DTU Library

Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information

Cordua, Knud Skou; Hansen, Thomas Mejer; Mosegaard, Klaus

Link back to DTU Orbit

Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information

Knud Skou Cordua[1], Thomas Mejer Hansen[1], and Klaus Mosegaard[1]

ABSTRACT

We present a general Monte Carlo full-waveform inversion strategy that integrates a priori information described by geostatistical algorithms with Bayesian inverse problem theory. The extended Metropolis algorithm can be used to sample the a posteriori probability density of highly nonlinear inverse problems, such as full-waveform inversion. Sequential Gibbs sampling is a method that allows efficient sampling of a priori probability densities described by geostatistical algorithms based on either two-point (e.g., Gaussian) or multiple-point statistics. We outline the theoretical framework for a full-waveform inversion strategy that integrates the extended Metropolis algorithm with sequential Gibbs sampling such that arbitrary complex geostatistically defined a priori information can be included. At the same time we show how temporally and/or spatially correlated data uncertainties can be taken into account during the inversion. The suggested inversion strategy is tested on synthetic tomographic crosshole ground-penetrating radar full-waveform data using multiple-point-based a priori information. This is, to our knowledge, the first example of obtaining a posteriori realizations of a full-waveform inverse problem. Benefits of the proposed methodology compared with deterministic inversion approaches include: (1) The a posteriori model variability reflects the states of information provided by the data uncertainties and a priori information, which provides a means of obtaining resolution analysis. (2) Based on a posteriori realizations, complicated statistical questions can be answered, such as the probability of connectivity across a layer. (3) Complex a priori information can be included through geostatistical algorithms. These benefits, however, require more computing resources than traditional methods do. Moreover, an adequate knowledge of data uncertainties and a priori information is required to obtain meaningful uncertainty estimates. The latter may be a key challenge when considering field experiments, which will not be addressed here.

INTRODUCTION

Albert Tarantola was one of the pioneers of seismic full-waveform inversion (see Tarantola, 1984, 1986, 1988). Using a steepest descent algorithm, he obtained the update gradient in each iteration by correlating a forward-propagated wavefield with the residual wavefield propagated backward in time from the receiver positions. This approach has later been referred to as the adjoint method (Talagrand and Courtier, 1987). The first numerical tests based on finite-difference simulations of the seismic signal showed promising results (Gauthier et al., 1986). Since then, several full-waveform inversion algorithms, based on Tarantola's pioneering work, have been developed and applied to seismic data (e.g., Mora, 1987; Crase et al., 1990; Pica et al., 1990; Djikpéssé and Tarantola, 1999).

Ground-penetrating radar (GPR) crosshole tomography is a popular method used to obtain tomographic images of near-surface geological structures and geophysical parameters. The crosshole GPR experiment involves a transmitting radar antenna (20 MHz–1 GHz; see Reynolds, 1997) lowered into a borehole and a receiving antenna placed in an adjacent borehole. The boreholes are typically separated by a distance of 5 m–20 m and are 5 m–100 m deep (e.g., Ernst et al., 2007a; Looms et al., 2008). An antenna is kept fixed in one borehole, while the other antenna is moved between multiple locations in the opposite borehole. The fixed antenna is moved to a new position and the procedure is repeated. At each combination of antennae positions a signal is transmitted between the antennae. In this way an arbitrary dense tomographic data set that covers the interborehole region can be obtained (Peterson, 2001).

Ernst et al. (2007a, 2007b) introduced and applied the adjoint-based optimization algorithm for inversion of tomographic crosshole GPR full-waveform data. They demonstrate that subwavelength features can be resolved when the full-waveform information of the GPR signal is accounted for during the inversion, which was a considerable improvement, compared with traditional ray-based inversion strategies. The inversion scheme of Ernst et al. (2007b) was later adopted for inversion of seismic waveform data by Belina et al. (2009). Recently, the algorithm of Ernst et al. (2007b) has been improved such that it exploits the full vector field of the electromagnetic wave propagation and allows for arbitrary antennae geometry (Meles et al., 2010). Klotzsche et al. (2010) demonstrate an application of the improved code to crosshole GPR waveform data acquired in a gravel aquifer. Lately, Meles et al. (2011) demonstrated that the adjoint-based method of Ernst et al. (2007b) becomes less sensitive to the starting model when the inversion is conditioned to the long wavelengths of the signal in the initial part of the inversion and higher frequencies are gradually incorporated.

The adjoint-based approach is desirable because, in relatively few iterations, it is able to obtain a model of geophysical parameters that minimize a misfit (i.e., objective) function between the observed and modeled waveforms. However, this inversion strategy has some limitations. First, the method is based on linearization of the inverse problem and, therefore, it cannot be guaranteed that the global minimum is found (e.g., Tarantola, 2005). Second, the convergence criterion is chosen somehow subjectively, which may adversely result in noise propagating into the model estimate (e.g., Ernst et al., 2007b). Finally, the method is based on a linear assumption of the forward relation and is limited to Gaussian data uncertainty and a priori information, which results in a Gaussian approximation of the a posteriori uncertainty estimate (Tarantola, 1984; Pratt and Worthington, 1990). Here we propose an algorithm that naturally deals with these limitations.

In a Bayesian formulation, the solution to the inverse problem is given as an a posteriori probability density (Tarantola and Valette, 1982). The a posteriori probability is based on the independent states of information provided by the data (related to the model parameters through a physical law), an associated data uncertainty model (that takes into account data noise and data simulation inadequacies), and the a priori information on the model. The combined states of information contained in the a posteriori probability density are reflected in the model variability of the a posteriori sample. Hence, data uncertainties will not cause artifacts in the solution, but rather influence the degree of a posteriori model variability if the nature of the uncertainties is appropriately accounted for in the uncertainty model. Resolution analysis is naturally obtained from the a posteriori model variability. One may simply be interested in calculating the covariance of the a posteriori model parameters, but more sophisticated questions, such as the probability of geological connectivity or the residence time of a fluid, may also be answered by the a posteriori statistics (Mosegaard, 1998).

The extended Metropolis algorithm (Mosegaard and Tarantola, 1995) can be used to sample the a posteriori probability density, even for highly nonlinear inverse problems. This algorithm does not need a closed form mathematical expression of the a priori information, but a "black box" algorithm that is able to sample the a priori probability density is sufficient. Hansen, et al., (2008; 2012) suggested a method that provides a means of controlling the perturbation step size and efficiency when sampling a priori information defined through any geostatistical algorithms that is based on sequential simulation (e.g. Gomez-Hernandez and Journel, 1993). This method is referred to as sequential Gibbs sampling (Hansen et al., 2012). They demonstrate that this method could serve as a black box algorithm in the extended Metropolis algorithm. Sequential Gibbs sampling can be used to sample a priori models based on either relatively simple two-point based statistical models, such as Gaussian-based a priori models, or more complex multiple-point-based statistical models. This provides a means of using complex a priori statistical models that allow reproduction of geologically plausible structures, such as channels and tortuosity. Such complex patterns (i.e., spatial autocorrelation) can be learned from so-called training images and reproduced by simulation algorithms based on multiple-point statistics (e.g., Strebelle, 2002). Multiple-point algorithms offer the flexibility of simulating realizations with high entropy (e.g., Gaussian distributions) as well as low entropy (i.e., a few facies) structures or a combination of both (Journel and Zhang, 2006). See for example Remy et al. (2008) for different examples of such complex statistical models. In this way, the extended Metropolis algorithm becomes very flexible with regard to the choice of a priori model.

Since his seminal work on the full-waveform inverse problem, Albert Tarantola had the vision that realistic a priori information for inversion could be learned from a large collection of "training images" of the subsurface (Mosegaard, 2011). In this paper we demonstrate how this is made possible by using the extended Metropolis algorithm in conjunction with a priori information defined by a geostatistical algorithm using sequential Gibbs sampling. Initially, the theoretical background for this inversion strategy is outlined. Subsequently, the theory is applied to a tomographic crosshole GPR full-waveform inverse problem in which the a priori information is inferred (i.e., learned) from a training image and realized through the geostatistical algorithm Single Normal equation SIMulation (Snesim) (Strebelle, 2002). Full-waveform data traces are often contaminated by noise and are, in addition, subject to uncertainties related to inadequacies in the data simulation algorithm. These components of data uncertainty do often exhibit some degree of temporal correlation. In this study, we consider a Gaussian-distributed data uncertainty component with a temporal autocorrelation. We show how this uncertainty is accounted for in the inversion through the data uncertainty model.

Tarantola (2005) was a proponent of the movie strategy, in which "movies" of multiple realizations from the a priori and a posteriori probability densities are compared, to understand the additional information provided by the data compared with the a priori information. Features that occur frequently in the a posteriori movie are regarded as well resolved. We show that the movie strategy provides a means of obtaining resolution analysis of the full-waveform inverse problem. Moreover, we demonstrate how a posteriori realizations can be used to quantify the probability of lithological connectivity. The term resolution is, in this paper, used such that high a posteriori model variability refers to low resolution and vice versa (Mosegaard, 1998). To our knowledge this is the first example of sampling the a posteriori probability density of a tomographic full-waveform inverse problem. Resolution analysis of the problem reveals that the combined states of information from the full-waveform data and the a priori information provides a high-resolution subsurface image, even in the case of considerably sparse data

coverage. Finally, we demonstrate how the geostatistically formulated a priori information serves as a guide in the initial burn-in part of the inversion procedure. In this way, convergence of the suggested full-waveform inversion algorithm becomes independent of the initial model.

The present method is completely explicit with regard to the data uncertainty model and a priori information. Establishment of an adequate a priori information and data uncertainty model (comprising data noise and modeling inadequacies) demands effort from the user to obtain trustworthy a posteriori probability density. These issues are far from trivial when considering field experiments and are beyond the scope of this work.

METHODOLOGY

Consider that the subsurface can be represented by a discrete set of model parameters, $\mathbf{m}$ (referred to as the model) and that a data set, $\mathbf{d}$, of indirect observations of the model parameters is provided. The model parameters describe some physical properties of the subsurface that influence the data. Hence, the forward relation between the model and the data can be expressed as the relation (e.g., Tarantola, 2005)

$$\mathbf{d} = g(\mathbf{m}), \qquad (1)$$

in which g is a linear or nonlinear mapping operator that often relies on a physical law. In this particular study the model represents a tomographic image of dielectric permittivity of the subsurface material, and data are waveforms of the vertical electrical component of electromagnetic waves propagated across the model between two boreholes. The forward relation in equation 1 is given as a finite-difference time-domain (FDTD) solution of Maxwell's equations (Ernst et al., 2007b). However, any numerical wave propagation modeling strategy (for GPR or seismic signals) can be applied. The inverse problem is to infer information about the model parameters on the basis of a set of data (and their uncertainties), a priori information about the model, and the forward relation between the model and the data.

In a Bayesian formulation the solution to the inverse problem is given as an a posteriori probability density, which can be formulated as (e.g., Tarantola, 2005)

$$\sigma_M(\mathbf{m}) = c\rho_M(\mathbf{m})L(\mathbf{m}), \qquad (2)$$

where c is a normalization constant, $\rho_M(\mathbf{m})$ is the a priori probability density, and $L(\mathbf{m})$ is the likelihood function. $\rho_M(\mathbf{m})$ describes the probability that the model satisfies the a priori information. $L(\mathbf{m})$ describes how well the modeled (i.e., simulated) data explain the observed data, given a statistical description of the data noise and modeling inadequacies (from here on referred to as the data uncertainty model). Hence, the a posteriori probability density describes the resulting state of information on the model parameters provided by the independent states of information given by the data (related to the model through the forward relation) and an a priori state of information on the model parameters. Hence, an adequate specification of the data uncertainty model (through the likelihood function) as well as the a priori information are crucial in order to ensure a correct a posteriori state of information (i.e., solution of the inverse problem).

Sampling the a posteriori probability density

A highly nonlinear inverse problem refers to the case in which the a priori probability density is far from being Gaussian or the likelihood function is highly non-Gaussian (typically) due to the nonlinear forward relation between model and data. According to equation 2 the product of these non-Gaussian probability densities signifies a highly non-Gaussian a posteriori probability density. In the case of full-waveform inversion, the forward relation is expected to be highly nonlinear. Moreover, realistic a priori information, as we introduce here, is typically far from being Gaussian.

The extended Metropolis algorithm is a versatile tool which, in particular, is useful to sample the a posteriori probability density of nonlinear inverse problems using arbitrarily complex a priori information. This algorithm is convenient in that it does not need an explicit expression of the a priori probability density. A black box algorithm that is able to perform a random walk in the a priori probability density is sufficient (Mosegaard and Tarantola, 1995). In this study we use sequential Gibbs sampling, which will be described below, as the black box algorithm for sampling the a priori probability density.

The extended Metropolis Algorithm consists of two randomized steps:

1) Exploration: one proposes a candidate model, $\mathbf{m}_{\text{propose}}$, which is a perturbation of a current model, $\mathbf{m}_{\text{current}}$, and at the same time is a realization of the a priori probability density.
2) Exploitation: one decides if the proposed model should be accepted or rejected. The proposed model is accepted with the Metropolis acceptance probability (referred to as the Metropolis rule)

$$P_{\text{accept}} = \min\left(1, \frac{L(\mathbf{m}_{\text{propose}})}{L(\mathbf{m}_{\text{current}})}\right), \qquad (3)$$

where $L(\mathbf{m}_{\text{propose}})/L(\mathbf{m}_{\text{current}})$ is the ratio between the likelihood evaluated in the proposed and the current model, respectively. If accepted, the proposed model becomes the current model and is a realization of the a posteriori probability density. Otherwise the proposed model is rejected and the current model counts again. Thus, in each iteration, the sample size of the a posteriori probability density increases.

The exploration step constitutes the strategy by which proposed models are drawn from the a priori probability density. For small exploration steps the proposed model will be relatively highly correlated with the current model, compared with large exploration steps. Thus, according to equation 3, small exploration steps will result in a high acceptance probability and vice versa. Recall that each evaluation of the likelihood function involves a computationally expensive (FDTD) forward calculation. It is, therefore, important to choose an appropriate exploration step size that does not explore the a posteriori probability density too slowly, but on the other hand does not waste too many expensive evaluations of the likelihood that are very unlikely to be accepted. In particular, the exploration strategy becomes very important when dealing with high-dimensional probability densities, since traditional sampling strategies may lead to a very inefficient exploration strategy (Hansen et al., 2008; 2012). Recently, Hansen, et al., (2008; 2012) introduced a flexible sampling strategy to sample high-dimensional a priori probability densities

defined by geostatistical algorithms. This sampling strategy provides a means of using a variable exploration step size when sampling such high-dimensional probability densities. This method is briefly outlined in the following.

Sampling geostatistically defined a priori information

Hansen et al. (2012) suggest a strategy referred to as sequential Gibbs sampling, which is a method that is capable of sampling probability densities defined by geostatistical algorithms based on sequential simulation. Originally, sequential simulation was used to generate realizations of two-point based statistical models, such as Gaussian models (e.g., Gomez-Hernandez and Journel, 1993). Two-point statistical models are limited to specify the spatial variability between pairs of data (hence the name two-point statistics) defined by a covariance model. Two-point statistics does not provide enough information to model plausible geological structures, such as channels and tortuosity. This can be overcome by using a model based on multiple-point statistics, as suggested by Guardiano and Srivastava (1993). In multiple-point-based statistical models, the a priori information is learned from a training image. The training image is scanned by a template that jointly considers spatial variability among a number of (more than two) pixel values in the image to obtain a joint probability distribution that holds information about these spatial correlations. The model parameter values are subsequently sequentially simulated from conditional probabilities on the basis of the jointly considered pixel values (e.g., Strebelle, 2002). The algorithm originally suggested by Guardiano and Srivastava (1993) was, however, computationally unfeasible. It originally was not until an efficient way of storing multiple-point statistics in machine memory was proposed that the use of multiple-point based models became computationally feasible (Strebelle, 2002). See e.g., Remy et al. (2008) for numerous examples of the application of sequential simulation from both two-point and multiple-point-based a priori models.

In this study sequential Gibbs sampling serves as the black box algorithm that samples the a priori probability density $\rho_M(\mathbf{m})$, described by a geostatistical simulation algorithm, with a controllable exploration step size. The flow of the sequential Gibbs sampling is as follows:

1) An initial unconditional realization of the a priori probability density, $\mathbf{m}_{\text{current}}$, defined in a 2D regular grid of model parameters is provided.
2) A square subarea, corresponding to model parameters $\mathbf{m}_{\text{subarea}}$, with side-length E_{step} (exploration step size) of the current model $\mathbf{m}_{\text{current}}$ is randomly chosen.
3) A realization of the conditional probability density $\rho_M(\mathbf{m}_{\text{subarea}}|\tilde{\mathbf{m}}_{\text{current}})$ is obtained using sequential simulation. $\tilde{\mathbf{m}}_{\text{current}}$ are the current model parameters outside the subarea. $\rho_M(\mathbf{m})$ is an a priori probability density that may be described by either two- or multiple-point statistics. This step is an application of the Gibbs sampler where sequential simulation is used to efficiently generate a realization from the conditional probability density (hence the name sequential Gibbs sampling). In practice this is performed simply by running the sequential simulation algorithm conditional to the model parameters outside the subarea. In this way a new, perturbed model, $\mathbf{m}_{\text{propose}}$ is obtained.
4) The proposed model becomes the current model and steps 2 and 3 are repeated to obtain multiple realizations of the a priori probability density.

Recall that, when applying sequential Gibbs sampling as an a priori model sampler in the extended Metropolis algorithm, the current model $\mathbf{m}_{\text{current}}$ is reused if the proposed model $\mathbf{m}_{\text{propose}}$ is rejected by the Metropolis rule. Moreover, the resemblance (i.e., correlation) between the current and proposed model, and thus the average Metropolis acceptance probability (cf. the Metropolis rule), can be controlled by the explorations step size E_{step} (i.e., the side length of the resimulated area).

The likelihood function and correlated data uncertainties

GPR or seismic full-waveform data are often contaminated with temporally correlated uncertainties along the individual waveform traces or spatially correlated (i.e., static) errors among data related to certain transmitter (source) or receiver positions. The probabilistic formulation of the inverse problem allows for an arbitrary data uncertainty model and it is, therefore, possible to account for these correlations in the data uncertainties. In the present study we consider that data uncertainties are Gaussian distributed with a temporal correlation along the individual waveform traces, but are uncorrelated among the individual traces. This type of uncertainty influences the state of information on the model parameters provided by the data and is, therefore, accounted for through the likelihood function. This particular likelihood function takes on the following form

$$L(\mathbf{m}) = c \prod_{k=1}^{K} \exp\left[-\frac{1}{2} (g(\mathbf{m})^k - \mathbf{d}_{\text{obs}}^k)^{\mathrm{T}} \mathbf{C}_{\mathrm{D}}^{-1} (g(\mathbf{m})^k - \mathbf{d}_{\text{obs}}^k) \right],$$
(4)

where $g(\mathbf{m})^k$ and $\mathbf{d}_{\text{obs}}^k$ are vectors that contain the simulated and observed waveform traces related to the kth transmitter-receiver pair. K is the total number of waveform traces (i.e., transmitter-receiver pairs). The factor c is a normalization constant. The term $\mathbf{C}_{\mathrm{D}}$ is the data covariance matrix that defines the variances and covariances of the data uncertainties. The temporal correlation of the data uncertainties is described by an exponential correlation function (e.g., Goovaerts, 1997)

$$\mathbf{C}_D(i, j) = c \, \exp\left(\frac{-3 \, s(i,j)}{a} \right),$$
(5)

where c is the sill (i.e., variance) and a is the range (i.e., correlation length) of the uncertainties. The term $s(i,j)$ is the temporal distance between the ith and jth sample point along the waveforms. $\mathbf{C}_D$ is a symmetric $N \times N$ matrix, where N is the number of samples in the individual waveform traces.

Cordua et al. (2009) quantified the influence of static-like errors in crosshole GPR experiments, which are data uncertainties that are spatially correlated among data related to the individual transmitter and receiver positions. This kind of data uncertainty may also be accounted for through the data covariance matrix, but this is not considered here. For a description of how to set up a data covariance matrix that accounts for static (i.e., spatially correlated) errors, see Cordua et al. (2008).

The burn-in period

If the probability that a sampling algorithm at any time enters an infinitesimal neighborhood N_j, that surrounds the model $\boldsymbol{m}_j$, is

equal to the probability that it leaves this neighborhood, the algorithm is said to satisfy microscopic reversibility. It can be shown that if a sampling algorithm satisfies microscopic reversibility, then the a posteriori probability density is the only equilibrium distribution for the algorithm, and the algorithm will converge toward this equilibrium distribution independently of the starting model. This property is satisfied by the extended Metropolis algorithm (see Mosegaard and Sambridge [2002] for a detailed description of the Metropolis algorithm). In practice, the extended Metropolis algorithm has reached the equilibrium distribution when the likelihood values start to fluctuate around a constant level (referred to as the equilibrium level) and the data are fitted within their uncertainties. When this phase is reached the algorithm is said to be burned-in. In case of a nonlinear full-waveform inverse problem the structure of the a posteriori probability density is completely unknown. If this distribution is multimodal, several modes may exist that fit data with the uncertainty. Hence, no guarantee can be given, no matter the choice of sample strategy, that the a posteriori probability density is appropriately represented (i.e., that all modes are represented) by a finite sample size.

The a posteriori probability density is defined over a high-dimensional space and leaves only small areas of significant probability. Therefore, the burn-in period may be long because the algorithm during this period randomly walks across large parts of this high-dimensional space searching for a small area of significant probability. As a consequence, we suggest performing large exploration steps in the initial part of the burn-in period and then gradually decreasing the exploration step size as the algorithm approaches the equilibrium level. In this way, the large-scale structures of the model are relatively quickly established in the initial part of the burn-in period, after which the algorithm gradually establishes the smaller-scale structures. However, microscopic reversibility cannot be guaranteed when the exploration step size is not kept constant while running the algorithm. Therefore, the algorithm is stopped when "apparent" burn-in has been reached and the exploration step size is subsequently set to a constant value.

The algorithm is started in an unconditional realization of the a priori model, which is expected to be far away from volumes of significant a posteriori probability density. We suggest adaptively adjusting the exploration step size during the burn-in period such that the Metropolis acceptance probability (equation 3) controls the exploration step size. This is controlled by updating the exploration step size after every M successive exploration steps of the Metropolis algorithm. The exploration step size update is performed just after the exploitation step (step 1) of the algorithm. Consider that the algorithm is at iteration number K, where K is a multiple of M. Then the exploration step size E_{step}^{i+1} during iteration number $K + 1$ to $K + M + 1$ (i.e., the next M iterations) is given as

$$E_{step}^{i+1} = E_{step}^i \frac{P_{average}}{P_{control}}, \qquad (6)$$

where E_{step}^i is the exploration step size during iteration number K-M to K (i.e., the M preceding iterations). $P_{average}$ is the average Metropolis acceptance probability (equation 3) during iteration number K-M to K and $P_{control}$ is a subjectively chosen constant acceptance ratio that the algorithm tries to adhere to by adjusting the exploration step size according to equation 6. For larger values of $P_{control}$, the exploration step size decreases faster and converges to a lower level, than it does for smaller values. The minimum possible

exploration step size involves resimulation of one model parameter, whereas the maximum step size is an (unconditional) simulation of all the model parameters that are statistically independent of the previous model.

A SYNTHETIC CROSSHOLE GPR FULL-WAVEFORM INVERSE PROBLEM

The methodology outlined above is tested on a synthetic tomographic crosshole GPR full-waveform data set. Wavefield simulations of the GPR signals (i.e., the forward relation) are obtained using FDTD calculations of Maxwell's equations in transverse electric mode (Ernst et al., 2007b). This FDTD method provides grid-based time-domain calculations of the electromagnetic wavefield propagation. The transmitting and receiving antennae are simulated as vertically orientated dipole-type antennae and are aligned parallel with the vertical boreholes. Transmitted and received signals concern the vertical component of the electrical field (Holliger and Bergmann, 2002). The FDTD algorithm by Ernst et al. (2007b) yields second-order accuracy in both time and space, and performs the calculations in 2D Cartesian coordinates. The edges of the FDTD grid are surrounded by a generalized perfectly matched layer (GPML) to absorb artificial boundary reflections.

The a priori information of the inverse problem is provided by the Snesim algorithm. Snesim is a fast geostatistical simulation algorithm that produces realizations (conditional or unconditional to point data) from a high-dimensional probability density that contains the spatial relations (i.e., patterns) learned from a training image for a relatively low number of categorical values (Strebelle, 2002). Figure 1 shows a training image that mimics a matrix of unconsolidated sand with embedded channels of gravel situated in an unsaturated environment. The geological information contained in the training image may have been obtained from outcrops in a nearby gravel pit and/or a natural cliff. The training image applied here does not necessarily represent a realistic near-surface environment, but is rather used to demonstrate the principle that geologically realistic features, such as complex channel structures, can be
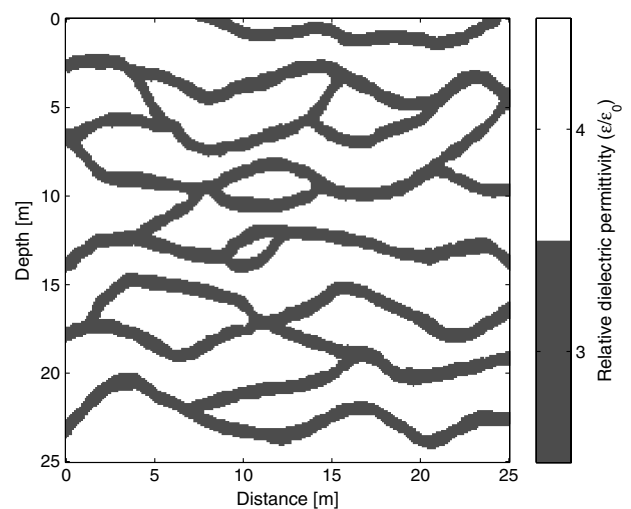

Figure 1. Training image containing geological information of the environment at which the crosshole GPR experiment is conducted. This information is used as a priori information in the waveform inversion.

represented by a multiple-point-based a priori model in the inversion procedure. Electromagnetic signals in the near-surface sediments are sensitive to the dielectric permittivity, the electrical conductivity, and the magnetic permeability of the materials. In this study we limit ourselves to considering only the influence of the dielectric permittivity, which is primarily governing the phase velocity of the signal. The relative dielectric permittivity of the sand and the gravel channel is set to $\varepsilon_r = 4.0$ (0.150 m/ns) and $\varepsilon_r = 3.0$ (0.173 m/ns), respectively (e.g., Ernst et al., 2006). The relative dielectric permittivity is given as $\varepsilon_r = \varepsilon/\varepsilon_0$, where ε is the absolute dielectric permittivity and ε_0 is the dielectric permittivity of free space.

Figure 2 shows the synthetic reference model to be considered and is, at the same time, an unconditional realization of the training image obtained using Snesim. The electrical conductivity is set to a constant value of 3 mS/m and in the following it is assumed known. Near-surface materials are considered nonmagnetic and the magnetic permeability is set to the magnetic permeability of free space (e.g., Davis and Annan, 1989).

A full-waveform synthetic data set is calculated using the FDTD algorithm. A Ricker wavelet with a central frequency of 100 MHz is used as the source pulse. The source pulse is assumed known during the inversion. The recorded synthetic GPR full-waveform data are the vertical component of the electrical field. The experiment is conditioned by data from four transmitters (two in each borehole) at depths of 3 m and 9 m. The receivers are equidistantly distributed in the two boreholes with a separation distance of 1.5 m (see Figure 2). Data acquired with a transmitter-receiver angle larger than 45° from horizontal are omitted since, in practice, these data are violated by effects of wave guiding in the boreholes (e.g., Peterson, 2001) and travel paths between the antenna tips instead of the center of the antennae (Irving and Knight, 2005). These effects are, among several other sources of uncertainty, a result of inadequate forward modeling which has to be seriously considered in field experiments. Such effects either have to be handled through a refined forward modeling approach, or accounted for in the likelihood function through a statistical description of the data uncertainties imposed by both data noise and modeling inadequacies. See
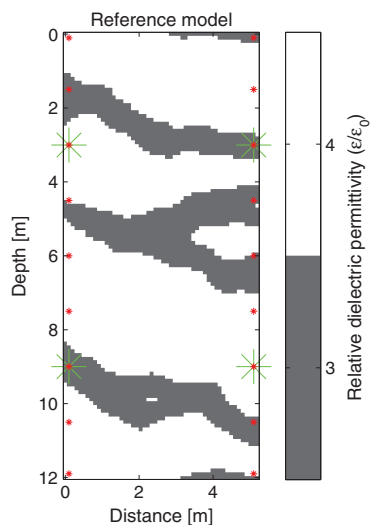
the discussion for a further treatment of these issues. The data geometry leads to a total of 20 recorded waveform traces. The resulting transmitter-receiver positions are connected with dotted lines in Figure 3 on top of the grid of the 6240 unknown model parameters.

Gaussian-distributed data uncertainties with a temporal autocorrelation described by the exponential correlation function in equation 5 are added to the waveform data. The temporal correlation length a (i.e., the range) is set to 12.7 ns. Figure 4 shows the five waveform traces related to the uppermost transmitter position in the left borehole (see Figure 2), which are related to the transmitter-receiver pairs marked by the numbers 1 to 5 in Figure 3. The noise-free simulated waveforms are plotted as dotted blue curves and the uncertain waveforms (noisy waveforms) are plotted as red curves. The 20 uncertain waveform traces are used as observed data in this study and have an average signal-to-noise ratio of 13.6. The uncertainty imposed on the "noise-free" data mimic the total contribution of data noise and modeling inadequacies. In the next section, full-waveform inversion will be performed on the uncertain waveform data with a priori information based on a geostatistical model inferred from the training image in Figure 1.

RESULTS

Burn-in

In the present example the algorithm is started in a realization of the multiple-point a priori model learned from the training image, unconditional to any information from the data. In this way the starting model is independent of data and relies only on the a priori information. The initial exploration step size has a side length of $E_{step} = 12$ m, which corresponds to the maximum dimension of the model size. Hence, the exploration step size cannot exceed this side length and at this point the algorithm produces statistically
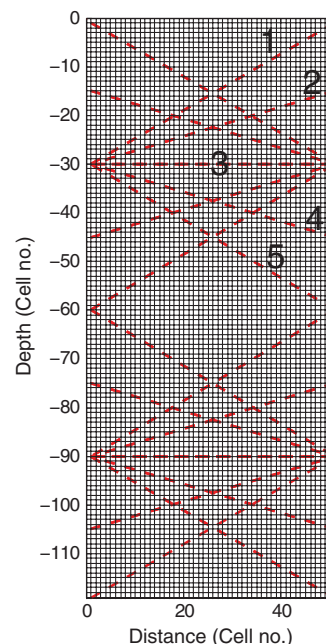

Figure 3. The transmitter-receiver positions connected by dotted lines on top of the 2D grid of $(120 \times 52 = 6240)$ unknown model parameters. The lines marked by the numbers 1 to 5 are related to the waveform data shown in Figure 4.

Figure 2. Synthetic reference model. Green asterisks show transmitter positions and the red dots show receiver positions.

independent realizations. The control acceptance, $P_{control}$, is set to 10% to obtain a relatively large exploration step size during the burn-in period. The exploration step size is evaluated after each 20th iteration (i.e., $M = 20$) according to equation 6. The evolution of the adaptive exploration step size during the first 1000 iterations is seen in Figure 5. It is observed that the exploration step size is constant and high in the very first part, after which it gradually decreases and stabilizes at a constant level of approximately $E_{step} = 2$ m. The associated development of the model is demonstrated in Figure 6. It is seen that the large-scale structures are very quickly brought into place, whereupon only fine-scale features of the model are accepted by the Metropolis rule (equation 3). After approximately 17,000 iterations the likelihood values start to fluctuate around an equilibrium level and the data residuals resemble a normal distribution with approximately the same standard deviation as the distribution of the noise. At this stage it is assumed that the algorithm has reached burn-in and produces representative realizations of the a posteriori probability density.

A posteriori statistics

The last model accepted by the Metropolis rule in the burn-in period is used as the starting model when the Metropolis algorithm is subsequently restarted with a constant exploration step size of 1 m. The results obtained in this study demonstrate that the equilibrium level does not change after the Metropolis algorithm is restarted with a constant exploration step size. In addition the data residuals of this period resemble a normal distribution with a standard deviation of $2.17 \cdot 10^{-4}$. As a comparison the standard deviation of the normal distributed noise is $2.24 \cdot 10^{-4}$, which demonstrates that the data are fitted within the data uncertainties.

Hence, this shows that the algorithm has reached burn-in and that the adaptive exploration step size may be appropriate when the step size converges toward a constant level (which is the case in our study). This burn-in strategy may serve as an approximate way of determining which exploration step size should be used to obtain a certain average acceptance probability. See Gelman et al. (1996) for an investigation on the choice of acceptance probability.

After 300,000 iterations the algorithm was stopped. During the sample period the algorithm had an average acceptance probability of 40%. Figure 7 shows the autocorrelation between the first model after burn-in and its correlation to the next150,000 models obtained from the a posteriori probability density. These models are not statistically independent because any proposed model in the Metropolis algorithm is a perturbation of a current model or when a proposed model is rejected the current model counts again. Therefore, the autocorrelation analysis of the a posteriori sample shows some correlation length between successive models. In the example shown in Figure 7, statistical independence is obtained after approximately 5800 iterations. This point is approximated as the point at which the autocorrelation curve intercepts the average level of the correlation curve after it has converged to a constant level. The average is shown as a dotted line in Figure 7 and is calculated as the average correlation coefficient between iteration 20,000 and 300,000. A similar analysis is performed on 10 models picked at different iteration numbers, equally distributed across the a posteriori sample, which gives an average of 6745 iterations of separation to obtain statistically independent realizations from the a posteriori sample.

In a probabilistic formulation, the solution to the inverse problem is not a single model estimate, but a sample of multiple model realizations drawn from the a posteriori probability density. Each realization is a tomographic image of the subsurface. Displaying multiple images together corresponds to a movie. The strategy of displaying and studying the solution to the inverse problem using such movies is referred to as the movie strategy (Tarantola, 2005). Roughly speaking, the a priori probability density is filtered by the likelihood function that results in the a posteriori probability density. Hence, displaying a "movie" of a priori realizations together with a movie of a posteriori realizations helps one to understand the
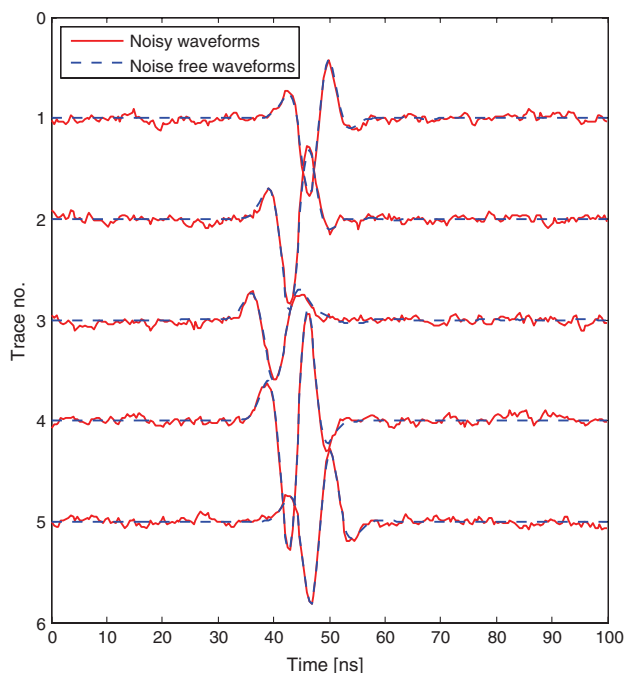

Figure 4. Five waveform traces with and without added noise. The signal-to-noise ratios of the plotted waveform traces are (from the top) 15.4, 20.8, 8.8, 41.3, and 22.2, respectively. The waveforms are related to the transmitter-receiver pairs marked by the numbers 1–5 in Figure 3.
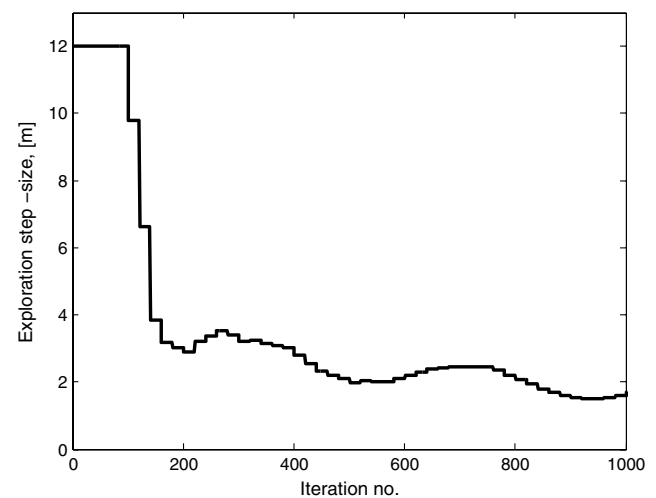

Figure 5. Evolution of the adaptive exploration step size during the first 1000 iterations of the burn-in period.

characteristics of the a priori information imposed on the inverse problem and to understand the state of information provided by the data (and their uncertainties). In particular, when the a priori information is provided through a black box algorithm, and no closed-form mathematical expression of the a priori probability density exists, an a priori movie may be important to understanding whether the state of information provided by the a priori probability density is commensurate with the a priori expectation of the user. See Koren et al. (1991) for a seminal example of using the movie strategy for a seismic inverse problem.

Figure 8 shows eight statistically independent realizations drawn from the a priori probability density (i.e., an a priori movie). This movie shows the state of information provided by the a priori probability density inferred from the training image by the Snesim algorithm. This movie shows a reproduction of the channel structures inferred from the training image, but with no resemblance between the location of the channels in individual realizations, as these models are unconditioned by any data. Figure 9 shows eight statistically independent realizations from the a posteriori probability density. The result clearly demonstrates a high degree of resemblance between the individual a posteriori realizations as a result of the conditioning to the full-waveform data. The high resemblance reveals that the data provides a high resolution of the inverse problem,

despite sparse data coverage (see Figure 3). Moreover, the individual realizations only slightly deviate from the reference model (see Figure 2), which confirms the good resolution provided by the full-waveform data, their uncertainties, and the training image. The a posteriori realizations in Figure 9 show some isolated small-scale features, which are not seen in the training image. These noncontinuous effects are caused by the Snesim simulation technique because the algorithm occasionally reduces the number of conditioning data events (i.e., pixel values) to avoid singularities during calculation of the conditional probabilities. A total of 45 statistically independent realizations are obtained from the 300,000 a posteriori realizations based on the model autocorrelation analysis (see Figure 7). These realizations can be used to ask higher-order statistical questions such as what the probability of connectivity is between a channel observed in the left borehole and a channel observed in the right borehole. Figure 9d, 9f, and 9h shows some examples of missing connectivity between the boreholes marked by red circles. For example it is found that in five out of the 45 realizations there is no connection between the points A and B as marked in Figure 9a. Hence, this gives an approximate a posteriori probability of connectivity between points A and B of $(45 - 5)/45 = 89\%$.

Figure 10 shows the mean and variance calculated from the a posteriori sample. It should be noted that the mean model is no longer a realization of the a posteriori probability density, but is only a statistical representation of the sample. The mean tells, in this particular case, the relative a posteriori probability of the presence of a channel at a certain position in the subsurface. The variance reveals that the uncertainty of the spatial location of the channels increases toward the edges of the channels and declines significantly when moving away from the edges.

Waveform data associated with the 45 statistically independent realizations from the a posteriori probability density are calculated for the receiver related to the uppermost transmitter position in the left borehole (see transmitter-receiver pairs marked by the numbers 1–5 in Figure 3). This data variability associated with the model a posteriori variability is plotted in Figure 11a (blue curves) together with the observed data (red curves). The (a posteriori) simulated waveforms show a high degree of similarity and appear in the plot almost as a single fat curve, but are in fact composed of 45 independent waveform curves. Hence, the a posteriori data variability demonstrates that the model a posteriori variability is only associated with very little variability in the waveforms. Moreover, Figure 11a shows that the simulated waveforms fit the observed data very well. Figure 11b shows the data residuals (blue curves) of the simulated waveform data together with the uncertainty component of the observed data (red curves). This plot reveals that the residuals (i.e., misfits) approximately fluctuate around the uncertainty component and resemble the noise statistics (variance and temporal autocorrelation) satisfactorily.
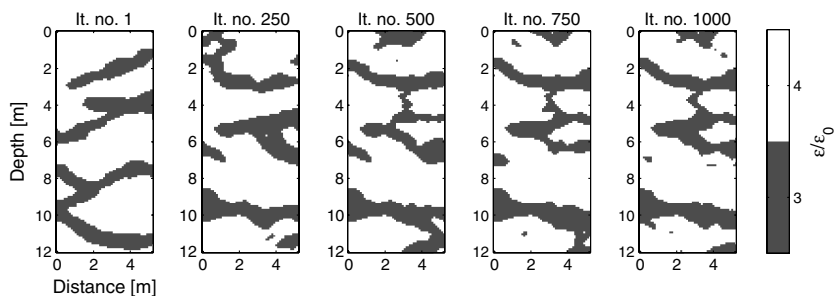

Figure 6. Development of the model during the first 1000 iterations. In this period, the relatively large-scale structures of the model are established. Compare with the reference model in Figure 2. Iteration no. 1 (It. no. 1) is the starting model, which is an unconditional realization of the multiple-point-based a priori model.
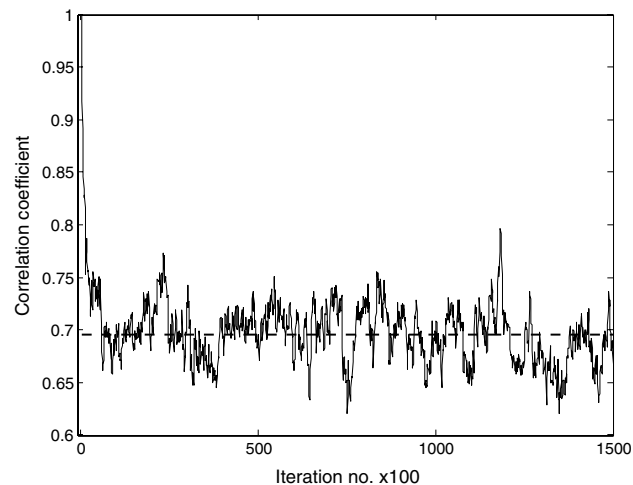

Figure 7. Autocorrelation analysis of the a posteriori sample to determine the number of iterations needed to obtain statistically independent realizations.

DISCUSSION

Hansen et al. (2006) demonstrated that linear inverse Gaussian theory and simple kriging (i.e., two-point statistics) can be merged

into one single theory. The method of sequential Gibbs sampling used with the extended Metropolis algorithm has provided a further step toward bridging probabilistic inverse problem theory and the field of geostatistics, even for highly nonlinear and non-Gaussian inverse problems (Hansen et al., 2008; 2012). Hansen et al., (2009; 2012) discussed how the complexity of inverse problems (i.e., the time needed to obtain an a posteriori sample) is reduced when a statistical a priori model with some degree of spatial correlation between the model parameters is considered. They showed that the effective dimension of the solution space of the inverse problem, using two-point-based a priori models, is considerably decreased when the a priori expected range of spatial correlation is increased. On the other hand, when no spatial correlation was considered, the high-dimensional inverse problem became unsolvable. Further, it was seen that whether the model parameters take values from a set of real or binary numbers had insignificant influence on the effective dimension compared with the degree of correlation between the model parameters. Hence, we consider the reduction of the effective dimension due to a priori defined spatial autocorrelation to be one of the keystones that makes Monte Carlo inversion of a computationally hard, full-waveform inverse problem feasible, even with thousands of (here 6240) model parameters.

In this study we wish to demonstrate that we are able to freely choose multiple-point-based a priori information, defined by, e.g., the Snesim algorithm, for Monte Carlo-based inversion. We considered a model with categorical parameters only, but this choice of using a multiple-point-based a priori model was a first step away from using simplified a priori models like the Gaussian model. We believe that such models are often chosen out of mathematical convenience rather than for the sake of geophysically and geologically based a priori expectations.

On the basis of the studies of Hansen et al. (2009), we believe that the solution space of the full-waveform inverse problem (i.e., the complexity of the inverse problem), considered in this study, is primarily reduced due to the multiple-point-based spatial correlations, and not significantly due to the binary model parameters. Accordingly, full-waveform inversion may also be tractable when making

use of two-point-based statistical a priori models with both continuous and categorical model parameter values, as long as some degree of spatial correlation can be considered a priori. This is also confirmed by our preliminary studies on this topic. The above discussion is encouraging with regard to the possibility of using a priori models based on either two-point statistics, multiple-point statistics, or a combination of both, as long as the chosen a priori model imposes some degree of spatial autocorrelation of the model parameters (see Journel and Zhang [2007] and further discussions at the end of this section). This provides a flexible tool for defining an appropriate a priori model that actually captures our a priori expectations.

The suggested inversion strategy is very general and may be equally applicable for tomographic inversion of any kind of data (e.g., GPR, seismic, x-ray, or electroencephalography data) or for reflection seismic inversion. In the example presented here the algorithm only inverts for the dielectric permittivity, whereas the electrical conductivity is kept fixed. The method could, however, be extended to invert for both parameters by introducing a step, just before the exploration step, that randomly chooses in which of the two fields the exploration should be performed.

A pseudo-full-waveform inversion approach for tomographic GPR data was proposed by Gloaguen et al. (2007). In their approach, multiple model realizations were obtained using a stochastic ray-based inversion strategy. Full-waveform simulations were subsequently calculated in these multiple models. Models related to waveform data that showed the best fit to the observed data were regarded as estimates of the waveform inversion. However, their approach does not guarantee a data fit within the uncertainties and the accepted models are not realizations from an a posteriori probability density function.

In this study the sequential Gibbs sampler is applied such that a continuous block of model parameters are resimulated in each step. Irving and Singha (2010) also used a type of sequential Gibbs sampling, but resimulated a subset of model parameters scattered randomly across the model. Note that sequential Gibbs sampling will correctly sample the a priori probability density function regardless
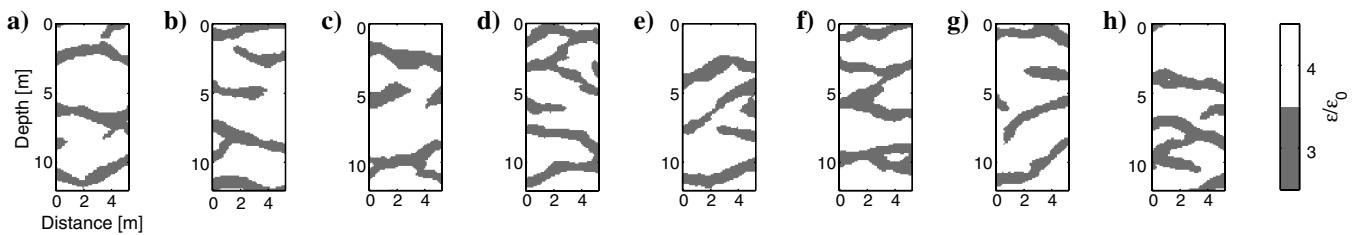

Figure 8. Eight statistically independent realizations of the a priori probability density (a priori movie).
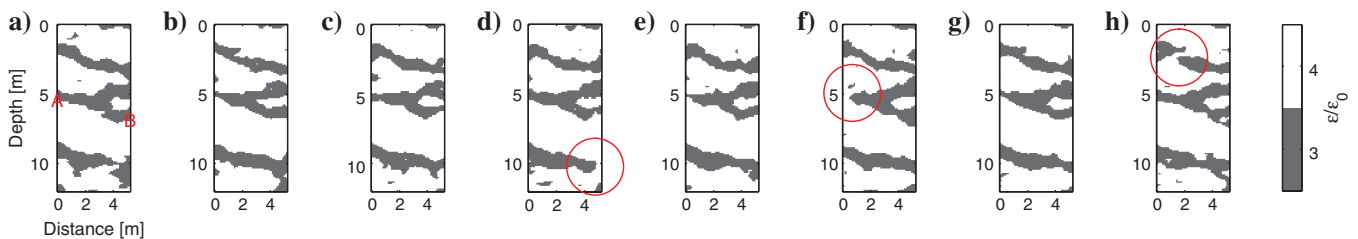

Figure 9. Eight statistically independent realizations of the a posteriori probability density (a posteriori movie).

of how that subset of model parameters to be resimulated is chosen (Hansen et al., 2012). Whether the continuous or the scattered exploration strategy should be used is still to be investigated, as such a choice may affect the computational efficiency when the sequential Gibbs sampler is used with the Metropolis algorithm. To understand
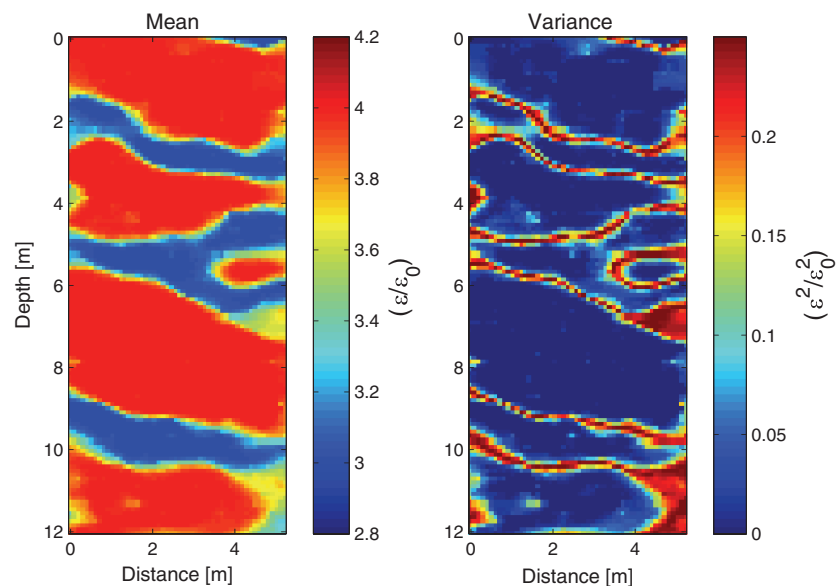

Figure 10. The mean and variance of the model parameters calculated on the basis of the a posteriori sample.
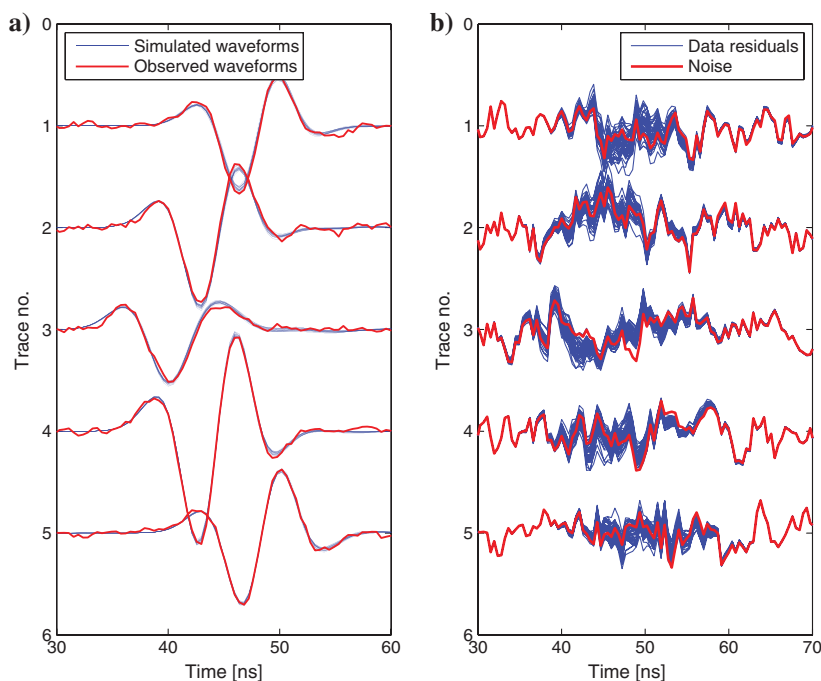

Figure 11. (a) A posteriori data variability. Blue curves: Simulated waveforms calculated for 45 statistically independent a posteriori models. Red curves: The observed waveform data. (b) Data residuals. Blue curves: The data residuals for 45 statistically independent a posteriori models. Red curves: The uncertainty component of the observed data. The waveforms are associated with the transmitter-receiver pairs indicated by the numbers 1–5 in Figure 3.

exploration strategies of high-dimensional spaces, more analytical approaches are required.

Ray-based inversion experiments on the same scale as the setup used in this study typically involve in the order of 700 to 1600 transmitter-receiver pairs to obtain a reasonable resolution (e.g., Tronicke and Holliger, 2005; Looms et al., 2008; Nielsen et al., 2010). In this study a high degree of resolution was obtained with as few as 20 transmitter–receiver pairs. However, in field experiments modeling inadequacies may lead to considerably more uncertainty in the data than considered in this study, which in turn leads to a lower resolution. In crosshole GPR full-waveform inversion the long spatial wavelengths of the model are typically used as a starting model. This model is obtained through ray-based inversion of first-arrival traveltimes and amplitudes of the waveform data (Ernst et al., 2007a; Meles et al., 2010). The data set considered in our study is expected to be too sparse to provide any useful information for a ray-based starting model. We, therefore, choose to start the full-waveform inversion in an unconditional realization of the a priori probability density, and burn-in is obtained anyway. Hence, the role of the a priori model is more than simply finding a posteriori model realizations that jointly honor data and the a priori model. It turns out that the use of a consistent a priori model acts as a guide in the burn-in process that allows the initial model to be far away from the true solution. The successful convergence from the data-independent starting model observed in the present study may be explained as a reduction of the complexity of the problem through the informative a priori information defined through the geostatistical algorithm (see Hansen et al., 2009; 2012). This encouraging observation suggests that future effort should be toward incorporating complex statistical a priori information into (adjoint) optimization based inversion approaches.

The most commonly used method for full-waveform inversion today is based on adjoint methods, as suggested by Tarantola (1984). This approach has some limitations: (1) Uncertainty estimates may theoretically be obtained through an a posteriori covariance, but only for a linear approximation of the forward relation limited to a Gaussian description of the data uncertainty and a priori model (Tarantola, 1984). (2) The method is based on simple Gaussian a priori information (if any at all); and (3) it relies on a subjective convergence criterion that may adversely result in data uncertainties propagating into the model estimate. The method we propose overcomes many of the limitations of using the adjoint-based approach. (1) It allows for arbitrary data geometry and density. (2) Complex a priori inversion can be included using any geostatistical

algorithm that can be used for sequential simulation or any other probabilistically defined a priori information (see Mosegaard and Tarantola, 1995). (3) The full uncertainty of the inverse problem can be quantified by analyzing the a posteriori sample. (4) There are no complications related to the incorporation of refined and complex forward algorithms. (5) Temporally and spatially correlated data uncertainties can be accounted for either through the data covariance matrix or any other mathematical description of the uncertainties (Cordua et al., 2008, 2009). (6) Finally, if the correct data uncertainty model is known there is no risk of uncertainties propagation into a single model estimate, but data uncertainties are instead reflected in the a posteriori model variability.

The adjoint-based approach has some benefits that the presented method lacks: In the case of dense data coverage, full-waveform inversion based on the adjoint-based approach may not need any a priori model at all, as the data themselves are sufficient to allow inference of a solution (Tarantola, 1984). The adjoint-based method has been successfully tested on field data (Ernst et al., 2007a; Klotzsche et al., 2010). In addition the suggested probabilistic inversion strategy needs substantially more (computationally expensive) forward calculations than does the traditional adjoint-based approach.

In the synthetic test performed here, each iteration involves (1) a model perturbation using Snesim which takes approximately 0.5 s, and (2) a forward calculation of the four transmitter positions, which is run in parallel (the number of parallel computations equals the number of transmitter positions). Each forward calculation takes approximately 12.5 s on a standard desktop computer with an Intel Core i7 processor. Thus, the total computation time of 300,000 iterations becomes approximately 45 days. As a comparison, the adjoint-based method needs approximately 20 iterations to converge to a solution estimate (Ernst et al., 2007a), which involves 20 FDTD forward calculations (for comparison with observed data), 20 backward (in time) calculations (for update direction), and 20 forward calculations (for step length) (Ernst et al., 2007b). Hence, obtaining one estimate using the adjoint-based method is approximately 5000 times faster than the time needed to obtain 45 independent realizations from the a posteriori probability density using the Monte Carlo method. However, the adjoint-based method and the Monte Carlo strategy are not directly comparable. The adjoint method is an optimization method that searches for one (in some sense) optimal solution. Additionally, the adjoint method assumes the full-waveform inverse problem to be a global optimization problem and, therefore, the method is always in risk of getting trapped in a local minimum. The Monte Carlo strategy, on the other hand, aims at characterizing the a posteriori probability density. Moreover, it can mathematically be shown that the Metropolis algorithm will converge toward the correct equilibrium distribution independently of the complexity of the data uncertainty and the a priori model (e.g., Mosegaard, 1998). Roughly speaking, the adjoint-based method serves as a fast way of obtaining an approximate maximum a posteriori estimate based on a limited data uncertainty and a priori model. The strategy suggested here could subsequently be started in this estimate to obtain a characterization of the correct a posteriori probability density (i.e., resolution analysis) based on a realistic data uncertainty and a priori model and the full nonlinearity of the forward relation. The significant time expense of the Monte Carlo strategy may in the future be mitigated through parallelization of the individual FDTD calculation on, for

example, clusters or graphical processing units (GPU) and by running several algorithms in parallel.

In this study the suggested algorithm was applied on a synthetic test case with a simple reference model. Perfect knowledge about the data uncertainty model, the a priori information, and the source pulse was assumed. Moreover, the electrical conductivity was assumed known. Hence, the robustness of the algorithm has to be further tested by considering more complex synthetic models of both the dielectric permittivity and electrical conductivity and larger data sets. According to equation 2, knowledge about the statistical properties of the data uncertainties and the a priori information evenly influences the a posteriori probability density function. Hence, to ensure a trustworthy a posteriori probability density the user needs to specify a realistic statistical description of the data uncertainties and the a priori information. This may, however, be a very challenging task when dealing with field experiments. On the other hand, if one does not choose the a priori information the inversion algorithm will implicitly make this choice. For example, least-squares and adjoint-based inversion algorithms implicitly consider a Gaussian a priori and data uncertainty model. In the method that we propose here, one is free to choose a Gaussian description of the a priori and data uncertainty model, but the user has the flexibility to choose other more complex descriptions.

In field experiments the list of sources of uncertainties involves both a background noise component and (typically) a major component related to modeling errors (i.e., discrepancies between observed data and simulated data due to shortcomings of the forward relation), such as: (1) source pulse uncertainty, (2) 2D assumption, (3) local antennae coupling effects, (4) effects of discretization, (5) small-scale near borehole heterogeneities, (6) numerical attenuation, (7) media dispersion, (8) point-dipole antenna assumption, (9) effects of high angle travel paths due to wave guiding and antennae spread, (10) unknown antennae positions.

Some of the above mentioned modeling inadequacies can be accounted for through more sophisticated FDTD algorithms (see e.g., Bergmann et al., 1998; Holliger and Bergmann, 2002; Ernst et al., 2006; Irving and Knight, 2006). Alternatively, these modeling uncertainties should be accounted for through the data uncertainty model (although it results in higher a posteriori model variability). Forward modeling inadequacies may lead to a combination of correlated and uncorrelated data uncertainties that may be accounted for through a data covariance matrix (e.g., Cordua et al., 2008; 2009) or a more complex description. See e.g., Peterson (2001), Cordua et al. (2008, 2009), and Irving and Knight (2005) for descriptions and quantifications of some of these uncertainties.

Statistics of the background noise component could be calculated in the signal recorded before the first arriving signal at the receiver. In this period only noise is recorded at the receiver. This may be obtained by performing an experimental covariance analysis of this part of the signal and then subsequently fitting an analytic covariance model to the experimental covariance. The analytic covariance model can then be used to set up an appropriate likelihood function that accounts for the inferred noise component. For more details on how to determine an experimental covariance model (i.e., semivariogram) and how to fit an analytic covariance model see e.g., Journel and Huijbregts (1978) or Goovaerts (1997).

Ernst et al. (2007a) suggested a method based on deconvolution for source pulse estimation, which may also be applied together with our inversion strategy. Then uncertainty related to this process

needs to be quantified or a Bayesian formulation of the source pulse determination could be an integral part of the inversion (see Buland and Omre [2003] for an example of Bayesian wavelet estimation).

Modeling error statistics related to the 2D assumption may be determined through simulation. Consider that a 3D a priori model is established: Forward simulations from, say, 100 realizations of the 3D a priori model could be compared with forward simulations obtained in the associated 2D profile in the transmitter-receiver plane of the 3D model. By subtracting 3D and 2D simulations an estimate of the modeling error statistics associated with the 2D assumptions could be obtained. In the same way modeling error statistics related to other inadequacies in the forward algorithms may be simulated by comparing the results from simple algorithms with more adequate algorithms. In this way adequate computationally expensive forward calculations can (in principle) be substituted with an appropriate data uncertainty model and a faster approximate forward simulation.

While the use of training images to describe a priori information allows complex a priori information to be quantified, the establishment of a priori information through a training image for field data experiments may be challenging. The training image represents a concept of the patterns expected a priori independently of the observed data. These expectations may be based on previous studies in the area based on, for example, kriging and inversion conditional to soft (e.g., seismic) and hard (borehole) data and/or studies of outcrops (e.g., Zhang, 2008). To compromise with the distinct facies (i.e., low entropy) seen in the training image applied here, the a priori model could constitute a combination of information from a training image and a high entropy (e.g., two-point statistically) based a priori model. In any case, we suggest that the movie strategy should be used to ensure that the algorithm or the mathematical expression (or a combination of both) that describes the a priori information also reflects the a priori expectations of the user. See Zhang (2008) for a description of how to convert geological information into training images that can be used with multiple-point statistical simulation algorithms. Journel and Zhang (2006) demonstrated that even for a training image with combined high- and low-entropy information that only approximately captures the complex low entropy structures of the true model, a better conditional modeling result was obtained than for a pure high entropy Gaussian a priori model.

To apply the proposed full-waveform inversion strategy to field data, the major challenge concerning a way of handling modeling uncertainties (probably through a combination of a statistical description and a more adequate modeling algorithm) and to obtaining realistic a priori information that involves training images (from previous studies) still needs to be addressed. Finally, challenges regarding mitigating the computational expense of the Monte Carlo strategy are left for future research.

CONCLUSION

We have outlined the theoretical background for a Monte Carlo-based full-waveform inversion strategy based on the extended Metropolis algorithm in conjunction with complex geostatistical based a priori information. The use of geostatistical algorithms for the description of a priori information can be accomplished in an efficient way through the method of sequential Gibbs sampling, which allows for inclusion of a priori information described by any geostatistical algorithm based on sequential simulation.

This, in turn, provides a means of using a priori information described by both two-point and multiple-point statistical a priori models. Inclusion of such statistical a priori information reduces the complexity of the inverse problem, which is a keystone in the feasibility of performing Monte Carlo sampling of the computationally hard full-waveform inverse problem. We have demonstrated the potential of this inversion strategy by sampling the a posteriori probability density of a tomographic full-waveform inverse problem using complex a priori information inferred from a training image using the geostatistical algorithm Snesim. The methodology provides a means of evaluating the a posteriori uncertainty, which is not provided using traditional adjoint-based optimization strategies for full-waveform inversion. However, it should be noted that, if the goal is a single inverse estimate based on pure Gaussian statistics, the adjoint-based optimization approach for inversion of full-waveform data is computationally considerably faster than the suggested inversion strategy. Establishment of adequate (e.g., a multiple-point-based) a priori information and a data uncertainty model for field data are critical to obtain meaningful a posteriori uncertainty estimates. Moreover, the major computational expenses of the Monte Carlo strategy have to be mitigated, which are all challenging tasks that need future research.

ACKNOWLEDGMENTS

REFERENCES

Belina, F. A., J. R. Ernst, and K. Holliger, 2009, Inversion of crosshole seismic data in heterogeneous environments: Comparison of waveform and ray-based approaches: Journal of Applied Geophysics, **68**, no. 1, 85–94, doi: 10.1016/j.jappgeo.2008.10.012.

Bergmann, T., J. O. A. Robertsson, and K. Holliger, 1998, Finite-difference modeling of electromagnetic wave propagation in dispersive and attenuating media: Geophysics, **63**, 856–867, doi: 10.1190/1.1444396.

Buland, A., and H. Omre, 2003, Bayesian wavelet estimation from seismic and well data: Geophysics, **68**, 2000–2009, doi: 10.1190/1.1635053.

Cordua, K. S., M. C. Looms, and L. Nielsen, 2008, Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data: Vadose Zone Journal, **7**, 263–271, doi: 10.2136/vzj2007.0008.

Cordua, K. S., L. Nielsen, M. C. Looms, T. M. Hansen, and A. Binley, 2009, Quantifying the influence of static-like errors in least-squares-based inversion and sequential simulation of cross-borehole ground penetrating radar data: Journal of Applied Geophysics, **68**, no. 1, 71–84, doi: 10.1016/j.jappgeo.2008.12.002.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538, doi: 10.1190/1.1442864.

Davis, J. L., and A. P. Annan, 1989, Ground-penetrating radar for high-resolution mapping of soil and rock stratigraphy: Geophysical Prospecting, **37**, 531–551, doi: 10.1111/GPR.1989.37.issue-5.

Djikpéssé, H. A., and A. Tarantola, 1999, Multiparameter l_1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, **64**, 1023–1035, doi: 10.1190/1.1444611.

Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007a, Application of a new 2D time-domain full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, no. 5, J53–J64, doi: 10.1190/1.2761848.

Ernst, J. R., K. Holliger, H. R. Maurer, and A. G. Green, 2006, Realistic FDTD modelling of borehole georadar antenna radiation: Methodology and application: Near Surface Geophysics, 19–31, doi: 10.3997/1873-0604.2005028.

Ernst, J. R., H. Maurer, A. G. Green, and K. Holliger, 2007b, Full-waveform inversion of crosshole radar data based on 2-D finite-difference time-domain solutions of Maxwell's equations: IEEE Transactions on Geoscience and Remote Sensing, **45**, 2807–2828, doi: 10.1109/TGRS.2007.901048.

Gauthier, O., J. Virieux, and A. Tarantola, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Numerical results: Geophysics, **51**, 1387–1403, doi: 10.1190/1.1442188.

Gelman, A., G. O. Roberts, and W. R. Gilks, 1996, Efficient Metropolis jumping rules: *in* J. Bernardo, J. Berger, A. Dawid, and A. Smith, eds., Bayesian statistics, **5**, Oxford University Press599–607.

Gloaguen, E., B. Giroux, D. Marcotte, and R. Dimitrakopoulos, 2007, Pseudo-full-waveform inversion of GPR data using stochastic tomography: Geophysics, **72**, J43–J51, doi: 10.1190/1.2755929.

Gomez-Hernandez, J., and A. Journel, 1993, Joint sequential simulation of multi-Gaussian fields: *in* A. Soares, ed., Geostatistics Troia, **92**, Proceedings of the 4th International Geostatics Congress, Kluwer Academic Publishers, 85–94.

Goovaerts, P., 1997, Geostatistics for natural resources evaluation: Oxford Univsity Press.

Guardiano, F., and R. Srivastava, 1993, Multivariate geostatistics: Beyond bivariate moments, *in* A. Soares, ed., Geostatistics Tróia '92, v. 1: Kluwer, 133–144.

Hansen, T. M., A. G. Journel, A. Tarantola, and K. Mosegaard, 2006, Linear inverse Gaussian theory and geostatistics: Geophysics, **71**, R101–R111, doi: 10.1190/1.2345195.

Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2012, Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling: Computational Geosciences, doi: 10.1007/s/10596-011-9271-1.

Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2008, Using geostatistics to describe complex a priori information for inverse problems, *in* J. M. Ortiz, and X. Emery, eds., Geostatistics 2008: Chile, vol. **1**.

Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2009, Reducing complexity of inverse problems using geostatistical priors: IAMG 2009: Stanford, Ca.

Holliger, K., and T. Bergmann, 2002, Numerical modeling of borehole georadar data: Geophysics, **67**, 1249–1257, doi: 10.1190/1.1500387.

Irving, J. D., and R. J. Knight, 2005, Effect of antennas on velocity estimates obtained from crosshole GPR data: Geophysics, **70**, no. 5, K39–K42, doi: 10.1190/1.2049349.

Irving, J. D., and R. J. Knight, 2006, Numerical simulation of antenna transmission and reception for crosshole ground-penetrating radar: Geophysics, **71**, no. 2, K37–K45, doi: 10.1190/1.2187768.

Irving, J. D., and K. Singha, 2010, Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities: Water Resources Research, **46**, no. 11, W11514, doi: 10.1029/2009WR008340.

Journel, A., and T. Zhang, 2006, The necessity of a multiple-point prior model: Mathematical Geology, **38**, no. 5, 591–610, doi: 10.1007/s11004-006-9031-2.

Journel, A. G., and C. J. Huijbregts, 1978, Mining geostatistics: Academic Press.

Klotzsche, A., J. van der Kruk, G. A. Meles, J. A. Doetsch, H. Maurer, and N. Linde, 2010, Full-waveform inversion of cross-hole ground-penetrating radar data to characterize a gravel aquifer close to the Thur River, Switzerland: Near Surface Geophysics, **8**, 635–649, doi: 10.3997/1873-0604.2010054.

Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, **96**, 20289–20299, doi: 10.1029/91JB02278.

Looms, M. C., K. H. Jensen, A. Binley, and L. Nielsen, 2008, Monitoring unsaturated flow and transport using cross-borehole geophysical methods: Vadose Zone Journal, **7**, no. 1, 227–237, doi: 10.2136/vzj2006.0129.

Meles, G., S. Greenhalgh, J. van der Kruk, A. Green, and H. Maurer, 2011, Taming the non-linearity problem in GPR full-waveform inversion for high contrast media: Journal of Applied Geophysics, **73**, no. 2, 174–186, doi: 10.1016/j.jappgeo.2011.01.001.

Meles, G. A., J. van der Kruk, S. A. Greenhalgh, J. R. Ernst, H. Maurer, and A. G. Green, 2010, A new vector waveform inversion algorithm for simultaneous updating of conductivity and permittivity parameters from combination crosshole/borehole-to-surface GPR data: IEEE Transactions on Geoscience and Remote Sensing, **48**, 3391–3407, doi: 10.1109/TGRS.2010.2046670.

Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data: Geophysics, **52**, 1211–1228, doi: 10.1190/1.1442384.

Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo sampling: Inverse Problems, **14**, 405–426, doi: 10.1088/0266-5611/14/3/004.

Mosegaard, K., 2011, Quest for consistency, symmetry, and simplicity — The legacy of Albert Tarantola: Geophysics, **76**, no. 5, W51–W61, doi: 10.1190/geo2010-0328.1.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 431–447, doi: 10.1029/94JB03097.

Mosegaard, K., and M. Sambridge, 2002, Monte Carlo analysis of inverse problems: Inverse Problems, **18**, R29–R54, doi: 10.1088/0266-5611/18/3/201.

Nielsen, L., M. C. Looms, T. M. Hansen, K. S. Cordua, and L. Stemmerik, 2010, Estimation of chalk heterogeneity from stochastic modelling conditioned by crosshole GPR travel times and LOG data: *in* R. Miller, J. Bradford, and K. Holliger, eds., Advances in near-surface seismology and ground-penetrating radar: SEG Geophysical Development Series 15, 379–398.

Peterson, J. E., 2001, Pre-inversion correction and analysis of radar tomographic data: Journal of Environmental and Engineering Geophysics, **6**, no. 1, 1–18, doi: 10.4133/JEEG6.1.1.

Pica, A., J. P. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 284–292, doi: 10.1190/1.1442836.

Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multisource cross-hole tomography. Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310, doi: 10.1111/GPR.1990.38.issue-3.

Remy, N., A. Boucher, and J. Wu, 2008, Applied geostatistics with SGeMS: A user's guide: Cambridge University Press.

Reynolds, J. M., 1997, An introduction to applied and environmental geophysics: John Wiley and Sons.

Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics: Mathematical Geology, **34**, no. 1, 1–21, doi: 10.1023/A:1014009426274.

Talagrand, O., and P. Courtier, 1987, Variational assimilation of meteorological observations with the adjoint vorticity equation. I. Theory: Quarterly Journal of the Royal Meteorological Society, **113**, 478, 1311–1328, doi: 10.1256/smsqj.47811.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.

Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: Geophysics, **51**, 1893–1903, doi: 10.1190/1.1442046.

Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation: Pure and Applied Geophysics, **128**, 1-2, 365–399, doi: 10.1007/BF01772605.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society of Industrial and Applied Mathematics.

Tarantola, A., and B. Valette, 1982, Inverse Problems = Quest for Information: Journal of Geophysics, **50**, 159–170.

Tronicke, J., and K. Holliger, 2005, Quantitative integration of hydrogeophysical data: Conditional geostatistical simulation for characterizing heterogeneous alluvial aquifers: Geophysics, **70**, no. 3, H1–H10, doi: 10.1190/1.1925744.

Zhang, T., 2008, Incorporating geological conceptual models and interpretations into reservoir modeling using multiple-point geostatistics: Earth Science Frontiers, **15**, no. 1, 26–35, doi: 10.1016/S1872-5791(08)60016-0.