v}$. The values of ϵ and τ are determined by the chi-square criteria $\epsilon = Ndata^2$ and $\tau = Nvar^2$ where *Ndata* and *Nvar* are the number of observed data and the number of target semivariogram values respectively. Thus, when the model is appropriately fit, $\frac{\Phi d}{\epsilon} = 1$ and $\frac{\Phi_v}{\rho} = 1$.

2.1. On Solution Probability

[23] In order to accurately represent the solution space, each model in the ensemble must be unbiased and equally probable. Given the inversion objective stated in equation (2), the conditional probability distribution is of the form [*Oliver et al.*, 1997]

$$P(\mathbf{m_{est}}) \propto \exp\left(-\frac{1}{2}\Phi_d - \frac{1}{2}\Phi_v\right), \tag{5}$$

where Φ_d and Φ_m are the data and semivariogram norms given above. As discussed in the previous section, $\Phi_d \approx \epsilon$ and $\Phi_v \approx \tau$ for each solution in the ensemble. Therefore each solution is equally probable. When there is uncertainty in the semivariogram, that uncertainty is reflected in the ensemble by probabilistically varying the value of the target semivariogram term $\mathbf{v_{obs}}$ contained within Φ_v. Solutions with a more probable structure occur more often than solutions with a less probable structure. Thus the semivariogram uncertainty is contained implicitly within the ensemble and reflected in the ensemble statistics.

2.2. Computing the Semivariogram Sensitivities $\mathbf{J_\Gamma}$

[24] The experimental semivariogram value at a separation distance (or lag) of h is given by [*Isaaks and Srivastava*, 1989]

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left(m_{i,k} - m_{i,j}\right)^2, \tag{6}$$

where i is the index that spans all $N(h)$ parameter cell pairs separated by lag h. The pair index i is not included in traditional mathematical descriptions of the semivariogram, but is added to the notation here in order to facilitate the forthcoming sensitivity derivation. Note also that 'cell' refers to a discrete region of the model where the parameter being estimated is homogeneous (e.g., a grid cell), and it is assumed that each cell dimension is much smaller than the target semivariogram range. The indices j and k span the individual cells. Thus $j \neq k$ for all pairs i and $m_{1,j} = m_{2,j} = m_{3,j} = \ldots$ That is, $m_{1,j}$, $m_{2,j}$, m_{3j}, .. all represent the same parameter (i.e., parameter m_j), but they belong to different pairs (e.g., pairs 1, 2, 3, ..).

[25] The variation in $\gamma(h)$ due to a perturbation in parameter m_j is given by

$$\delta\gamma(h)_j = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left(m_{i,k} - \left(m_{i,j} + \delta m_j\right)\right)^2 - \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left(m_{i,k} - m_{i,j}\right)^2. \tag{7}$$

Note that if cell j does not belong to pair i then $\delta m_j = 0$ and pair i does not contribute to $\delta\gamma(h)_j$. Thus we may write

$$\delta\gamma(h)_j = \frac{1}{2N(h)} \sum_{i=1}^{nlag(h)j} \left(m_{i,k} - \left(m_{i,j} + \delta m_j\right)\right)^2 - \frac{1}{2N(h)} \sum_{i=1}^{nlag(h)j} \left(m_{i,k} - m_{i,j}\right)^2, \tag{8}$$

where $nlag(h)_j$ is the number of pairs at lag h that include cell j, and i spans all of these pairs. Rearranging constants, factoring the $(m_{i,k} - m_{i,j})^2$ out of the first term, and expanding the result gives

$$2N(h)\delta\gamma(h)_j = \sum_{i=1}^{nlag(h)j} \left(m_{i,k} - m_{i,j}\right)^2 \left(1 - \frac{2\delta m_j}{\left(m_{i,k} - m_{i,j}\right)} + \frac{\delta m_j^2}{\left(m_{i,k} - m_{i,j}\right)^2}\right) - \sum_{i=1}^{nlag(h)j} \left(m_{i,k} - m_{i,j}\right)^2. \tag{9}$$

When δm_j is small the second-order term in δm_j is insignificant. Neglecting the second-order term in δm_j, collecting the results and simplifying gives

$$2N(h)\delta\gamma(h)_j = -2 \sum_{i=1}^{nlag(h)j} \left(m_{i,k} - m_{i,j}\right)\delta m_j. \tag{10}$$

Dividing equation (10) by $2N(h)\delta m_j$ gives

$$\frac{\delta\gamma(h)_j}{\delta m_j} = \frac{1}{N(h)} \sum_{i=1}^{nlag(h)j} \left(m_{i,j} - m_{i,k}\right) \tag{11}$$

or equivalently

$$J_{\Gamma:h,j} = \frac{\delta\gamma(h)_j}{\delta m_j} = \frac{1}{N(h)} \left(m_j \cdot nlags(h)_j - \sum_{i=1}^{nlags(h)j} m_{i,k} \right), \tag{12}$$

where $J_{\Gamma:h,j}$ is row h and column j of $\mathbf{J_\Gamma}$. We compute $\mathbf{J_\Gamma}$ using a modified version of the *gamv* program included within the GSLIB library of geostatistical routines [*Deutsch and Journel*, 1998].

2.3. Choosing the Parameters α and β

[26] Each term of the objective function in equation (1) is nonunique, and any solution $\mathbf{m_{est}}$ that minimizes the objective function is a single realization that appropriately fits the data and the semivariogram as given by ϵ and τ. The path a particular solution takes through the model space as the inversion algorithm progresses, and ultimately the final solution, depends both upon the starting model and upon the way in which β is chosen at each iteration of the inversion. For a regularized inversion in which $\Phi_v = 0$ has a unique solution, β is chosen so that $\Phi_d = \epsilon$. If Φ_d drops below ϵ (i.e., the data are overfit) the next iteration will increase β which smoothes the updated model and increases Φ_d.

[27] In our case, consideration must be taken because $\Phi_v = 0$ does not have a unique solution. If at some iteration in the inversion Φ_d drops below ϵ, increasing β will not necessarily increase Φ_d because the inversion algorithm may find an updated solution that reduces both Φ_d and Φ_v. The algorithm may ultimately find a solution that fits the data 'exactly' (or overfits the data) and also fits the semivariogram. Thus it is possible for the inversion routine to encounter the situation where $\Phi_d < \epsilon$ cannot be corrected by increasing β. If $\Phi_d < \epsilon$ the solution honors data noise which causes artifacts in the solution.

[28] Our approach to ensuring $\Phi_d \geq \epsilon$ for all iterations is based on the three conditions that can exist as the inversion progresses.

2.3.1. Condition 1: Both the Data and Semivariogram(s) Are Underfit ($\Phi_d > \epsilon$ and $\Phi_v > \tau$)

[29] In this case β is set so that equal importance (or weight) is placed on minimizing Φ_d and Φ_v,

$$\beta^{k+1} = \frac{\Phi_d^k}{\Phi_v^k}. \tag{13}$$

The damping parameter α is chosen on the basis of the total misfit of the objective function as follows:

$$\alpha^{k+1} = C\left[\left(\Phi_d^k - \epsilon\right)^2 + \beta^k\left(\Phi_v^k - \tau\right)^2\right]. \tag{14}$$

Here C is a user-chosen constant that controls the rate of decrease in α as the objective function is minimized. The damping parameter α must decrease with the objective function value. If not, the damping term will dominate the inversion as the objective function decreases, and progress will cease. For now, C is a user-chosen value based on trial and error. A C value that is too large will result in slow progress, and a C value that is too small will result in an oscillating solution. Once a suitable C value is determined, that C value may be used for each solution in the ensemble.

2.3.2. Condition 2: Data Are Fit and Semivariogram(s) are Underfit ($\Phi_d \leq \epsilon$ and $\Phi_v > \tau$)

[30] In this case $\mathbf{J_G}$ is not computed and row 1 of equation (4) is removed so that the inversion focuses only on fitting the semivariogram. Thus

$$\begin{aligned} \beta^{k+1} &= \beta^k \\ \alpha^{k+1} &= C\left[\beta^k\left(\Phi_v^k - \tau\right)^2\right]. \end{aligned} \tag{15}$$

2.3.3. Condition 3: Data Are Underfit and Semivariogram(s) Are Fit ($\Phi_d > \epsilon$ and $\Phi_v \leq \tau$)

[31] In this case $\mathbf{J_\Gamma}$ is not computed and row 2 of equation (4) is removed so that the inversion focuses only on fitting the data. Thus the value of β is of no consequence, and α is given by

$$\alpha^{k+1} = C\left[\left(\Phi_d^k - \epsilon\right)^2\right]. \tag{16}$$

[32] The fourth possible condition ($\Phi_d \leq \epsilon$ and $\Phi_v \leq \tau$) represents convergence. Although other choices of α and β are possible, the choices given above disallow a significant overfit to Φ_d and provide a method to optimize convergence speed by adjusting the C value.

2.4. Obtaining Slowness Semivariograms From Porosity Semivariograms

[33] Although in this work we present results and discussion in terms of velocity for the synthetic example, the actual estimated parameter in velocity tomography is slowness, which is the reciprocal of velocity. Therefore we use the term slowness and velocity interchangeably to indicate the estimated parameter. In the field example shown in the Results section, we obtain slowness semivariograms from porosity semivariograms using Topp's equation [*Topp et al.*, 1980] which is an empirical petrophysical model giving the electromagnetic wave slowness as a function of saturated porosity. *Ferre et al.* [1996] showed that over a wide range of porosities Topp's equation may be approximated by the equation

$$S(x) = \frac{\Theta(x) + 0.1841}{0.1181 C_e}. \tag{17}$$

Here $S(x)$ and $\Theta(x)$ are slowness and porosity at position x, respectively, and C_e is the free-space electromagnetic wave velocity. The difference between Topp's equation and equation (17) is less than 3 percent for $\Theta(x) \leq 0.35$. Substituting equation (17) into the slowness semivariogram equation gives

$$\gamma_s(h) = \frac{1}{2N} \sum_{i=1}^{N} \left(\frac{\Theta\left(x_{i,k}\right) + 0.1841}{0.1181 C_e} - \frac{\Theta\left(x_{i,j}\right) + 0.1841}{0.1181 C_e} \right)^2. \tag{18}$$

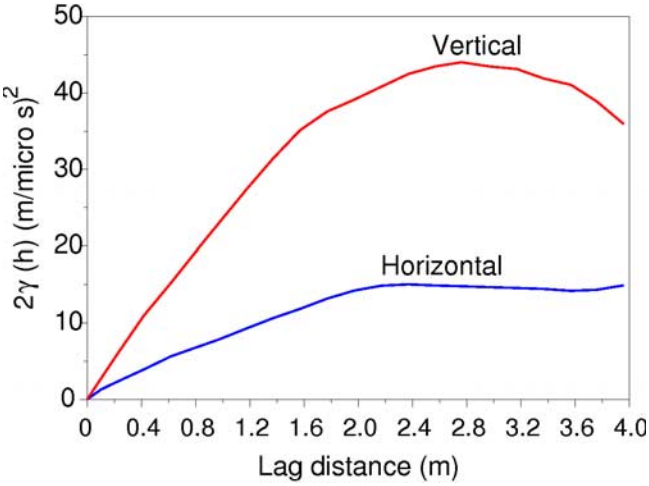

Figure 2. Optimal (or target) horizontal and vertical velocity semivariograms used in unconditional simulation and synthetic velocity tomography examples.

Equation (18) can be simplified to

$$\gamma_s(h) = \frac{1}{(0.1181 C_e)^2}\gamma_\Theta(h), \tag{19}$$

where $\gamma_\Theta(h)$ is the porosity semivariogram given by

$$\gamma_\Theta(h) = \frac{1}{2N}\sum_{i=1}^{N}\left(\Theta(x_{i,k}) - \Theta(x_{i,j})\right)^2. \tag{20}$$

In our field example we use equation (19) to convert $\gamma_\Theta(h)$ semivariograms to $\gamma_s(h)$ semivariograms which are in turn used to constrain the tomographic inversion.

3. Results

3.1. Examining and Removing Solution Bias

[34] In order to accurately estimate solution statistics by sampling from the solution space, we must ensure that the solutions are equally probable and unbiased. In this section we discuss biasing issues related to fitting the spatial covariance and we present a practical methodology for eliminating spatial covariance biasing.

[35] To evaluate biasing issues we use the concept that, for an unconditional solution, unbiased cells display no tendency to vary or assume a value that is different than any other unbiased cell. That is, each unbiased cell has the same probability distribution (e.g., mean and variance) in the unconditional case, and the ensemble mean and variance of unbiased cells will approach the expected mean and variance as the number of solutions in the ensemble increases. Because the semivariogram gives no information concerning absolute values, the expected mean is insensitive to the semivariogram and depends solely on the starting models in the unconditional case. If starting-cell values are chosen from distributions with differing means, then cell ensemble means will also have differing values resulting in a bias caused by the choice of starting models. We avoid starting-model bias by populating each starting-model cell with values sampled from the same normal distribution. The expected mean in this case is the mean of the starting-model probability distribution, and the ensemble mean of unbiased cells will approach this expected mean (for the unconditional case) as more solutions are added to the ensemble.

[36] In contrast to the expected mean, the expected variance of unbiased cells is dependent upon the semivariogram and independent of the starting-model distribution. The starting-model variance of each cell has no effect on the expected variance of that cell, because the expected variance is specified by the semivariogram sill (in the unconditional case only). When the spatial covariance is stationary, directional semivariograms have the same sill and the expected variance for each cell is the sill value. In this study however, we show a synthetic example using semivariograms having different sill values in the horizontal and vertical directions. In this case the expected variance lies between the sill values and the ensemble variance of all unbiased cells will approach this expected variance. We also consider a field example for a system which is nonstationary overall. The system includes subregions (layers and lenses) which have means and variances that are consistent within a given subregion but different between subregions [*Barrash and Clemo*, 2002]. As we will show, ensemble variance estimates can be biased because of the decreasing number of cell pairs near model boundaries. If unaccounted for, this spatial covariance biasing decreases the utility of using ISS to estimate solution uncertainty measures such as ensemble variance near boundaries.

[37] Spatial covariance biasing arises because of unequal sampling of the model grid at different scales (i.e., lag separations) when computing the semivariogram sensitivities by equation (12). For example, consider the horizontal and vertical semivariograms shown in Figure 2. These semivariograms represent a nonstationary model (layered system with different means and variances per layer; note the resulting different horizontal and vertical sill values [*Kupfersberger and Deutsch*, 1999; *Barrash and Clemo*, 2002]) and were chosen to demonstrate the indifference of the ISS method to stationarity. That is, ISS is equally useful for any specified semivariogram or system of semivariograms regardless of stationarity.

[38] Figures 3a, 3b, and 3c show three unconditional solutions generated by ISS that conform to the semivariograms in Figure 2 up to a lag of 3 m. The division of these solutions into zones will be explained shortly. To demonstrate spatial covariance biasing, consider a single cell located at position (−3 m, 5 m) on the left edge of the model. From that cell the model extends vertically to the maximum lag distance of the vertical semivariogram (3 m) both above and below. Thus this cell is associated with an equal number of pairs at each lag in the vertical direction (assuming a uniform grid). In the horizontal direction there are no values to the left of the cell, and consequently a lesser number of cell pairs compared to the vertical direction. Thus the horizontal semivariogram is less sensitive to this cell than is the vertical semivariogram. Because of this sensitivity difference, the inversion algorithm uses this cell more strongly to fit the vertical semivariogram than the horizontal semivariogram, resulting in a bias toward the vertical semivariogram. If we move the cell to the right the length of one horizontal lag to (−2.8 m, 5 m), the horizontal semivariogram value at the first lag is no longer biased because there

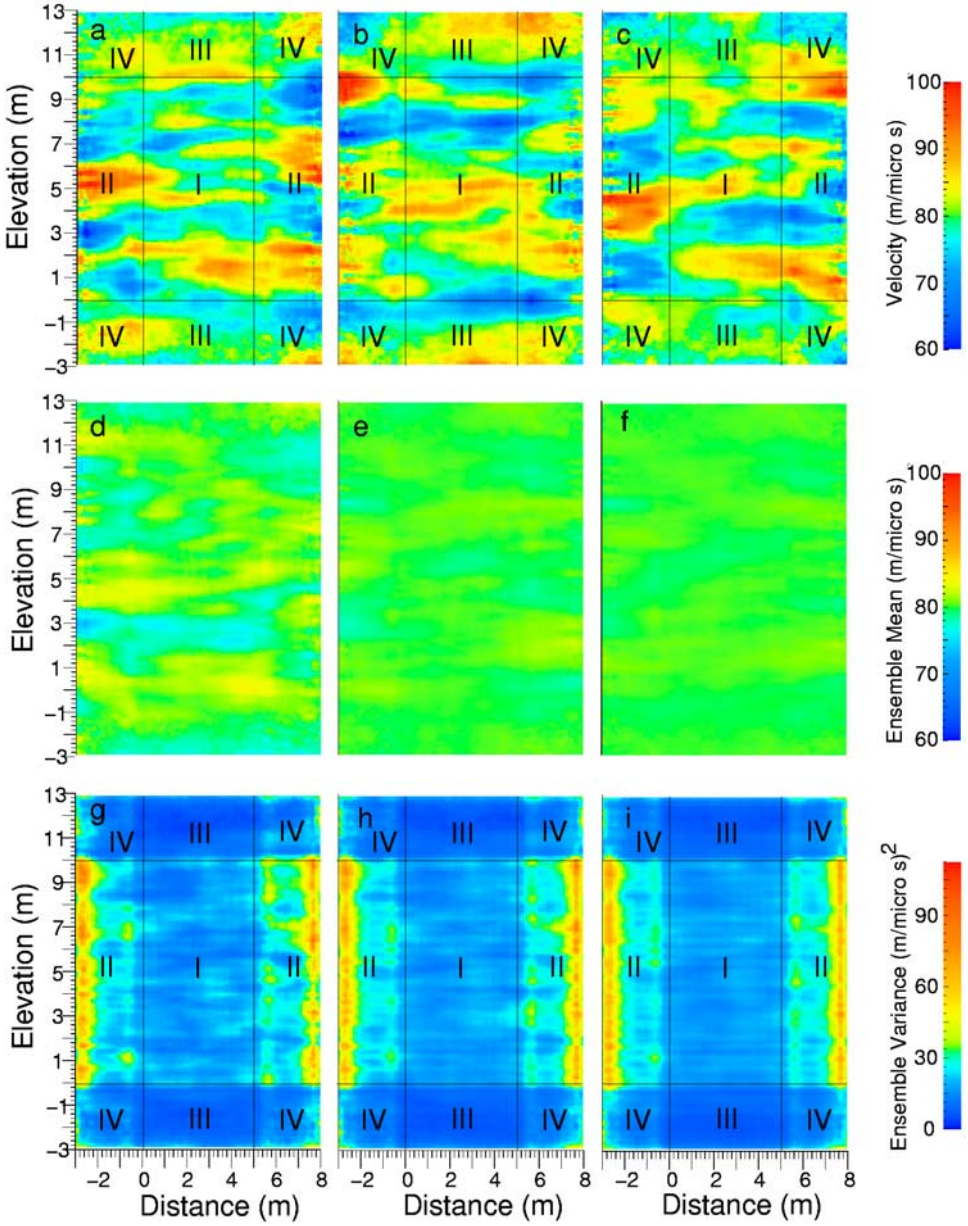

Figure 3. ISS unconditional simulation results demonstrating spatial covariance biasing of the velocity model using semivariograms in Figure 2. (a, b, and c) Sample ISS unconditional solutions. (d, e, and f) Ensemble mean of 10, 50, and 100 solutions, respectively. (g, h, i) Ensemble variance of 10, 50, and 100 solutions, respectively. The ensemble mean is unbiased in all regions I–IV. For ensemble variance, region I is unbiased. Region II is biased toward the vertical semivariogram. Region III is biased toward the horizontal semivariogram. Region IV is biased toward both the horizontal and vertical semivariogram.

are equal numbers of pairs in the horizontal and vertical directions with respect to the first lag. As we continue to move the cell to the right, the vertical bias is removed at increasing scale lengths because the sampling discrepancy is eliminated. When the cell is the maximum horizontal semivariogram lag distance away from the left edge (3 m), the vertical bias at all scale lengths is eliminated and the solution at that cell is no longer biased. Horizontal bias can be demonstrated in the same manner by considering, for example, a cell at (2.5 m, 13 m) and progressively moving the cell downward until the horizontal bias at all scales is removed at (2.5 m, 10 m).

[39] To show the effects of spatial covariance biasing on ensemble statistics, we again consider the models shown in Figures 3a, 3b, and 3c. Each of these models represents one unconditional solution taken from an ensemble of many solutions. To generate each solution, we create a random, normally distributed starting model with a mean velocity of 80 m/μs, a variance of 4 (m/μs)^2, and no spatial correlation. Because the final model depends in part on the starting model, we begin each inversion with a random starting model so that the final solution is not biased by the starting model. The value of the starting-model variance is of no consequence in the ensemble variance because the ensemble

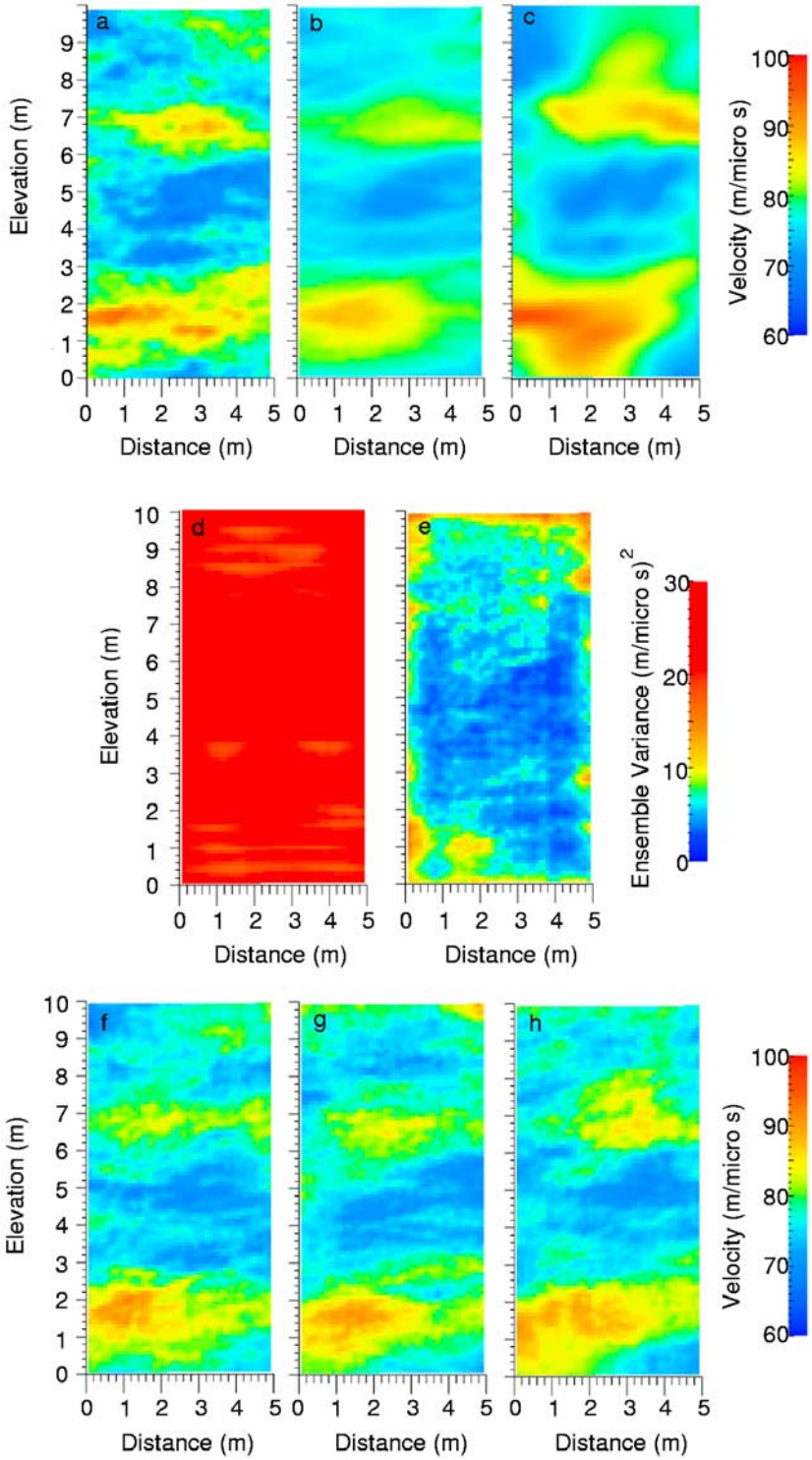

Figure 4. Synthetic velocity tomography results using ISS. (a) Optimal model. (b) ISS ensemble mean (100 solutions). (c) Regularized solution. (d) Unconditional ensemble variance. (e) Conditional ensemble variance. (f, g, and h) Sample solutions taken from the ensemble.

variance of each cell depends solely on the semivariogram sill values. Figures 3d, 3e, and 3f show the ensemble means of 10, 50 and 100 unconditional solutions respectively. As the number of solutions in the ensemble increases, the ensemble mean of each cell approaches the expected mean of 80 m/μs demonstrating that the ensemble mean is not subject to spatial covariance biasing, and is not biased. The fact that the unconditional ensemble mean is only dependent upon the starting-model mean becomes important when analyzing the conditional ensemble mean, because differences in the conditional ensemble mean arise to satisfy the conditional data (e.g., traveltime data in this

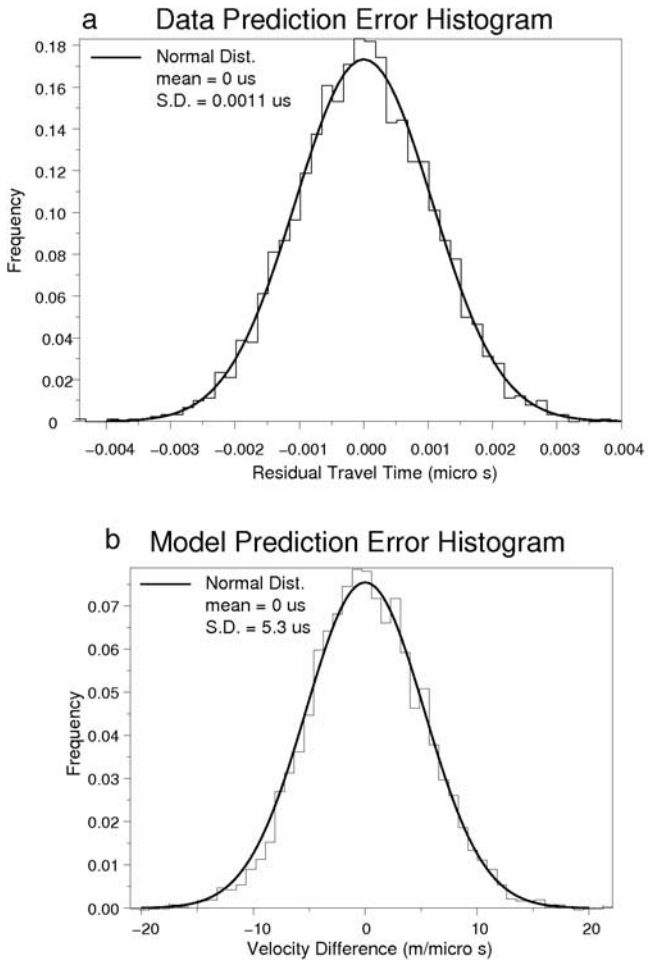

Figure 5. Typical histogram of (a) data prediction residuals and (b) model prediction residuals. Each histogram displays Gaussianity, suggesting an unbiased solution.

case) and thus give valuable insight into the resolving capability of conditional data.

[40] The effects of spatial covariance biasing on ensemble variance are demonstrated in Figures 3g, 3h, and 3i which show the ensemble variance of 10, 50, and 100 solutions respectively. The ensemble variance of each cell in the unbiased region (region I) approaches the expected value of approximately 27 $(m/\mu s)^2$, the average value of the horizontal and vertical semivariogram sills. In region II the ensemble variance increases toward the boundaries because solutions are vertically biased and the vertical semivariogram has a greater sill value. The effects of the vertical bias are also evident at the solution boundaries of region II (Figures 3a, 3b, and 3c) which display greater relative variability. In region III, where the horizontal semivariogram dominates at the upper and lower boundaries, the ensemble displays lower variance which is also evident in each solution. Region IV is both vertically and horizontally biased. Region I is the only region which displays the unconditional expected variance due to the combined constraints imposed by the vertical and horizontal semivariograms.

[41] At this point in the development of ISS we remove spatial covariance biasing by expanding the model boundaries beyond the region of interest to a distance of the maximum horizontal and vertical lag distance in the horizontal and vertical directions respectively. Although removing the bias by expanding the boundaries increases the size of the model grid, it may not be necessary in many circumstances. For instance, it is often necessary to expand the boundaries of numerical models beyond the region of interest to reduce boundary effects. In such cases, the model will not be biased if the boundaries are at least the maximum lag distance away from the region of interest.

[42] We found that, in general, expanding the boundaries does not cause a large increase in computation time for two reasons. First, the expanded part of the model only applies to the semivariogram constraints. The data are not sensitive to (and data sensitivities are not computed for) the expanded portion of the model, which essentially becomes unconditional. Second, the semivariogram sensitivities are efficient to compute. However, expanding the boundaries will have a more pronounced effect on memory requirements, particularly for very large 3D models. We recognize that the necessity of expanding model boundaries is a drawback to the ISS method, but only if model boundaries are closer than one semivariogram range to the region of interest. We are currently investigating more efficient methods of removing spatial covariance biasing. For this work, we remove spatial covariance biasing by expanding the boundaries in the examples that follow, but only the region of interest is shown.

3.2. Synthetic Data Example

[43] In this section we demonstrate ISS using a synthetic data example. Figure 4a shows the synthetic radar velocity model used to construct the semivariograms shown in Figure 2. We generated traveltime data by solving the eikonal equation [*Aldridge and Oldenburg*, 1993] with 41 sources placed at 0.25 m vertical increments in the source well at the left boundary of the model. Receivers were placed at the right boundary of the model with the same configuration as the source well for a total of 1681 source-receiver pairs (i.e., 1681 traveltime data). The horizontal distance between the source well and receiver well is 5 m. The traveltime data were contaminated with normally distributed noise having a mean of zero and a standard deviation of 0.5% of the maximum traveltime. Each grid cell is 0.1 m by 0.1 m for a total of 5000 grid cells. The semivariograms were generated with the kt3d program of the GSLIB library [*Deutsch and Journel*, 1998] using a search angle of 25 degrees and a maximum search width of 0.8 m. Semivariogram values were estimated at approximately 0.2 m lags.

[44] We generated an ensemble of conditional solutions using the synthetic traveltimes and velocity semivariograms. Variations in the ensemble mean and variance typically stabilized to an approximately constant value after about 60 solutions. We generated 100 solutions for this example. To determine each starting model, we computed the mean and standard deviation of the apparent velocity for each source-receiver pair. The apparent velocity is the source-receiver separation distance divided by the traveltime. The starting models were then populated with normally distributed velocity values having no spatial correlation and a mean and variance equal to the mean and variance of the apparent velocity.

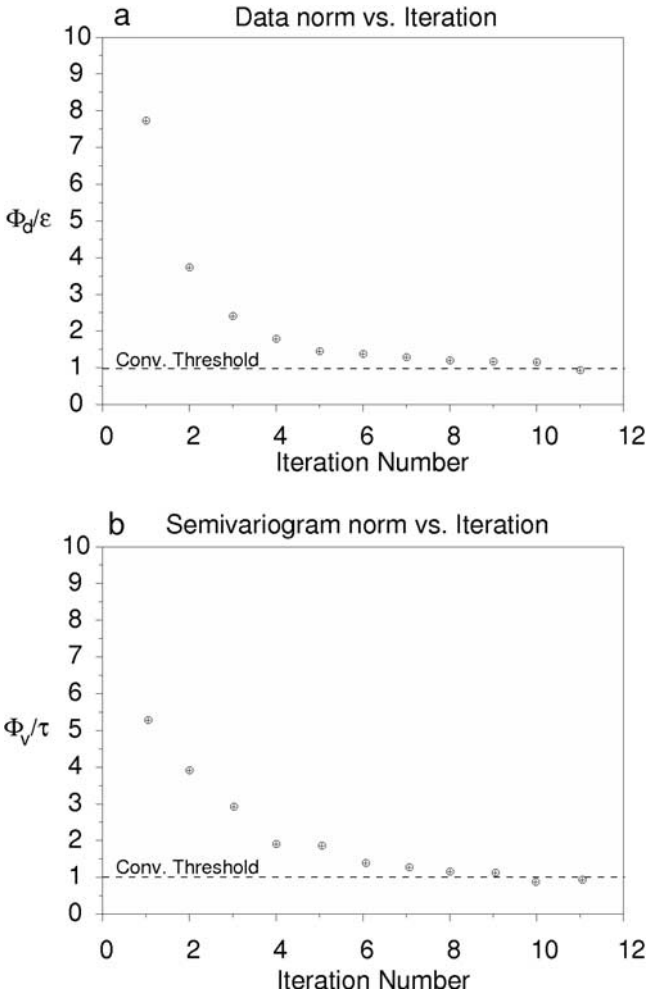

Figure 6. Typical convergence of normalized (a) data norm Φ_d and (b) semivariogram norm Φ_v. As discussed in section 2.3, the weighting parameter β is chosen so that equal weight is placed on Φ_d and Φ_v until iteration 10 when Φ_v drops below the convergence threshold. At iteration 11, the inversion focuses only on reducing Φ_d, and Φ_d drops below the convergence threshold, causing the solution to meet the convergence criteria.

[45] The optimal velocity distribution, ensemble mean, and regularized solution are shown in Figures 4a, 4b, and 4c, respectively. Figure 4c is regularized using Tikhonov type regularization [*Tikhonov*, 1963] by constraining the inversion to minimize the first spatial derivative between cells with equal weighting in the horizontal and vertical directions. The regularization weighting is chosen so that the normalized chi-square value of squared data residuals is equal to 1 (i.e., the data are appropriately fit given the noise level).

[46] The unconditional and conditional ensemble variances are shown in Figures 4d and 4e. The unconditional variance is also shown in region I of Figure 3i, but with a different color scale. Figures 4f, 4g, and 4h are example conditional solutions taken from the ensemble. Each solution displays the general larger-scale structure of the optimal model. Each solution also includes smaller-scale features which are in the null-space of the traveltime data, (i.e., are not resolved by the traveltime data), but are included in the model in order to satisfy the short-lag semivariogram constraints.

[47] We can investigate the conditional solution bias for the synthetic example using residuals analysis. Assuming the data noise are normally distributed (a condition imposed in this synthetic case), we would expect the traveltime residual (predicted minus observed data) to be normally distributed for an unbiased model. Figure 5a shows a typical data residual histogram and, for comparison, a normal distribution with zero mean and the same standard deviation as the residual histogram, for a typical model in the ensemble. The residuals display a normal distribution with zero mean, suggesting unbiased velocity estimates. In this synthetic case, we can also investigate bias in the conditional solution by examining the distribution of the difference between predicted and optimal model values, which should also be zero mean and normally distributed for unbiased velocity estimates. Figure 5b shows the model residual distribution for a typical solution in the ensemble. The model residuals also display normality which further indicates that the velocity estimates are unbiased.

[48] Figure 6 shows the typical convergence sequence for a solution. Each term of the objective function decreases toward the convergence threshold at approximately the same rate because the weighting term β is chosen such that equal weight is placed on each term. At iteration number 10, Φ_v drops below the convergence threshold so that at iteration 11, the inversion focuses only on fitting Φ_d, causing Φ_d to drop slightly below the convergence threshold and the solution to converge. Note that at iteration 11, Φ_v increases slightly (because the inversion is neglecting the semivariogram term), but does not increase above the convergence threshold. If Φ_v did increase above the convergence threshold, then the inversion would extend to iteration 12 which would focus only on minimizing the semivariogram norm as discussed in section 2.3.

[49] Figure 7 shows the typical velocity semivariogram fit for a given solution in the conditional ensemble (see

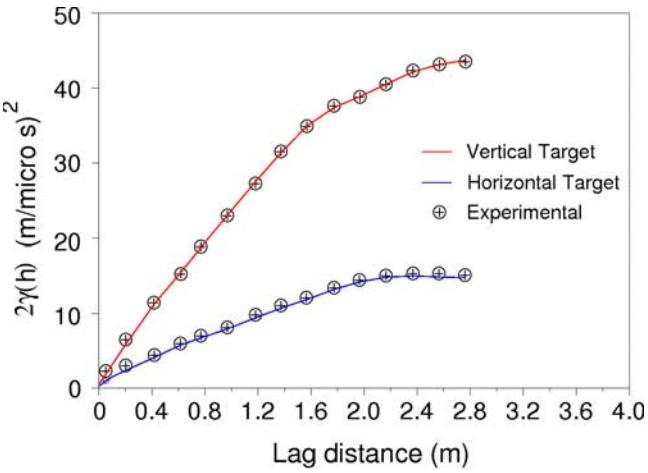

Figure 7. Optimal velocity semivariograms and typical experimental velocity semivariogram values for the synthetic conditional solutions. The optimal (or target) semivariograms are equivalent to those shown in Figure 2 up to a lag of 3 m.

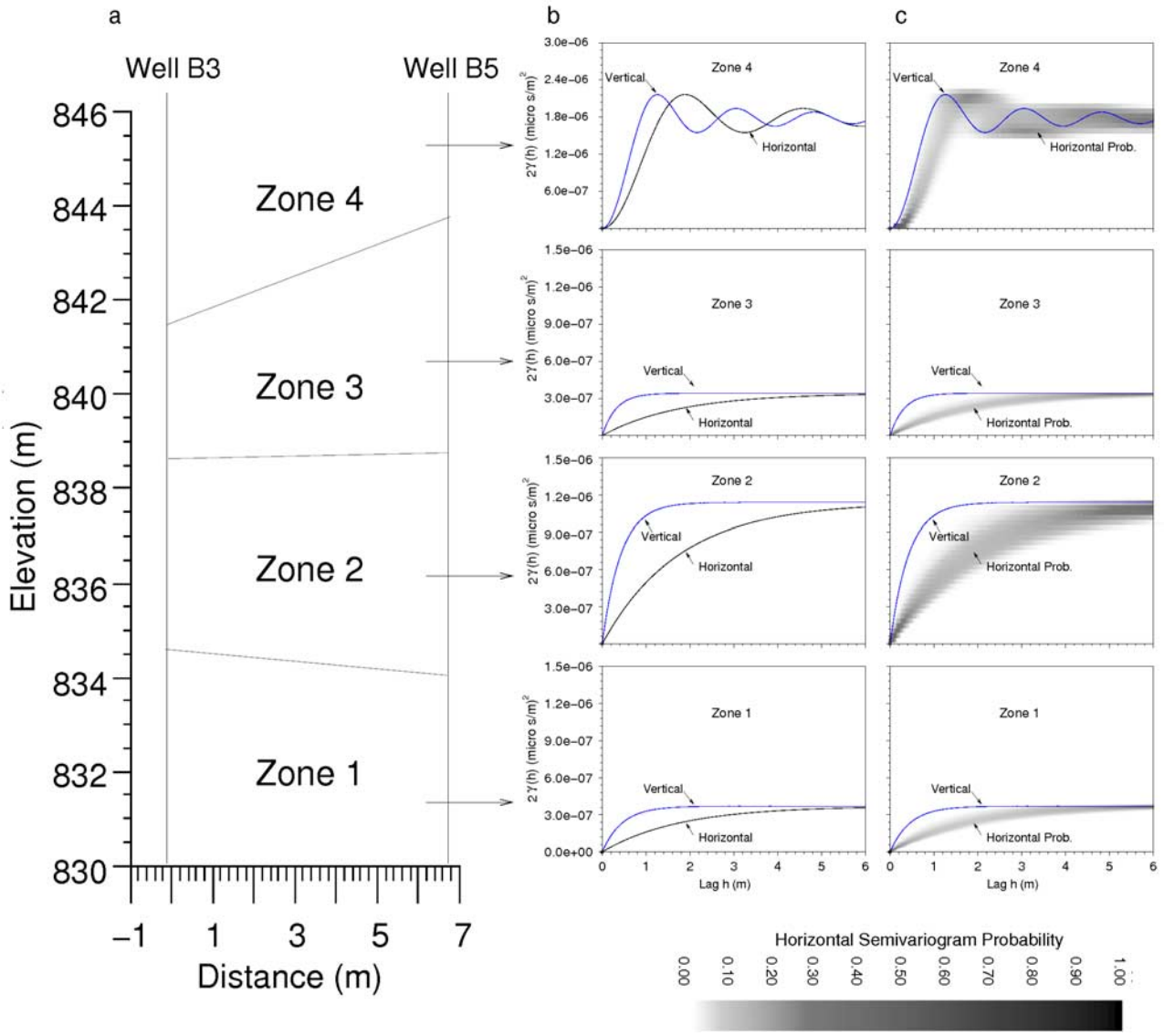

Figure 8. (a) Zones 1 through 4 between wells B3 and B5 as identified by borehole porosity logs [*Barrash and Clemo*, 2002]. (b) Zonal slowness semivariograms for each zone. The vertical range and sill and horizontal sill were identified by *Barrash and Clemo* [2002]. The horizontal range is assumed to be approximately 6 m in each zone in this case. (c) Zonal slownes semivariograms for each zone with uncertainty in the horizontal range. In this case, the horizontal range for each zone is assumed to have a normal probability distribution with a mean of 6 m and a standard deviation of 2 m, giving the shaded horizontal semivariogram probability distribution with respect to lag distance.

Figures 4f, 4g, and 4h, for example). The semivariogram fit is specified by $\mathbf{W}_v$, whose diagonal elements are 1% of the corresponding semivariogram magnitude in this case.

3.3. Field Example

[50] The Boise Hydrogeophysical Research Site (BHRS) is a research well field in a shallow unconfined aquifer consisting of heterogeneous coarse-grained fluvial deposits. The arrangement of wells at the BHRS is designed to capture the three-dimensional distributions of geologic, hydrogeologic and geophysical parameters [*Barrash and Knoll*, 1998]. *Barrash and Clemo* [2002] adopted a hierarchical (multiscale) approach to characterize the geostatistical structure of porosity at the BHRS on the basis of neutron porosity logs. They examined four stratigraphic layers within the saturated zone with different geostatistical characteristics, including the region between wells B3 and B5 (Figure 8a) which we consider for a field example here. Zones 1 through 3 are modeled with exponential semivariograms and zone 4 with a periodic semivariogram. Zones 1 and 3 display lower porosity means and variances than zones 2 and 4. The semivariograms for each zone are well constrained in the vertical direction. In the horizontal direction, sill values are relatively well constrained but the ranges for each zone are uncertain because of limited horizontal sampling [*Barrash and Clemo*, 2002].

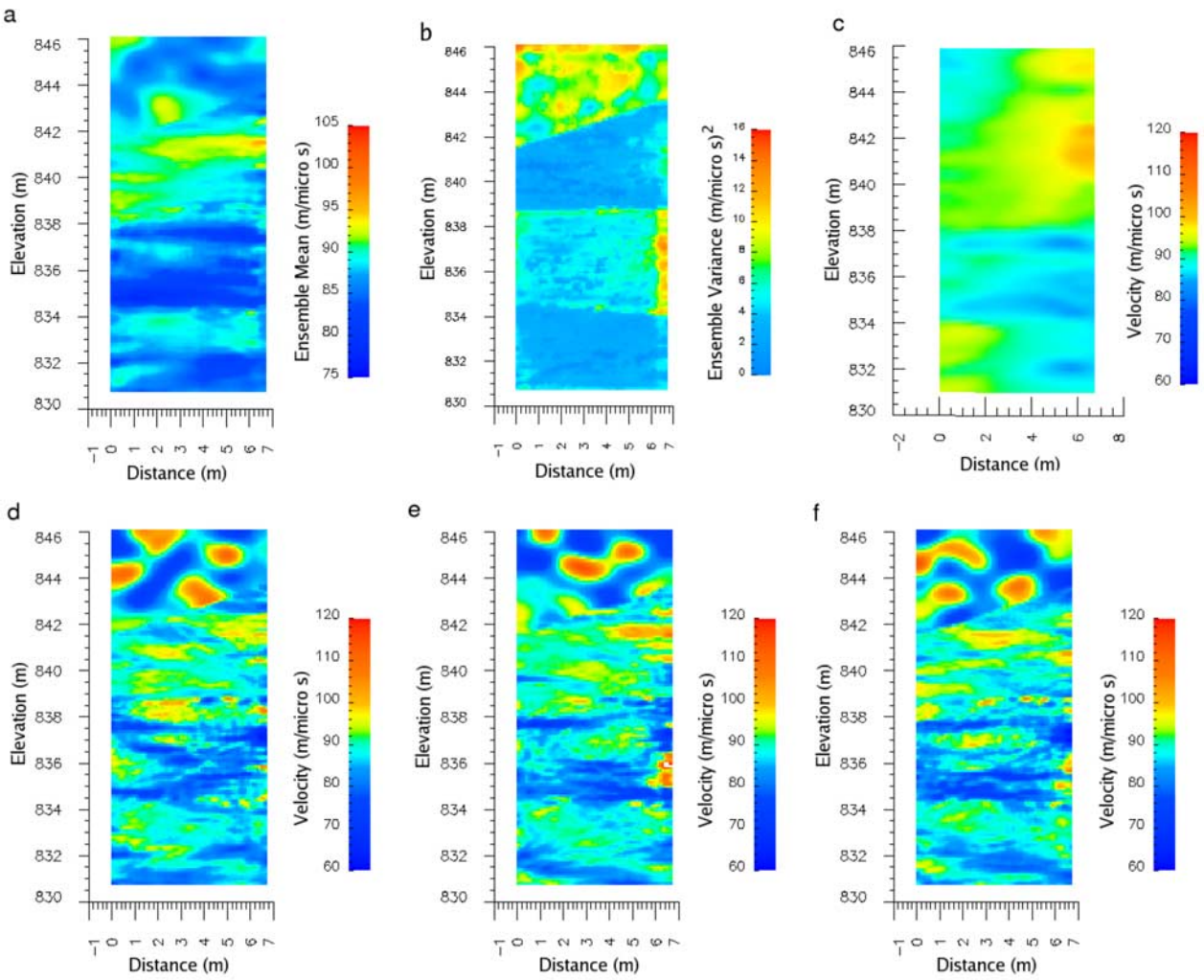

Figure 9. (a) B3–B5 ensemble mean. (b) B3–B5 ensemble variance. (c) Regularized solution. The regularization constraints for this solution minimize the first spatial derivative of the velocity values with 3 to 1 horizontal to vertical weighting. (d, e, and f) Sample solutions taken from the ensemble.

[51] In this field example we generate ensembles under two different assumptions concerning the horizontal range in each zone in order to demonstrate the effects of semivariogram uncertainty on the ensemble. In the first case, we assume the range is known and is approximately 6 m for each zone as shown in Figure 8b. (Note that in this example the semivariograms are displayed in terms of slowness which is the reciprocal of velocity). In the second case, we represent the uncertainty in the horizontal semivariogram by assuming each zone has a normally distributed horizontal range with a mean of 6 m and a standard deviation of 2 m, giving the horizontal semivariogram probability distribution for each zone shown in Figure 8c. Before each solution is constructed, a range value for each zone is randomly sampled from the horizontal range distribution described above. Once the ranges are sampled, the corresponding horizontal target semivariograms are constructed and used to constrain the spatial covariance in each zone. Thus, in the second case, the uncertainty in the horizontal range is incorporated into the ensemble and is expressed in the ensemble statistics. We expect less resolution in the ensemble mean and a larger ensemble variance for the second case where uncertainty in the horizontal semivariogram is taken into account.

[52] We used equation (19) to convert porosity semivariograms to radar slowness semivariograms and we used radar traveltime data with sources in well B3 and receivers in well B5. Data were collected at 20 cm receiver increments and 5 cm source increments for a total of 12304 traveltimes. On the basis of an independent noise analysis, we assume normally distributed noise in the traveltime data with a standard deviation of 0.5% of the maximum traveltime. For each inverse solution the spatial covariance structure is constrained by the corresponding zonal semivariograms, with no spatial correlation constraints across zone boundaries. Starting models are constructed in the same manner as discussed in the synthetic example. To invert the data we use a 6.8 m by 15.5 m grid of 0.01 m by 0.01 m cells for a total of 10695 cells. For comparison, we also compute a regularized solution using the same data, same noise assumption, and same grid configuration. The model constraints for the regularized solution specify first spatial

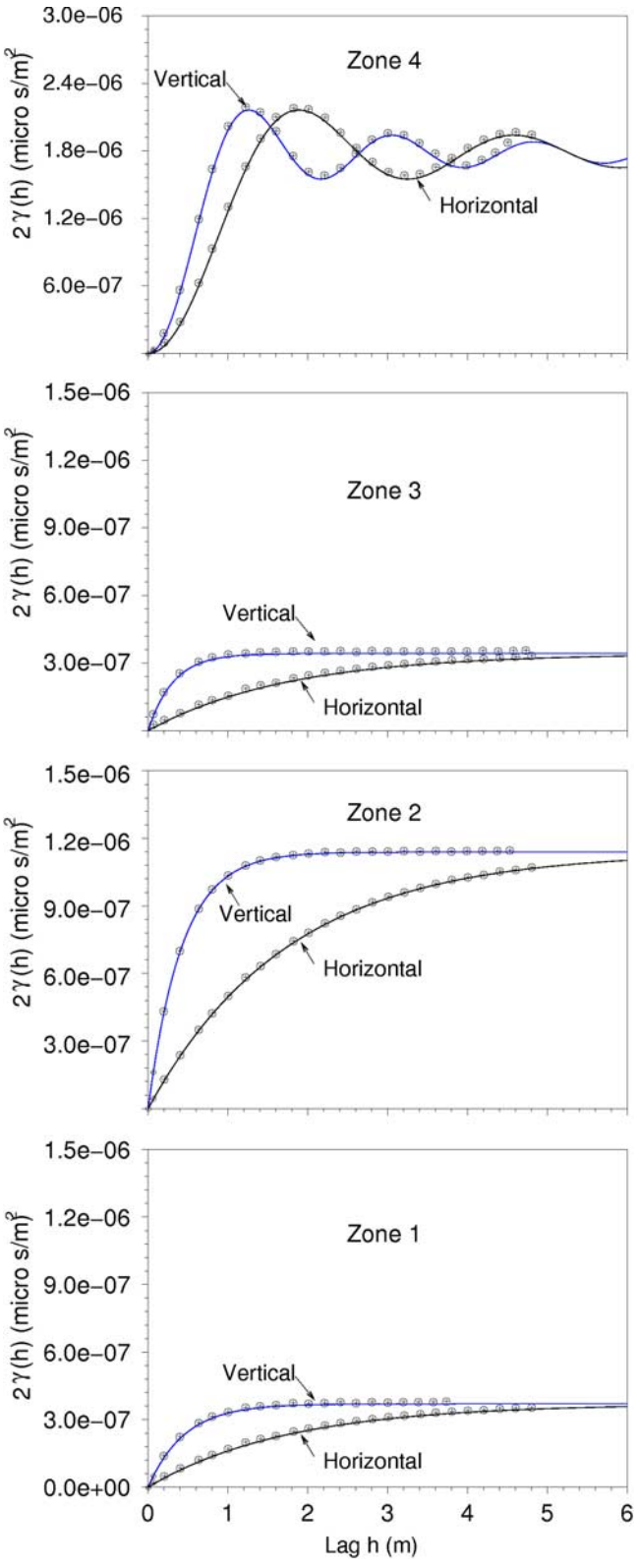

Figure 10. Typical slowness experimental semivariogram for each zone. The lines represent the optimal or target semivariogram values as shown in Figure 8. The circled pluses represent the predicted semivariogram values (i.e., the experimental semivariogram of a typical solution).

derivative minimization with a ratio of 3 to 1 horizontal to vertical weighting.

[53] For the first case (e.g., semivariograms in Figure 8b), we generated an ensemble of 100 solutions which we used to compute the ensemble mean (Figure 9a) and variance (Figure 9b). We recognize that the ensemble variance is biased to some degree because of the boundaries between zones. However, we expect the bias to be small because the horizontal and vertical semivariograms have the same sill value in any given zone. Figure 9b shows the regularized solution, and Figures 9d, 9e, and 9f show typical ISS solutions taken from the ensemble. Figure 10 shows the typical semivariogram fit for each zone.

[54] We also generated an ensemble of 100 ISS solutions for the second case where uncertainty is included in the horizontal semivariogram (see Figure 8c). The ensemble mean and variance are shown in Figures 11a and 11b, respectively. Figure 12 shows a typical example of the sampled semivariograms, semivariogram fit, and corresponding solution in the ensemble where uncertainty is included in the horizontal semivariogram range.

4. Discussion

[55] We have shown synthetic and field examples of the ISS method and we have discussed spatial covariance biasing and how to account for it. Each solution matches the semivariograms, explains the data, and represents one equally probable possibility of capturing the model that best describes the subsurface given the uncertainty in the data and in the spatial covariance structure. Although the solutions are used as a means of computing ensemble statistics in this work, each solution, generally speaking, can also provide a useful predictive model, particularly in cases where predictions are sensitive to small-scale variability such as in contaminant transport problems.

[56] The ensemble mean captures the large-scale variability well and demonstrates the resolving capability of the data. Consider that the semivariogram constraints give no information concerning the ensemble mean. In the limiting case of no data, the ensemble mean of many solutions will converge to the mean of the starting models, as shown in Figures 3d, 3e, and 3f. Smaller-scale structures in the nullspace of the data will be averaged out and hence will not be evident in the ensemble mean. Deviations from the mean of the starting models that are persistent in the ensemble mean are required only to fit the data, and thus show the information content or resolving capability of the traveltime data.

[57] Resolution is often investigated through examination of the model resolution matrix [*Menke*, 1989]. However, in nonlinear problems the model resolution matrix is dependent upon the predicted model which is itself an estimate. Thus the resolution matrix may provide a misleading estimate of the resolution in nonlinear problems [*Oldenborger*, 2006]. We believe that assessing model resolution through the ISS ensemble mean holds promise and warrants further research.

[58] The ensemble mean also identifies the larger-scale structure of the subsurface. This is particularly evident when comparing the ensemble mean of both the synthetic and field examples to the corresponding regularized solutions. In the synthetic example the ensemble mean appears to be a

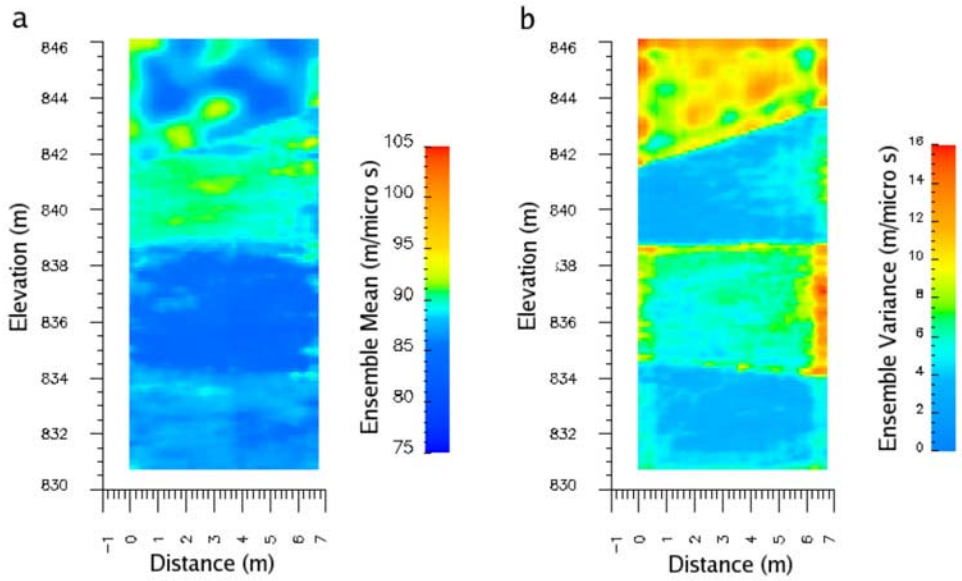

Figure 11. ISS ensemble statistics using semivariograms in Figure 8b. (a) Ensemble mean and (b) ensemble variance for 100 solutions including uncertainty in the horizontal semivariogram range. Note that the mean displays less resolution, and the overall variance is greater than in the case where uncertainty is not included in the semivariograms (see Figures 9a and 9b).

spatially averaged version of the optimal solution. The regularized solution resembles the optimal model but exhibits smearing artifacts typical of regularized solutions. For example, the bottom boundary of the regularized model exhibits a large velocity artifact that is smeared downward from higher in the model. The regularized solution in the field example exhibits these same features and we presume that the ensemble mean more accurately characterizes the primary structures of this region of the BHRS.

[59] The ensemble variance provides a useful measure of the worth of data in reducing solution uncertainty. For the unconditional synthetic case, the ensemble variance of each cell approaches the expected unconditional variance as the number of solutions in the ensemble increases. In ISS the unconditional variance, which is determined by the semivariogram sill values, is the greatest possible variance that any cell can assume. When data are added to the solution, those cells that are sensitive to the data are constrained and the ensemble variances of those cells are reduced. Cells that are less sensitive to the data are free to vary over a wider range and will exhibit greater ensemble variance values up to the limit of the expected unconditional ensemble variance (Figures 3d and 3e). The unconditional ensemble variance describes the model uncertainty given only the spatial covariance information. The conditional ensemble variance describes the model uncertainty given both spatial covariance and traveltime data. In this case, the traveltime data display great worth in reducing solution uncertainty with respect to what is known given spatial covariance information only.

[60] Although the ensemble mean and variance are useful for characterizing the subsurface, one of the more useful products of the ISS method is the ensemble of solutions itself. The ensemble provides a mechanism whereby a range of possible predicted conditions can be investigated, subject to our knowledge about the subsurface as given by the data and spatial covariance structure and associated uncertainty. Just as the solution ensemble describes the probability distribution of subsurface states given the data and spatial covariance information, predictions based on the ensemble describe the probability distribution of predicted conditions given the same information.

[61] In order for the ensemble statistics to accurately represent the posterior probability distribution, solutions in the ensemble must be unbiased, must be properly represented probabilistically, and there must be enough solutions in the ensemble to accurately describe the solution space. We have shown that the ISS method provides unbiased solutions in the unconditional case, (so long as model boundaries are expanded to remove spatial covariance biasing), demonstrating that the solutions are not biased by the spatial covariance constraints. We have also shown that solutions are unbiased in the conditional case (i.e., model and data residuals have zero mean and are normally distributed as in Figure 5) demonstrating that the solutions are not biased by the inversion procedure (i.e., the choice of α and β).

[62] We have shown how uncertainty in the spatial covariance structure can be included in the ensemble. Models with a more likely structure will occur more often in the ensemble such that the solution probability is correctly represented, given the prior uncertainty in the semivariogram. We believe that including semivariogram uncertainty is important because, in practice, the spatial structure is rarely well known. For example, in geostatistical inversion techniques a few semivariogram parameters are often fitted to many data in order to estimate the spatial structure of the subsurface. However, it is known (and is shown in this paper) that the same data set can be fit equally well by many different semivariograms. Then, even though

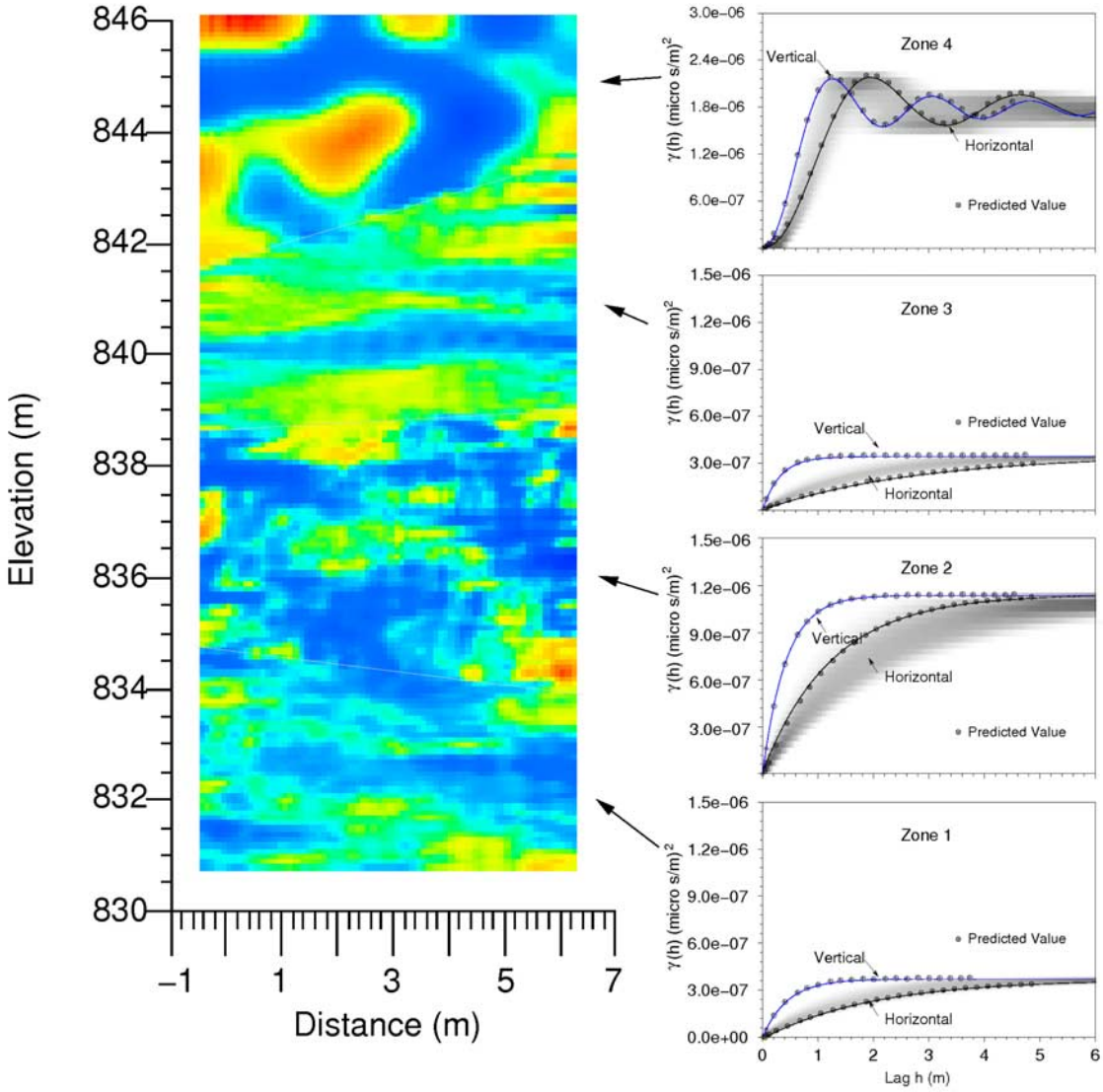

Figure 12. Example target semivariograms and corresponding solution for the case when uncertainty is included in the horizontal range. Lines represent the target semivariograms, and circled pluses represent the predicted experimental semivariogram values. Note the horizontal range and corresponding horizontal spatial covariance structure for each zone. For instance, zone 2 has a short range resulting in a more patchy structure, while zone 3 has a longer range resulting in a more layered structure.

many data are fit to a few semivariogram parameters in the geostatistical inversion approach, the estimation problem is still nonunique. If one particular estimated semivariogram is used as prior information, the resulting uncertainty estimates will likely be overly optimistic because many other semivariograms will work equally well. This point is illustrated by comparing Figures 9a and 9b to Figures 11a and 11b. When the semivariograms are assumed to be known, the ensemble mean exhibits greater resolution (9a) and smaller overall variance (9b) than when uncertainty is expressed in the horizontal range (11a and b).

[63] The primary computational expense of implementing ISS is generating an ensemble of solutions. Computing the semivariogram sensitivities is a relatively efficient process. For instance, our BHRS model consists of 10695 cells and computing the semivariogram sensitivities at each nonlinear iteration requires approximately 15 seconds s. Generating 100 solutions in this case required approximately 6 hours on a computer running a 2.4 GHz processor. Note that the convergence scheme we have presented is robust but is not optimized. While the processing demands of ISS are not prohibitive, we believe they could be reduced significantly by implementing inversion optimization techniques. In addition, the memory required to implement ISS is relatively small. One row is added to $\mathbf{J}_{\boldsymbol{\Gamma}}$ for each lag in each semivariogram. For instance, in the synthetic example $\mathbf{J}_{\boldsymbol{\Gamma}}$ has 30 rows compared to the 5000 rows in $\mathbf{J}_{\mathbf{G}}$. For the field example, $\mathbf{J}_{\boldsymbol{\Gamma}}$ has 200 rows and $\mathbf{J}_{\mathbf{G}}$ has 10695 rows.

[64] With respect to flexibility, the ISS method will find solutions honoring any semivariogram or number of semivariograms without regard to stationarity or anisotropy, so long as there is an overlapping space of solutions that also

honor the data. In addition, implementing ISS into existing deterministic codes can be accomplished in a relatively straightforward fashion by augmenting the existing Jacobian matrix with $\mathbf{J}_\Gamma$ and adding the α and β parameters to the inversion. For instance, in our field example an existing ray-based tomography code [*Aldridge and Oldenburg*, 1993] was modified to include $\mathbf{J}_\Gamma$, α and β in order to implement ISS. As we have shown, ISS can also be used as a method of unconditional simulation by disregarding the data term of the objective function. In that case, the ensemble mean will be a function of the mean of the starting models, and the ensemble variance will be controlled by the semivariogram(s) sill value.

5. Summary

[65] We have presented a method of sampling equally probable models from a space containing nonunique models that explain a particular data set and follow specified spatial covariance properties as expressed by one or more semivariograms. Because the models generated are geostatistically accurate they are useful as predictive models where small-scale variability is important. Because the models are equally probable, the ensemble statistics provide an accurate method of estimating (1) what features the data are able to resolve and (2) the worth of data for decreasing uncertainty pertaining to the existence of those features. We have also shown a realistic example of how uncertainty in spatial covariance of the subsurface may be represented in the ensemble. Assuming the uncertainty in the spatial covariance structure is adequately expressed in the inverse formulation, the ensemble of models and resulting statistics describe what is known about the subsurface given a particular data set and covariance structure.

[66] **Acknowledgments.** We appreciate the careful and thorough reviews of four anonymous reviewers. We would like to thank Mike Knoll for facilitating collection of the radar data used in the field example. This work was supported by a grant from the Inland Northwest Research Alliance (Ph.D. fellowship for T.J.) and by U.S. EPA grant X-97008501-0.

References

Aldridge, D., and D. Oldenburg (1993), Two dimensional tomographic inversion with finite-difference traveltimes, *J. Seismic Explor.*, *2*, 257–274.

Anderman, E., M. Hill, and E. Poeter (1996), Two-dimensional advective transport in ground-water flow parameter estimation, *Groundwater*, *34*(6), 1001–1009.

Anderson, M. (1997), Characterization of geologic heterogeneity, in *Subsurface Flow and Transport: A Stochastic Approach*, *Int. Hydrol. Ser.*, edited by G. Dagan and S. Neuman, chap. 2.1, pp. 23–43, Cambridge Univ. Press, New York.

Barrash, W., and T. Clemo (2002), Hierarchical geostatistics and multifacies systems: Boise Hydrogeophysical Research Site, Boise, Idaho, *Water Resour. Res.*, *38*(10), 1196, doi:10.1029/2002WR001436.

Barrash, W., and M. Knoll (1998), Design of a research wellfield for calibrating geophysical methods against hydrologic parameters, paper presented at Conference on Hazardous Waste Research, Great Plains/Rocky Mt. Hazardous Substance Res. Cent., Kans. State Univ., Snowbird, Utah.

Capilla, J., J. Gomez-Hernandez, and S. Andres (1998), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data—2. Demonstration on a synthetic aquifer, *J. Hydrol.*, *203*, 175–188.

Carrera, J., A. Alcolea, A. Medina, J. Hildalgo, and L. Slooten (2005), Inverse problem in hydrogeology, *Hydrogeol. J.*, *13*, 206–222.

Clifton, P., and S. Neuman (1982), Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona, *Water Resour. Res.*, *18*(4), 1215–1234.

Copty, N., and Y. Rubin (1995), A stochastic approach to the characterization of lithofacies from surface seismic and well data, *Water Resour. Res.*, *31*(7), 1673–1686.

Day-Lewis, F., J. Harris, and S. Gorelick (2002), Time-lapse inversion of crosswell radar data, *Geophysics*, *69*(6), 1740–1752.

Deutsch, C., and A. Journel (1998), *GSLIB: Geostatistical Software Library and Users Guide*, 2nd ed., Oxford University Press, New York.

Doherty, J. (2003), Ground water model calibration using pilot points and regularization, *Groundwater*, *41*(2), 170–177.

Ferre, P., D. Rudolph, and R. Kachanoski (1996), Spatial averaging of water content by time domain reflectometry: Implications for twin rod probes with and without dielectric coatings, *Water Resour. Res.*, *32*(2), 271–279.

Gilks, W., S. Richardson, and D. J. Spiegelhalter (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hill, London.

Gomez-Hernandez, J., A. Sahuquillo, and J. Capilla (1997), Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data—I. Theory, *J. Hydrol.*, *203*, 162–174.

Hill, M. (1998), Methods and guidelines for effective model calibration, *U.S. Geol. Surv. Water Resour. Invest. Rep.*, *98-4005*.

Hubbard, S., and Y. Rubin (2000), Hydrogeological parameter estimation using geophysical data: A review of selected techniques, *Contam. Hydrol.*, *45*, 3–34.

Hubbard, S., Y. Rubin, and E. Majer (1999), Spatial correlation structure estimation using geophysical and hydrogeological data, *Water Resour. Res.*, *35*(6), 1809–1826.

Hyndman, D., J. Harris, and S. Gorelick (1994), Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resour. Res.*, *30*(7), 1965–1977.

Isaaks, E., and R. Srivastava (1989), *Applied Geostatistics*, Oxford Univ. Press, New York.

Journel, A., and C. Huijbregts (1978), *Mining Geostatistics*, Academic, New York.

Jussel, P., F. Stauffer, and T. Dracos (1994), Transport modeling in heterogeneous aquifers: 2. Three-dimensional transport model and stochastic numerical tracer experiments, *Water Resour. Res.*, *30*(6), 1819–1831.

Kitanidis, P., and E. Vomvoris (1983), A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one dimensional simulations, *Water Resour. Res.*, *19*(3), 677–690.

Kupfersberger, H., and C. Deutsch (1999), Methodology for integrating analog geologic data in 3-D variogram modeling, *AAPG Bull.*, *83*(8), 1262–1278.

McKenna, S., and E. Poeter (1995), Field example of data fusion in site characterization, *Water Resour. Res.*, *31*(12), 3229–3240.

Medina, A., and J. Carrera (2003), Geostatistical inversion of coupled problems: Dealing with the computational burden and different types of data, *J. Hydrol.*, *281*, 251–264.

Menke, W. (1989), *Geophysical Data Analysis: Discrete Inverse Theory*, Academic, San Diego, Calif.

Neuman, S. (1973), Calibration of distributed parameter groundwater flow models viewed as a multiple objective decision process under uncertainty, *Water Resour. Res.*, *9*(4), 1006–1021.

Neuman, S. (1994), Generalized scaling of permeabilities: Validation and effect of support scale, *Geophys. Res. Lett.*, *21*(5), 349–352.

Oldenborger, G. (2006), Advances in electrical resistivity tomography: Modeling, electrode position errors, time-lapse monitoring of an injection/withdrawal experiment, and solution appraisal, Ph.D. dissertation, Boise State Univ., Boise.

Oliver, D., L. Cunha, and A. Reynolds (1997), Markov chain Monte Carlo methods for conditioning a permeability field to pressure data, *Math. Geol.*, *29*(1), 61–91.

Paige, C., and M. Saunders (1982), LSQR: An algorithm for sparse linear equations and sparse least squares, *Trans. Math. Software*, *8*(1), 43–71.

Parker, R. (1994), *Geophysical Inverse Theory*, Princeton Univ. Press, Princeton, N. J.

Sudicky, E., and P. Huyakorn (1991), Contaminant migration in imperfectly known heterogeneous groundwater systems, *Rev. Geophys.*, *29*, 240–253.

Tikhonov, A. (1963), Regularization of incorrectly posed problems, *Sov. Math. Dokl., Engl. Transl.*, *4*, 1624–1627.

Tonkin, M., T. Clemo, and J. Doherty (2003), Computationally efficient regularized inversion for highly parameterized modflow models, in *MODFLOW and More 2003: Understanding through Modeling*, pp. 595–599, Int. Groundwater Modeling Cent., Golden, Colo.

Topp, G., J. Davis, and A. Annan (1980), Electromagnetic determination of soil water content: Measurements in coaxial transmission lines, *Water Resour. Res.*, *16*(3), 574–582.

Zhou, C., L. Lui, and J. Land (2001), Nonlinear inversion of borehole-radar tomography data to reconstruct velocity and attenuation distribution in Earth materials, *J. Appl. Geophys.*, *47*, 271–284.

Zimmerman, D., et al. (1998), A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resour. Res.*, *34*(6), 1373–1414.

T. C. Johnson, Energy Resource Recovery and Management, Idaho National Laboratory, P.O. Box 1625, Idaho Falls, ID 83415-2107, USA. (timothy.johnson@inl.gov)

W. Barrash, W. P. Clement, T. Clemo, and P. S. Routh, Center for Geophysical Investigation of the Shallow Subsurface, Department of Geosciences, Boise State University, 1910 University Dr., Boise, ID 83725-1536, USA.