Inference of multi-Gaussian relative permittivity fields by probabilistic inversion of crosshole ground-penetrating radar data

Jürg Hunziker[1], **Eric Laloy**[2], **Niklas Linde**[1]

[1] *Applied and Environmental Geophysics Group, Institute of Earth Sciences, University of Lausanne, Lausanne, Switzerland*

[2] *Belgian Nuclear Research Center, Mol, Belgium*

E-mail: jurg.hunziker@unil.ch, eric.laloy@sckcen.be, niklas.linde@unil.ch

(October 20, 2017)

Running head: **Probabilistic inversion of crosshole GPR data**

ABSTRACT

In contrast to deterministic inversion, probabilistic Bayesian inversion provides an ensemble of solutions that can be used to quantify model uncertainty. In this work, we present a probabilistic inversion approach that uses crosshole first-arrival traveltimes to estimate an underlying geostatistical model, the subsurface structure and the standard deviation of the data error simultaneously. The subsurface is assumed to be represented by a multi-Gaussian field, which allows us to reduce the dimensionality of the problem significantly. Compared to previous applications in hydrogeology, novelties of this study include an improvement of the dimensionality reduction algorithm to avoid streaking artifacts, it is the first application to geophysics and the first application to field data. The results of a synthetic example show that the model domain enclosed by one borehole pair is generally too small to provide reliable estimates of geostatistical variables. A real-data example based on two borehole

pairs confirms these findings and demonstrates that the inversion procedure also works under realistic conditions with, for example, unknown measurement errors.

INTRODUCTION

Crosshole tomography of ground-penetrating radar (GPR) data is often deterministic using, for example, the first-arrivals only (e.g., Hubbard et al., 2001) or the full wavefield (e.g., Ernst et al., 2007; Klotzsche et al., 2013). Similar algorithms are available for seismic crosshole tomography (e.g., Bregman et al., 1989; Thompson Reiter and Rodi, 1996; Zhou et al., 2008). Deterministic inversion algorithms have the advantage that they find an optimal solution after a few to some dozens of iterations and are, therefore, comparatively fast. A drawback of these algorithms is that they only find one solution. This solution is strongly dependent on the regularization operator used and standard uncertainty estimates based on this solution are only valid for a linearized theory. In contrast, probabilistic inversion algorithms using global sampling methods require many more steps, but provide a set of solutions, drawn from the so-called posterior distribution. Probabilistic inversion algorithms require as input a prior distribution that encompasses the prior probability of all parameter combinations from which possible solutions can be drawn.

Different probabilistic inversion algorithms have been proposed. Below we discuss a few representative examples in the context of geostatistical prior models and crosshole ground-penetrating radar (GPR). The linear stochastic approach to geophysical inverse problems was first proposed by Asli et al. (2000) and then applied to stochastic GPR tomography by Gloaguen et al. (2005). Hansen et al. (2006) used linear theory and an assumed underlying geostatistical model. Cordua et al. (2012) used a training image representing the expected geological structure to find the posterior distribution of relative permittivity (also referred to as dielectric constant). A hybrid approach of deterministic and probabilistic inversion (Johnson et al., 2007, 2012) aims at finding a subsurface model explaining the

data while honoring spatial statistics given by a variogram. Linde and Vrugt (2013) proposed a probabilistic inversion algorithm using a discrete cosine transform to reduce the dimensionality of the inverse problem. Stochastic inversion algorithms can also be used to integrate different types of data. For example, simulated annealing was used to integrate inverted GPR crosshole traveltimes, borehole-logs and a geostatistical model (Tronicke and Holliger, 2005; Dafflon et al., 2009a,b). Stochastic data integration using Bayes rule was also done by Dubreuil-Boisclair et al. (2011). A 3D timelapse study for mapping a water plume was published by Laloy et al. (2012). Hansen et al. (2013a,b) presented and demonstrated an algorithm that is able to estimate the geostatistical model along with the subsurface structure. A common issue in probabilistic inversion is how to choose an appropriate prior distribution. To address this, Hansen et al. (2008) proposed a mechanism, which was also used by Looms et al. (2010), to determine how consistent the prior distribution is with respect to the posterior distribution.

In this study, we rely on a probabilistic inversion approach that uses Markov chain Monte Carlo (MCMC) to infer an underlying geostatistical model along with the subsurface structure that is represented as a multi-Gaussian field and the standard deviation of the data error of crosshole GPR first-arrival traveltime data. We build on the work by Laloy et al. (2015) that combined efficient geostatistical resimulation by circulant embedding and dimensionality reduction to infer hydraulic conductivity fields from synthethic tracer test data. Four novelties are introduced with respect to Laloy et al. (2015): (1) we consider geophysical data; (2) the method is applied for the first time to field data; (3) methodological adaptations are made to avoid streaking artifacts without significant extensions of the model domain; (4) a simple approach to account for model errors is introduced to avoid over-fitting and inversion artifacts when considering field data.

METHOD

Our inversion procedure builds on the method by Laloy et al. (2015), which combines a geostatistical model parameterization with model reduction using circulant embedding and the DREAM$_{(ZS)}$ algorithm (Laloy and Vrugt, 2012; ter Braak and Vrugt, 2008). DREAM$_{(ZS)}$ is a Markov chain Monte Carlo (MCMC) type algorithm that makes use of differential evolution updating from an archive of past states for faster convergence. In recent years, this method and related variants have been used widely for moderately (100s of parameters) high-dimensional geophysical inverse problems (e.g., Linde and Vrugt, 2013; Lochbühler et al., 2014; Rosas-Carbajal et al., 2013, 2015). A more complete list of applications is provided by Vrugt (2016).

Despite the efficiency of this MCMC algorithm, estimating the posterior distribution of the relative permittivity for each individual pixel (5000 in our synthetic examples) would take an unfeasibly long time with currently available hardware and software. To reduce the parameter dimensionality, we assume herein that the relative permittivity field can be described as a multi-Gaussian field. This assumption allows us to compress the image representing the medium using a slightly altered version of the circulant-embedding algorithm described by Laloy et al. (2015). In our case, 5000 unknowns corresponding to pixels with a size of 0.1 times 0.1 m are reduced to 257 unknowns (250 dimension-reduction variables and seven variables describing the statistical variability of the subsurface and data errors).

These seven variables are: (1) the standard deviation of the data error, (2) the mean relative permittivity, (3) the standard deviation of the relative permittivity, (4) the integral scale along the major axis of anisotropy, (5) the anisotropy angle (rotation anticlockwise from the vertical axis), (6) the ratio of the integral scale along the minor axis of anisotropy

to the integral scale along the major axis of anisotropy and (7) the shape parameter ν in the Matérn function (Matérn, 1960). This last parameter ν determines the shape of the variogram (e.g., $\nu = 0.5$: exponential model, $\nu = 1$: Whittle model and $\nu \to \infty$: Gaussian model).

Likelihood function, model proposal and Metropolis ratio

The inversion uses the following likelihood function to determine how well a model $\mathbf{m}$ explains the measured data $\mathbf{d}$:

$$L(\mathbf{m}|\mathbf{d}) = \frac{1}{(2\pi)^{N/2}\sigma_e^N} \exp\left(-\frac{1}{2\sigma_e^2}\sum_{j=1}^{N}(G_j(\mathbf{m}) - d_j)^2\right), \tag{1}$$

where the operator G creates simulated data based on the model $\mathbf{m}$, N is the number of datapoints and σ_e is the standard deviation of the data error. We estimate the latter using hierarchical Bayes (Malinverno and Briggs, 2004), which implies that it is inferred in the inversion. Equation 1 is valid for residuals that are independent and identically distributed following a Gaussian distribution with zero mean (i.e., no bias).

In DREAM$_{(\mathrm{ZS})}$, a new model proposal $\mathbf{z}$ of length l is generated from the current model $\mathbf{m}_{t-1}$ by differential evolution (Laloy and Vrugt, 2012; Vrugt, 2016) using δ pairs of model realizations from the archive $\mathbf{X}$:

$$\mathbf{z} = \mathbf{m}_{t-1} + (\mathbf{1}_l + \mathbf{e}_l)\gamma(\delta, l)\sum_{j=1}^{\delta}\left(\mathbf{X}^{r_1(j)} - \mathbf{X}^{r_2(j)}\right) + \boldsymbol{\epsilon}_l, \tag{2}$$

with r_1 and r_2 being two different rows of the archive of previous models. The archive contains models from all chains. The step-length of the update is given by γ. The two parameters $\mathbf{e}_l$ and $\boldsymbol{\epsilon}_l$ are drawn from a uniform and a normal distribution with a mean of zero, respectively. Given that the model proposal in equation 2 is symmetric, the Metropolis

acceptance probability α can be used:

$$\alpha = \min\left\{1, \frac{L(\mathbf{z})}{L(\mathbf{m}_{t-1})}\frac{\rho(\mathbf{z})}{\rho(\mathbf{m}_{t-1})}\right\}. \tag{3}$$

The prior distribution is given by ρ and L is the likelihood function as defined in equation 1. If $\mathbf{z}$ is accepted, $\mathbf{m}_t = \mathbf{z}$. Otherwise, the new model remains the same as the previous model. Thus $\mathbf{m}_t = \mathbf{m}_{t-1}$.

The circulant-embedding algorithm

Circulant embedding allows for an efficient parameter dimension reduction (for details see Dietrich and Newsam, 1997; Laloy et al., 2015). The first step consists of computing the covariance matrix $\mathbf{S}$ of size $(2m-1)\times(2n-1)$ or larger, where m and n specify the number of pixels in the two directions of the 2D subsurface model. $\mathbf{S}$ is a covariance kernel obtained from the Matérn function (Matérn, 1960), which describes the geostatistical properties of the medium using the following five variables: standard deviation of the relative permittivity, the integral scale, the anisotropy angle, the anisotropy ratio and the shape parameter ν in the Matérn function. If a larger size for $\mathbf{S}$ is used, then the following expressions need to be adapted accordingly. The eigenvalues $\mathbf{\Omega}$ of the covariance matrix $\mathbf{S}$ are obtained via a Fourier transformation:

$$\mathbf{\Omega} = \frac{\text{real}[\text{FFT2}(\text{FFTSHIFT}\{\mathbf{S}\})]}{(2m-1)\times(2n-1)}, \tag{4}$$

where FFT2 performs a two-dimensional fast Fourier Transform and FFTSHIFT swaps opposite quadrants of a matrix.

In addition, a complex Gaussian matrix $\mathbf{Z} = \mathbf{Z}_1 + i\mathbf{Z}_2$ is constructed, where $\mathbf{Z}_1$ and $\mathbf{Z}_2$ have a standard Gaussian distribution and $i = \sqrt{-1}$ is the imaginary unit. We obtain $\mathbf{Z}_1$ and $\mathbf{Z}_2$ by linearly interpolating (in the frequency domain) the vector containing the

aforementioned dimension-reduction variables onto a vector of length $(2m-1) \times (2n-1)$, followed by a random permutation to avoid short lag autocorrelations. This random permutation scheme is kept fixed during the inversion. Finally, $\mathbf{Z}$ needs to be reshaped to have the same dimension as the matrix $\mathbf{\Omega}$.

As a last step, the square root of the eigenvalues $\mathbf{\Omega}$ are component-wise multiplied ($\odot$) with the complex Gaussian matrix $\mathbf{Z}$ and Fourier transformed:

$$\mathbf{f} = \mathrm{FFT2}\left(\sqrt{\mathbf{\Omega}} \odot \mathbf{Z}\right) \tag{5}$$

The real and the imaginary part of $\mathbf{f}$ are two independent zero-mean multi-Gaussian realizations of size $(2m-1) \times (2n-1)$. To obtain a possible representation of the subsurface, we extract a matrix of the size $m \times n$ from the real part of $\mathbf{f}$ and add the mean of the relative permittivity to $\mathbf{f}$.

The integral scale influences the average size of isolated structures that a multi-Gaussian model realization consists of. A larger integral scale results in larger patches. The deviation of these patches from circles is a consequence of geostatistical anisotropy. The minimum allowable size of the circulant-embedding covariance matrix $\mathbf{S}$ is proportional to the size of the integral scale (Dietrich and Newsam, 1997). If the size of the circulant-embedding covariance matrix is too small for a certain integral scale, streaking artifacts appear in the resulting model realization (Figure 1a), which negatively affect the inversion results. Such streaking artifacts related to similar algorithms have been described by various authors (e.g., Ripley, 1987; Chilès and Delfiner, 1999).

The origin of the streaking artifacts lies in the computation of the matrix of eigenvalues $\mathbf{\Omega}$ from the covariance matrix $\mathbf{S}$ (equation 4). If the integral scale increases, then the portion of the covariance matrix with significant values increases as well. When the area of

significant values in the covariance matrix theoretically exceeds the size of the actual matrix (i.e., we have significant values at the edge of the matrix), artifacts are introduced during the FFT2, which eventually lead to the streaking in the subsurface model realizations. Dietrich and Newsam (1997) investigated this effect. They showed for a Gaussian variogram model that the size of the covariance matrix has to be more than five times larger than the integral scale. For a Whittle model this factor is even larger, while it is smaller for an exponential model.

The simplest way of avoiding streaking artifacts (Figure 1a) is to compute the covariance matrix on a larger domain to prevent truncation of significant values in the covariance matrix. Unfortunately, a larger covariance matrix $\mathbf{S}$ leads to a larger model domain $\mathbf{f}$, from which we need to extract a model of the desired size. This implies that the same number of dimension-reduction variables describes a larger model. In other words, the compression factor increases, potentially removing details of the subsurface model. Therefore, this is not a desired procedure.

Another option to avoid this streaking is to apply a taper to the covariance matrix before the Fourier Transform is computed, thereby forcing the values of the covariance matrix to approach zero towards the edge of the matrix. This taper avoids streaking as is shown in Figure 1b, but it also alters the resulting property fields (e.g., by decreasing connectivity between the high permittivity zones).

A third option is introduced and pursued in this paper. In this approach, the covariance matrix is computed on a larger domain such that the area of large values in the covariance matrix is completely captured and the eigenvalues $\mathbf{\Omega}$ (see equation 4) can be reliably computed by FFT2. A larger domain results after the FFT2 in a more densely sampled Fourier

domain. This domain is downsampled before equation 5 is applied, such that the matrix of eigenvalues has the same number of elements as the desired model domain. Because of this, the dimension-reduction variables do not need to encode a larger model domain. Thus, the compression is optimal and no information contained in the covariance matrix is altered or deleted. This method produces a multi-Gaussian field with improved continuity (Figure 1c) with respect to the one created with a taper. In this paper, we use a downsampling rate of 4 in the horizontal direction. No covariance matrix extension and downsampling is applied in the vertical direction, because we consider test cases with strong horizontal anisotropy.

To quantify the effect of the streaking artifacts and to quantify how well the taper and the domain extension with subsequent downsampling work, we have computed 1000 subsurface multi-Gaussian realizations based on the same geostatistical model (1) without dimensionality reduction, (2) with dimensionality reduction but without treatment of the artifacts, (3) with dimensionality reduction and a tapered covariance matrix and (4) with dimensionality reduction with domain extension and using subsequent downsampling. The mean variograms of these models and their standard deviation are shown in Figure 2. If the artifacts are left untreated, the mean variogram of the models with dimensionality reduction differs at all offsets from the mean variogram of the models created without dimensionality reduction (Figure 2a). The corresponding relative error of the mean variogram is very large at small offsets (blue curve in Figure 2d). Tapering the covariance matrix improves the situation at small offsets, but creates larger differences between the two mean variograms at larger offsets (Figure 2b and red curve in Figure 2d). Using the method of domain extension and subsequent downsampling reduces the difference further at all offsets (Figure 2c and yellow curve in Figure 2d). Note, that with the domain extension and subsequent downsampling, the relative error is at all offsets below the relative error of the untreated

model realizations. Nevertheless, relative errors persist as a consequence of the dimensionality reduction (250 dimension-reduction variables and 6 additional variables to describe 5000 unknown pixel values).

SYNTHETIC EXAMPLE

For our initial synthetic example, we consider a "true" relative permittivity model (Figure 3) which is 10 m deep and 5 m wide. Two vertical boreholes are located 0.25 m from the vertical edges of the model domain, which implies a borehole separation of 4.5 m. 33 receivers are located in the left borehole between a depth of 0.25 m and 9.85 m at intervals of 0.3 m. The same number of sources is located at the same depth levels in the right borehole. For each shot position, receivers higher or lower than 50° from the horizontal are not considered as high-angle raypaths are often problematic in field data (Peterson, 2001). The 879 first-arrival traveltimes for the remaining source-receiver pairs are computed using the finite-difference code by Podvin and Lecomte (1991). This code, which is also used as a forward solver in the MCMC inversion, is able to compute the traveltimes in 2D accurately even if the model features large and arbitrarily shaped velocity contrasts. The obtained data are contaminated with uncorrelated Gaussian distributed noise with a standard deviation of 0.5 ns.

Using more chains in DREAM$_{(ZS)}$, compared to the default value of three, allows for a broader sampling of the solution space during the burn-in phase, in which the algorithm searches for the global minimum. However, distributing a fixed amount of forward simulations over more chains results evidently in shorter chains, limiting the exploration of the posterior distribution after the burn-in phase. After trying different options we found that 16 chains and a total of 2 million samples (125'000 samples per chain) was a good choice for

the given dataset. For efficiency, DREAM$_{(ZS)}$ works in parallel with as many processors as parallel chains. The prior distribution for the dimension reduction variables is a standard Gaussian distribution, while the prior distribution for the seven auxiliary variables is a uniform or a log-uniform distribution. Table 1 lists the inversion variables with their ranges, the values used to generate the "true" model, the prior distribution types and units.

To determine the end of the burn-in phase, we visually inspect the development of the weighted root mean square error for each chain (Figure 4a). When the weighted root mean square error starts to oscillate around 1, the burn-in phase is completed. This implies that the algorithm starts to explore and sample proportionally to the posterior distribution. In our case, the burn-in phase is completed after approximately 3×10^4 steps corresponding to 24% of the chain length. To ensure that the burn-in phase is not considered when estimating the posterior distribution, we use the last 75% samples of the chains and consider the first 25% of the samples of the chains as the burn-in phase. The acceptance rate is approximately 30% when the algorithm is exploring the posterior distribution.

To judge if convergence has also been reached formally (i.e., the posterior distribution is sufficiently sampled), we compute the potential scale reduction factor $\hat{R}$ (Gelman and Rubin, 1992) for the last 75% of the samples of each chain (i.e., the part of the chain after the burn-in phase). This factor compares the variance within a chain to the variance of all chains together. Formal convergence is commonly considered to be achieved when $\hat{R} < 1.2$ for each parameter. This criterion is fulfilled in this example.

Figures 4b to 4h depict the prior distributions of the seven variables describing the data error and the geostatistical variability of the model (dashed black line), the posterior distributions (blue bars) and the values chosen to generate the subsurface structure in

Figure 3 (vertical red line). As generating a subsurface structure is a statistical process and the resulting subsurface realizations have finite sizes, differences between the properties of specific geostatistical realizations and the underlying geostatistical model can arise (ergodic variations). This implies that the value indicated by the vertical red line is not necessarily the most appropriate value for this specific test model, but it is the true value describing the underlying geostatistical model (it is sampled on average if generating a sufficient number of finite model realizations). The most appropriate geostatistical parameters of our true subsurface model is nevertheless expected to be close to the value indicated by the vertical red line.

The inferred posterior distribution of the standard deviation of the data error is very narrow (Figure 4b), which suggests that it is well resolved. The peak of the distribution is located at 0.5 ns, which is the value used to generate the noise-contaminated synthetic data. The algorithm assumes that the noise is homoscedastic, uncorrelated and Gaussian distributed. As we generated the noise following these prerequisites, the assumptions are fulfilled and it is straightforward to estimate this parameter.

The mean of the relative permittivity is also well resolved (Figure 4c). This is not surprising, as an error in this parameter strongly influences the GPR velocities and, thus, the first-arrival traveltimes. In other words, small variations in this parameter lead immediately to large data residuals.

The remaining parameters are generally well retrieved, but with higher uncertainties. Although we invert for the variance of the relative permittivity, we plot the standard deviation (Figure 4d), because it is more straightforward to interpret. The true value is part of the posterior distribution. However, as the standard deviation is overestimated, the re-

sulting subsurface model realizations will tend to feature larger variability than in the test model. The estimate of this parameter would be improved if the true integral scale were reduced or, conversely, the distance between the boreholes were increased. This has been confirmed numerically (not shown herein).

For the integral scale (Figure 4e), we find that the peak of the posterior distribution is close to the true value. For the anisotropy angle (Figure 4f), the peak is just seven degrees from the true value. The ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy is overestimated (Figure 4g), making the retrieved subsurface structure more isotropic than it actually is. The shape parameter ν is underestimated, meaning that there is more small-range correlation in the sampled model realizations than in the "true" model. However, as the smallest source-receiver offset is 4.5 m, information about the variability of the medium at short distances is hardly present in the data. It is also likely that this under-estimation is due to the choice of the log-uniform prior that tends to favor smaller parameter values.

For each sampled solution, which is part of the MCMC chain, an image of the subsurface can be produced. To illustrate the variety of model realizations that the algorithm can visit during the inversion process, a selection of eight subsurface structures randomly drawn from the prior distribution is shown in Figure 5. The rather large variability is illustrated by comparing Figures 5f and 5g. While the former consists of patches of the size of a pixel (0.1 times 0.1 m), the latter features very large patches covering several meters. This illustrates the prior variability in the integral scale. Comparing Figures 5b and 5h helps visualizing the prior variability in anisotropy.

Figures 6a to 6d and 6f to 6i show subsurface structures corresponding to eight randomly

chosen posterior realizations after burn-in. The standard deviation and the mean of 128 posterior realizations that were sampled at regular intervals are shown in Figures 6e and 6j, respectively. The values of the variables describing the statistical variability corresponding to these eight model realizations are given in Table 2. Note that the large-scale structure is similar, but the degree of small-scale variability is highly varying. In other words, all solutions show the same main features, such as, the two low relative permittivity zones (blue) at a depth of 2.5 m and a horizontal offset of 3 m and at a depth of 6 m and a horizontal offset of less than 1 m. Also all solutions show a high relative permittivity trend ranging from the top left to the lower right. These features are also found in the true subsurface structure (Figure 3). There are also differences. For example, model realization 8 (Figure 6i) features multiple small patches, whereas model realization 1 (Figure 6a) consists of fewer and larger patches resulting in an overall smoother appearance. Consulting Table 2 reveals that the variable that differs the most significantly between the two solutions is the shape parameter ν. While the variogram of model realization 8 follows approximately a Whittle model, the one of model realization 1 has more of a Gaussian structure.

As the boreholes are 0.25 m from the edges of the models and the GPR signal samples only the region between the boreholes, there is no information in the GPR signal about the medium to the left side of the left borehole and to the right side of the right borehole. Therefore, the subsurface structure in these areas is only constrained by the geostatistical model. This results in a larger standard deviation in these regions (Figure 6e). Otherwise, the standard deviation is relatively low and seems to be uncorrelated with the subsurface structure. The mean of the 128 selected model realizations (Figure 6j) resembles the true subsurface (Figure 3), which underlines how well the inversion retrieves the main elements of the subsurface.

The inversion results presented so far are based on noise-contaminated data from one true multi-Gaussian model realization (Figure 3). Creating subsurface realizations is a stochastic process and the same set of values describing the underlying geostatistical model can lead to very different model realizations. The ability of the inversion to infer geostatistical properties might depend on the model realization that is treated as the true model. In other words, the finite size of the subsurface domain implies that the global statistics of the underlying geostatistical model is not necessarily reproduced at the scale of the borehole pair. More reliable estimates of these global statistics require, thus, the usage of several borehole pairs. To investigate the variability caused by this spatial sub-sampling, we considered MCMC inversions for ten different subsurface model realizations based on the same underlying geostatistical properties as those given in Table 1. The corresponding simulated forward responses were contaminated with different noise realizations, but from the same underlying distribution as before.

To further ensure that the posterior density is well-sampled, the data of each of the ten models were inverted three times. These three runs were combined into one posterior distribution, resulting in ten marginal posterior distributions given the ten test models considered. If the geostatistical inference is independent of the true subsurface model (e.g., the geostatistical realization in Figure 3 that is treated as the true model), all ten distributions would be similar. Figure 7 shows the ten posterior distributions of the statistical parameters.

In terms of the standard deviation of the data error (Figure 7a), as well as for the mean of the relative permittivity (Figure 7b), the inferred values for all the ten runs correspond relatively well with the true values. This implies, that these variables can be resolved independently of the subsurface structure. For the other five variables, the spread is consid-

erably larger, but the values used to generate the true subsurface model realizations are for all runs part of the posterior distribution. The spread of the posterior distributions of some of the statistical variables shows that the resolvability of a statistical variable is affected by the actual subsurface structure. For example, depending on the subsurface structure, the integral scale (Figure 7d) can be overestimated (solid red and dashed green distribution), more or less correctly estimated (solid blue or dashed red distribution) or underestimated (black, cyan, solid green and dashed blue distributions).

Averaging all ten distributions results in a posterior distribution that is visualized in Figure 7 with a thick solid black line. Except for the standard deviation of the relative permittivity and the integral scale, this distribution peaks very closely to the true geostatistical parameters. This suggests that the inversion method provides consistent and unbiased results. This procedure of randomizing the true model is not commonly done in geophysical MCMC applications, but it is strongly recommended when the necessary computing resources are available as it allows for more robust conclusions.

REAL-DATA EXAMPLE

We now consider field data from the South Oyster Bacterial Transport Study Site near Oyster, Virginia, USA (Hubbard et al., 2001; Scheibe et al., 2011). This site hosts an uncontaminated sandy Pleistocene aquifer, which was studied intensively using various subsurface characterization methods to evaluate the relative importance of hydrogeological and chemical heterogeneities in controlling bacterial transport at the field scale (DeFlaun et al., 1997). Understanding bacterial transport is important for bioremediation, which is a technology to degrade organic wastes or immobilize inorganic contaminants in the subsurface using microorganisms. We chose this site for testing our algorithm with real data, because

it provides one of the most convincing examples to date of how geophysics may lead to improved predictions of contaminant transport (Scheibe and Chien, 2003). This site displays an unusually high correlation between radar velocity and the logarithm of hydraulic conductivity. In a joint inversion that considered a geostatistical regularization inferred from borehole logs, Linde et al. (2008) found a correlation coefficient as high as 0.78. Furthermore, this study was used in a previous and related MCMC inversion study by Linde and Vrugt (2013). Finally, flowmeter borehole logs and previous geophysical tomographic studies (Hubbard et al., 2001) have been used to propose a multi-Gaussian geostatistical model of subsurface heterogeneity.

The setup considered first consists of two boreholes (M3 and S14) separated by 7 m. The 57 source locations and the 57 receiver locations are vertically separated by 12.5 cm and distributed over a depth range of 7 m. The boreholes, whose positions are known in 3D, have been projected in a 2D-plane. From the $57 \times 57 = 3249$ source-receiver traveltime pairs only one datum was missing. As in the synthetic example, we sample the solution space using 16 chains. Each chain consists of 1 million samples which is eight times more than for the synthetic example.

The first trials of our method with this dataset resulted in geologically unrealistic structures and, for this type of dataset, an unrealistically low data error estimate (0.25 ns). To avoid this overfitting of the data, we had to account for modeling errors, as for example wrong antenna positions. We assumed these modeling errors to be uncorrelated and Gaussian distributed with a standard deviation of 0.1 ns. We subtracted accordingly at each datapoint a randomly generated estimate of the modeling error in the forward step of our algorithm. Our high-dimensional estimate of the modeling error is updated by a model proposal procedure that uses a 10° gradual deformation (Hu, 2000) of the modeling error

estimate in the previous step. This approach of accounting for modeling errors falls within the classical framework of Kennedy and O'Hagan (2001), even if they considered lower-order descriptions of model errors (i.e., using less parameters to describe model errors) based on a Gaussian process model. It is well-recognized that ignoring model errors (e.g., by assuming that the Eikonal solver provides a perfect representation of the underlying physics) leads to MCMC results in which larger data sets only serve to be increasingly certain about the wrong model (e.g., Brynjarsdóttir and O'Hagan, 2014). We find that adding this very simplistic model error description avoids some of the issues we experience with overfitting, but we are well aware that it is simplistic and that developing a more suitable parameterization of model errors is an important area of future research (e.g., Hansen et al., 2014).

Using the same parameter values in DREAM$_{(ZS)}$ as for the synthetic example, all the Markov chains converge to essentially the same model and show very low variability within each chain. To avoid this collapse of the estimated posterior distribution, we require the differential evolution update to use three pairs from the archive instead of one ($\delta = 3$ in equation 2) and we no longer allow for a jump rate γ of one (Laloy and Vrugt, 2012). This avoids direct jumps from one model realization to another already visited model realization. To further increase the diversity of model updates, we additionally allow Snooker updates with a probability of 10% (ter Braak and Vrugt, 2008).

Furthermore, the following changes compared to the synthetic example were introduced: The prior distribution of the integral scale along the major axis of anisotropy was changed from uniform to log-uniform to favor small values and the prior distribution of the standard deviation of the data error was changed from log-uniform to uniform to counteract an underestimation. Compared to the synthetic example we invert for the standard deviation of the relative permittivity and not the variance, because the former is more straightforward

to interpret. As Hubbard et al. (2001) found that the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy is 0.2, we restrict this parameter to vary between 0.1 and 0.3 to enforce a similar layering. To avoid fractal-like subsurface realizations with small scale variations, we change the lower bound of the shape parameter ν from 0.1 to 0.5. The MCMC settings are summarized in Table 3. Values different from the synthetic example are marked with an asterisk (*).

Figure 8a shows the weighted RMSE misfit. The spikes in the weighted RMSE misfit, which were absent in the synthetic example, are caused by the Snooker update. Using the weighted RMSE misfit, we determined the burn-in phase to be completed after 30% of the chain length. Although the weighted RMSE misfit is decreasing slightly afterwards, we consider the burn-in to be completed after 30% as the remaining decrease is very subtle. The potential scale reduction factor $\hat{R}$ is computed accordingly for the last 70% of the chain. As mentioned in the synthetic example, convergence is formally confirmed if $\hat{R} < 1.2$ for all parameters. This is fulfilled for 209 of 257 parameters. The remaining 48 parameters (all of them dimension reduction variables) feature an $\hat{R}$ between 1.2 and 1.4. Thus, the retrieved posterior distribution for these parameters might be slightly deteriorated.

A major difference with respect to the synthetic example is the acceptance rate, which is after burn-in on average only 0.2%. This low acceptance rate is caused by the way we account for modeling errors to avoid overfitting.

The posterior distribution for the seven statistical variables describing the subsurface are plotted in Figures 8b to 8h. Of the two parameters that feature a very narrow distribution (relative permittivity and standard deviation of the permittivity) a zoom is plotted as well. The standard deviation of the data error is estimated to be approximately 0.34 ns (Figure

8b). This is lower than the value used by Linde and Vrugt (2013) for the same dataset.

The posterior distribution of the mean and the standard deviation of the relative permittivity feature a clear peak (Figures 8c and 8d). The distribution of the integral scale along the major axis of anisotropy (Figure 8e) is broader than the preceding distributions. The most likely value is estimated to be approximately 1.9 m. This is reasonably close to the estimate of Hubbard et al. (2001), which found a value of approximately 1.5 m. Note, that they included data from all the boreholes at the site in their analysis, while we are only using one borehole pair. As we expect a horizontally layered medium, the anisotropy angle is supposed to be approximately 90°, which is what we retrieve (Figure 8f).

Our algorithm finds a value of around 0.27 for the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy (Figure 8g). Hubbard et al. (2001) found with a value of 0.2 a slightly stronger anisotropy. As the distribution is not squeezed towards the upper limit of the prior range, we can assume that the algorithm was not hindered by the limited prior range and that the result would be similar for a wider prior range. The estimate of the shape parameter ν (Figure 8h) suggests a Gaussian-like model. Note, that we have excluded band-limited fractal models by setting the lower bound of the prior range to 0.5.

To illustrate the range of models that the algorithm can choose from during the burn-in phase, nine randomly selected model realizations are plotted in Figure 9. To facilitate a comparison with the results presented by Linde and Vrugt (2013), we converted the models from relative permittivity ε_r to velocity $v = 1/\sqrt{\varepsilon_r \varepsilon_0 \mu_0}$ with ε_0 and μ_0 being the relative permittivity and magnetic permeability of free space. The tested models can for example consist of relatively small structures (Figure 9d) as well as larger structures (Figure 9g) or

feature a roughly horizontal anisotropy (Figure 9b), inclined anisotropy (Figures 9a or 9e) or almost no anisotropy at all (Figure 9d).

We selected 16 different model realizations from the posterior distribution at regular intervals (Figures 10a to 10p). Thereby, model 1 represents an earlier sample of the Markov chains than model 2 and so on. The chain, which provided the model, was selected randomly for each model. The standard deviation of 128 uniformly sampled posterior realizations is shown in Figure 10q and the mean model for the same 128 realizations in Figure 10r. Our subsurface structures agree well with those shown by Linde et al. (2008) and Linde and Vrugt (2013). However, the subsurface structures retrieved by Linde et al. (2008) and Linde and Vrugt (2013) feature less details.

Given the synthetic examples suggested that one borehole pair does not include enough subsurface area to estimate the geostatistical parameters reliably, we also inverted a second borehole pair of the Oyster site. These data between T1 and S9 are perpendicular to the previously considered borehole pair (see Plate 5 in Hubbard et al. (2001)). We used the same input parameters as before. However, from the 16 Markov chains, two got stuck at high weighted RMS values and they were discarded for the further analysis. Formal convergence for the remaining 14 chains was achieved as $\hat{R} < 1.2$ for all parameters.

The weighted RMSE misfit of the inversion of this second borehole pair stabilized a bit below 1.5, which is a higher value than for the first borehole pair. A likely reason for this is, that estimates of the borehole deviations were not at our disposal for this borehole pair, thereby resulting in larger uncertainties related to the position of the antennas.

The posterior distributions of the geostatistical parameters using data from the T1-S9 borehole pair are shown in Figures 11a to 11g (red plots) along with the corresponding

posterior distributions of the M3-S14 borehole pair (blue plots). Similarly to the synthetic example (Figure 7) the two distributions are overlapping (except for the standard deviation of the data error, which is extremely low). Given that the two borehole pairs are perpendicular to each other, it is unclear if differences in the geostatistical parameters are caused by spatial subsampling (as studied in Figure 7) or by horizontal anisotropy. Nevertheless, for practical purposes we find that the estimates obtained from the two borehole pairs are satisfactory.

The mean and the standard deviation of the velocity field based on 128 randomly selected models from the posterior distribution are plotted in Figures 11h and 11i for the T1-S9 borehole pair. The subsurface structure reveals a layered structure that is very similar to the structure found between the M3-S14 borehole pair. It correlates also well with the hydraulic conductivity structure retrieved by deterministic inversion shown in Plate 5 of Hubbard et al. (2001).

DISCUSSION

Our synthethic and field-based results suggest that the inference of a geostatistical parameter, for example the integral scale, may be adversely affected when considering data from one borehole pair only. This is important to keep in mind when interpreting inversion results. In one case, a parameter might be estimated well, in other words, the true value is close to the peak of the posterior distribution, while in another case the estimate is worse and the true value is located towards the edges of the posterior distribution. The finite size of the model domain implies that randomization of the true model in synthetic studies (see Figure 7) is necessary to evaluate the quality of the geostatistical inference.

To avoid overfitting and geologically unrealistic subsurface structures in our real-data example, we subtracted an uncorrelated Gaussian estimate of the modeling error in the forward step (de facto creating a stochastic forward operator). This is a simple approach to deal with model errors, but it is consistent with the classical theory outlined by Kennedy and O'Hagan (2001). Instead of using a correlated Gaussian process model as they do, we simply consider the case of uncorrelated model errors. This implies that the 257 variables describing the subsurface are inferred as before, while the N model errors are inversion variables that are simply sampled from their prior distribution. A more sophisticated scheme is likely to improve the results and the uncertainty estimates, but we leave this for future research. Advantages of our approach are its simplicity and straightforward implementation. A disadvantage is the resulting low acceptance rate, although this can partly be mitigated by creating the random estimates of the modeling error through gradual deformation. Another disadvantage is that the user has to specify two more parameters, which are the standard deviation of the Gaussian modeling error estimates and the degree of gradual deformation. Despite these limitations, we argue that this approach is more satisfying than ignoring modeling errors altogether.

The subsurface structures retrieved for the M3-S14 borehole pair of our real-data example are in good agreement with previous deterministic (Linde et al., 2008) and probabilistic inversion results (Linde and Vrugt, 2013). The primary advantages of the presented approach compared to the one by Linde and Vrugt (2013) is that the model parameterization is grounded on classical geostatistical theory and that the underlying geostatistical model is inferred along with the actual subsurface structure. In the future, we plan to use this inversion framework to address the feasibility of fully probabilistic inversion of full waveform data.

As the topics covered by Hansen et al. (2013a,b) are similar to those in this article, we would like to highlight the main differences here. First of all, in this study, the number of unknowns is reduced using a dimensionality reduction and the associated approach to avoid streaking effects. This speeds up the convergence and allows for inverting even larger model domains. Furthermore, with the DREAM$_{(ZS)}$ algorithm, we use a different MCMC sampler allowing for an efficient sampling of the posterior distribution. Second, we estimate also all geostatistical parameters describing the geostatistical model and discuss the problem of estimating these parameters from a finite-size model domain, for example from one borehole pair only. Although estimating all geostatistical parameters might be possible with the SIPPI-toolbox described by Hansen et al. (2013a,b) as well, only an example that estimates the horizontal and the vertical integral scale is presented therein. Third, we also estimate the standard deviation of the data error, which is not the case for the algorithm presented by Hansen et al. (2013a,b). Fourth, we present a pragmatic and extremely simple method of dealing with unknown modeling errors.

CONCLUSIONS

We have presented a probabilistic approach to invert crosshole GPR first-arrival travel-times. We assume that the subsurface can be represented by a multi-Gaussian field. Both, the subsurface structure as well as the geostatistical parameters describing the model are estimated during the inversion process. Using a dimensionality reduction and circulant embedding, the number of unknowns are reduced to 250 unknown dimension reduction variables and seven variables describing the statistical variability of the subsurface and the data. We successfully removed streaking artifacts for subsurface models featuring a large integral scale by performing a domain extension and subsequent subsampling in the Fourier

domain. A synthetic example demonstrated the possibilities of this algorithm. It also showed that the retrieval of geostatistical variables depends on the subsurface structure, because the inter-borehole interval is not large enough to make an unbiased estimate of the underlying statistical model. Ergodicity can only be achieved by considering several subsurface model realizations described by the same set of geostatistical parameters or by extending the model domain by considering multiple borehole pairs. The same observation was made when considering field data from the South Oyster Bacterial Transport Site, Virginia. To avoid overfitting due to unaccounted model errors in the real-data example, uncorrelated Gaussian estimates of the modeling error were subtracted in the forward step. The subsurface models retrieved in this way agree with previous studies.

ACKNOWLEDGMENTS

We are grateful to John Peterson and Susan Hubbard, who provided the crosshole GPR data from the South Oyster Bacterial Transport Site. We thank the editor John Etgen, the associate editor Jan van der Kruk, the reviewer Thomas Mejer Hansen and two anonymous reviewers for their constructive comments.

REFERENCES

Asli, M., D. Marcotte, and M. Chouteau, 2000, Direct Inversion of Gravity Data by Cokriging: Geostat2000, Cape Town, South Africa.

Bregman, N. D., R. C. Bailey, and C. H. Chapman, 1989, Crosshole seismic tomography: Geophysics, **54**, 200 – 215.

Brynjarsdóttir, J., and A. O'Hagan, 2014, Learning about physical parameters: The importance of model discrepancy: Inverse Problems, **30**, 114007.1–24.

Chilès, J.-P., and P. Delfiner, 1999, Geostatistics, modeling spatial uncertainty: Wiley.

Cordua, K. S., T. M. Hansen, and K. Mosegaard, 2012, Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information: Geophysics, **77**, H19 – H31.

Dafflon, B., J. Irving, and K. Holliger, 2009a, Simulated-annealing-based conditional simulation for the local-scale characterization of heterogeneous aquifers: Journal of Applied Geophysics, **68**, 60 – 70.

———, 2009b, Use of high-resolution geophysical data to characterize heterogeneous aquifers: Influence of data integration method on hydrological predictions: Water Resources Research, **45**, W09407–1 – W09407–12.

DeFlaun, M. F., C. J. Murray, W. Holben, T. Scheibe, A. Mills, T. Ginn, T. Griffin, E. Majer, and J. L. Wilson, 1997, Preliminary observations on bacterial transport in a coastal plain aquifer: FEMS Microbiol. Rev., **20**, 473–487.

Dietrich, C. R., and G. N. Newsam, 1997, Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix: SIAM Journal of Scientific Computation, **18**, 1088 – 1107.

Dubreuil-Boisclair, C., E. Gloaguen, D. Marcotte, and B. Giroux, 2011, Heterogeneous

aquifer characterization from ground-penetrating radar tomography and borehole hydrogeophysical data using nonlinear Bayesian simulations: Geophysics, **76**, J13 – J25.

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

Gelman, A. G., and D. B. Rubin, 1992, Inference from iterative simulation using multiple sequences: Statistical Science, **7**, 457 – 472.

Gloaguen, E., D. Marcotte, M. Chouteau, and H. Perroud, 2005, Borehole radar velocity inversion using cokriging and cosimulation: Journal of Applied Geophysics, **57**, 242–259.

Hansen, T. M., K. S. Cordua, B. H. Jacobsen, and K. Mosegaard, 2014, Accounting for imperfect forward modeling in geophysical inverse problems - Exemplified for crosshole tomography: Geophysics, **79**, H1 – H21.

Hansen, T. M., K. S. Cordua, M. C. Looms, and K. Mosegaard, 2013a, SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information, Part 1 - Methodology: Computers & Geosciences, **52**, 470–480.

———, 2013b, SIPPI: A Matlab toolbox for sampling the solution to inverse problems with complex prior information, Part 2 - Application to crosshole GPR tomography: Computers & Geosciences, **52**, 481–492.

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

Hansen, T. M., M. C. Looms, and L. Nielsen, 2008, Inferring the subsurface structural covariance model using cross-borehole ground penetrating radar tomography: Vadose Zone Journal, **7**, 249 – 262.

Hu, L. Y., 2000, Gradual Deformation and Iterative Calibration of Gaussian-Related

Stochastic Models: Mathematical Geology, **32**, 87–108.

Hubbard, S. S., J. Chen, J. Peterson, E. L. Majer, K. H. Williams, D. J. Swift, B. Mailloux, and Y. Rubin, 2001, Hydrogeological characterization of the South Oyster Bacterial Transport Site using geophysical data: Water Resources Research, **37**, 2431 – 2456.

Johnson, T. C., P. S. Routh, T. Clemo, W. Barrash, and C. W. P., 2007, Incorporating geostatistical constraints in nonlinear inverse problems: Water Resources Research, **43**, W10422.

Johnson, T. C., R. J. Versteeg, M. Rockhold, L. D. Slater, D. Ntarlagiannis, W. J. Greenwood, and J. Zachara, 2012, Characterization of a contaminated wellfield using 3D electrical resistivity tomography implemented with geostatistical, discontinuous boundary, and known conductivity constraints: Geophysics, **77**, EN85 – EN96.

Kennedy, M. C., and A. O'Hagan, 2001, Bayesian calibration of computer models: Journal of the Royal Statistical Society, **63**, 425 – 464.

Klotzsche, A., J. van der Kruk, N. Linde, J. Doetsch, and H. Vereecken, 2013, 3-D characterization of high-permeability zones in a gravel aquifer using 2-D crosshole GPR full-waveform inversion and waveguide detection: Geophysical Journal International, **195**, 932 – 944.

Laloy, E., N. Linde, D. Jacques, and J. A. Vrugt, 2015, Probabilistic inference of multi-Gaussian fields from indirect hydrological data using circulant embedding and dimensionality reduction: Water Resources Research, **51**, 4224–4243.

Laloy, E., N. Linde, and J. A. Vrugt, 2012, Mass conservative three-dimensional water tracer distribution from Markov chain Monte Carlo inversion of time-lapse ground-penetrating radar data: Water Resources Research, **48**, W07510.

Laloy, E., and J. A. Vrugt, 2012, High-dimensional posterior exploration of hydrologic mod-

els using multiple-try DREAM$_{(ZS)}$ and high-performance computing: Water Resources Research, **48**, WO1526.

Linde, N., A. Tryggvason, J. E. Peterson, and S. S. Hubbard, 2008, Joint inversion of crosshole radar and seismic traveltimes acquired at the South Oyster Bacterial Transport Site: Geophysics, **73**, G29 – G37.

Linde, N., and J. A. Vrugt, 2013, Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion: Vadose Zone Journal, **12**, no. 1, 1 – 16.

Lochbühler, T., S. J. Breen, R. L. Detwiler, J. A. Vrugt, and N. Linde, 2014, Probabilistic electrical resistivity tomography of a CO2 sequestration analog: Journal of Applied Geophysics, **107**, 80 – 92.

Looms, M. C., T. M. Hansen, K. S. Cordua, L. Nielsen, K. H. Jensen, and A. Binley, 2010, Geostatistical inference using crosshole ground-penetrating radar: Geophysics, **75**, J29 – J41.

Malinverno, A., and V. A. Briggs, 2004, Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes: Geophysics, **69**, 1005 – 1016.

Matérn, B., 1960, Spatial Variation: Meddelanden fran Statens Skogsforskningsinstitut, Stockholm, **49**.

Peterson, J. E., 2001, Pre-inversion correction and analysis of radar tomographic data: Journal of Environmental and Engineering Geophysics, **6**, 1 – 18.

Podvin, P., and I. Lecomte, 1991, Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools: Geophysical Journal International, **105**, 271 – 284.

Ripley, B. D., 1987, Stochastic simulation: Wiley.

Rosas-Carbajal, M., N. Linde, T. Kalscheuer, and J. A. Vrugt, 2013, Two-dimensional probabilistic inversion of plane-wave electromagnetic data: Methodology, model constraints and joint inversion with electrical resistivity data: Geophysical Journal International, **196**, 1508 – 1524.

Rosas-Carbajal, M., N. Linde, J. Peacock, F. Zyserman, T. Kalscheuer, and S. Thiel, 2015, Probabilistic three-dimensional time-lapse inversion of magnetotelluric data: Application to an enhanced geothermal system: Geophysical Journal International, **203**, 1946 – 1960.

Scheibe, T. D., and Y. J. Chien, 2003, An evaluation of conditioning data for solute transport prediction.: Ground Water, **41**, 128 – 141.

Scheibe, T. D., S. S. Hubbard, T. C. Onstott, and M. F. DeFlaun, 2011, Lessons learned from bacterial transport research at the South Oyster Site: Ground Water, **49**, 745 – 763.

ter Braak, C. J. F., and J. A. Vrugt, 2008, Differential Evolution Markov Chain with snooker updater and fewer chains: Statistics and Computing, **18**, 435–446.

Thompson Reiter, D., and W. Rodi, 1996, Nonlinear waveform tomography applied to crosshole seismic data: Geophysics, **61**, 902 – 913.

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

Vrugt, J. A., 2016, Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation: Environmental Modelling & Software, **75**, 273 – 316.

Zhou, B., S. Greenhalgh, and A. Green, 2008, Nonlinear traveltime inversion scheme for crosshole seismic tomography in tilted transversely isotropic media: Geophysics, **73**, D17

– D33.

Table 1: Bounds of the prior distribution, original value used to generate the model, type of prior distribution and units of the inversion variables.

variable	lower bound	upper bound	original value	prior distribution type	unit
Dimension reduction variables	-6	6		standard Gaussian	-
Standard deviation of the data error	$\log(0.01) = -4.6$	standard deviation of data	$\log(0.5) = -0.7$	log-uniform	ns
Mean relative permittivity	10	30	17	uniform	-
Variance of the relative permittivity	0.1	4	2	uniform	-
Integral scale along the major axis of anisotropy	0.5	3.5	2	uniform	m
Anisotropy angle	60	120	69	uniform	degrees
Ratio of the integral scales	0.1	1	0.5	uniform	-
Shape parameter ν	$\log(0.1) = -2.3$	$\log(20) = 3$	$\log(2.1) = 0.74$	log-uniform	-

Table 2: Statistical variability parameters of the posterior models shown in Figure 6.

Statistical variability parameter	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8
Mean relative permittivity [-]	16.86	16.76	16.79	16.87	16.77	16.97	16.64	16.79
Standard deviation of the relative permittivity [-]	1.57	1.56	1.59	1.62	1.77	1.5	1.79	1.8
Integral scale along the major axis of anisotropy [m]	1.7	2.3	2.18	1.71	2.26	1.84	1.97	1.96
Anisotropy angle [°]	82.67	80.94	72.91	62.32	72.52	80.74	75.37	66.69
Ratio of the integral scales [-]	0.76	0.7	0.8	0.66	0.69	0.56	0.71	0.65
Shape parameter ν [-]	4.69	1.14	1.62	1.56	1.08	1.51	1.07	1.01

Table 3: Bounds of the prior distribution, type of prior distribution and units of the inversion variables for the real-data example. Entries marked with an asterisk * are different compared to the synthetic example.

variable	lower bound	upper bound	prior distribution type	unit
Dimension reduction variables	-6	6	standard Gaussian	-
Standard deviation of the data error	0.25*	1*	uniform*	ns
Mean relative permittivity	10	30	uniform	-
Standard deviation of the relative permittivity*	1*	10*	uniform	-
Integral scale along the major axis of anisotropy	0.5	3.5	log-uniform*	m
Anisotropy angle	60	120	uniform	degrees
Ratio of the integral scales	0.1	0.3*	uniform	-
Shape parameter ν	0.5*	20	log-uniform	-

LIST OF FIGURES

5 A selection of eight randomly chosen model realizations from the prior distribution illustrates the range of model realizations that the algorithm is considering in the synthetic inversion studies.

6 (a-d) and (f-i) Randomly chosen realizations of the posterior distribution. (e) and (j) The standard deviation and the mean of 128 model realizations. The colorbar of subfigure (j) applies for all plots, except (e). The true model is shown in Figure 3.

7 The colored dashed and solid lines depict the marginal posterior distributions based on independent inference of ten "true" models with the same underlying geostatistical model (Table 1). (a) The standard deviation of the data error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the integral scale along the major axis of anisotropy, (e) the anisotropy angle, (f) the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy and (g) the shape parameter in the Matérn function. The distribution plotted with a thick black line is the average of all 10 distributions. The vertical black line indicates the value used to generate the subsurface structures. Note that the horizontal axis of Figures 7a and 7b does not span the whole prior range.

8 (a) The weighted root mean square error for each of the 16 chains used to invert the field data between boreholes M3 and S14. Posterior distribution for the real-data example of (b) the standard deviation of the data error, (c) the mean of the relative permittivity, (d) the standard deviation of the relative permittivity, (e) the integral scale along the major axis of anisotropy, (f) the anisotropy angle, (g) the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy and (h) the shape parameter in the Matérn function. The dashed line shows the prior distribution.

9 A selection of nine randomly chosen model realizations from the prior distribution

illustrates the range of model realizations that the algorithm is considering in the field inversion studies.

10 (a-p) 16 model realizations from the posterior distribution of the real-data example between boreholes M3 and S14 at the South Oyster Bacterial Transport Site. The model realizations are taken from various chains randomly selected at regular intervals. Instead of the relative permittivity, the GPR velocity is plotted. (q) shows the standard deviation and (r) the mean of 128 model realizations. The colorbar of the mean model is valid for all subfigures, except (q). The asterisks and the triangles represent the source and receiver positions, respectively.

11 Posterior distributions for the M3-S14 and T1-S9 borehole pairs of the field example: (a) the standard deviation of the data error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the integral scale along the major axis of anisotropy, (e) the anisotropy angle, (f) the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy and (g) the shape parameter in the Matérn function. Note that the x-axis does not span the complete prior range. The (h) mean and the (i) standard deviation of 128 model realizations for the T1-S9 borehole pair.

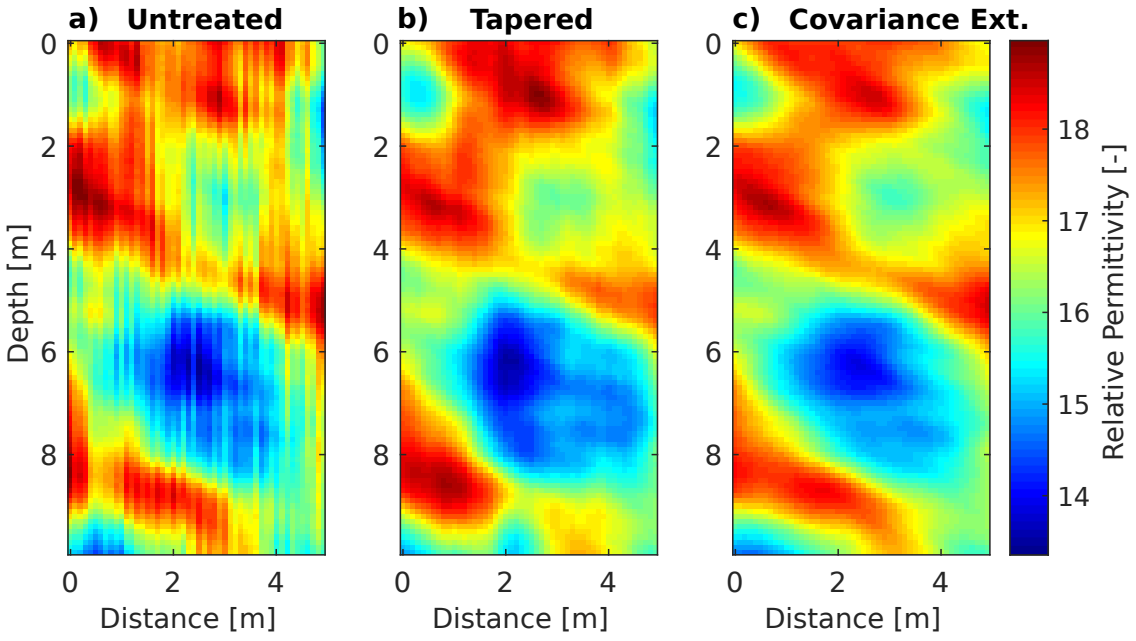

Figure 1: An isotropic multi-Gaussian model realization with an integral scale of 3.5 m. The other geostatistical parameters are given in Table 1. (a) Vertically aligned artifacts appear when the circulant-embedding covariance matrix is too small in the horizontal direction. These artifacts can partly be removed with (b) a taper or completely by (c) implementing domain extension and subsequent downsampling.

Hunziker et al. –

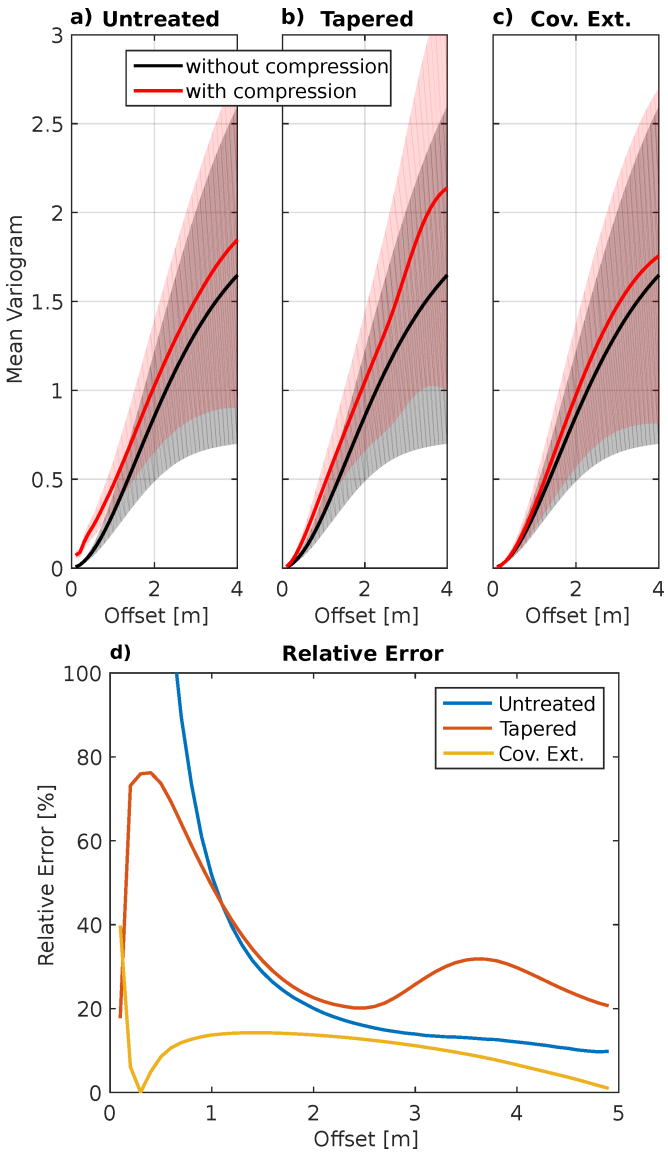

Figure 2: Variograms based on 1000 randomly generated multi-Gaussian model realizations with an integral scale of 3.5 m using dimensionality reduction (red) and without any compression (black): (a) without treatment of streaking artifacts, (b) tapered covariance matrix and (c) domain extension and subsequent downsampling. The shaded areas indicate the standard deviations of the variograms. (d) Relative error in percent of the mean variograms using dimensionality reduction compared with the mean variogram of the uncompressed model realizations.

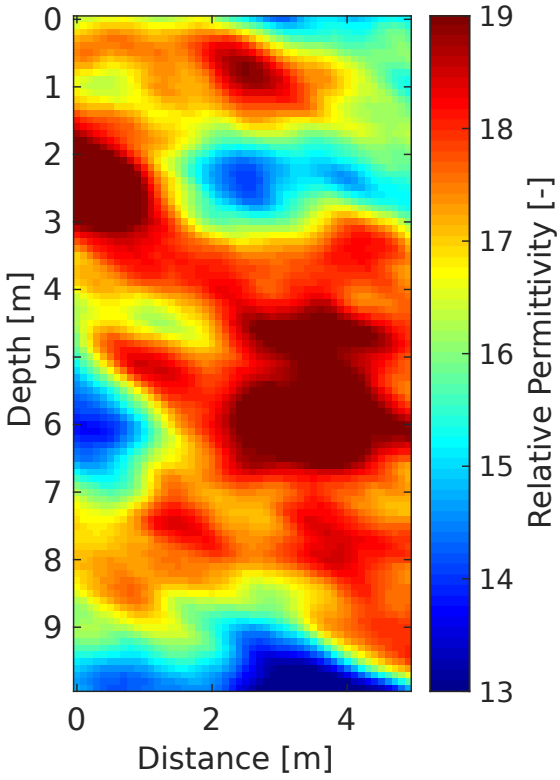

Figure 3: A realization of a multi-Gaussian relative permittivity field that is treated as the true model in the synthetic experiments reported in Figures 4-6.

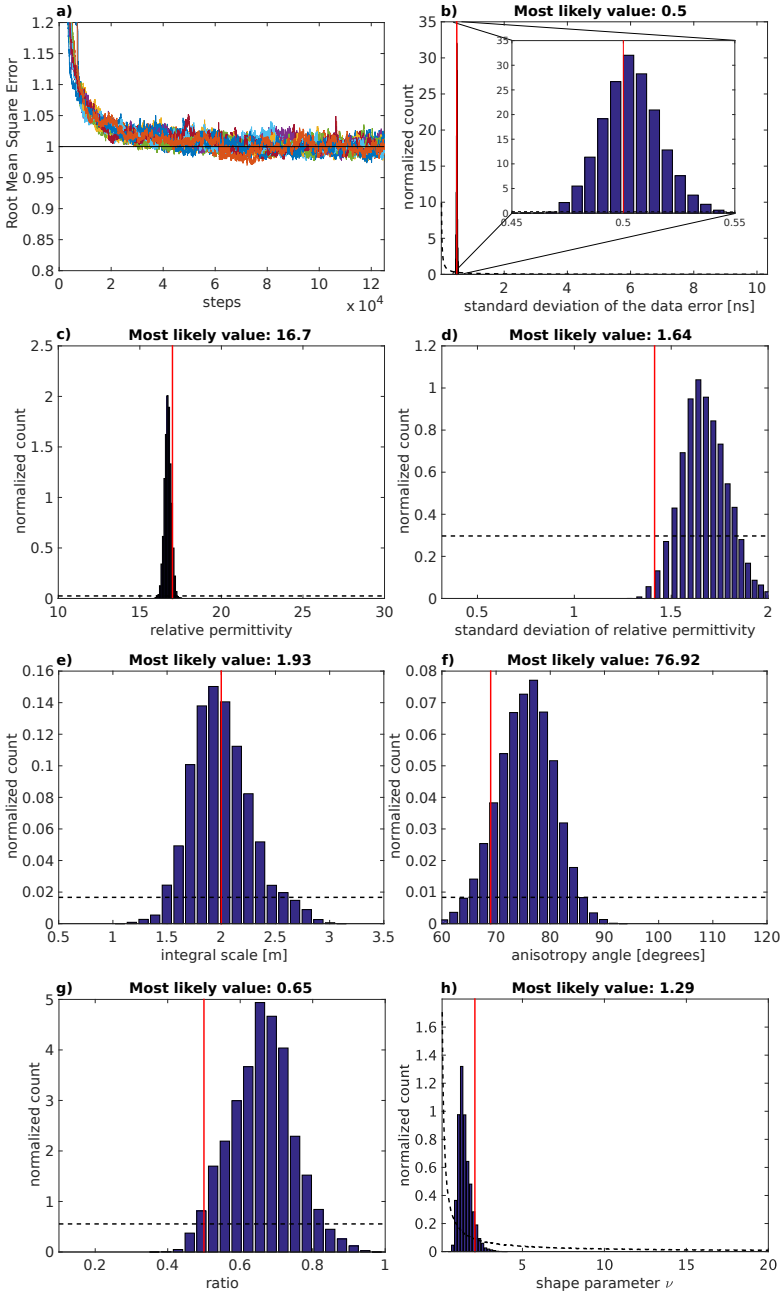

Figure 4: (a) The root mean square error for each of the 16 chains weighted with the estimated standard deviation of the data error. Marginal posterior distribution of: (b) the standard deviation of the data error (inlet shows a zoom), (c) the mean of the relative permittivity, (d) the standard deviation of the relative permittivity, (e) the integral scale along the major axis of anisotropy, (f) the anisotropy angle, (g) the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy

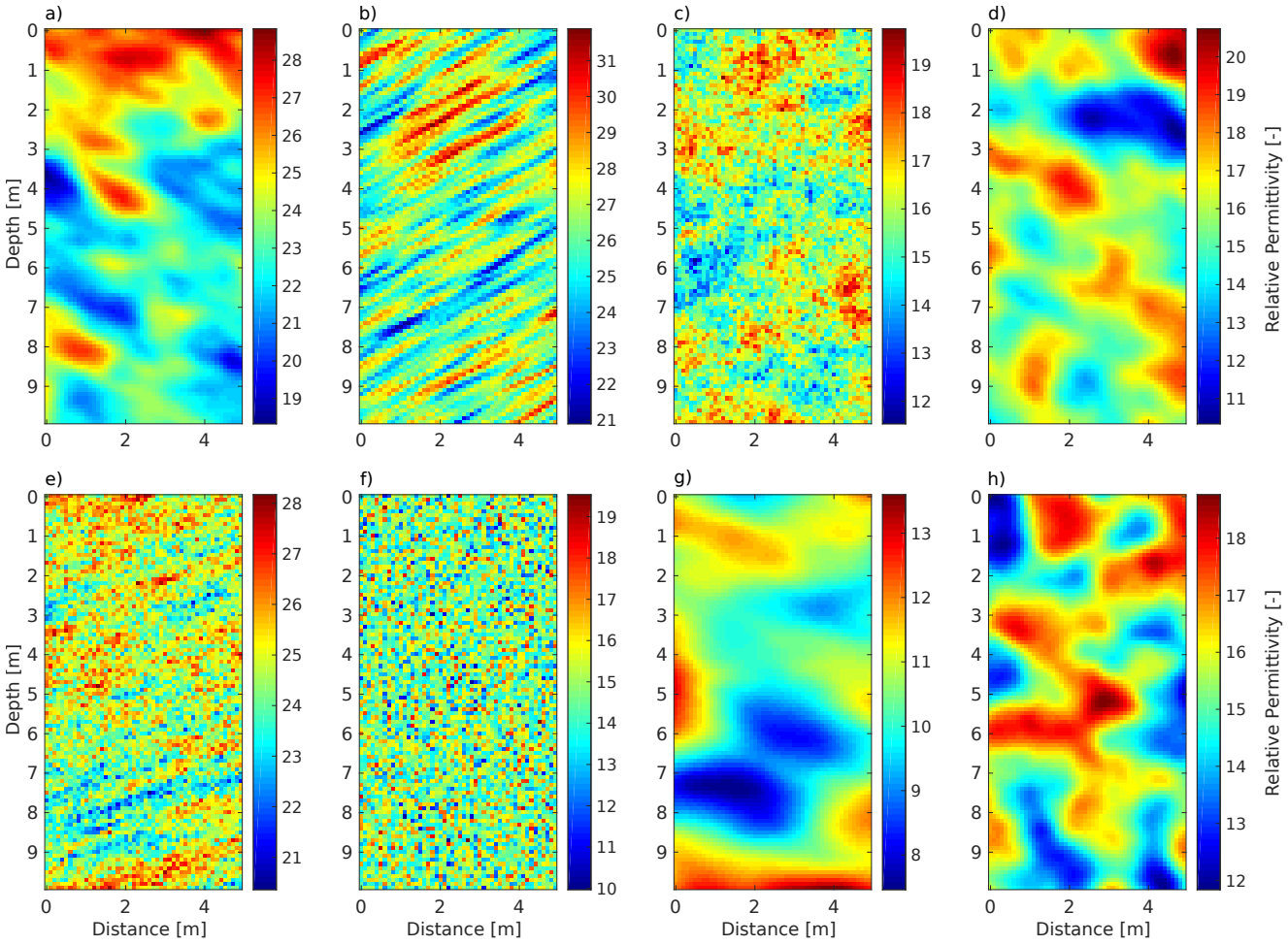

Figure 5: A selection of eight randomly chosen model realizations from the prior distribution illustrates the range of model realizations that the algorithm is considering in the synthetic inversion studies.

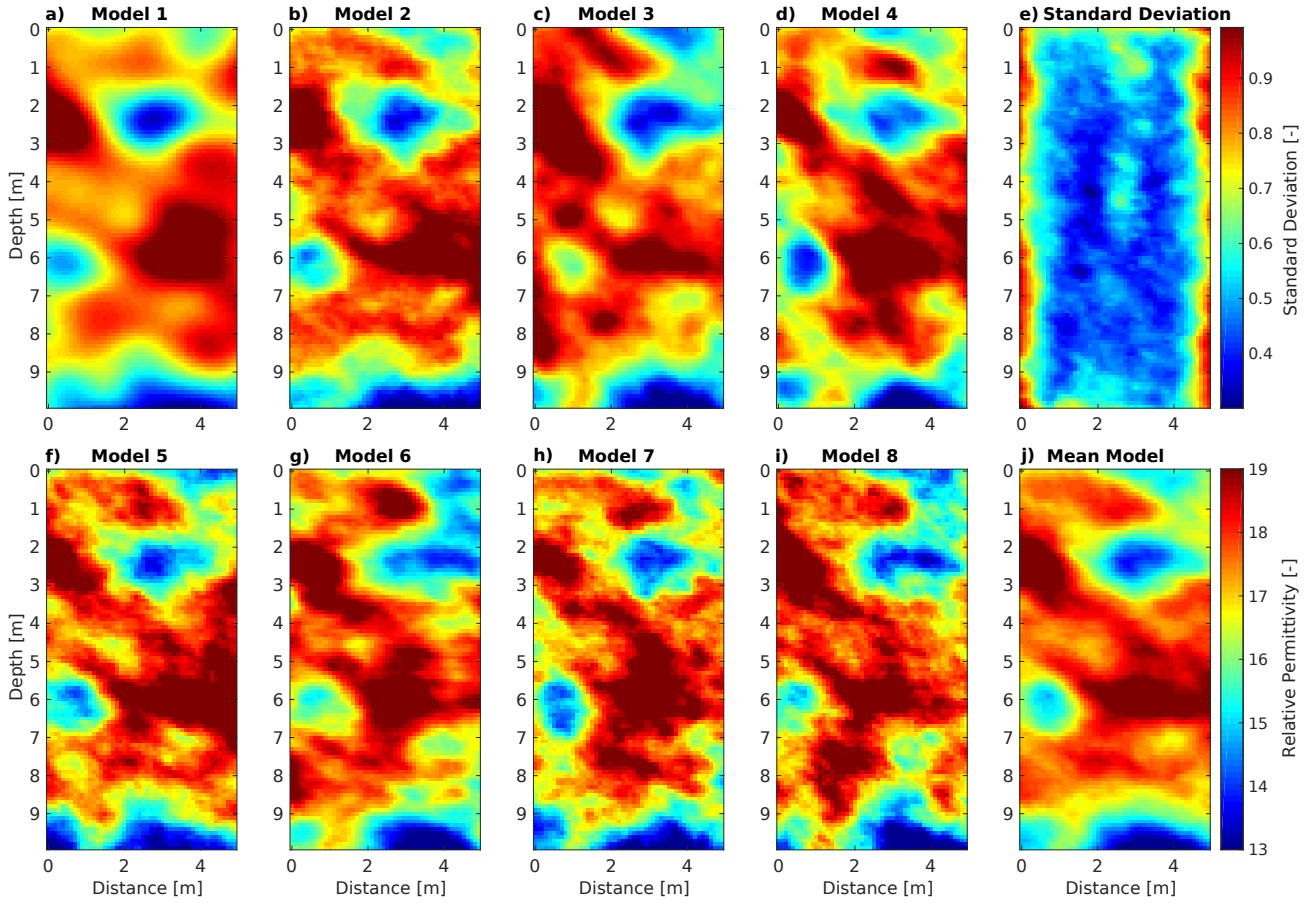

Figure 6: (a-d) and (f-i) Randomly chosen realizations of the posterior distribution. (e) and (j) The standard deviation and the mean of 128 model realizations. The colorbar of subfigure (j) applies for all plots, except (e). The true model is shown in Figure 3.

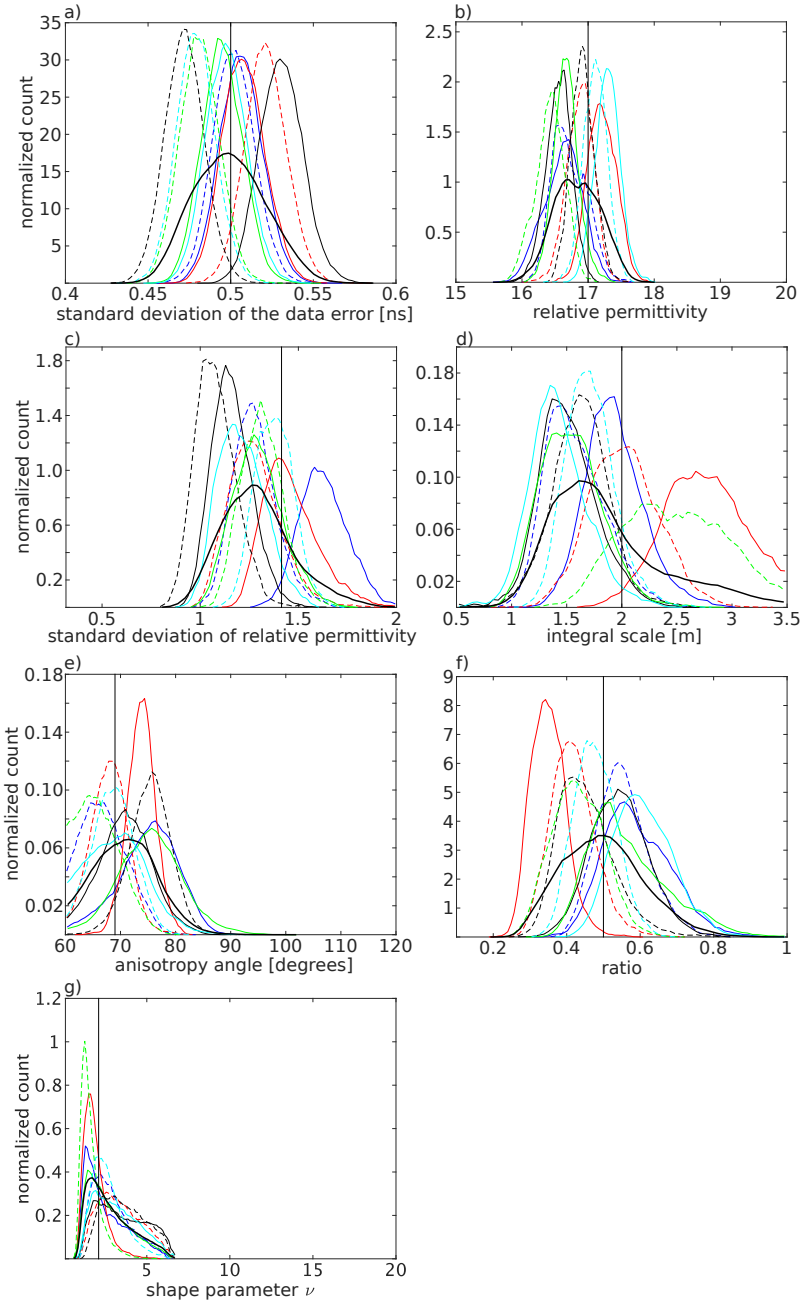

Figure 7: The colored dashed and solid lines depict the marginal posterior distributions based on independent inference of ten "true" models with the same underlying geostatistical model (Table 1). (a) The standard deviation of the data error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the integral scale along the major axis of anisotropy, (e) the anisotropy angle, (f) the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy

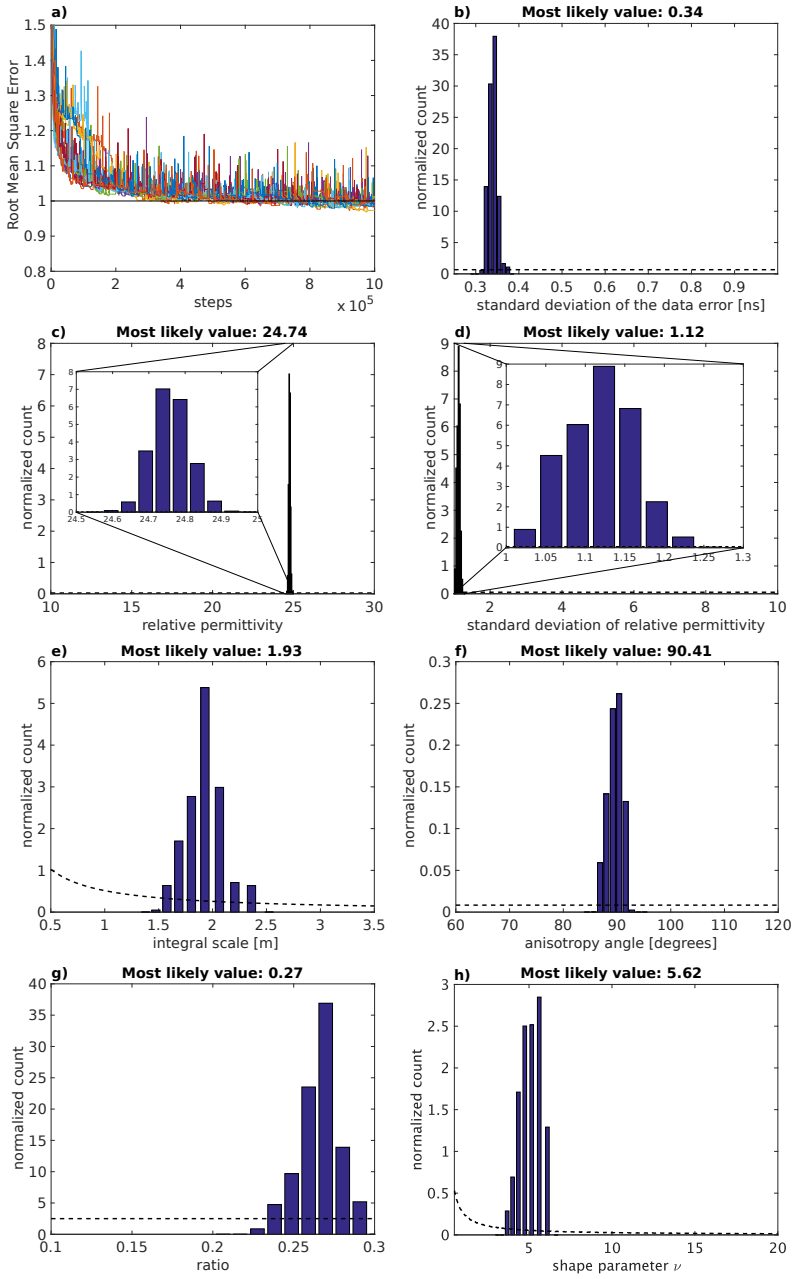

Figure 8: (a) The weighted root mean square error for each of the 16 chains used to invert the field data between boreholes M3 and S14. Posterior distribution for the real-data example of (b) the standard deviation of the data error, (c) the mean of the relative permittivity, (d) the standard deviation of the relative permittivity, (e) the integral scale along the major axis of anisotropy, (f) the anisotropy angle, (g) the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy and (h)

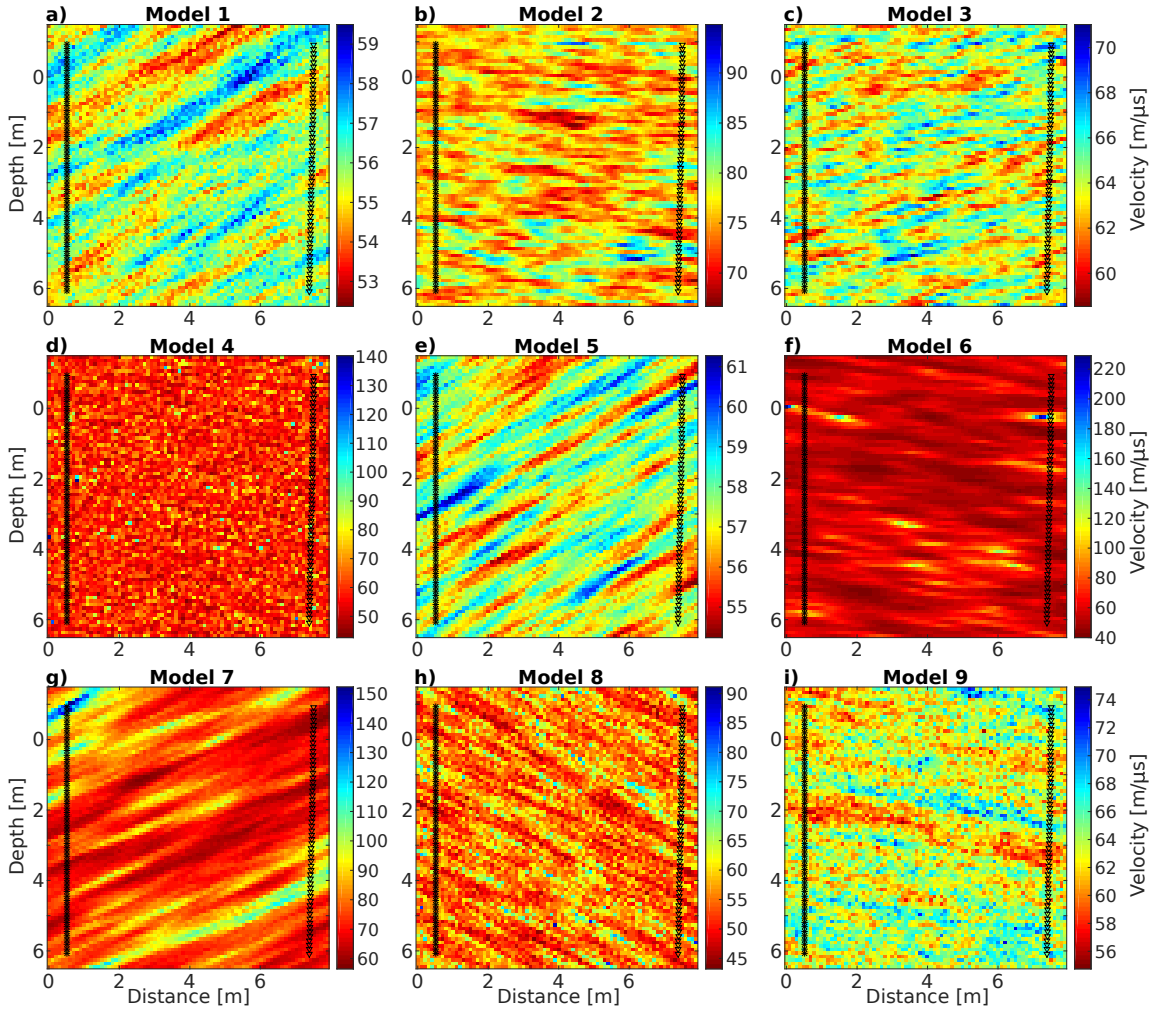

Figure 9: A selection of nine randomly chosen model realizations from the prior distribution illustrates the range of model realizations that the algorithm is considering in the field inversion studies.

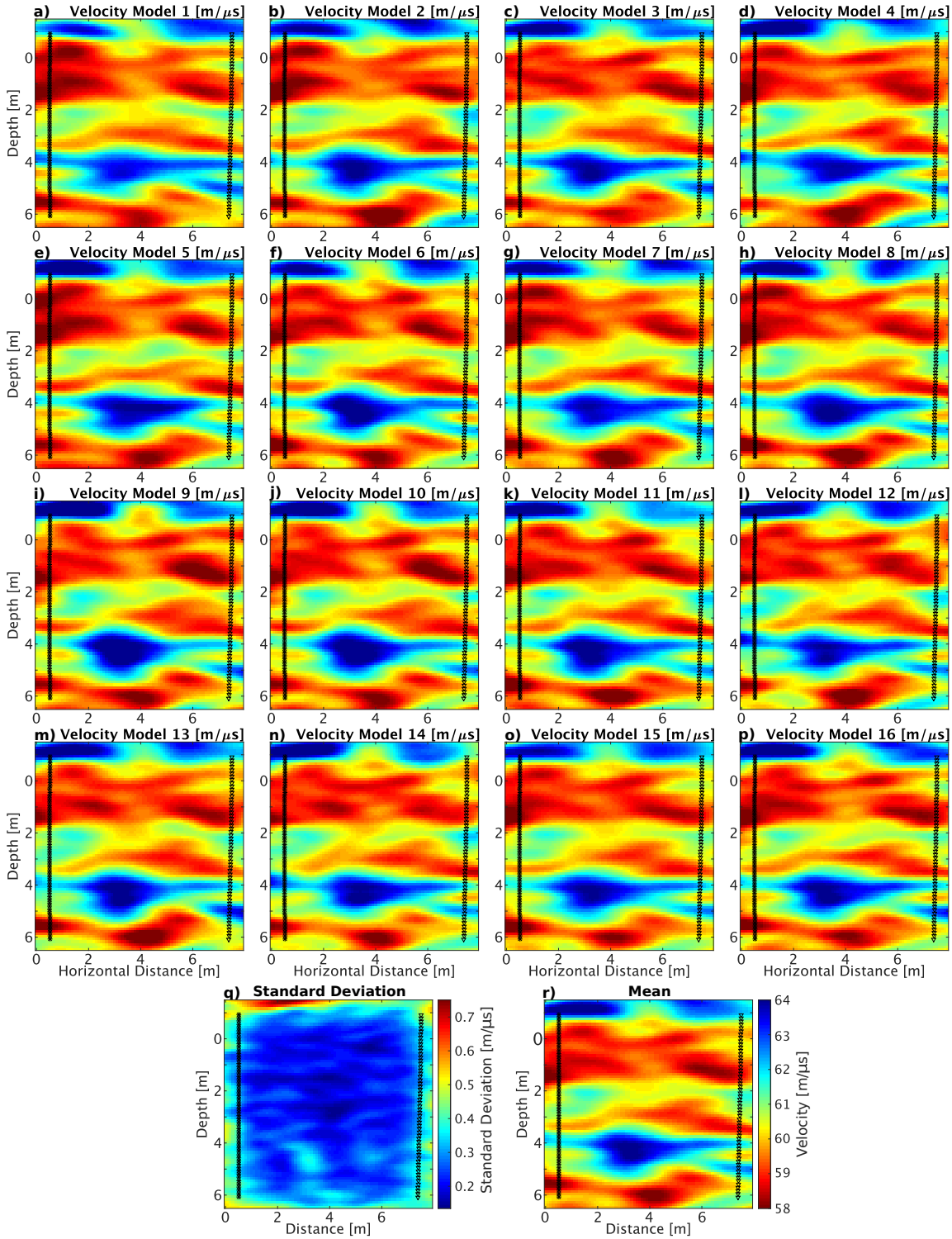

Figure 10: (a-p) 16 model realizations from the posterior distribution of the real-data example between boreholes M3 and S14 at the South Oyster Bacterial Transport Site. The model realizations are taken from various chains randomly selected at regular intervals.

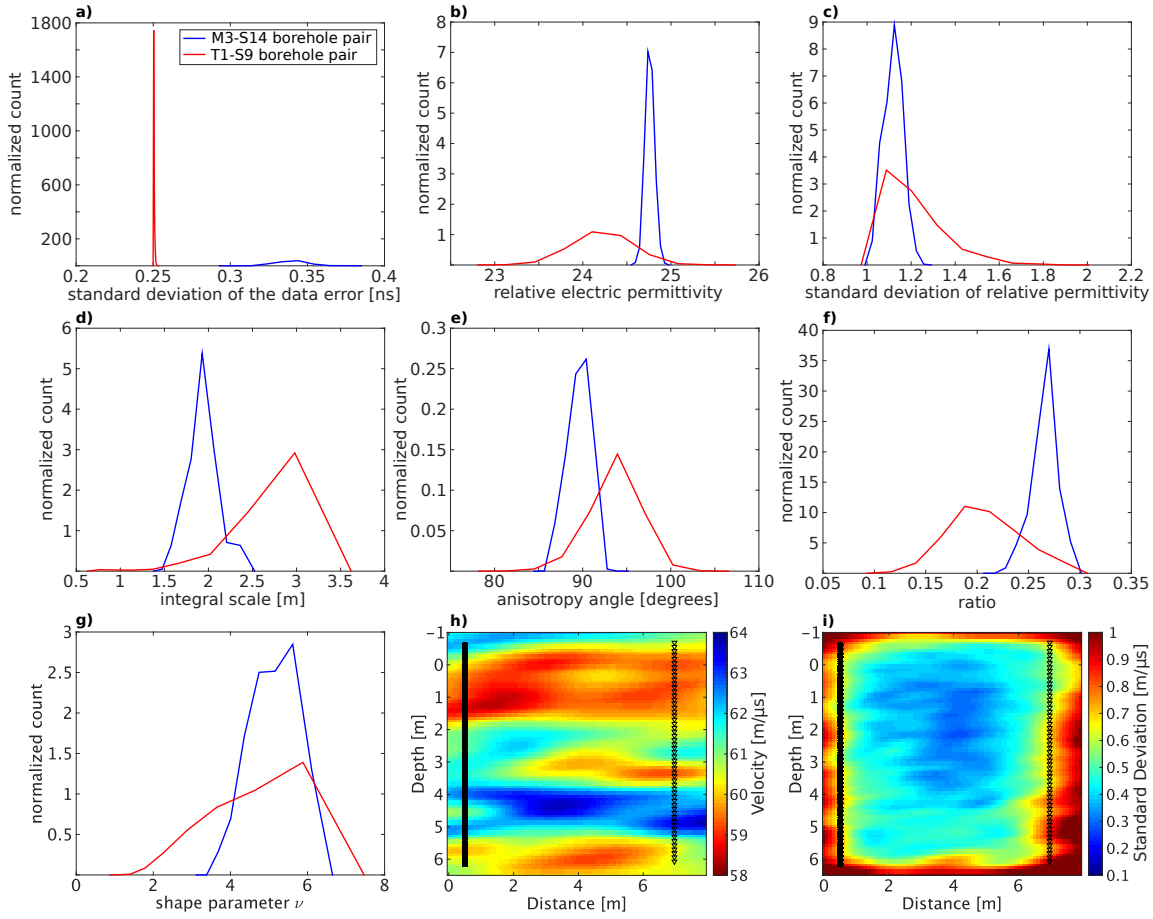

Figure 11: Posterior distributions for the M3-S14 and T1-S9 borehole pairs of the field example: (a) the standard deviation of the data error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the integral scale along the major axis of anisotropy, (e) the anisotropy angle, (f) the ratio of the integral scale along the minor axis of anisotropy to the integral scale along the major axis of anisotropy and (g) the shape parameter in the Matérn function. Note that the x-axis does not span the complete prior range. The (h) mean and the (i) standard deviation of 128 model realizations for the T1-S9 borehole pair.