



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

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

Published in:
Geophysics

Link to article, DOI:
[10.1190/geo2011-0170.1](https://doi.org/10.1190/geo2011-0170.1)

Publication date:
2012

[Link back to DTU Orbit](#)

Citation (APA):

Cordua, K. S., Hansen, T. M., & Mosegaard, K. (2012). Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information. *Geophysics*, 72(2), H19-H3. DOI: 10.1190/geo2011-0170.1

DTU Library

Technical Information Center of Denmark

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

1 **TITLE**

2

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

5

6 **Right Running Head:** Monte Carlo full waveform inversion

7

8 Knud Skou Cordua¹, Thomas Mejer Hansen¹, and Klaus Mosegaard¹

9

10 ¹Technical University of Denmark, Department of Informatics and Mathematical Modeling,
11 Richard Petersens Plads, DTU – building 321, DK-2800 Lyngby, Denmark

12

13 Corresponding author: K. S. Cordua (kcor@imm.dtu.dk).

14 Other e-mail addresses:

15 T. M. Hansen (tmeha@imm.dtu.dk); K. Mosegaard (kmos@imm.dtu.dk).

16

17 Original paper; date of submission: May 11, 2011

18 Revised paper; date of submission: September 12, 2011

19 Revised paper (after second review); date of resubmission: November 26, 2011

20 Accepted for publication; November 26, 2011

21

22 Preliminary results of the work presented in this manuscript were presented at SEG 2010, Denver
23 (see Cordua et al., 2010).

24

ABSTRACT

This paper presents 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 non-linear 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 2-point (e.g. Gaussian) or multiple-point statistics. In this paper 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 to 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. Moreover, an adequate knowledge of data uncertainties and *a priori* information is required in order to obtain meaningful uncertainty estimates. The latter may be a key challenge when considering field experiments, which will not be addressed here.

1 **Abbreviations:** GPR, Ground penetrating radar; Snesim, Single normal equation simulation;
2 FDTD, Finite-difference time-domain.

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

INTRODUCTION

Albert Tarantola was one of the pioneers of seismic full waveform inversion (see Tarantola, 1984, 1986, 1988). Using a steepest descent algorithm, the update gradient was in each iteration obtained by correlating a forward propagated wave field with the residual wave field propagated backwards 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. Looms et al., 2008; Ernst et al., 2007a). 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 inter-borehole region can be obtained (Peterson, 2001).

Ernst et al. (2007a, b) introduced and applied the adjoint-based optimization algorithm for inversion of tomographic crosshole GPR full waveform data. They demonstrate that sub-wavelength 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 to traditional

1 ray-based inversion strategies. The inversion scheme by Ernst et al. (2007b) was later adopted for
2 inversion of seismic waveform data by Belina et al. (2009). Recently, the algorithm by Ernst et al.
3 (2007b) has been improved such that it exploits the full vector field of the electromagnetic wave
4 propagation and allows for arbitrary antennae geometry (Meles et al., 2010). Klotzsche et al. (2010)
5 demonstrate an application of the improved code to crosshole GPR waveform data acquired in a
6 gravel aquifer. Lately, Meles et al. (2011) demonstrated that the adjoint-based method of Ernst et al.
7 (2007b) becomes less sensitive to the starting model when the inversion is conditioned to the long
8 wavelengths of the signal in the initial part of the inversion and higher frequencies are gradually
9 incorporated.

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

20 In a Bayesian formulation the solution to the inverse problem is given as an *a*
21 *posteriori* probability density (Tarantola and Valette, 1982). The *a posteriori* probability is based on
22 the independent states of information provided by the data (related to the model parameters through
23 a physical law), an associated data uncertainty model (that takes into account data noise and data
24 simulation inadequacies) and the *a priori* information on the model. The combined states of

1 information contained in the *a posteriori* probability density are reflected in the model variability of
2 the *a posteriori* sample. Hence, data uncertainties will not cause artifacts in the solution, but rather
3 influence the degree of *a posteriori* model variability if the nature of the uncertainties is
4 appropriately accounted for in the uncertainty model. Resolution analysis is naturally obtained from
5 the *a posteriori* model variability. One may simply be interested in calculating the covariance of the
6 *a posteriori* model parameters, but more sophisticated questions, such as the probability of
7 geological connectivity or the residence time of a fluid, may also be answered by the *a posteriori*
8 statistics (Mosegaard, 1998).

9 The extended Metropolis algorithm (Mosegaard and Tarantola, 1995) can be used to
10 sample the *a posteriori* probability density, even for highly nonlinear inverse problems. This
11 algorithm does not need a closed form mathematical expression of the *a priori* information, but a
12 “black box” algorithm that is able to sample the *a priori* probability density is sufficient. Hansen et
13 al. (2008; accepted) suggested a method that provides a means of controlling the perturbation step-
14 size and efficiency when sampling *a priori* information defined through any geostatistical
15 algorithms that are based on sequential simulation (e.g. Gomez-Hernandez and Journel, 1993). This
16 method is referred to as Sequential Gibbs sampling (Hansen et al., accepted). They demonstrate that
17 this method could serve as a “black box” algorithm in the extended Metropolis algorithm.
18 Sequential Gibbs sampling can be used to sample *a priori* models based on either relatively simple
19 2-point based statistical models, such as Gaussian based *a priori* models, or more complex
20 multiple-point based statistical models. This provides a means of using complex *a priori* statistical
21 models that allow reproduction of geologically plausible structures, such as channels and tortuosity.
22 Such complex patterns (i.e. spatial autocorrelation) can be learned from so-called training images
23 and reproduced by simulation algorithms based on multiple-point statistics (e.g. Strebelle, 2002).
24 Multiple-point algorithms offer the flexibility of simulating realizations with high entropy (e.g.

1 Gaussian distributions) as well as low entropy (i.e. a few facies) structures or a combination of both
2 (Journel and Zhang, 2006). See for example Remy et al. (2008) for different examples of such
3 complex statistical models. In this way, the extended Metropolis algorithm becomes very flexible
4 with regards to the choice of *a priori* model.

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

19 Tarantola (2005) was a proponent of the movie strategy, in which “movies” of
20 multiple realizations from the *a priori* and *a posteriori* probability densities are compared in order
21 to understand the additional information provided by the data compared to the *a priori* information.
22 Features that occur frequently in the *a posteriori* movie are regarded as well resolved. We show that
23 the movie strategy provides a means of obtaining resolution analysis of the full waveform inverse
24 problem. Moreover, we demonstrate how *a posteriori* realizations can be used to quantify the

1 probability of lithological connectivity. The term resolution is, in this paper, used such that high *a*
2 *posteriori* model variability refers to low resolution and vice versa (Mosegaard, 1998). To our
3 knowledge this is the first example of sampling the *a posteriori* probability density of a
4 tomographic full waveform inverse problem. Resolution analysis of the problem reveals that the
5 combined states of information from the full waveform data and the *a priori* information provides a
6 high resolution subsurface image, even in the case of considerably sparse data coverage. Finally, we
7 demonstrate how the geostatistically formulated *a priori* information serves as a guide in the initial
8 burn-in part of the inversion procedure. In this way, convergence of the suggested full waveform
9 inversion algorithm becomes independent of the initial model.

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

15

16

METHODOLOGY

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

$$22 \quad \mathbf{d} = g(\mathbf{m}), \quad (1)$$

23 where g is a linear or non-linear mapping operator that often relies on a physical law. In this
24 particular study the model represents a tomographic image of dielectric permittivity of the

1 subsurface material, and data are waveforms of the vertical electrical component of electromagnetic
2 waves propagated across the model between two boreholes. The forward relation in equation 1 is
3 given as a finite-difference time-domain (FDTD) solution of Maxwell's equations (Ernst et al.,
4 2007b). However, any numerical wave propagation modeling strategy (for GPR or seismic signals)
5 can be applied. The inverse problem is to infer information about the model parameters based on a
6 set of data (and their uncertainties), *a priori* information about the model, and the forward relation
7 between the model and the data.

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

$$10 \quad \sigma_M(\mathbf{m}) = c \rho_M(\mathbf{m}) L(\mathbf{m}), \quad (2)$$

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

21

22 **Sampling the *a posteriori* probability density**

23 A highly nonlinear inverse problem refers to the case where the *a priori* probability
24 density is far from being Gaussian or the likelihood function is highly non-Gaussian (typically) due

1 to the nonlinear forward relation between model and data. According to equation 2 the product of
 2 these non-Gaussian probability densities signify a highly non-Gaussian *a posteriori* probability
 3 density. In the case of full waveform inversion, the forward relation is expected to be highly
 4 nonlinear. Moreover, realistic *a priori* information, as we introduce here, is typically far from being
 5 Gaussian.

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

13 The extended Metropolis Algorithm consists of two randomized steps:

14 1) Exploration; propose a candidate model, $\mathbf{m}_{propose}$, which is a perturbation of a current model,
 15 $\mathbf{m}_{current}$, and at the same time is a realization of the *a priori* probability density.

16 2) Exploitation; decide if the proposed model should be accepted or rejected. The proposed model is
 17 accepted with the Metropolis acceptance probability (referred to as the Metropolis rule):

$$18 \quad P_{accept} = \min \left(1, \frac{L(\mathbf{m}_{propose})}{L(\mathbf{m}_{current})} \right), \quad (3)$$

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

1 The exploration step constitutes the strategy by which proposed models are drawn
2 from the *a priori* probability density. For small exploration steps the proposed model will be
3 relatively highly correlated with the current model, compared to large exploration steps. Thus,
4 according to equation 3, small exploration steps will result in a high acceptance probability and vice
5 versa. Recall that each evaluation of the likelihood function involves a computationally expensive
6 (FDTD) forward calculation. It is, therefore, important to choose an appropriate exploration step-
7 size that does not explore the *a posteriori* probability density too slowly, but on the other hand does
8 not waste too many expensive evaluations of the likelihood that are very unlikely to be accepted. In
9 particular, the exploration strategy becomes very important when dealing with high dimensional
10 probability densities, since traditional sampling strategies may lead to a very inefficient exploration
11 strategy (Hansen et al., 2008; accepted). Recently, Hansen et al. (2008; accepted) introduced a
12 flexible sampling strategy to sample high dimensional *a priori* probability densities defined by
13 geostatistical algorithms. This sampling strategy provides a means of using a variable exploration
14 step-size when sampling such high dimensional probability densities. This method is briefly
15 outlined in the following.

16

17 **Sampling geostatistically defined *a priori* information**

18 Hansen et al. (accepted) suggest a strategy referred to as sequential Gibbs sampling,
19 which is a method that is capable of sampling probability densities defined by geostatistical
20 algorithms based on sequential simulation. Originally, sequential simulation was used to generate
21 realizations of 2-point based statistical models, such as Gaussian models (e.g. Gomez-Hernandez
22 and Journel, 1993). 2-point statistical models are limited to specify the spatial variability between
23 pairs of data (hence the name 2-point statistics) defined by a covariance model. 2-point statistics
24 does not provide enough information to model plausible geological structures, such as channels and

1 tortuosity. This can be overcome by using a model based on multiple-point statistics, as suggested
2 by Guardiano and Srivastava (1993). In multiple-point based statistical models, the *a priori*
3 information is learned from a training image. The training image is scanned by a template that
4 jointly consider spatial variability between a number of (larger than two) pixel values in the image
5 in order to obtain a joint probability distribution that hold information about these spatial
6 correlations. The model parameter values are subsequently sequentially simulated from conditional
7 probabilities based on the jointly considered pixel values (e.g. Strebelle, 2002). The algorithm
8 suggested by Guardiano and Srivastava (1993) was, however, computationally unfeasible. It was
9 not until an efficient way of storing multiple-point statistics in machine memory was proposed that
10 the use of multiple-point based models became computationally feasible (Strebelle, 2002). See e.g.
11 Remy et al. (2008) for numerous examples of the application of sequential simulation from both 2-
12 point and multiple-point based *a priori* models.

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

- 16 1) An initial unconditional realization of the *a priori* probability density, $\mathbf{m}_{current}$, defined in a 2-
17 dimensional regular grid of model parameters is provided.
- 18 2) A square subarea, corresponding to model parameters $\mathbf{m}_{subarea}$, with side-length E_{step} (exploration
19 step-size) of the current model $\mathbf{m}_{current}$ is randomly chosen.
- 20 3) A realization of the conditional probability density $\rho_M(\mathbf{m}_{subarea} | \tilde{\mathbf{m}}_{current})$ is obtained using
21 sequential simulation. $\tilde{\mathbf{m}}_{current}$ are the current model parameters outside the subarea. $\rho_M(\mathbf{m})$ is an *a*
22 *priori* probability density that may be described by either 2- or multiple-point statistics. This step is
23 an application of the Gibbs sampler where sequential simulation is used to efficiently generate a

1 realization from the conditional probability density (hence the name sequential Gibbs sampling). In
 2 practice this is performed simply by running the sequential simulation algorithm conditional to the
 3 model parameters outside the subarea. In this way a new, perturbed model, $\mathbf{m}_{propose}$ is obtained.
 4 4) The proposed model becomes the current model and steps 2 and 3 are repeated in order to obtain
 5 multiple realizations of the *a priori* probability density.

6 Recall that, when applying sequential Gibbs sampling as an *a priori* model sampler in
 7 the extended Metropolis algorithm, the current model $\mathbf{m}_{current}$ is reused if the proposed model
 8 $\mathbf{m}_{propose}$ is rejected by the Metropolis rule. Moreover, the resemblance (i.e. correlation) between the
 9 current and proposed model, and thus the average Metropolis acceptance probability (cf. the
 10 Metropolis rule), can be controlled by the explorations step-side E_{step} (i.e. the side-length of the
 11 resimulated area).

12

13 **The likelihood function and correlated data errors**

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

$$23 \quad L(\mathbf{m}) = c \prod_{k=1}^K \exp\left(-\frac{1}{2} \left(g(\mathbf{m})^k - \mathbf{d}_{obs}^k\right)^T \mathbf{C}_D^{-1} \left(g(\mathbf{m})^k - \mathbf{d}_{obs}^k\right)\right), \quad (4)$$

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

$$6 \quad \mathbf{C}_D(i, j) = c \exp\left(\frac{-3s(i, j)}{a}\right), \quad (5)$$

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

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

16

17 **The burn-in period**

18 If the probability that a sampling algorithm, at any time, enters an infinitesimal
 19 neighborhood N_j , that surrounds the model \mathbf{m}_j , is equal to the probability that it leaves this
 20 neighborhood, the algorithm is said to satisfy microscopic reversibility (e.g. Mosegaard and
 21 Sambridge, 2002). It can be shown that if a sampling algorithm satisfies microscopic reversibility,
 22 then the *a posteriori* probability density is the only equilibrium distribution for the algorithm, and
 23 the algorithm will converge toward this equilibrium distribution independently of the starting model

1 (e.g. Kaipio et al., 2000). This property is satisfied by the extended Metropolis algorithm (see
2 Mosegaard and Sambridge (2002) for a detailed description of the Metropolis algorithm). In
3 practice, the extended Metropolis algorithm has reached the equilibrium distribution when the
4 likelihood values starts to fluctuate around a constant level (referred to as the equilibrium level) and
5 the data are fitted within their uncertainties. When this phase is reached the algorithm is said to be
6 burned-in. In case of a non-linear full waveform inverse problem the structure of the *a posteriori*
7 probability density is completely unknown. If this distribution is multimodal, several modes may
8 exist that fit data with the uncertainty. Hence, no guarantee can be given, no matter the choice of
9 sample strategy, that the *a posteriori* probability density is appropriately represented (i.e. all modes
10 are represented) by a finite sample size.

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

22 The algorithm is started in an unconditional realization of the *a priori* model, which is
23 expected to be far away from volumes of significant *a posteriori* probability density. We suggest
24 adaptively adjusting the exploration step-size during the burn-in period such that the Metropolis

1 acceptance probability (equation 3) controls the exploration step-size. This is controlled by updating
 2 the exploration step-size after each M successive exploration steps of the Metropolis algorithm. The
 3 exploration step-size update is performed just after the exploitation step (step 1) of the algorithm.
 4 Consider that the algorithm is at iteration number K , where K is a multiple of M . Then the
 5 exploration step-size E_{step}^{i+1} during iteration number $K+1$ to $K+M+1$ (i.e. the next M iterations) is
 6 given as:

$$7 \quad E_{step}^{i+1} = E_{step}^i \frac{P_{average}}{P_{control}}, \quad (6)$$

8 where E_{step}^i is the exploration step-size during iteration number $K-M$ to K (i.e. the M preceding
 9 iterations). $P_{average}$ is the average Metropolis acceptance probability (equation 3) during iteration
 10 number $K-M$ to K and $P_{control}$ is a subjectively chosen constant acceptance ratio that the algorithm
 11 tries to adhere to by adjusting the exploration step-size according to equation 6. For larger values of
 12 $P_{control}$, the exploration steps-size decreases faster and converges to a lower level, than for smaller
 13 values. The minimum possible exploration step-size involves resimulation of one model parameter,
 14 whereas the maximum step-size is an (unconditional) simulation of all the model parameters that
 15 are statistically independent of the previous model.

16

17 **A SYNTHETIC CROSSHOLE GPR FULL WAVEFORM INVERSE PROBLEM**

18 The methodology outlined above is tested on a synthetic tomographic crosshole GPR
 19 full waveform data set. Wave field simulations of the GPR signals (i.e. the forward relation) are
 20 obtained using FDTD calculations of Maxwell's equations in transverse electric mode (Ernst et al.,
 21 2007b). This FDTD method provides grid-based time-domain calculations of the electromagnetic
 22 wave field propagation. The transmitting and receiving antennae are simulated as vertically
 23 orientated dipole-type antennae and are aligned parallel with the vertical boreholes. Transmitted and

1 received signals concern the vertical component of the electrical field (Holliger and Bergmann,
2 2002). The FDTD algorithm by Ernst et al. (2007b) yields second order accuracy in both time and
3 space, and performs the calculations in 2D Cartesian coordinates. The edges of the FDTD grid are
4 surrounded by a generalized perfectly matched layer (GPML) in order to absorb artificial boundary
5 reflections.

6 The *a priori* information of the inverse problem is provided by the Snesim algorithm.
7 Snesim is a fast geostatistical simulation algorithm that produces realizations (conditional or
8 unconditional to point data) from a high dimensional probability density that contains the spatial
9 relations (i.e. patterns) learned from a training image for a relatively low number of categorical
10 values (Strebelle, 2002). Figure 1 shows a training image that mimics a matrix of unconsolidated
11 sand with embedded channels of gravel situated in an unsaturated environment. The geological
12 information contained in the training image may have been obtained from outcrops in a nearby
13 gravel pit and/or a natural cliff. The training image applied here does not necessarily represent a
14 realistic near surface environment, but is rather used to demonstrate the principle that geologically
15 realistic features, such as complex channel structures, can be represented by a multiple-point based
16 *a priori* model in the inversion procedure. Electromagnetic signals in the near surface sediments are
17 sensitive to the dielectric permittivity, the electrical conductivity, and the magnetic permeability of
18 the materials. In this study we limit ourselves to considering only the influence of the dielectric
19 permittivity, which is primarily governing the phase velocity of the signal. The relative dielectric
20 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
21 m/ns), respectively (e.g. Ernst et al., 2006). The relative dielectric permittivity is given as $\varepsilon_r = \varepsilon / \varepsilon_0$
22 , where ε is the absolute dielectric permittivity and ε_0 is the dielectric permittivity of free space.

23 Figure 2 shows the synthetic reference model to be considered and is, at the same
24 time, an unconditional realization of the training image obtained using Snesim. The electrical

1 conductivity is set to a constant value of 3 mS/m and in the following it is assumed known. Near
2 surface materials are considered non-magnetic and the magnetic permeability is set to the magnetic
3 permeability of free space (e.g. Davis and Annan, 1989).

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

20 Gaussian distributed data uncertainties with a temporal autocorrelation described by
21 the exponential correlation function in equation 5 are added to the waveform data. The temporal
22 correlation length a (i.e. the range) is set to 12.7 ns. Figure 4 shows the five waveform traces
23 related to the uppermost transmitter position in the left borehole (see Figure 2), which are related to
24 the transmitter-receiver pairs marked by the numbers 1 to 5 in Figure 3. The noise free simulated

1 waveforms are plotted as a dotted blue curves and the uncertain waveforms (noisy waveforms) are
2 plotted as the red curves. The 20 uncertain waveform traces are used as observed data in this study
3 and have an average signal-to-noise ratio of 13.6. The uncertainty imposed on the “noise free” data
4 mimic the total contribution of data noise and modeling inadequacies. In the next section, full
5 waveform inversion will be performed on the uncertain waveform data with *a priori* information
6 based on a geostatistical model inferred from the training image in Figure 1.

7

8

RESULTS

9

Burn-in

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

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 independent realizations. The control acceptance, $P_{control}$, is set to 10 % in order 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 17000 iterations the likelihood values start to fluctuate around an equilibrium level and the data residuals resemble a

1 normal distribution with approximately the same standard deviation as the distribution of the noise.
2 At this stage it is assumed that the algorithm has reached burn-in and produces representative
3 realizations of the *a posteriori* probability density.

5 ***A posteriori* statistics**

6 The last model accepted by the Metropolis rule in the burn-in period is used as the
7 starting model when the Metropolis algorithm is subsequently restarted with a constant exploration
8 step-size of 1 m. The results obtained in this study demonstrate that the equilibrium level does not
9 change after the Metropolis algorithm is restarted with a constant exploration step-size. In addition
10 the data residuals of this period resembles a normal distribution with a standard deviation of
11 $2.17 \cdot 10^{-4}$. As a comparison the standard deviation of the normal distributed noise is $2.24 \cdot 10^{-4}$,
12 which demonstrates that the data are fitted within the data uncertainties. Hence, this shows that the
13 algorithm has reached burn-in and that the adaptive exploration step-size may be appropriate when
14 the step-size converges towards a constant level (which is the case in our study). This burn-in
15 strategy may serve as an approximate way of determining which exploration step-size should be
16 used in order to obtain a certain average acceptance probability. See Gelman et al. (1996) for an
17 investigation on the choice of acceptance probability.

18 After 300000 iterations the algorithm was stopped. During the sample period the
19 algorithm had an average acceptance probability of 40%. Figure 7 shows the autocorrelation
20 between the first model after burn-in and its correlation to the first 150000 realizations obtained
21 from the *a posteriori* probability density. These models are not statistically independent because
22 any proposed model in the Metropolis algorithm is a perturbation of a current model or when a
23 proposed model is rejected the current model counts again. Therefore, the autocorrelation analysis
24 of the *a posteriori* sample shows some correlation length between successive models. In the

1 example shown in Figure 7, statistical independence is obtained after approximately 5800 iterations.
2 This point is approximated as the point at which the autocorrelation curve intercepts the average
3 level of the correlation curve after it has converged to a constant level. The average is shown as a
4 dotted line in Figure 7 and is calculated as the average correlation coefficient between iteration
5 20000 and 300000. A similar analysis is performed on 10 models picked at different iteration
6 numbers, equally distributed across the *a posteriori* sample, which gives an average of 6745
7 iterations of separation in order to obtain statistically independent realizations from the *a posteriori*
8 sample.

9 In a probabilistic formulation, the solution to the inverse problem is not a single model
10 estimate, but a sample of multiple model realizations drawn from the *a posteriori* probability
11 density. Each realization is a tomographic image of the subsurface. Displaying multiple images
12 together corresponds to a movie. The strategy of displaying and studying the solution to the inverse
13 problem using such movies is referred to as the movie strategy (Tarantola, 2005). Roughly
14 speaking, the *a priori* probability density is filtered by the likelihood function which results in the *a*
15 *posteriori* probability density. Hence, displaying a “movie” of *a priori* realizations together with a
16 “movie” of *a posteriori* realizations helps to understand the characteristics of the *a priori*
17 information imposed on the inverse problem and to understand the state of information provided by
18 the data (and their uncertainties). In particular, when the *a priori* information is provided through a
19 “black box” algorithm, and no closed form mathematical expression of the *a priori* probability
20 density exists, an *a priori* movie may be important in order to understand if the state of information
21 provided by the *a priori* probability density is commensurate with the *a priori* expectation of the
22 user. See Koren et al. (1991) for a seminal example of using the movie strategy for a seismic
23 inverse problem.

1 Figure 8 shows eight statistically independent realizations drawn from the *a priori*
2 probability density (i.e. an *a priori* movie). This “movie” shows the state of information provided
3 by the *a priori* probability density inferred from the training image by the Snesim algorithm. This
4 movie shows a reproduction of the channel structures inferred from the training image, but with no
5 resemblance between the location of the channels in individual realizations, as these models are
6 unconditioned by any data. Figure 9 shows eight statistically independent realizations from the *a*
7 *posteriori* probability density. The result clearly demonstrates a high degree of resemblance
8 between the individual *a posteriori* realizations as a result of the conditioning to the full waveform
9 data. The high resemblance reveals that the data provides a high resolution of the inverse problem,
10 despite sparse data coverage (see Figure 3). Moreover, the individual realizations only slightly
11 deviate from the reference model (see Figure 2), which confirms the good resolution provided by
12 the full waveform data, their uncertainties, and the training image. The *a posteriori* realizations in
13 Figure 9 shows some isolated small scale features, which are not seen in the training image. These
14 non-continuous effects are caused by the Snesim simulation technique, because the algorithm
15 occasionally reduces the number of conditioning data events (i.e. pixel values) in order to avoid
16 singularities during calculation of the conditional probabilities. A total of 45 statistically
17 independent realizations are obtained from the 300000 *a posteriori* realizations based on the model
18 autocorrelation analysis (see Figure 7). These realizations can be used to ask higher order statistical
19 questions such as; what is the probability of connectivity between a channel observed in the left
20 borehole and a channel observed in the right borehole? Figures 9 d, f, and h show some examples of
21 missing connectivity between the boreholes marked by black circles. For example it is found that in
22 five out of the 45 realizations there is no connection between the points A and B as marked in
23 Figure 8a. Hence, this gives an approximate *a posteriori* probability of connectivity between points
24 A and B of $(45-5)/45 = 89\%$.

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

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

20

21

DISCUSSION

22

23

24

Hansen et al. (2006) demonstrated that linear inverse Gaussian theory and simple kriging (i.e. 2-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 towards

1 bridging probabilistic inverse problem theory and the field of geostatistics, even for highly
2 nonlinear and non-Gaussian inverse problems (Hansen et al., 2008; accepted). Hansen et al. (2009;
3 accepted) discussed how the complexity of inverse problems (i.e. the time needed to obtain an *a*
4 *posteriori* sample) is reduced when a statistical *a priori* model with some degree of spatial
5 correlation between the model parameters is considered. They showed that the effective dimension
6 of the solution space of the inverse problem, using 2-point based *a priori* models, is considerably
7 decreased when the *a priori* expected range of spatial correlation is increased. On the other hand,
8 when no spatial correlation was considered, the high-dimensional inverse problem became
9 unsolvable. Further, it was seen that whether the model parameters take values from a set of real or
10 binary numbers had insignificant influence on the effective dimension as compared to the degree of
11 correlation between the model parameters. Hence, we consider the reduction of the effective
12 dimension due to *a priori* defined spatial autocorrelation to be one of the keystones that makes
13 Monte Carlo inversion of a computationally hard, full waveform inverse problem feasible, even
14 with thousands of (here 6240) model parameters.

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

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

1 when making use of 2-point-based statistical *a priori* models with both continuous and categorical
2 model parameter values as long as some degree of spatial correlation can be considered *a priori*.
3 This is also confirmed by our preliminary studies on this topic. The above discussion is encouraging
4 with regard to the possibility of using *a priori* models based on either 2-point statistics, multiple-
5 point statistics, or a combinations of both, as long as the chosen *a priori* model imposes some
6 degree of spatial autocorrelation of the model parameters (See Journel and Zhang (2006) and further
7 discussions at the end of this section). This provides a flexible tool for defining an appropriate *a*
8 *priori* model that actually captures our *a priori* expectations.

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

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

23 In this study the sequential Gibbs sampler is applied such that a continuous block of
24 model parameters are resimulated in each step. Irving and Singha (2010) also used a type of

1 sequential Gibbs sampling, but considered to resimulate a subset of model parameters scattered
2 randomly across the model. Note that sequential Gibbs sampling will correctly sample the *a priori*
3 probability density function regardless of how that subset of model parameters to be resimulated is
4 chosen (Hansen et al., accepted). Whether the continuous or scattered exploration strategy should be
5 used is still to be investigated, as such a choice may affect the computational efficiency when the
6 sequential Gibbs sampler is used with the Metropolis algorithm. In order to understand exploration
7 strategies of high dimensional spaces, more analytical approaches are required.

8 Ray-based inversion experiments on the same scale as the setup used in this study
9 typically involve in the order of 700 to 1600 transmitter – receiver pairs in order to obtain a
10 reasonable resolution (e.g. Looms et al., 2008; Nielsen et al., 2010; Tronicke and Holliger, 2005). In
11 this study a high degree of resolution was obtained with as few as 20 transmitter – receiver pairs.
12 However, in field experiments modeling inadequacies may lead to considerably more uncertainty in
13 the data than considered in this study, which in turn leads to a lower resolution. In crosshole GPR
14 full waveform inversion the long spatial wavelengths of the model are typically used as a starting
15 model. This model is obtained through ray-based inversion of first arrival travel times and
16 amplitudes of the waveform data (Ernst et al., 2007a; Meles et al., 2010). The data set considered in
17 our study is expected to be too sparse to provide any useful information for a ray-based starting
18 model. We, therefore, choose to start the full waveform inversion in an unconditional realization of
19 the *a priori* probability density and burn-in is obtained anyway. Hence, the role of the *a priori*
20 model is more than simply finding *a posteriori* model realizations that jointly honor data and the *a*
21 *priori* model. It turns out that the use of a consistent *a priori* model acts as a guide in the burn-in
22 process that allows the initial model to be far away from the true solution. The successful
23 convergence from the data independent starting model observed in the present study may be
24 explained as a reduction of the complexity of the problem through the informative *a priori*

1 information defined through the geostatistical algorithm (see Hansen et al., 2009; accepted). This
2 encouraging observation suggests that future effort should be towards incorporating complex
3 statistical *a priori* information into (adjoint) optimization based inversion approaches.

4 The most commonly used method for full waveform inversion today is based on
5 adjoint methods, as suggested by Tarantola (1984). This approach has some limitations: 1)
6 Uncertainty estimates may theoretically be obtained through an *a posteriori* covariance for a linear
7 approximation of the forward relation limited to a Gaussian description of the data uncertainty and
8 *a priori* model (Tarantola, 1984). 2) The method is based on simple Gaussian *a priori* information
9 (if any at all), and 3) it relies on a subjective convergence criterion that may adversely result in data
10 uncertainties propagating into the model estimate. The method we propose overcomes many of the
11 limitations of using the adjoint-based approach: 1) It allows for arbitrary data geometry and density.
12 2) Complex *a priori* inversion can be included using any geostatistical algorithm that can be used
13 for sequential simulation or any other probabilistic defined a priori information (see Mosegaard and
14 Tarantola, 1995). 3) The full uncertainty of the inverse problem can be quantified by analyzing the
15 *a posteriori* sample. 4) There are no complications related to the incorporation of refined and
16 complex forward algorithms. 5) Temporally and spatially correlated data uncertainties can be
17 accounted for either through the data covariance matrix or any other mathematical description of the
18 uncertainties (Cordua et al., 2008, 2009). 6) Finally, if the correct data uncertainty model is known
19 there is no risk of uncertainties propagation into a single model estimate, but data uncertainties are
20 instead reflected in the *a posteriori* model variability.

21 The adjoint-based approach has some benefits that the presented method lacks: In case
22 of dense data coverage, full waveform inversion based on the adjoint-based approach may not need
23 any *a priori* model at all, as the data itself are sufficient to allow inference of a solution (Tarantola,
24 1984). The adjoint-based method has been successfully tested on field data (Ernst et al., 2007a;

1 Klotzsche et al., 2010). In addition the suggested probabilistic inversion strategy needs substantially
2 more (computationally expensive) forward calculations than compared to the traditional adjoint-
3 based approach.

4 In the synthetic test performed here each iteration involves: 1) A model perturbation
5 using Snesim which takes approximately 0.5 s. 2) A forward calculation of the four transmitter
6 positions, which is run in parallel (the number of parallel computations equals the number of
7 transmitter positions). Each forward calculation takes approximately 12.5 s on a standard desktop
8 computer with an Intel Core i7 processor. Thus, the total computation time of 300000 iterations
9 becomes approximately 45 days. As a comparison, the adjoint-based method needs approximately
10 20 iterations in order to converge to a solution estimate (Ernst et al., 2007a), which involves 20
11 FDTD forward calculations (for comparison with observed data), 20 backward (in time)
12 calculations (for update direction), and 20 forward calculations (for step length) (Ernst et al.,
13 2007b). Hence, obtaining one estimate using the adjoint-based method is approximately 5000 times
14 faster than compared to the time needed to obtain 45 independent realizations from the *a posteriori*
15 probability density using the Monte Carlo method. However, the adjoint-based method and the
16 Monte Carlo strategy are not directly comparable: The adjoint method is an optimization method
17 that searches for one (in some sense) optimal solution. Additionally, the adjoint method assumes the
18 full waveform inverse problem to be a global optimization problem and, therefore, the method is
19 always in risk of getting trapped in a local minimum. The Monte Carlo strategy, on the other hand,
20 aims at characterizing the *a posteriori* probability density. Moreover, it can, mathematically, be
21 shown that the Metropolis algorithm will converge towards the correct equilibrium distribution
22 independent of the complexity of the data uncertainty and *a priori* model (e.g. Mosegaard, 1998).
23 Roughly speaking, the adjoint-based method serves as a fast way of obtaining an approximate
24 maximum *a posteriori* estimate based on a limited data uncertainty and *a priori* model. The strategy

1 suggested here could subsequently be started in this estimate in order to obtain a characterization of
2 the correct *a posteriori* probability density (i.e. resolution analysis) based on a realistic data
3 uncertainty and *a priori* model and the full non-linearity of the forward relation. The significant
4 time expense of the Monte Carlo strategy may in the future be mitigated through parallelization of
5 the individual FDTD calculation on e.g. clusters or graphical processing units (GPU) and by
6 running several algorithms in parallel.

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

22 In field experiments the list of sources of uncertainties involves both a background
23 noise component and (typically) a major component related to modeling errors (i.e. discrepancies
24 between observed data and simulated data due to shortcomings of the forward relation) such as: 1)

1 Source pulse uncertainty: 2) 2D assumption: 3) Local antennae coupling effects: 4) Effects of
2 discretization: 5) Small-scale near borehole heterogeneities: 6) Numerical attenuation: 7) Media
3 dispersion: 8) Point-dipole antenna assumption: 9) Effects of high angle travel paths due to wave
4 guiding and antennae spread: 10) Unknown antennae positions.

5 Some of the above mentioned modeling inadequacies can be accounted for though
6 more sophisticated FDTD algorithms (see e.g. Bergmann et al., 1998; Ernst et al., 2006; Holliger
7 and Bergmann, 2002; Irving and Knight, 2006). Alternatively, these modeling uncertainties should
8 be accounted for through the data uncertainty model (although it results in higher *a posteriori*
9 model variability). Forward modeling inadequacies may lead to a combination of correlated and
10 uncorrelated data uncertainties that may be accounted for through a data covariance matrix (e.g.
11 Cordua et al., 2008, 2009) or a more complex description. See e.g. Peterson (2001), Cordua et al.
12 (2008, 2009) and Irving and Knight (2005) for descriptions and quantifications of some of these
13 uncertainties.

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

22 Ernst et al. (2007a) suggested a method based on deconvolution for source pulse
23 estimation, which may also be applied together with our inversion strategy. Then uncertainty related
24 to this process needs to be quantified or a Bayesian formulation of the source pulse determination

1 could be an integral part of the inversion (see Buland and Omre (2003) for an example of Bayesian
2 wavelet estimation).

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

12 While the use of training images to describe *a priori* information allows complex *a*
13 *priori* information to be quantified, the establishment of *a priori* information through a training
14 image for field data experiments may be challenging. The training image represents a concept of the
15 patterns expected *a priori* independent of the observed data. These expectations may be based on
16 previous studies in the area based on e.g. kriging and inversion conditional to soft (e.g. seismic) and
17 hard (borehole) data and/or studies of outcrops (e.g. Zhang, 2008). In order to compromise with the
18 distinct facies (i.e low entropy) seen in the training image applied here, the *a priori* model could
19 constitute a combination of information from a training image and a high entropy (e.g. two-point
20 statistical) based *a priori* model. In any case, we suggest that the movie strategy should be used in
21 order to ensure that the algorithm or mathematical expression (or a combination of both) that
22 describes the *a priori* information also reflects the *a priori* expectations of the user. See Zhang
23 (2008) for a description of how to convert geological information into training images that can be
24 used with multiple-point statistical simulation algorithms. Journel and Zhang (2006) demonstrated

1 that even for a training image with combined high and low entropy information that only
2 approximately captures the complex low entropy structures of the true model, a better conditional
3 modeling result was obtained than for a pure high entropy Gaussian *a priori* model.

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

11 CONCLUSION

12 We have outlined the theoretical background for a Monte Carlo based full waveform
13 inversion strategy based on the extended Metropolis algorithm in conjunction with complex
14 geostatistical based *a priori* information. The use of geostatistical algorithms for the description of
15 *a priori* information can be accomplished in an efficient way through the method of sequential
16 Gibbs sampling, which allows for inclusion of *a priori* information described by any geostatistical
17 algorithm based on sequential simulation. This, in turn, provides a means of using *a priori*
18 information described by both 2-point and multiple-point statistical *a priori* models. Inclusion of
19 such statistical *a priori* information reduces the complexity of the inverse problem, which is a
20 keystone in the feasibility of performing Monte Carlo sampling of the, computationally hard, full
21 waveform inverse problem. We have demonstrated the potential of this inversion strategy by
22 sampling the *a posteriori* probability density of a tomographic full waveform inverse problem using
23 complex *a priori* information inferred from a training image using the geostatistical algorithm
24 Snesim. The methodology provides a means of evaluating the *a posteriori* uncertainty, which is not

1 provided using traditional adjoint-based optimization strategies for full waveform inversion.
2 However, it should be noted that, if the goal is a single inverse estimate based on pure Gaussian
3 statistics, the adjoint-based optimization approach for inversion of full waveform data is
4 computationally considerably faster than the suggested inversion strategy. Establishment of
5 adequate (e.g. a multiple-point based) *a priori* information and a data uncertainty model for field
6 data are critical in order to obtain meaningful *a posteriori* uncertainty estimates. Moreover, the
7 major computational expenses of the Monte Carlo strategy have to be mitigated, which are all
8 challenging tasks that need future research.

10 ACKNOWLEDGEMENTS

11 We thank DONG E&P, Denmark, for financial support. We wish to thank James Irving and two
12 anonymous reviewers who provided thoughtful and valuable comments that helped improve the
13 manuscript.

15 REFERENCES

- 16 Belina, F. A., J. R. Ernst, and K. Holliger, 2009, Inversion of crosshole seismic data in
17 heterogeneous environments: Comparison of waveform and ray-based approaches: Journal of
18 Applied Geophysics, **68**, 85 – 94.
- 19
- 20 Bergmann, T., J. O. A. Robertsson, and K. Holliger, Finite-difference modeling of electromagnetic
21 wave propagation in dispersive and attenuating media: Geophysics, **63**, 856 – 867.
- 22
- 23 Buland, A. and H. Omre, 2003, Bayesian wavelet estimation from seismic and well data:
24 Geophysics, **68**, 2000 – 2009.

1

2 Cordua, K. S., T. M. Hansen, and K. Mosegaard, 2010, Monte Carlo full waveform inversion of
3 tomographic crosshole data using complex geostatistical a priori information, SEG, Expanded
4 Abstracts, 29, 4291 – 4296.

5

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

8

9 Cordua, K. S., L. Nielsen, M. C. Looms, T. M. Hansen, and A. Binley, 2009, Quantifying the
10 influence of static-like errors in least-squares-based inversion and sequential simulation of cross-
11 borehole ground penetrating radar data: Journal of Applied Geophysics, **68**, 71 – 84.

12

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

15

16 Davis, J. L., and A. P. Annan, 1989, Ground-penetrating radar for high-resolution mapping of soil
17 and rock stratigraphy: Geophysical Prospecting, **37**, 531 – 551.

18

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

21

22 Ernst, J. R., A. G. Green, H. Maurer, and K. Holliger, 2007a, Application of a new 2D time-domain
23 full-waveform inversion scheme to crosshole radar data: Geophysics, **72**, J53 – J64.

24

1 Ernst, J. R., K. Holliger, H. R. Maurer, and A. G. Green, 2006, Realistic FDTD modelling of
2 borehole georadar antenna radiation: Methodology and application: *Near Surface Geophysics*, 19 –
3 31.

4

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

8

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

11

12 Gelman, A., G. O. Roberts, and W. R. Gilks, 1996, Efficient Metropolis jumping rules: *Bayesian*
13 *statistics*, **5**, 599 – 607.

14

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

17

18 Gomez-Hernandez, J. and A. Journel, 1993, Joint sequential simulation of multi-Gaussian fields,
19 *Geostatistics Troia*, **92**, 85 – 94.

20

21 Goovaerts, P., 1997, *Geostatistics for natural resources evaluation*. Oxford Univ. Press, Oxford,
22 UK.

23

1 Guardiano, F., and R. Srivastava, 1993, Multivariate geostatistics: Beyond bivariate moments: *in*
2 A. Soares ed., *Geostatistics Tróia 1992, Proceedings of the 4th International Geostatistics Congress:*
3 *Kluwer Academic Publishers*, 133 – 144.

4

5 Hansen, T. M., A. G. Journel, A. Tarantola, and K. Mosegaard, 2006, Linear inverse Gaussian
6 theory and Geostatistics: *Geophysics*, **71**, R101 - R111.

7

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

11

12 Hansen, T. M., K. Mosegaard, and K. S. Cordua, 2009, Reducing complexity of inverse problems
13 using geostatistical priors, *IAMG 2009*, August 23 - 28, Stanford, Ca.

14

15 Hansen, T. M., K. Mosegaard, and K. S. Cordua, accepted, Inverse problems with non-trivial priors:
16 Efficient solution through Sequential Gibbs Sampling: *Computational Geosciences*, xx, xxx – xxx.

17

18 Hicks, G. J. and R. G. Pratt, 2001, Reflection waveform inversion using local descent methods:
19 Estimating attenuation and velocity over a gas–sand deposit: *Geophysics*, **66**, 598–612.

20

21 Holliger, K. and T. Bergmann, 2002, Numerical modeling of borehole georadar data: *Geophysics*,
22 **67**, 1249 – 1257.

23

1 Irving, J. D. and R. J. Knight, 2005, Effect of antennas on velocity estimates obtained from
2 crosshole GPR data: *Geophysics*, **70**, K39 – K42.

3

4 Irving, J. D. and R. J. Knight, 2006, Numerical simulation of antenna transmission and reception for
5 crosshole ground-penetrating radar: *Geophysics*, **71**, K37 – K45.

6

7 Irving, J. D. and K. Singha, 2010, Stochastic inversion of tracer test and electrical geophysical data
8 to estimate hydraulic conductivities: *Water resources research*, 46, W11514.

9

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

11

12 Journel, A. and T. Zhang, 2006, The Necessity of a Multiple-Point Prior Model: *Mathematical*
13 *Geology*, **38**, 591 – 610.

14

15 Kitanidis, P. K., 1995, Quasi-linear geostatistical theory for inversing: *Water resources research*, **31**,
16 2411 – 2419.

17

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

21

22 Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and
23 resolution analysis of seismic background velocities: *Journal of Geophysical Research*, **96**, 20289 –
24 20299.

1

2 Looms, M.C., K. H. Jensen, A. Binley, and L. Nielsen, 2008, Monitoring unsaturated flow and
3 transport using cross-borehole geophysical methods: *Vadose Zone journal*, **7**, 227 – 237.

4

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

9

10 Meles, G., S. Greenhalgh, J. van der Kruk, A. Green, and H. Maurer, 2011, Taming the non-
11 linearity problem in GPR full-waveform inversion for high contrast media: *Journal of Applied
12 Geophysics*, **73**, 174 – 186.

13

14 Mora, P., 1987, Nonlinear two-dimensional elastic inversion of multioffset seismic data:
15 *Geophysics*, **52**, 1211 – 1228.

16

17 Mosegaard, K., 1998, Resolution analysis of general inverse problems through inverse Monte Carlo
18 sampling: *Inverse Problems*, **14**, 405 – 426.

19

20 Mosegaard, K., 2011, Quest for consistency, symmetry, and simplicity – The legacy of Albert
21 Tarantola: *Geophysics*, **76**, W51 – W61.

22

23 Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems:
24 *Journal of Geophysical Research*, **100**, 431 – 447.

1
2 Nielsen, L., M. C. Looms, T. M. Hansen, K. S. Cordua, and L. Stemmerik, 2010, Estimation of
3 chalk heterogeneity from stochastic modelling conditioned by crosshole GPR travel times and LOG
4 data: in *Advances in Near-Surface Seismology and Ground-Penetrating Radar*, eds. Miller, R.,
5 Bradford, J., and Holliger, K., Society of Exploration Geophysicists, Tulsa, Oklahoma.
6
7 Peterson, J.E., 2001, Pre-inversion correction and analysis of radar tomographic data: *Journal of*
8 *Environmental and Engineering Geophysics*, **6**, 1 – 18.
9
10 Pica, A., J. P. Diet, and A. Tarantola, 1990, Nonlinear inversion of seismic reflection data in a
11 laterally invariant medium: *Geophysics*, **55**, 284 – 292.
12
13 Pratt, R. G. and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole
14 tomography. Part 1: Acoustic wave-equation method: *Geophysical Prospecting*, **38**, 287-310.
15
16 Remy N., A. Boucher A, and J. Wu, 2008, *Applied Geostatistics with SGeMS: A User's Guide*
17 Cambridge University Press.
18
19 Reynolds, J.M., 1997, *An introduction to applied and environmental geophysics*: John Wiley and
20 Sons Ltd. Pp.
21
22 Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point
23 statistics: *Mathematical Geology*, **34**, 1 – 21.
24

1 Talagrand, O., and P. Courtier, 1987, Variational assimilation of meteorological observations with
2 the adjoint vorticity equation. I. Theory: Quarterly journal of the Royal Meteorological Society,
3 **113**, 1311-1328.

4
5 Tarantola, A., and B. Valette, 1982, Inverse Problems = Quest for Information: Journal of
6 Geophysics, **50**, 159 – 170.

7
8 Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation:
9 Geophysics, **49**, 1259 – 1266.

10
11 Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection
12 data: Geophysics, **51**, 1893 – 1903.

13
14 Tarantola, A., 1988, Theoretical Background for the Inversion of Seismic Waveforms, Including
15 Elasticity and Attenuation: Pure and Applied Geophysics **128**, 1/2, 365 - 399.

16
17 Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society
18 of Industrial and Applied Mathematics, Philadelphia, PA.

19
20 Topp, G. C., J. L. Davis, and A. P. Annan, 1980, Electromagnetic determination of soil water
21 content: Measurements coaxial transmission lines: Water Resources Research, **16**, 574–582.

22

1 Tronicke, J. and K. Holliger, 2005, Quantitative integration of hydrogeophysical data: Conditional
2 geostatistical simulation for characterizing heterogeneous alluvial aquifers: *Geophysics*, **70**, H1 –
3 H10.

4
5 Zhang, T., 2008, Incorporating Geological Conceptual Models and Interpretations into Reservoir
6 Modeling Using Multiple-Point Geostatistics: *Earth Science Frontiers*, **15**, 26 – 35.

7

8 **FIGURE CAPTIONS**

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

12

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

15

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

19

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

23

1 Figure 5. Evolution of the adaptive exploration step-size during the first 1000 iterations of the burn-
2 in period.
3
4 Figure 6. Development of the model during the first 1000 iterations. In this period the relatively
5 large scale structures of the model are established. Compare with the reference model in Figure 2.
6 Iteration no. 1 (It. no. 1) is the starting model, which is an unconditional realization of the multiple-
7 point based *a priori* model.
8
9 Figure 7. Autocorrelation analysis of the *a posteriori* sample in order to determine the number of
10 iterations needed to obtain statistically independent realizations.
11
12 Figure 8. Eight statistically independent realizations of the *a priori* probability density (*a priori*
13 movie).
14
15 Figure 9. Eight statistically independent realizations of the *a posteriori* probability density (*a*
16 *posteriori* movie).
17
18 Figure 10. The mean and variance of the model parameters calculated on the basis of the *a*
19 *posteriori* sample.
20
21 Figure 11. a) *A posteriori* data variability. Blue curves: Simulated waveforms calculated for 45
22 statistically independent *a posteriori* models. Red curve: The observed waveform data. b) Data
23 residuals. Blue curves: The data residuals for 45 statistically independent *a posteriori* models. Red

- 1 curve: The uncertainty component of the observed data. The waveforms are associated with the
- 2 transmitter-receiver pairs indicated by the numbers 1 – 5 in Figure 3.

Figure 1 ↑

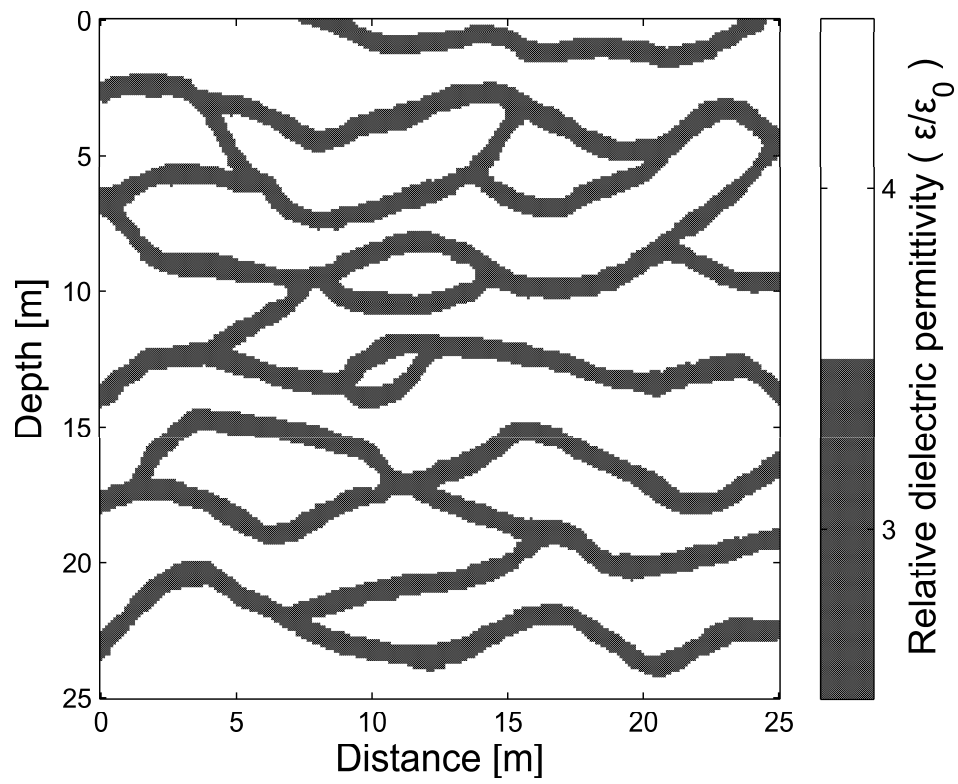


Figure 2 ↑

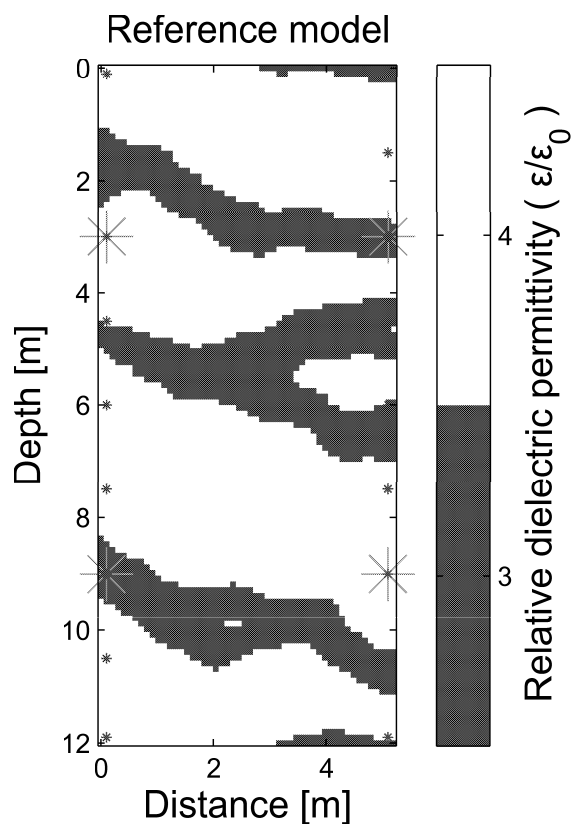


Figure 3 ↑

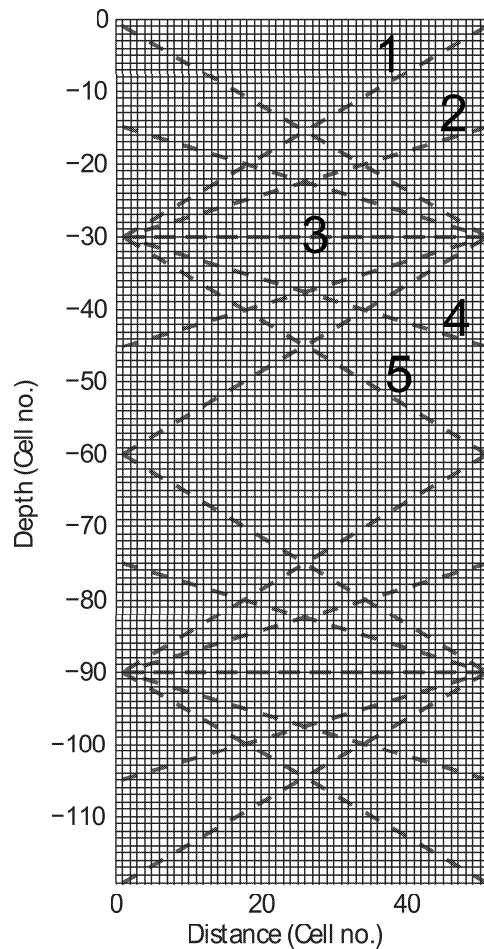


Figure 4 ↑

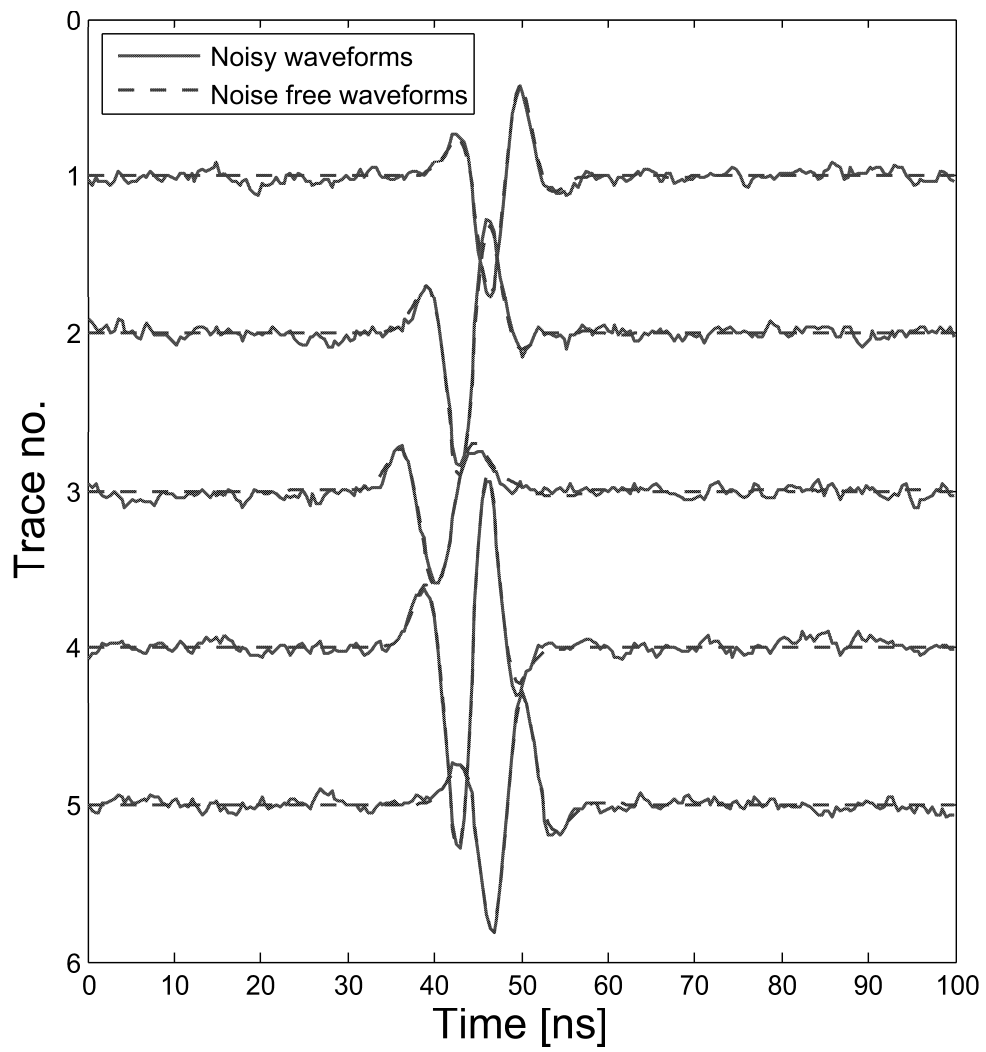


Figure 5 ↑

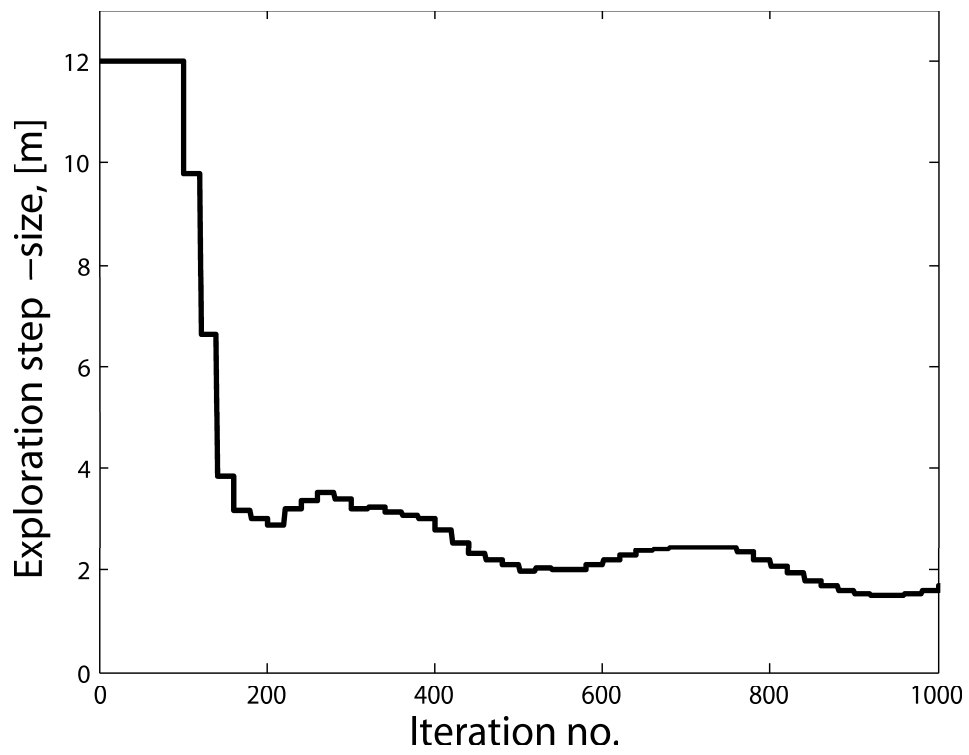


Figure 6 ↑

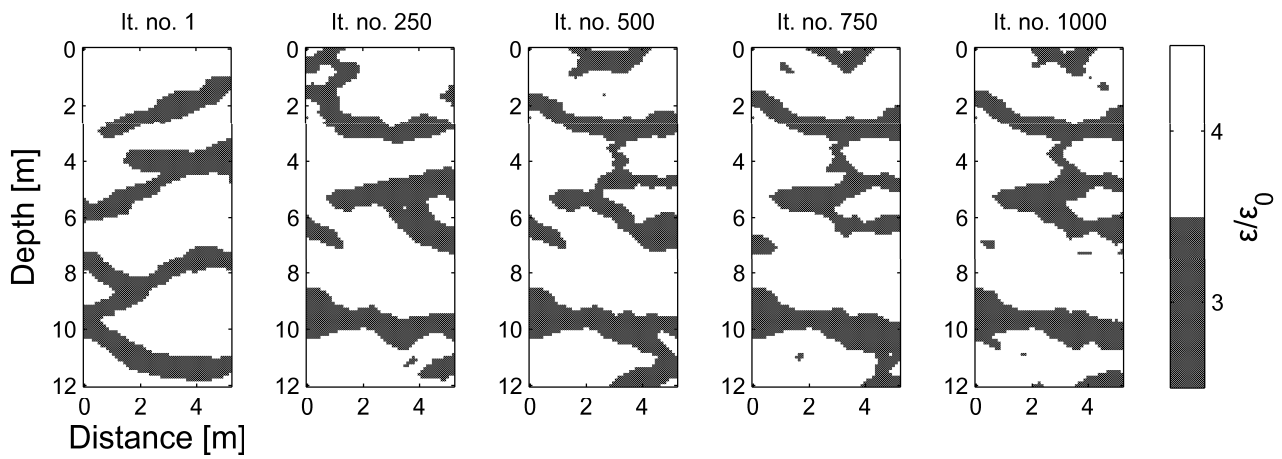


Figure 7 ↑

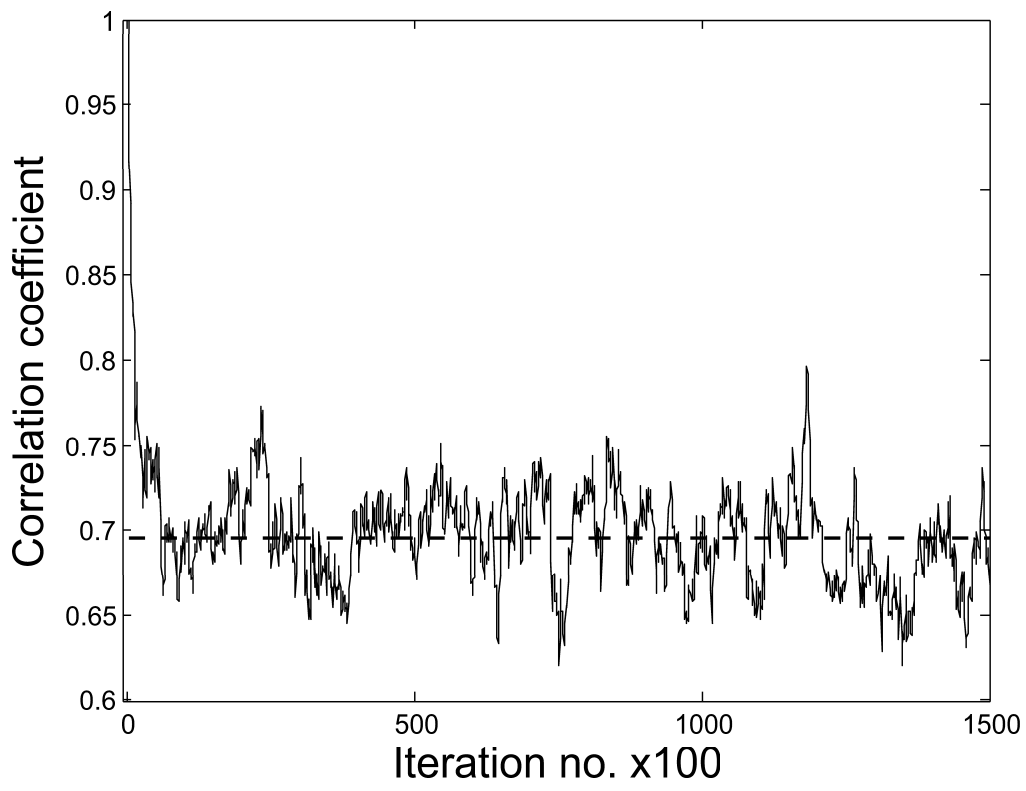


Figure 8 ↑

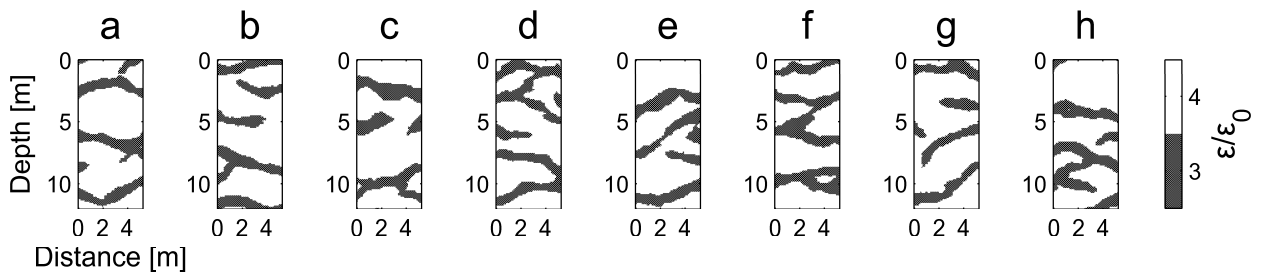


Figure 9 ↑

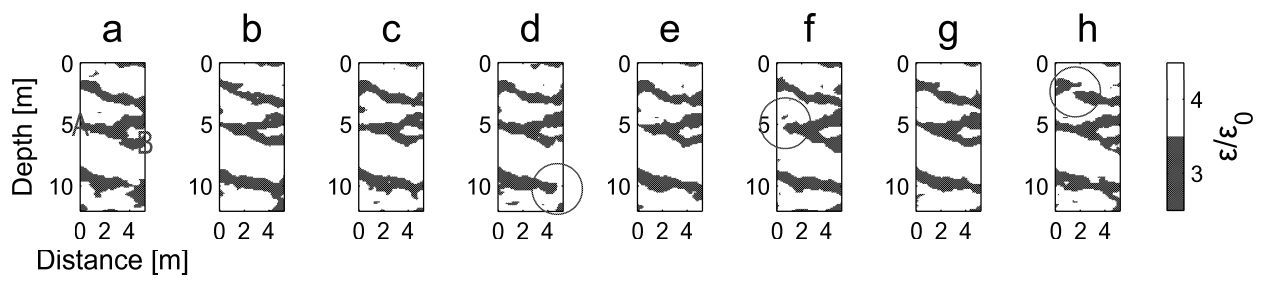


Figure 10 ↑

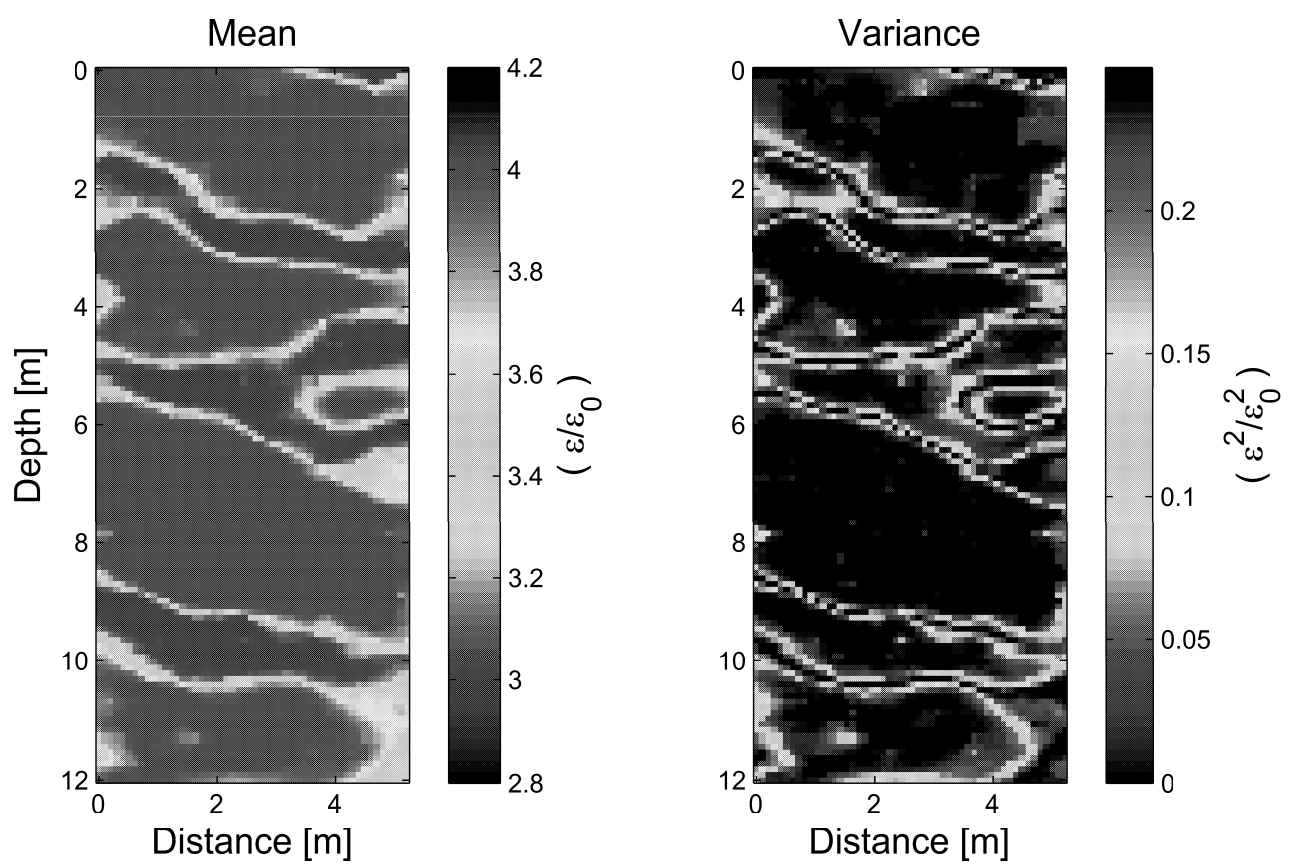


Figure 11 ↑

