submitted to *Geophys. J. Int.*

Bayesian full-waveform tomography with application to crosshole ground penetrating radar data

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

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

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

SUMMARY

We present a probabilistic full-waveform inversion (FWI) approach that infers a geostatistical model along with the subsurface structure. Probabilistic FWI with Markov chain Monte Carlo (MCMC) allows for uncertainty quantification and removes the requirement of having a starting model in the cone of attraction of the assumed correct global minimum. We demonstrate our approach on a synthetic and a field dataset. For the latter, we compare the results with deterministic FWI and a cone penetration test. Our results compare equally well with the cone penetration test as the deterministic results do. This is a positive result as for the deterministic inversion almost seven times more data were used than for the probabilistic inversion. Furthermore, the probabilistic FWI was able to converge to the posterior distribution starting from randomly drawn models. However, our uncertainty estimates are too narrow, because the necessarily short Markov chains implied by a computationally costly forward problem and the global nature of the model proposal scheme prevented a full exploration of the posterior probability density function. Without prior information such as borehole logs, the algorithm is only able to infer relative electric conductivity values, because the unknown amplitude of the wavelet and the mean of the conductivity are strongly correlated. This study clearly demonstrates the

feasibility of probabilistic FWI and highlights the advantages and disadvantages of the approach.

Key words: Waveform inversion, Statistical methods, Tomography, Ground penetrating radar, Hydrogeophysics

1 INTRODUCTION

Many successful applications of deterministic full-waveform inversion (FWI) of seismic data have been published. These studies include onshore surface seismic applications (e.g., Operto et al. 2004; Baeten et al. 2013), offshore projects (e.g., Vigh & Starr 2008; Prieux et al. 2011), time-lapse studies (e.g., Hicks et al. 2016) and engineering applications, for example, using a tunnel boring machine as a source (Bharadwaj et al. 2017). For ground penetrating radar (GPR), deterministic FWI has mainly been applied in crosshole settings (e.g., Ernst et al. 2007a; Klotzsche et al. 2010; Gueting et al. 2017; Keskinen et al. 2017), but also in surface configurations for soil moisture estimation (e.g., Minet et al. 2012; Tran et al. 2014).

Deterministic FWI has two major drawbacks. First, as deterministic FWI obtains, for a given starting model, just one regularized solution of the infinite set of solutions that can explain the data, it is not straightforward to obtain a meaningful estimate of uncertainty (Linde et al. 2017). Therefore, none of the studies mentioned above (except Minet et al. 2012) provide any information about the uncertainty of the inferred solution. Second, the inferred model strongly depends on the starting model. If the forward response of the starting model lies more than half a wavelength away from the correct solution, the algorithm will converge to a local minimum in the neighborhood of the starting model (Virieux & Operto 2009).

Nonlinear probabilistic FWI is potentially able to overcome these two drawbacks. Since nonlinear probabilistic FWI provides a selection of models that are sampled proportionally to the posterior distribution (Tarantola 2005), it is ideally suited for uncertainty quantification. In its simplest form, the mean model and some measure of variability (e.g., percentiles, standard deviation) can be used to determine, which parts of the model have been inferred with great certainty (i.e., small variability between the model realizations) and which parts are affected by large uncertainties (indicated by a large variability between the different model realizations). It is also straightforward to calculate marginal posterior probability density functions (PDF) for parameters of interest (Chen et al. 2000). In principle, it is, furthermore, not a problem if the marginal posterior PDF has several modes of probable values for a parameter (e.g., Laloy & Vrugt 2012). These general characteristics facilitate interpretation and help to draw useful and reliable conclusions from the inferred subsurface images. We highlight that in

a probabilistic framework, the solution is obtained by treating the subsurface as a spatial random function. This implies that the true values of the subsurface can be associated with low posterior probability as data are noisy, incomplete and band-limited.

Probabilistic FWI often relies on global sampling algorithms that can leave a local minimum, which implies, theoretically speaking, that there is no need of having a starting model that lies in the cone of attraction of the correct solution (i.e., with a forward response that differs less than half a wavelength from the true model (Virieux & Operto 2009)). Hence, probabilistic FWI might remove the human labor time needed to search for an adequate starting model.

The downside is that probabilistic full-waveform inversion is computationally extremely expensive as it requires tens of thousands to hundreds of thousands of forward-model calculations compared to tens or hundreds in deterministic FWI. This implies that a fast forward solver and ample computational resources are necessary. Additionally, a sampling algorithm that converges fast helps reducing the run-time. As inverse problems with fewer unknowns (i.e., a smaller dimensionality) usually allow for faster convergence than problems featuring a larger number of unknowns, it is often tempting to reduce the dimensionality of the problem. The most straightforward way of reducing the number of unknowns in geophysical tomography problems is by describing the Earth as a stack of horizontal layers. This approach has been implemented and applied to probabilistic FWI on field data for example by Mallick (1999), Hong & Sen (2009) and Aleardi & Mazzotti (2017). All these studies use a genetic algorithm combined with elements inspired by the Markov chain Monte Carlo (MCMC) method or a Gibbs sampler to quantify uncertainty.

However, the Earth is often not horizontally layered and more flexible, but still efficient, parameterizations are required. Datta & Sen (2016) and Ely et al. (2018) proposed to parameterize the subsurface with several layers with irregular topography. Each layer-interface consists of several nodes, which construct the interface when interpolated. Also the velocity within a layer is not constant, but can be obtained through interpolation between layer-interfaces. In this way, rather complicated subsurfaces can be inferred even though isolated regions of different material properties or small-scale fluctuations within layers cannot be represented.

Another straightforward way of decreasing model dimensionality, is to infer the subsurface parameters on a very coarse grid as, for example, proposed by Sajeva et al. (2016). Although the coarse model is interpolated on a fine mesh using a bilinear interpolater (Proakis & Manolakis 1996) to run the forward solver, the final models only capture the long-wavelength structure of the subsurface and the results might be biased (Linde & Vrugt 2013). Therefore, Sajeva et al. (2016) perform after probabilistic FWI a deterministic FWI to obtain a more detailed subsurface model on a finer grid.

More advanced parameterizations of the subsurface can be formulated by using for example the

discrete cosine transform (Linde & Vrugt 2013; Qin et al. 2016, 2018), a discrete wavelet transform (Ray et al. 2018), a geostatistical model (Gloaguen et al. 2007) or a training image (Cordua et al. 2012). In addition to an advanced subsurface parametrization, Ray et al. (2018) also use a state-of-the-art sampling method. Their reversible jump MCMC algorithm allows to alter the dimensionality of the problem during the inversion by increasing the number of parameters only if required. An early example of such a MCMC algorithm is provided by Dettmer & Dosso (2013). These advanced subsurface parameterizations in combination with a probabilistic FWI approach have so far, to the authors knowledge, only been applied to synthetic examples with the sole exception of Qin et al. (2018), who recently applied their method to a field dataset with the goal to detect defects in a concrete wall.

Other studies used a more problem-related parameterization. For example, Kotsi et al. (2018) perform probabilistic time-lapse FWI assuming that only a small part of the model is changing. This implies that the changes can be described with a few parameters only. Zhu et al. (2016) apply first a deterministic inversion and subsequently perturb the model to explore the posterior PDF. In this case, it is assumed that a global minimum has already been found and it only needs to be explored.

In this paper, we first briefly present our Bayesian FWI, which is a probabilistic inversion algorithm based on Bayes' theorem. To reduce the dimensionality, we use a geostatistical model in combination with the circulant embedding algorithm of Dietrich & Newsam (1997). We demonstrate our approach on crosshole ground penetrating radar (GPR) tomography problems considering, firstly, a synthetic and, secondly, a field-data example. Besides Qin et al. (2018), who applied a two-stage probabilistic FWI algorithm based on the discrete cosine transform to a non-destructive testing problem, this is, to our knowledge, the first geophysical application of a geostatistical subsurface parametrization to a field-data probabilistic FWI problem. Although we use GPR data, our findings are also applicable to similar seismic datasets, because both are based on wave propagation phenomena and are, therefore, modeled with similar partial differential equations.

2 METHOD

Bayesian inversion aims at obtaining the posterior PDF $p(\mathbf{m}|\mathbf{d})$, which describes the probability that a model $\mathbf{m}$ is correct given the observed data $\mathbf{d}$ and some prior knowledge, through the application of Bayes' theorem (Bayes 1763):

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{m})p(\mathbf{d}|\mathbf{m})}{p(\mathbf{d})}, \tag{1}$$

where $p(\mathbf{m})$ is the prior PDF containing prior knowledge about the model parameters and $p(\mathbf{d}|\mathbf{m})$ describes the probability that the observed data $\mathbf{d}$ is produced by the model $\mathbf{m}$. The latter PDF is

often expressed as a likelihood $L(\mathbf{m}|\mathbf{d})$. The remaining quantity in equation 1 is the evidence $p(\mathbf{d})$, which is a normalizing constant ensuring that the obtained posterior PDF is a true probability density distribution which integrates to one ($\int_{-\infty}^{\infty} p(\mathbf{m}|\mathbf{d})\partial\mathbf{m} = 1$). As $p(\mathbf{d})$ is usually difficult to obtain and we are mainly interested in relative probabilities, this quantity is normally ignored. Writing equation 1 using the likelihood $L(\mathbf{m}|\mathbf{d})$ and ignoring the evidence $p(\mathbf{d})$ leads to the following relation:

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{m})L(\mathbf{m}|\mathbf{d}). \tag{2}$$

The likelihood function $L(\mathbf{m}|\mathbf{d})$ quantifies how well a model $\mathbf{m}$ explains the data $\mathbf{d}$, for instance, through the following relation:

$$L(\mathbf{m}|\mathbf{d}) = \frac{1}{(2\pi)^{N/2}\sigma_e^N} \exp\left(-\frac{1}{2\sigma_e^2}\sum_{j=1}^{N}(G_j(\mathbf{m}) - d_j)^2\right), \tag{3}$$

where G is the forward operator simulating N data points for a given model $\mathbf{m}$. The forward operator G used here, solves Maxwell's equations in a two-dimensional space applying the finite-difference method in the time-domain on a staggered grid. It uses perfectly matched layer absorbing boundaries to minimize artificial reflections at the edges of the model space (Ernst et al. 2007a,b). The likelihood function also includes σ_e, which is the standard deviation of the data-error. We estimate σ_e during the inversion process using hierarchical Bayes (Malinverno & Briggs 2004). The likelihood function in equation 3 assumes that the data residuals are independent and identically distributed following a zero-mean Gaussian distribution.

As FWI is a non-linear process, we cannot infer the posterior distribution $p(\mathbf{m}|\mathbf{d})$ by analytical means. Instead, we need to sample it by sampling the prior PDF $p(\mathbf{m})$ and the likelihood function $L(\mathbf{m}|\mathbf{d})$ for a set of different models $\mathbf{m}$. To do so, we use an MCMC method, which describes a random walk that samples the solution space proportionally to the posterior distribution. We accept a proposed model $\mathbf{m}_{\text{prop}}$, drawn from a symmetric proposal distribution, with the Metropolis acceptance probability α, which depends on the prior density and the likelihood of that model proposal with respect to those of the current model $\mathbf{m}_{\text{cur}}$ (Metropolis et al. 1953):

$$\alpha = \min\left\{1, \frac{L(\mathbf{m}_{\text{prop}}|\mathbf{d})p(\mathbf{m}_{\text{prop}})}{L(\mathbf{m}_{\text{cur}}|\mathbf{d})p(\mathbf{m}_{\text{cur}})}\right\}. \tag{4}$$

An MCMC algorithm consists of two phases. In the first phase, the burn-in phase, the algorithm moves from the initial model to a region of high posterior density, while in the second phase the algorithm samples the solution space proportionally to the posterior density. Only the models sampled after the burn-in phase are considered when approximating the posterior PDF from the MCMC results.

The MCMC algorithm of our choice is the DREAM$_{\text{(ZS)}}$ algorithm (ter Braak & Vrugt 2008; Laloy & Vrugt 2012). This algorithm is able to rapidly converge to the posterior PDF, because of the use

of (1) parallel, interacting Markov chains and (2) a database of previously accepted models that are used when generating model proposals. At the core of DREAM$_{(ZS)}$ is the symmetric model proposal scheme that describes how a new model $\mathbf{m}_{\text{prop}}$, consisting of l model parameters, is proposed based on the current model $\mathbf{m}_{\text{cur}}$ (Vrugt 2016):

$$\mathbf{m}_{\text{prop}} = \mathbf{m}_{\text{cur}} + (\mathbf{1}_l + \mathbf{e}_l)\gamma(\delta, l) \sum_{j=1}^{\delta} \left(\mathbf{X}^{r_1(j)} - \mathbf{X}^{r_2(j)} \right) + \boldsymbol{\epsilon}_l. \tag{5}$$

The algorithm takes the difference of δ randomly selected pairs of previously accepted models from the database $\mathbf{X}$ to obtain the update direction and step size for each model parameter. Here, the superscripts r_1 and r_2 represent two different rows of the archive $\mathbf{X}$ and $\gamma(\delta, l)$ is a constant. To improve sampling efficiency, only a fraction of randomly drawn model parameters are updated per iteration, while the remaining ones stay unchanged. The vectors $\mathbf{e}_l$ and $\boldsymbol{\epsilon}_l$ are of length l and their values are drawn from a uniform and a normal distribution with a mean of zero, respectively, to add some randomness to the proposal process. The matrix $\mathbf{1}_l$ is the l by l identity matrix. The interaction of the chains is accomplished through the archive $\mathbf{X}$, which contains models from all chains. Thus, a proposed model $\mathbf{m}_{\text{prop}}$ for one chain, can be based on models previously sampled by other chains. Even more important, the adaptive nature of the model proposal ensures that model parameters with a large variance are sampled with large proposal steps, whereas well-defined parameters of small variance are sampled with small steps. For more details including theoretical arguments, we refer to Vrugt (2016). This algorithm has been widely used for various recent MCMC applications in geophysics (e.g., Hinnell et al. 2010; Bikowski et al. 2012; Rosas-Carbajal et al. 2013; Lochbühler et al. 2015).

To improve the convergence speed of the MCMC chain, we reduce the number of unknowns. We achieve this by subsampling in the frequency domain using the circulant embedding algorithm by Dietrich & Newsam (1997). This algorithm was developed to manipulate large covariance matrices and to create geostatistical realizations with a given covariance structure. Laloy et al. (2015) demonstrated how the circulant embedding algorithm can be used for dimensionality reduction. First, a covariance matrix based on a 2D Matérn geostatistical model is computed. This geostatistical model consists of the follwoing six parameters: (1) the mean and (2) the standard deviation of the considered physical property, (3) the anisotropy angle and (4) ratio, (5) the integral scale of the major axis and (6) a shape parameter. The latter determines if the covariance of the considered property is described by an exponential model, a Whittle model, a Gaussian model or a model in-between. Through a two-dimensional Fourier transform, we obtain the eigenvalues $\boldsymbol{\Omega}$ of the covariance matrix. Second, a complex Gaussian matrix $\mathbf{Z}$ is created through interpolation in the frequency domain and predetermined random permutation of the dimension reduction variables. In our case, we use 250 dimension reduction variables to create the matrix $\mathbf{Z}$, which consists of almost 20'000 elements. Next we multiply elementwise

the square-root of the eigenvalues $\mathbf{\Omega}$ with the complex matrix $\mathbf{Z}$ and apply again a two-dimensional Fourier transform:

$$\mathbf{f} = \text{FFT2}\left(\sqrt{\mathbf{\Omega}} \odot \mathbf{Z}\right), \tag{6}$$

where $\odot$ represents elementwise multiplication. The resulting matrix $\mathbf{f}$ is complex and circulant, which means that the edges of opposite sides match. To obtain a geostatistical model realization, we take the real part of $\mathbf{f}$ and extract a subdomain of 100 times 50 elements (i.e., 5000 pixels). Extracting a sub-domain is needed to avoid a circulant representation of the subsurface. With this approach, we reduce the model $\mathbf{m}$ from 5000 unknown pixels to 250 dimension reduction variables and six parameters describing a geostatistical model. A stochastic travel-time inversion example along with an improvement of this circulant embedding algorithm is presented in Hunziker et al. (2017). It is important to note that even though equation 5 does not update all unknowns to obtain a model proposal, the resulting subsurface image always changes throughout the complete spatial domain, because the Fourier transform in equation 6 is an integral over all elements of $\sqrt{\mathbf{\Omega}} \odot \mathbf{Z}$. Thus, even though only a few unknowns are updated, the proposed model changes globally.

Compared to Hunziker et al. (2017), we use the complete GPR wavefield as data and not just the first arrival travel-times, which implies that we replace an Eikonal forward solver with a finite-difference time-domain forward solver. This allows us to not only invert for the relative electric permittivity but also for the electric conductivity. As these two quantities are not fully independent, we link them by applying the same geostatistical model to both of them. This implies that the inferred conductivity field features the same anisotropy angle and ratio, the same integral scale and the same shape parameter as the inferred permittivity field, but it is described by a different mean and standard deviation. This is needed because conductivity and relative permittivity are very different in absolute values. The inferred conductivity and permittivity distributions are also described by different sets of dimension reduction parameters. Thus, while they share some common geostatistical parameter values, the inferred subsurface structures of conductivity and permittivity can be completely different. We chose this relatively weak coupling between permittivity and conductivity to provide some guidance and at the same time have sufficient freedom to find a suitable subsurface structure. In the present study, the term subsurface structure refers to a geostatistical model realization describing the location and orientation of specific subsurface features. Motivated by a given field site, it would be straightforward to apply the same dimension reduction parameters to both fields, permittivity and conductivity.

Another methodological difference between this study and Hunziker et al. (2017) is the necessity to estimate a source wavelet. To this end, we define six points in time with a specific amplitude, from which the wavelet is obtained by shape-preserving piecewise cubic interpolation. As each of the six points consists of a time-value and an amplitude-value, this description of the wavelet adds another

twelve unknowns to the problem. An additional 13th unknown is introduced to scale the wavelet. In total, the whole problem is described by 522 unknowns, as summarized in Table 1.

[Table 1 about here.]

3 SYNTHETIC EXAMPLE

For our synthetic example, we generated a true subsurface structure in terms of relative permittivity and electric conductivity (Figure 1). The relative permittivity structure is identical to the one considered by Hunziker et al. (2017) and the geostatistical model parameters used are given in Table 2.

[Figure 1 about here.]

[Table 2 about here.]

There are two boreholes in the setup. One at 0.25 m offset and one at 4.75 m offset, resulting in a borehole separation of 4.5 m. The first borehole features four sources while the second borehole is equipped with 33 receivers. We use only four source positions, in order to reduce the computational time of the forward problem. If large parallel computers are available, sources can be added without slowing down the inversion process. Sources and receivers are indicated in Figure 1 with asterisks and triangles, respectively. According to this geometry, full-waveform GPR data were generated with the algorithm of Ernst et al. (2007a) and contaminated with zero-mean, uncorrelated and homoscedastic Gaussian noise, whose standard deviation is 5% of the maximal amplitude of the data.

The prior PDF $p(\mathbf{m})$ of the geostatistical model parameters are either uniform or log-uniform. The corresponding bounds are given in Table 2. The prior PDF of the dimension reduction variables is a standard Gaussian distribution with the parameters assumed to be independent from each other. We ran the inversion of the synthetic data twice using three interacting Markov chains each time. Each chain has a length of 100'000 steps and is initiated by randomly drawing a geostatistical model and corresponding subsurface structure from the prior PDF. To illustrate the diversity of subsurface structures considered by our algorithm, the relative permittivity structures of the six starting models (two runs with each three chains) are shown in Figure 2. This set consists of structures with larger (Figure 2b) and smaller features (Figure 2c) as well as of structures with high anisotropy (Figures 2a, e, and f) and rather isotropic models (Figures 2b and c).

[Figure 2 about here.]

[Figure 3 about here.]

The conductivity structures corresponding to these relative permittivity structures are shown in Figure 3. Comparing Figures 2 and 3 illustrates that, although they share the same integral scale,

anisotropy and shape factor, the generated relative permittivity and associates conductivity fields are very different from each other. The difference in dimension reduction variables causes the locations of zones with high and low amplitudes to not correlate between the permittivity and corresponding conductivity structures.

To examine the convergence of the algorithm, we plot the root mean square error (RMSE) weighted with the estimate of the standard deviation of the data error σ_e (Figure 4a). As σ_e is estimated extremely well, this is equivalent to weighting the RMSE with the true σ_e. Clearly, the assumption that the data error is Gaussian distributed, uncorrelated and homoscedastic is true for this synthetic example as stated above. In an ideal case, the weighted RMSE converges to one. If the weighted RMSE stays above one, the data are underfitted or, in other words, the inferred model is not able to explain the data completely. A weighted RMSE below one indicates overfitting of the data, implying that the inferred model tries to explain the noise. In our synthetic example, the weighted RMSE converges nicely to one, which was expected as all assumptions are fulfilled. A weighted RMSE of one is reached after approximately 20'000 steps or 20% of the chains. This suggests that the last 80% of the chains are sampling the posterior PDF.

[Figure 4 about here.]

Convergence to an appropriately sampled posterior PDF can be assessed formally using the potential scale reduction factor $\hat{R}$ (Gelman & Rubin 1992), which compares after burn-in the variance of the individual Markov chains with the variance of all the Markov chains merged together. Convergence is usually considered if $\hat{R} < 1.2$ for all parameters. Using the last 80% of the chains, 58.4% of the parameters have a value of $\hat{R}$ smaller than 1.2 and 98.8% have a value of $\hat{R}$ smaller than 2. Thus, formal convergence has not been reached and longer chains are needed. However, this was not possible given the computational resources available to us. As the posterior PDF has not been sampled completely, we refrain from plotting marginal posterior PDFs. However, the inferred models are part of the posterior density distribution and provide useful information about the subsurface. That formal convergence has not been reached, can also be visually confirmed through an inspection of the Markov chains as a function of steps (Figure 4b-i), as some of the parameters (e.g., the standard deviation of permittivity or the mean conductivity) appear to sample around different mean values in the different runs.

The acceptance rate describes the percentage of the proposed models that are actually accepted (equation 4). If the acceptance rate is very high (i.e., the vast majority of the tested models is accepted), the algorithm is not exploring enough and remains in the neighborhood of accepted models. On the other hand, if the acceptance rate is very low (i.e., the vast majority of tested models is rejected), the proposed models lie to far apart and the algorithm fails to efficiently sample the posterior distribution.

In our case, we achieve an acceptance rate between 10 and 40% during the posterior exploration phase, which suggests an overall efficient exploration of the solution space (Gelman et al. 1996).

The mean relative permittivity (Figure 4b) reaches a stable value at an early stage of the MCMC chain. However, it underestimates slightly the actual value. The standard deviation of the permittivity (Figure 4c) needs longer to stabilize, but eventually the chains oscillate around the correct value. Oscillations of the chains around the true value can also be observed for the mean electric conductivity (Figure 4d), the standard deviation of the electric conductivity (Figure 4e) and the anisotropy ratio (Figure 4h).

The true value of the shape parameter (Figure 4f) is only sampled by a few chains and only for a short time. This is another indication of incomplete exploration of the posterior PDF. However, this parameter is difficult to estimate, because quite a broad range of values can produce reasonable subsurface structures. In other words, the algorithm is not so sensitive to this parameter. This result is in agreement with the travel-time inversion results by Hunziker et al. (2017), which are very similar for this parameter.

Another less well inferred parameter is the anisotropy angle (Figure 4g). Our experiment lacks sufficient data coverage to properly infer this parameter. As there are only four sources in the left borehole, the left part of the considered area has a very selective data coverage. This leaves a rather narrow part of only 2 meters with sufficient data coverage, which is insufficient to estimate the anisotropy angle correctly. Finally, there is the integral scale of the major axis (Figure 4i), which is overestimated. Thus, the structural elements of the subsurface are estimated to be larger than they actually are. As for the anisotropy angle, we believe that the data coverage is insufficient to properly estimate this parameter. Indeed, the travel-time inversion by Hunziker et al. (2017) was able to obtain better estimates for the integral scale and the anisotropy angle using 33 instead of only four sources.

[Figure 5 about here.]

Estimates of the wavelet, which are taken from the Markov chains at regular intervals during the posterior exploration phase, are plotted in Figure 5 along with the true wavelet for the two runs. The six squares define the uniform prior PDF for the anchor points, from which the wavelet is obtained by shape-preserving piecewise cubic interpolation. The obtained estimates of the wavelet vary moderately and match quite well the true wavelet, indicating that the algorithm can estimate the shape of the wavelet properly. Errors in the wavelet affect primarily the estimate of the electric conductivity. If the wavelet is too large in amplitude, the corresponding amplitude of the conductivity is too large as well. In our synthetic example, the amplitude of the wavelet is known and, therefore, this ambiguity can be assessed, which is not the case in the field-data example presented in the next section.

[Figure 6 about here.]

After completion of the MCMC inversion, we can draw realizations from the posterior PDF (after the burn-in phase). Eight permittivity structures drawn randomly from the posterior PDF of the two runs are shown in Figure 6. Furthermore, the mean and standard deviation for each pixel of the subsurface structure are calculated based on 256 permittivity structures drawn from the posterior. The standard deviation has been normalized by the overall standard deviation of the true subsurface permittivity structure (Figure 1a) to make it comparable with the standard deviation of the conductivity structures presented below.

Certain features appear in all the posterior realizations, as for example the high-permittivity areas around 1 m depth and between 3 and 8 m depth as well as the low-permittivity areas around 2 m depth and below 9 m depth. But there are also differences between the subsurface structures. For instance, some are smoother (Figures 6b, d and f) than others (Figures 6c and g). There are also local differences, for example, the high-permittivity zone below the second source observed in Figure 6i and c is absent in Figures 6a, b and f. Correspondingly, we observe a larger standard deviation below the second source (Figure 6e), which gives an indication about how certain one can be about a specific feature.

All of the high-permittivity and low-permittivity areas described in the previous paragraph appear also in the true subsurface structure (Figure 1a). In the true subsurface structure we also observe a high-permittivity area on the left side between 2 and 3 m depth. This area falls between two source positions and, therefore, has too poor data coverage for being properly sampled in our setup. In other words, the permittivity in this area is not conditioned by the data, which results in a high uncertainty. This is reflected in the high standard deviation in this area (Figure 6). However, this standard deviation is not large enough, as is indicated by the fact that the subsurface structures depicted in Figure 6 all show a low-permittivity area between the first two source positions, although the true model consists of a high-permittivity area at this location. This pinpoints that the algorithm has not fully explored the posterior PDF. As stated earlier, this might be due to the short Markov chains. In addition, our global model proposal scheme in the frequency domain does not allow for local changes of the subsurface structure. Indeed, changing the value of either a single dimension reduction variable or a single geostatistical parameter will in both cases affect the whole resulting model realization (see Laloy et al. 2015, for details). It is, thus, not possible to improve one specific part of the current subsurface structure while keeping the rest unchanged. Consequently, our global update mechanism may cause slow exploration of the posterior PDF.

[Figure 7 about here.]

In Figure 7, we show the conductivity structures corresponding to the same MCMC iterations as the permittivity structures shown in Figure 6. While permittivity is related to the wave-velocity, conductivity is related to the amplitude of the signal and, therefore, coupled with the amplitude of the wavelet (Busch et al. 2012). Although the amplitude of the source wavelet is assumed known in this synthetic example, the estimate of the electric conductivity is less reliable than the estimate of the permittivity. This has also been observed for deterministic FWI by Klotzsche et al. (2013). In our results, this is reflected in an elevated standard deviation in conductivity compared to the standard deviation of the permittivity (Figure 6e). As stated above, this comparison is possible, because the standard deviations of permittivity and conductivity have been normalized by the corresponding standard deviation of the true subsurface structure. While the standard deviation of the derived posterior permittivity field is high mainly in between and close to the source positions, the standard deviation of the derived posterior conductivity field features large values also in other areas. Still, the main features are retrieved when compared to the true conductivity structure (Figure 1b).

[Figure 8 about here.]

To enable a detailed comparison of the inferred models with the true subsurface structure, we plot the true and inferred relative permittivity and electric conductivity fields along three virtual boreholes at horizontal offsets of 0.75, 2.5 and 4.25 m in Figure 8. These offsets correspond to 0.5 m to the right of the source borehole, exactly in the middle between the source and receiver boreholes and 0.5 m to the left of the receiver borehole, respectively. For a more quantitative comparison, we also provide the correlation coefficient R and the mean deviation md between the mean subsurface structure and the true subsurface structure in Figure 8. We observe that the retrieved subsurface structure in the virtual borehole close to the sources does not agree so well with the true subsurface structure (low R, high md). The agreement in the two other boreholes is much better. The reason for this inequality stems from the unbalanced data coverage close to the source borehole. We also observe that in this leftmost virtual borehole, the spread of the 256 realizations plotted in gray is larger than in the other two virtual boreholes, which leads to a larger uncertainty. However, the true solution is not for every depth level part of the range of possible models.

As mentioned previously, conductivity is much more difficult to estimate than permittivity. This is highlighted by the lower correlation coefficients obtained for conductivity than for permittivity. We can conclude from the excellent fit between inferred and true virtual borehole logs shown in Figures 8b, c, e, and f, that if the data coverage is balanced and the medium is properly sampled, our FWI approach is able to infer the geostatistical model and the subsurface structure in terms of relative permittivity and, to a lesser extent, electric conductivity, while simultaneously providing uncertainty

information. However, uncertainty quantification appears limited in poorly resolved areas, because of the too short length of the Markov chains and the slow sampling induced by our global update scheme.

4 FIELD-DATA EXAMPLE

For our field-data example, we use data from the Krauthausen site in Germany, which was established by the research center Jülich in 1993. Various boreholes sample the uppermost aquifer composed of alluvial terrace sediments with sandy to gravely grain size and only small amounts of clay (Gueting et al. 2017). Details about the test site and deployed instruments are given by Döring (1997) and Vereecken et al. (2000). The GPR data used in this study were measured in boreholes B31 and B38, which are approximately 4.8 m apart. This borehole-pair spans a plane which is roughly orthogonal to the dominant groundwater flow direction. The receivers are distributed over a depth of 6.9 m, the sources over a depth of 4.5 m. In the middle between these two boreholes, a cone penetration test (CPT) has been carried out, which we use to assess the accuracy of our inversion results.

As for the synthetic example, we used only four source positions of the 27 actually measured. More sources could straightforwardly be included into the inversion process, but that requires additional CPUs in order to keep the runtime of the forward steps at the same level. These sources are all located in borehole B31, while the 70 receivers are located in borehole B38. In agreement with the synthetic example, we run three parallel Markov chains with each chain consisting of 100'000 steps and we draw the starting models randomly from the prior PDF.

[Figure 9 about here.]

The weighted RMSE of the three chains is shown in Figure 9a. In this run, only two out of the three Markov chains converged to a weighted RMSE of one. Chain 1 (plotted with a dotted blue line) got stuck in a local minimum. Luckily, the non-converging chain can easily be identified and disregarded in the calculation of the posterior PDF. For the other two chains, it takes more steps to converge than for the synthetic example. Therefore, we use only the last 50% of Markov chains 2 and 3 for approximating the posterior PDF (compared to 80% and all chains in the synthetic example). The inset of Figure 9a magnifies the weighted RMSE between inversion steps 40'000 and 100'000. In this range, the weighted RMSE decreases from 1.2 to 1. This slow decrease to 1 indicates that after 100'000 iterations, the sampling process has located the posterior distribution, but did not explore it enough. The acceptance rate for this phase resides at a low 4.5% also indicating poor exploration. As for the synthetic example, the potential scale reduction factor $\hat{R}$ indicates as well that formal convergence has not been reached yet and longer chains are required. Again, if more CPUs are available, the forward

step takes less time and more steps can be carried out to explore the posterior distribution without loosing time.

The Markov chains are plotted for the same geostatistical parameters as for the synthetic example in Figures 9b to i. For all geostatistical parameters, chains 2 and 3 converge to similar values, while those for chain 1 are different, because it got stuck in a local minimum. Similarly to the synthetic example, we observe that the mean relative permittivity (Figure 9b) converges more rapidly to a stable value than the other parameters. The algorithm finds a roughly horizontal anisotropy with a ratio between 0.4 and 0.5 (Figure 9g, h) suggesting horizontally-oriented elongated subsurface features. The standard deviation of the relative permittivity and the mean conductivity did not converge to a stable value, due to the narrow prior PDF. Therefore, the algorithm tries to optimize the inferred models otherwise, which is a slow and difficult process, which might lead to the slow decrease of the weighted RMSE observed in Figure 9a.

[Figure 10 about here.]

For each chain, 16 wavelet estimates drawn at regular intervals from the exploration phase of the Markov chains are shown in Figure 10. Considering the results shown in Figure 9, it is not surprising that the wavelet estimates for chain 1 are different than those of the other chains. The variability of the wavelet estimates is much smaller than for the synthetic example. Compared to the synthetic example, we have no indication about the true amplitude of the wavelet for this field-data example as the actual designs of commercial borehole antennas are proprietary. This results in an ambiguity between the amplitude of the wavelet and the electric conductivity, because these two are inherently coupled (Busch et al. 2012). A larger amplitude of the wavelet connected with an increased amplitude of the electric conductivity will fit the data similarly well as a smaller amplitude of the wavelet connected with a smaller electric conductivity. Therefore, our conductivity values are only meaningful in a relative sense. This has already been observed for ray-based attenuation tomography by Holliger et al. (2001).

[Figure 11 about here.]

To visualize how well our probabilistic inversion explains the observed data, we compare in Figure 11 forward-modeled data based on an inferred model (red lines, left column) with the observed data (black lines). As basis for the simulated data, any model taken from the Markov chains after burn-in can be considered. For convenience, we choose the last model of the third chain. The arrival time and shape of the transmission response is reproduced well for all sources. Thus, the inferred model explains the data appropriately. This good data-fit also confirms that the shape of the wavelet has been inferred correctly. Additionally, we also show in Figure 11 data based on the model obtained by Gueting et al. (2017) using a deterministic inversion (green lines, right column). The model found by the Bayesian

approach explains the data better as demonstrated by the weighted RMSE, which are smaller for all sources in the probabilistic inversion (see legends of Figure 11). Mainly the amplitude of the signal is overestimated in the deterministic approach, which affects primarily the conductivity model. However, as almost seven times more data was used in the deterministic inversion, it is not surprising that the data-fit is worse, because fitting all the data becomes more difficult with an increasing amount of data.

[Figure 12 about here.]

[Figure 13 about here.]

After having discussed the inferred geostatistical model, the wavelet and the match in the data-space, it is important to inspect the subsurface structures obtained by the inversion process. For this reason, eight subsurface structures drawn from the posterior PDF (considering only chains 2 and 3) are plotted in terms of relative permittivity in Figure 12 along with the source locations (red asterisks) and the receiver locations (black triangles). The models show moderate variability for example in the middle between the two boreholes at five meters depth, but also consistency as for example the high-permittivity band between one and three meters depth. The conductivity structures corresponding to these permittivity structures are shown in Figure 13.

[Figure 14 about here.]

For the field-data case, we do not have the true physical property fields to compare our solution with, but we can compare our inferred solutions with the obtained subsurface structure of the deterministic inversion of Gueting et al. (2017). While we used only four source positions and 70 receivers, Gueting et al. (2017) used 27 source positions and 129 receivers, resulting in a much denser data coverage. As deterministic inversion routines just converge to the closest minimum, a good starting model is essential to obtain a reasonable solution. Gueting et al. (2017) chose a permittivity model obtained from a ray-based travel-time inversion as their starting model. For the electric conductivity, a homogeneous starting model was used.

The permittivity and electric conductivity structures obtained by Gueting et al. (2017) through deterministic inversion are shown in Figures 14a and d, respectively, along with the locations of sources (red asterisks) and receivers (black triangles). Figures 14b and e show the mean permittivity and mean electric conductivity subsurface structure based on 32 samples drawn from the posterior probability density distribution inferred by our Bayesian FWI approach (considering only chains 2 and 3). The corresponding standard deviation is shown in Figures 14c and f, respectively.

Despite the different data coverage and amount of data, there are clear similarities between the two subsurface structures: (1) the high-permittivity zone between 1 and 3 m depth, (2) the intermediate-permittivity zone between 3 and 6 m depth, (3) the low-permittivity zone around 6 m depth and (4) the

dip from left to right of the top of the high-permittivity zone around 1 m depth to just name the most prominent similarities. While the Bayesian and the deterministic inversion agree on a low-permittivity zone close to the left borehole between 3 and 5 m depth, the Bayesian FWI does not support a low-permittivity zone close to the right borehole between 3 and 4 m depth. This is not surprising due to the low data coverage in the Bayesian approach close to the right borehole, which is also reflected in a slightly higher standard deviation around the source borehole compared to the receiver borehole (Figure 14c).

As stated earlier, we are only able to estimate relative values of conductivity. The deterministic inversion predicts a high-conductivity zone between 1 and 3 m depth and a low-conductivity zone below 3 m depth. Due to the low data cover in the top part, related to our source-receiver geometry, we consider the zone of elevated conductivity above 1 m depth in our model as erroneous and ignore it. The increased standard deviation above 1 m is another indication that the model parameters for this zone are not well determined. Our model vaguely indicates the presence of the high-conductivity zone found by the deterministic inversion and the low-conductivity zone below.

[Figure 15 about here.]

In between the two boreholes, a CPT and a neutron log have been acquired. As the neutron log gives us an estimate of porosity, we convert the inferred relative permittivity estimate to porosity following exactly the procedure applied by Gueting et al. (2017) at the same site. They used the CRIM equation with a value of 84 for the dielectric permittivity of the pore fluid (water at 10 °C) and a value of 4.5 for the dielectric permittivity of the solid grain matrix (quartz).

The CPT logs along with the FWI results at that offset are plotted in Figure 15. Both FWI results visually match nicely with the CPT neutron log. The correlation coefficient, given in the legend, shows that the deterministic and probabilistic inversion correlate similarly well with the log. However, the mean deviation between the CPT-log and the FWI results (normalized by the mean porosity of the CPT log to obtain a percentage) is larger for the deterministic inversion (md = 21.0 %) than for the probabilistic inversion (md = 14.5 %). Considering the mean and the standard deviation of the three curves, the probabilistic inversion result reproduces the statistics of the CPT log better than the deterministic inversion result using only a small fraction of the data used in the deterministic FWI.

5 DISCUSSION

The main drawback of a fully non-linear probabilistic FWI is the high computational cost. Our algorithm needed 35 days of computational time on 48 2.2 GHz AMD Opteron processors using three Markov chains with 100'000 steps per chain. Of this time, more than 34 days were spent on the forward

solver, which requires approximately 10 seconds for one forward solve. If more computer resources are available, the runtime can be shortened considerably. For example, one can run the three Markov chains in parallel, which would already reduce the runtime by a factor of three. The forward problem is parallelized over source positions and through a domain decomposition. Parallelization over source positions allows to evaluate the finite-difference simulation for each source on a separate set of processors. Domain decomposition allows to split the finite difference problem into several smaller problems running on different processors. Tests on the Swiss high-performance computing center revealed that the forward problem can be solved in one second if enough processors are available. If all these measures are implemented, the same Bayesian FWI can thus be solved in one or two days. Another alternative to enable faster computations would be to port the code to a GPU environment.

Using additional computational resources does not only save time, but also permits for extended simulations. Parallelization over sources or Markov chains allows to include additional source locations or run more parallel chains without loosing actual time if CPUs are available. Especially adding more source positions is crucial in order to obtain a more balanced data coverage. If 300'000 samples can be computed in a few days instead of a month, the Markov chains can also be extended to allow for a proper exploration of the posterior PDF, which was not the case in this study. Another negative aspect of the high computational cost is that it prohibits extensive testing of alternative choices pertaining to data, parameterization, likelihoods, initialization, etc.

In theory, MCMC algorithms always explore the posterior distribution fully for infinite chains, but for finite chains this might not be the case. In the inversion runs for the synthetic example, all chains converged to the same region of the posterior PDF. However, in the inversion run for the field-data example one chain stayed in a local minimum. In this case, that chain featured a larger RMSE than the chains that have found the area of the posterior PDF and, thus, could be easily identified. For the calculation of the posterior PDF, chains that did not sample the posterior distribution should be neglected. To ensure that the posterior distribution has been sampled, we recommend to run the same inversion process several times. We did so in this study for the synthetic experiment, but not for the field-data experiment because of the long runtime. However, as it is straightforward to identify chains that got stuck in a local minimum (increased RMSE) and as we have a CPT and deterministic inversion results for comparison, we can verify that the posterior distribution has actually been located in the field-data example as well. even if this aspect was not explored in this study, it might be that the risk of getting stuck in a local minima would greatly reduce if uniform starting models were used.

Even if all chains converge to the posterior distribution, it might not be sampled fully. One reason might be that the chains are too short. This can be simply resolved by running longer chains if time or computational resources are available. If this is not the case, one can aim for a faster convergence

through a reduction of the dimensionality of the problem, which can be achieved by considering only a subsample of the problem or an aggressive model compression strategy. In both cases, the number of dimension reduction variables is reduced, which scales the problem down. Laloy et al. (2015) have shown that the number of dimension reduction variables can be reduced considerably with respect to the original dimensionality of the problem without affecting the resulting multi-Gaussian fields too much. In our case, 250 dimension reduction variables turned out to be the optimum number to balance loss of statistical accuracy and model compression. Another alternative is to only consider a fraction of the available data.

Besides too short Markov chains, another reason for incomplete exploration of the posterior distribution might be related to the model proposal scheme. The scheme used herein produces a global model update, which means that the complete model is altered at every proposal step, although some parameters remain unchanged. However, if the current model is very good in one part, but rather bad in another part, a global update, while improving the bad part, will most likely deteriorate the already good part. In doing so, the chances for acceptance of the proposed model might not be increased. In other words, global model proposals tend to the current model if it fits the data. The solution here are local model updates, which are not possible with our scheme, because the model is created in the Fourier domain and the inverse transform affects the model globally.

To improve posterior exploration, different MCMC algorithms could be considered, for instance, parallel tempering is an approach that allows to explore the posterior PDF with several parallel Markov chains that operate at different energy levels. An example is provided by Laloy et al. (2016). Furthermore, various approaches are available to accelarate MCMC (Robert et al. 2018).

An incomplete exploration of the posterior distribution has implications on the uncertainty estimates. As not all acceptable solutions are explored, the variability of the inferred solutions is smaller than it should be. Therefore, it happens that the true solution falls outside the range of accepted solutions (e.g., Figures 8a, 8d, 8e and 15). In the same way, also the standard deviation of permittivity and conductivity shown in Figures 6, 7 and 14 is smaller than it is supposed to be. However, the standard deviation of permittivity and conductivity still provides useful information about which regions of the model are resolved well by the data and which regions are not.

In deterministic FWI, each pixel of the subsurface model is treated as an unknown whose value is independent of the neighboring pixels unless a regularization is applied accordingly. As the resolution close to the sources and receivers is low, pixels in this area usually cannot be resolved (see for example Figures 9 and 11 of Klotzsche et al. (2013)). In our Bayesian inversion, we treat the subsurface as a multi-Gaussian field governed by an underlying geostatistical model. In this case, the pixels are dependent of their neighbors. Therefore, our geostatistical approach does not suffer from this limitation.

Not only is our approach able to infer the permittivity and electric conductivity between the source and receiver locations, but we can even infer an estimate of the subsurface parameters outside the area enclosed by sources and receivers where no signal has probed the subsurface. However, this estimate is only conditioned by the geostatistical model and, therefore, subject to high uncertainties.

We have seen, that the algorithm converges much faster for the synthetic data than for the field data (Figures 4a and 9a). The main reason might be that, our modeling is in perfect agreement with the physics in the synthetic case, because we have used the same forward-solver in the inversion process as was used to generate the synthetic data. However, in the field-data case, we face issues like the spatial extent of sources, dispersion and 3D-effects, which are not represented properly in the forward-solver. Additionally, while we know the absolute amplitude of the wavelet in the synthetic case, we have to estimate it in the field case. This adds an important degree of freedom which slows down convergence. Another reason might be the amount of data. We use four source positions in both cases, but the field-data example has, with 70 receivers compared to the 33 receivers of the synthetic example, more than two times more measurements. Thus, the field-data case is much more constrained by the data than the synthetic-data case, which might also lead to a slower convergence. Eventually, we should also consider the noise. In the synthetic case, the noise perfectly fulfills the assumption of being Gaussian distributed, uncorrelated and homoscedastic. However, in the field-data example this might not be the case. Finally, our approach is based on the assumption that the subsurface is well-described by a stationary multi-Gaussian model, which is not necessarily the case in a field setting, thereby hindering the convergence further.

Lastly, an aspect that deserves further research is the estimate of the electric conductivity. Currently, the algorithm can only infer relative electric conductivity because the ambiguity between amplitude of the source wavelet and amplitude of the conductivity cannot be resolved without additional information. A possibility to resolve this ambiguity is the inclusion of resistivity logs to calibrate the absolute conductivity values. Maurer & Musil (2004) also showed that correcting for systematic errors improves the estimates of the electric conductivity.

6 CONCLUSIONS

In this study, Bayesian FWI is used to infer a rather complex subsurface structure together with an underlying multivariate Gaussian geostatistical model. Our approach is first tested using a synthetic example with a known subsurface structure, before a field-data example is presented. For this field-data application, data from a cone-penetration test are used to validate the inversion results. The observed match is overall very good, given the limited data coverage. Due to the limited amount of computational resources available, only four source locations were used for the forward simulations resulting

in an unbalanced data coverage. However, if more processors are available, additional sources can be included without slowing down the inversion process. A comparison of our Bayesian FWI results with deterministic FWI results for the field-data example reveals that, through the usage of a geostatistical model, Bayesian FWI retrieves images that are less smooth and feature more details. This is especially striking given that for the deterministic FWI almost seven times more source positions were used than in our Bayesian FWI. These results are promising, even if we were unable to sufficiently sample the posterior PDF to enable a full uncertainty quantification. This points to the need for faster forward solvers and exploration of more efficient MCMC samplers.

7 ACKNOWLEDGEMENTS

We thank Jan van der Kruk, Anja Klotzsche and Niels Gueting for sharing the forward GPR code and the data from the Krauthausen site. We also thank Jan Dettmer and an anonymous reviewer for their constructive reviews of the manuscript.

REFERENCES

Aleardi, M. & Mazzotti, A., 2017. 1D elastic full-waveform inversion and uncertainty estimation by means of a hybrid genetic algorithmGibbs sampler approach, *Geophysical Prospecting*, **65**, 64–85.

Baeten, G., de Maag, J. W., Plessix, R.-E., Klaassen, R., Qureshi, T., Kleemeyer, M., ten Kroode, F., & Rujie, Z., 2013. The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study, *Geophysical Prospecting*, **61**, 701–711.

Bayes, T., 1763. LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S, *Philosophical Transactions*, **53**, 370–418.

Bharadwaj, P., Drijkoningen, G., Mulder, W., Thorbecke, J. W., Neducza, B., & Jenneskens, R., 2017. A shear-wave seismic system using full waveform inversion to look ahead of a tunnel-boring machine, *Near Surface Geophysics*, **15**, 210–224.

Bikowski, J., Huisman, J. A., Vrugt, J. A., Vereecken, H., & van der Kruk, J., 2012. Inversion and sensitivity analysis of ground penetrating radar data with waveguide dispersion using deterministic and Markov chain Monte Carlo methods, *Near Surface Geophysics*, **10**, 641–652.

Busch, S., van der Kruk, J., Bikowski, J., & Vereecken, H., 2012. Quantitative conductivity and permittivity estimation using full-waveform inversion of on-ground GPR data, *Geophysics*, **77**(6), H79–H91.

Chen, M.-H., Shao, Q.-M., & Ibrahim, J. G., 2000. *Estimating Marginal Posterior Densities. In: Monte Carlo Methods in Bayesian Computation*, pp. 94–123, Springer New York.

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*, **77**, H19 – H31.

Datta, D. & Sen, M. K., 2016. Estimating a starting model for full-waveform inversion using a global optimization method, *Geophysics*, **81**(4), R211–R223.

Dettmer, J. & Dosso, S. E., 2013. Probabilistic two-dimensional water-column and seabed inversion with self-adapting parameterizations, *The Journal of the Acoustical Society of America*, **133**(5), 2612–2623.

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

Döring, U., 1997. Transport der reaktiven Stoffe Eosin, Uranin und Lithium in einem heterogenen Grundwasserleiter, *PhD thesis, Christian Albrecht Univ. Kiel, Germany*.

Ely, G., Malcolm, A., & Poliannikov, O. V., 2018. Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method, *Geophysics*, **83**(2), R63–R75.

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

Ernst, J. R., Maurer, H., Green, A. G., & Holliger, K., 2007b. Full-Waveform Inversion of Crosshole Radar Data Based on 2-D Finite-Difference Time-Domain Solutions of Maxwell's Equations, *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 2807 – 2828.

Gelman, A., Roberts, G. O., & Gilks, W. R., 1996. Efficient Metropolis Jumping Rules, *Bayesian Statistics*, **5**, 599–607.

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

Gloaguen, E., Giroux, B., Marcotte, D., & Dimitrakopoulos, R., 2007. Pseudo-full-waveform inversion of borehole GPR data using stochastic tomography, *Geophysics*, **72**(5), J43–J51.

Gueting, N., Vienken, T., Klotzsche, A., van der Kruk, J., Vanderborght, J., Caers, J., Vereecken, H., & Englert, A., 2017. High resolution aquifer characterization using crosshole GPR full-waveform tomography: Comparison with direct-push and tracer test data, *Water Resources Research*, **53**, 49–72.

Hicks, E., Hoeber, H., Houbiers, M., Lescoffit, S. P., Ratcliffe, A., & Vinje, V., 2016. Time-lapse full-waveform inversion as a reservoir-monitoring tool - A North Sea case study, *The Leading Edge*, pp. 850–858.

Hinnell, A. C., Ferré, T. P. A., Vrugt, J. A., Huisman, J. A., Moysey, S., Rings, J., & Kowalsky, M. B., 2010. Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion, *Water Resources Research*, **46**(4), W00D40.

Holliger, K., Musil, M., & Maurer, H. R., 2001. Ray-based amplitude tomography for crosshole georadar data: a numerical assessment, *Journal of Applied Geophysics*, **47**, 285–298.

Hong, T. & Sen, M. K., 2009. A new MCMC algorithm for seismic waveform inversion and corresponding uncertainty analysis, *Geophysical Journal International*, **177**, 14–32.

Hunziker, J., Laloy, E., & Linde, N., 2017. Inference of multi-Gaussian relative permittivity fields by probabilistic inversion of crosshole ground-penetrating radar data, *Geophysics*, **82**(5), H25–H40.

Keskinen, J., Klotzsche, A., Looms, M. C., Moreau, J., van der Kruk, J., Holliger, K., Stemmerik, L., & Nielsen, L., 2017. Full-waveform inversion of Crosshole GPR data: Implications for porosity estimation in

chalk, *Journal of Applied Geophysics*, **140**, 102–116.

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

Klotzsche, A., van der Kruk, J., Linde, N., Doetsch, J., & Vereecken, H., 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.

Kotsi, M., Malcolm, A., & Ely, G., 2018. 4D Full-Waveform Metropolis Hastings Inversion Using a Local Acoustic Solver, *SEG Expanded Abstracts*.

Laloy, E. & Vrugt, J. A., 2012. High-dimensional posterior exploration of hydrologic models using multiple-try DREAM$_{(ZS)}$ and high-performance computing, *Water Resources Research*, **48**, WO1526.

Laloy, E., Linde, N., Jacques, D., & Vrugt, J. A., 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., Linde, N., Jacques, D., & Mariethoz, G., 2016. Merging parallel tempering with sequential geostatistical resampling for improved posterior exploration of high-dimensional subsurface categorical fields, *Advances in Water Resources*, **90**, 57–69.

Linde, N. & Vrugt, J. A., 2013. Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion, *Vadose Zone Journal*, **12**, 1 – 16.

Linde, N., Ginsbourger, D., Irving, J., Nobile, F., & Doucet, A., 2017. On uncertainty quantification in hydrogeology and hydrogeophysics, *Advances in Water Resources*, **110**, 161–181.

Lochbühler, T., Vrugt, J. A., Sadegh, M., & Linde, N., 2015. Summary statistics from training images as prior information in probabilistic inversion, *Geophysical Journal International*, **201**, 157–171.

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

Mallick, S., 1999. Some practical aspects of prestack waveform inversion using a genetic algorithm: An example from the east Texas Woodbine gas sand, *Geophysics*, **64**(2), 326–336.

Maurer, H. R. & Musil, M., 2004. Effects and removal of systematic errors in crosshole georadar attenuation tomography, *Journal of Applied Geophysics*, **55**, 261–270.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E., 1953. Equation of State Calculations by Fast Computing Machines, *The Journal of Chemical Physics*, **21**(6), 1087–1092.

Minet, J., Bogaert, P., Vanclooster, M., & Lambot, S., 2012. Validation of ground penetrating radar full-waveform inversion for field scale soil moisture mapping, *Journal of Hydrology*, pp. 112–123.

Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A., & Dell'Aversana, P., 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study, *Geophysical Prospecting*, **52**, 625–651.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O. I., & Kommedal, J. H., 2011. On

the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study, *Geophysical Journal International*, **187**, 1495–1515.

Proakis, J. G. & Manolakis, D. G., 1996. *Digital signal processing: principles, algorithms, and applications*, Prentice Hall.

Qin, H., Xie, X., Vrugt, J. A., Zeng, K., & Hong, G., 2016. Underground structure defect detection and reconstruction using crosshole GPR and Bayesian waveform inversion, *Automation in Construction*, **68**, 156–169.

Qin, H., Vrugt, J. A., Xie, X., & Zhou, Y., 2018. Improved characterization of underground structure defects from two-stage Bayesian inversion using crosshole GPR data, *Automation in Construction*, **95**, 233–244.

Ray, A., Kaplan, S., Washbourne, J., & Albertin, U., 2018. Low frequency full waveform seismic inversion within a tree based Bayesian framework, *Geophysical Journal International*, **212**, 522–542.

Robert, C. P., Elvira, V., Tawn, N., & Wu, C., 2018. Accelerating MCMC Algorithms, *arXiv e-prints*, p. arXiv:1804.02719.

Rosas-Carbajal, M., Linde, N., Kalscheuer, T., & Vrugt, J. A., 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.

Sajeva, A., Aleardi, M., Stucchi, E., Bienati, N., & Mazzotti, A., 2016. Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the Marmousi model, *Geophysics*, **81**(4), R173–R184.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, Siam.

ter Braak, C. J. F. & Vrugt, J. A., 2008. Differential Evolution Markov Chain with snooker updater and fewer chains, *Statistics and Computing*, **18**(4), 435–446.

Tran, A. P., André, F., & Lambot, S., 2014. Validation of Near-Field Ground-Penetrating Radar Modeling Using Full-Wave Inversion for Soil Moisture Estimation, *IEEE Transactions on Geoscience and Remote Sensing*, **52**, 5483–5497.

Vereecken, H., Döring, U., Hardelauf, H., Jaekel, U., Hashagen, U., Neuendorf, O., Schwarze, H., & Seidemann, R., 2000. Analysis of solute transport in a heterogeneous aquifer: The Krauthausen field experiment, *Journal of Contaminant Hydrology*, **45**, 329–358.

Vigh, D. & Starr, E. W., 2008. 3D prestack plane-wave, full-waveform inversion, *Geophysics*, **73**, VE135–VE144.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**, WCC127–WCC152.

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.

Zhu, H., Li, S., Fomel, S., Stadler, G., & Ghattas, O., 2016. A Bayesian approach to estimate uncertainty for full-waveform inversion using a priori information from depth migration, *Geophysics*, **81**(5), R307–R323.

LIST OF FIGURES

1 Multi-Gaussian fields considered as (a) the true relative permittivity structure and (b) the true electric conductivity structure for the synthetic example. Asterisks denote the source positions in the left borehole and triangles the receiver positions in the right borehole.

2 Starting models for the relative electric permittivity randomly drawn from the prior PDF for each chain of the two MCMC runs of the synthetic example.

3 Electrical conductivity starting models corresponding to the permittivity models shown in Figure 2.

4 The inversion was run twice using three Markov chains in each run. For each run are shown: (a) the weighted root mean square error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the mean of the electric conductivity, (e) the standard deviation of the electric conductivity, (f) the shape parameter ν, (g) the anisotropy angle, (h) the anisotropy ratio and (i) the integral scale of the major axis.

5 32 wavelet estimates from the posterior exploration phase extracted from the Markov chains at constant intervals (gray) and the correct wavelet that was used to generate the synthetic dataset (black). The thin black boxes in both plots indicate the uniform prior PDF of the six points defining the wavelet. a) Run 1, b) run 2.

6 (a-d) and (f-i) Randomly chosen relative permittivity realizations of the posterior distribution (the last 80% of the Markov chains). (e) The standard deviation of each pixel of the subsurface structure based on 256 relative permittivity posterior realizations normalized by the overall standard deviation of the true permittivity structure (Figure 1a). (j) The mean of the same 256 relative permittivity posterior realizations. Asterisks and triangles represent sources and receivers, respectively. The colorbar of the mean permittivity structure is also valid for the individual posterior realizations.

7 (a-d) and (f-i) Electric conductivity realizations of the posterior distribution corresponding to the relative permittivity model realizations shown in Figure 6. (e) The standard deviation of each pixel of the subsurface structure based on 256 relative conductivity posterior realizations normalized by the overall standard deviation of the true conductivity structure (Figure 1b). (j) The mean of the same 256 relative conductivity posterior realizations. Asterisks and triangles represent sources and receivers, respectively. The colorbar of the mean conductivity structure is also valid for the individual posterior realizations.

8 (a,b,c) Relative permittivity logs and (d,e,f) electric conductivity logs in virtual boreholes at a horizontal offset of (a,d) 0.75 m, (b,e) 2.5 m and (c,f) 4.25 m. The offsets correspond to 0.5 m from the source borehole, in the middle between the two boreholes and 0.5 from the receiver borehole, respectively. The gray logs correspond to the 256 posterior realizations used to compute the mean and standard deviation plotted in Figures 6e, 6j, 7e and 7j. The blue logs correspond to the mean and the black logs to the true subsurface structure at those offsets. The correlation coefficient R and the mean deviation md between the mean subsurface structure and the true subsurface structure at that offset are given in each plot. The latter is given relative to the mean permittivity or conductivity of the true subsurface structure.

9 The inversion of the field-data example was run once using three Markov chains. For these three chains are shown: (a) the weighted root mean square error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the mean of the conductivity, (e) the standard deviation of the conductivity, (f) the shape parameter ν, (g) the anisotropy angle, (h) the anisotropy ratio and (i) the integral scale.

10 For each chain, 16 evenly-spaced wavelet estimates drawn from the posterior distribution are plotted.

11 Comparison between observed data (black lines), simulated data based on the last model of the third Markov chain of the Bayesian inversion (red lines) and simulated data based on the model obtained from the deterministic inversion (green lines). All four source locations used in the Bayesian inversion are displayed, where source 1 is the topmost source in the borehole and source 4 the bottommost source in the same borehole. For clarity, traces are only shown for every fifth receiver. The legend gives for each source the weighted RMSE between the observed and the simulated data.

12 Randomly chosen relative permittivity realizations of the posterior distribution (the last 50% of the Markov chains). Red asterisks and black triangles represent sources and receivers, respectively.

13 Electric conductivity structures corresponding to the relative permittivity model realizations shown in Figure 12. Red asterisks and black triangles represent sources and receivers, respectively.

14 Comparison of deterministic and probabilistic inversion results. (a) and (d): Permittivity and conductivity model obtained from the deterministic inversion; (b) and (e): Mean permittivity and mean conductivity model obtained from the probabilistic inversion; (c) and (f): Standard deviation of the permittivity and conductivity models obtained from the probabilistic inversion. The mean and the standard deviation for the probabilistic inversion results are calculated based on 32 posterior realizations. Red asterisks and black triangles represent sources and receivers, respectively, used for the different inversion approaches.

15 Comparison of the deterministic inversion results (green), the probabilistic inversion results (red and gray) and a borehole log (blue) for the porosity. The borehole is located in the middle between the two boreholes containing the GPR antennas. The permittivity model obtained from the GPR data has been converted to porosity using the CRIM equation as described by Gueting et al. (2017). The correlation coefficient R and the mean deviation md between the GPR inversion results and the borehole log is given in brackets in the legend.

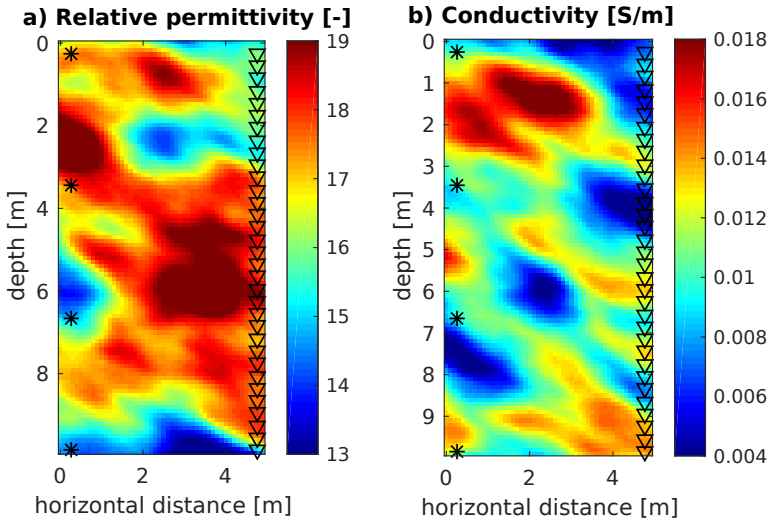

Figure 1. Multi-Gaussian fields considered as (a) the true relative permittivity structure and (b) the true electric conductivity structure for the synthetic example. Asterisks denote the source positions in the left borehole and triangles the receiver positions in the right borehole.

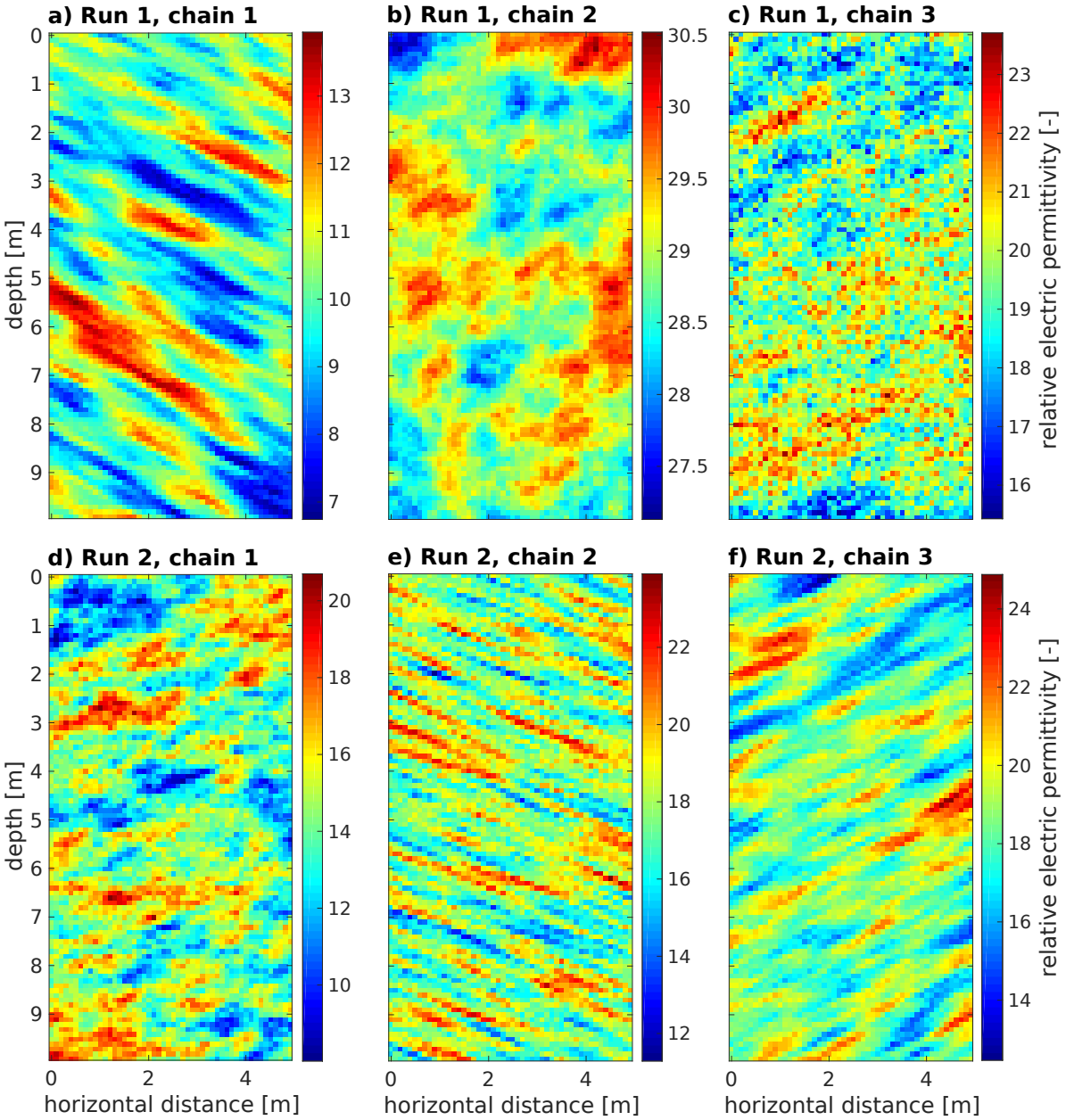

Figure 2. Starting models for the relative electric permittivity randomly drawn from the prior PDF for each chain of the two MCMC runs of the synthetic example.

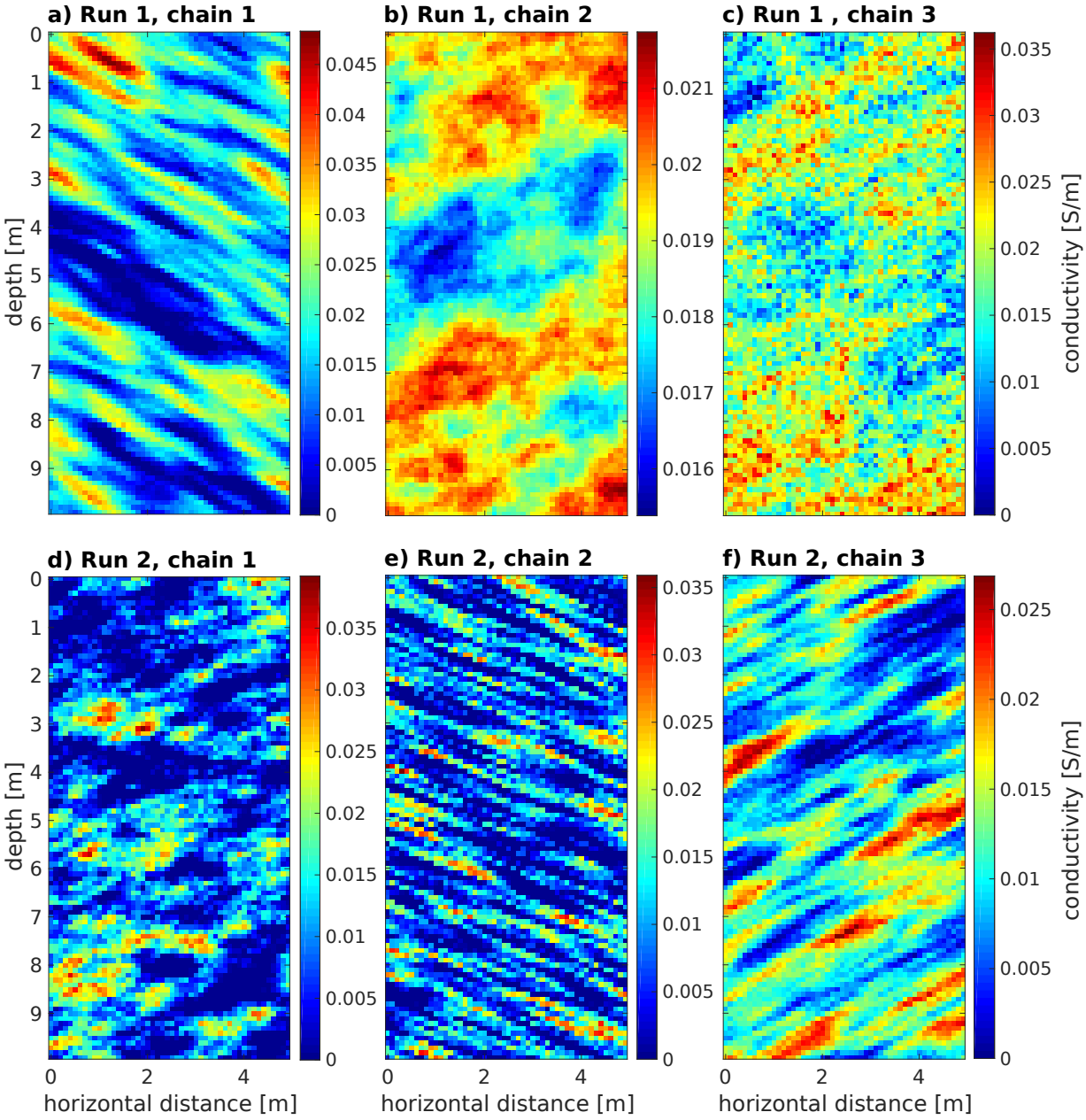

Figure 3. Electrical conductivity starting models corresponding to the permittivity models shown in Figure 2.

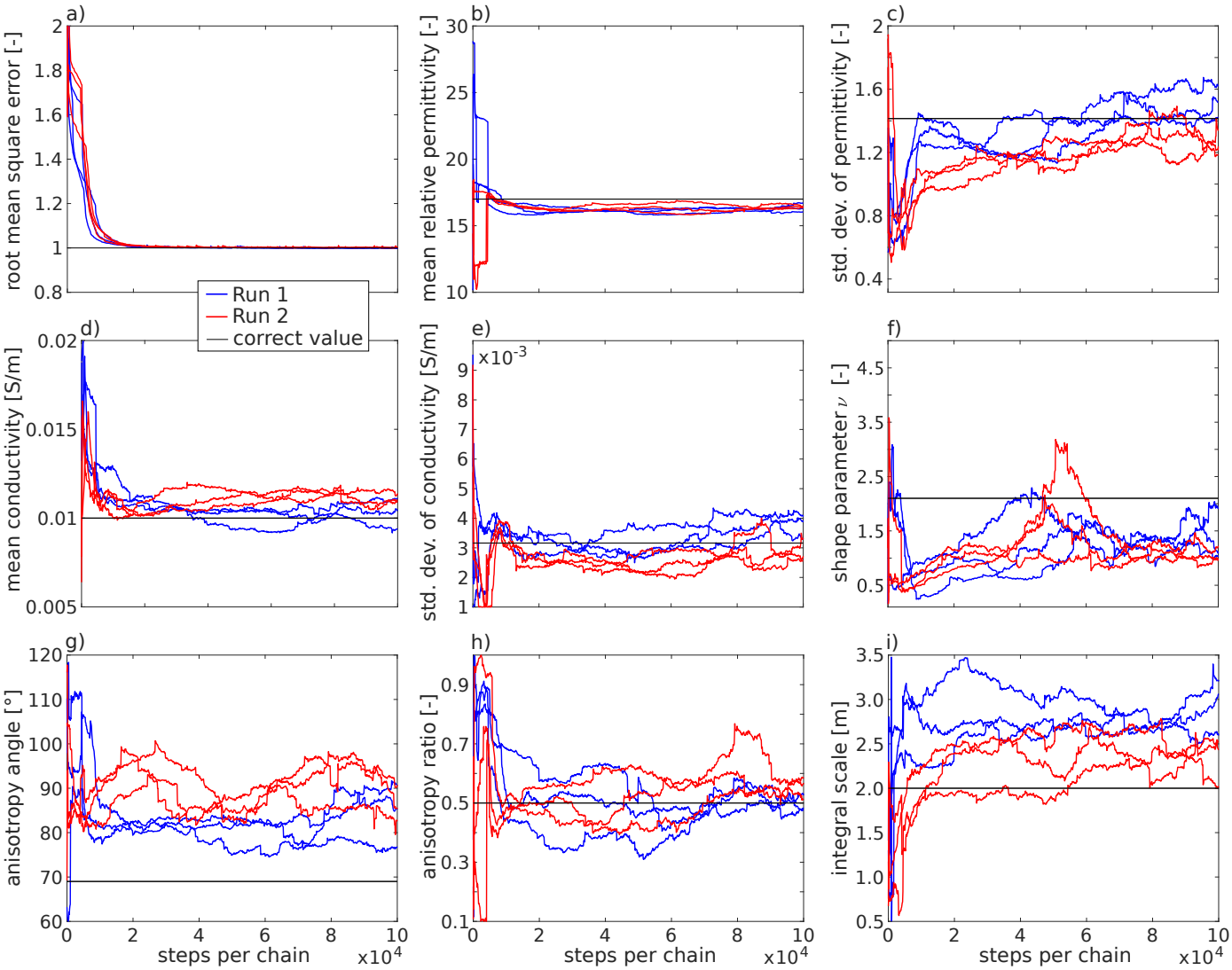

Figure 4. The inversion was run twice using three Markov chains in each run. For each run are shown: (a) the weighted root mean square error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the mean of the electric conductivity, (e) the standard deviation of the electric conductivity, (f) the shape parameter ν, (g) the anisotropy angle, (h) the anisotropy ratio and (i) the integral scale of the major axis.

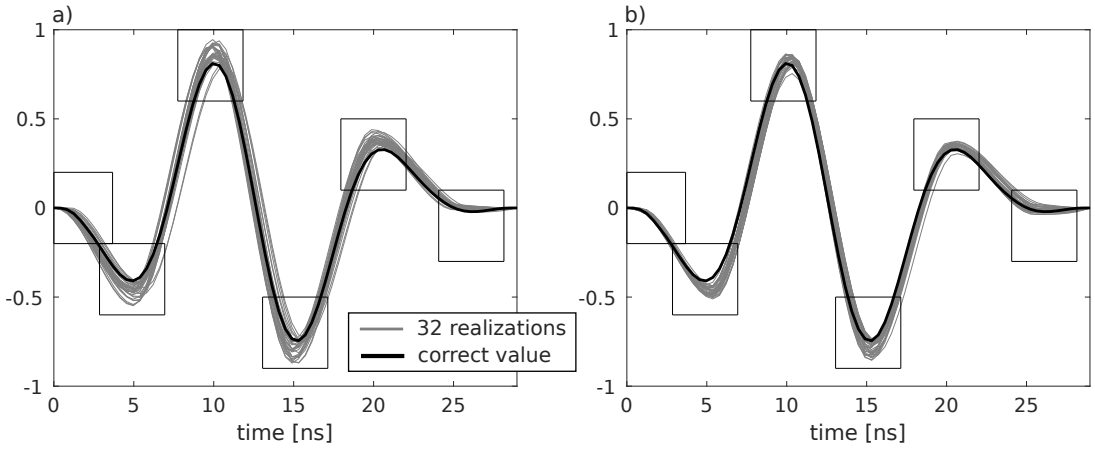

Figure 5. 32 wavelet estimates from the posterior exploration phase extracted from the Markov chains at constant intervals (gray) and the correct wavelet that was used to generate the synthetic dataset (black). The thin black boxes in both plots indicate the uniform prior PDF of the six points defining the wavelet. a) Run 1, b) run 2.

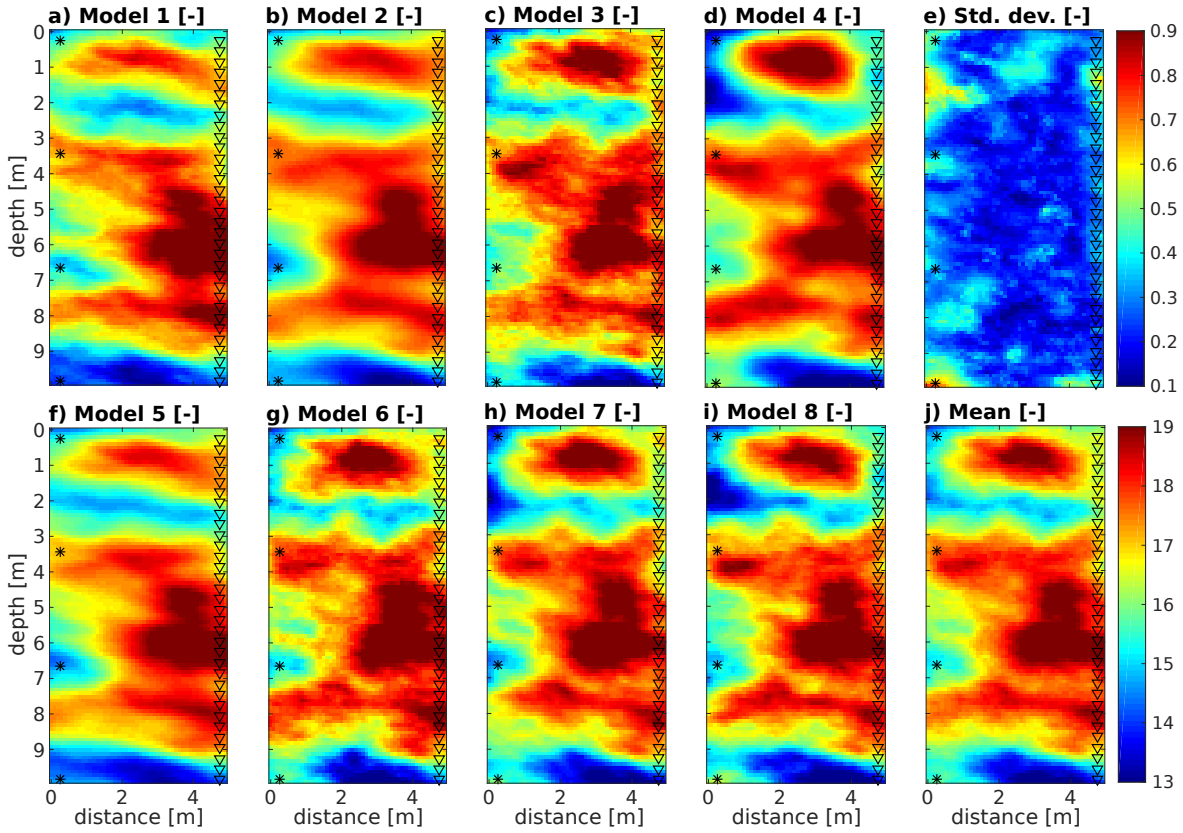

Figure 6. (a-d) and (f-i) Randomly chosen relative permittivity realizations of the posterior distribution (the last 80% of the Markov chains). (e) The standard deviation of each pixel of the subsurface structure based on 256 relative permittivity posterior realizations normalized by the overall standard deviation of the true permittivity structure (Figure 1a). (j) The mean of the same 256 relative permittivity posterior realizations. Asterisks and triangles represent sources and receivers, respectively. The colorbar of the mean permittivity structure is also valid for the individual posterior realizations.

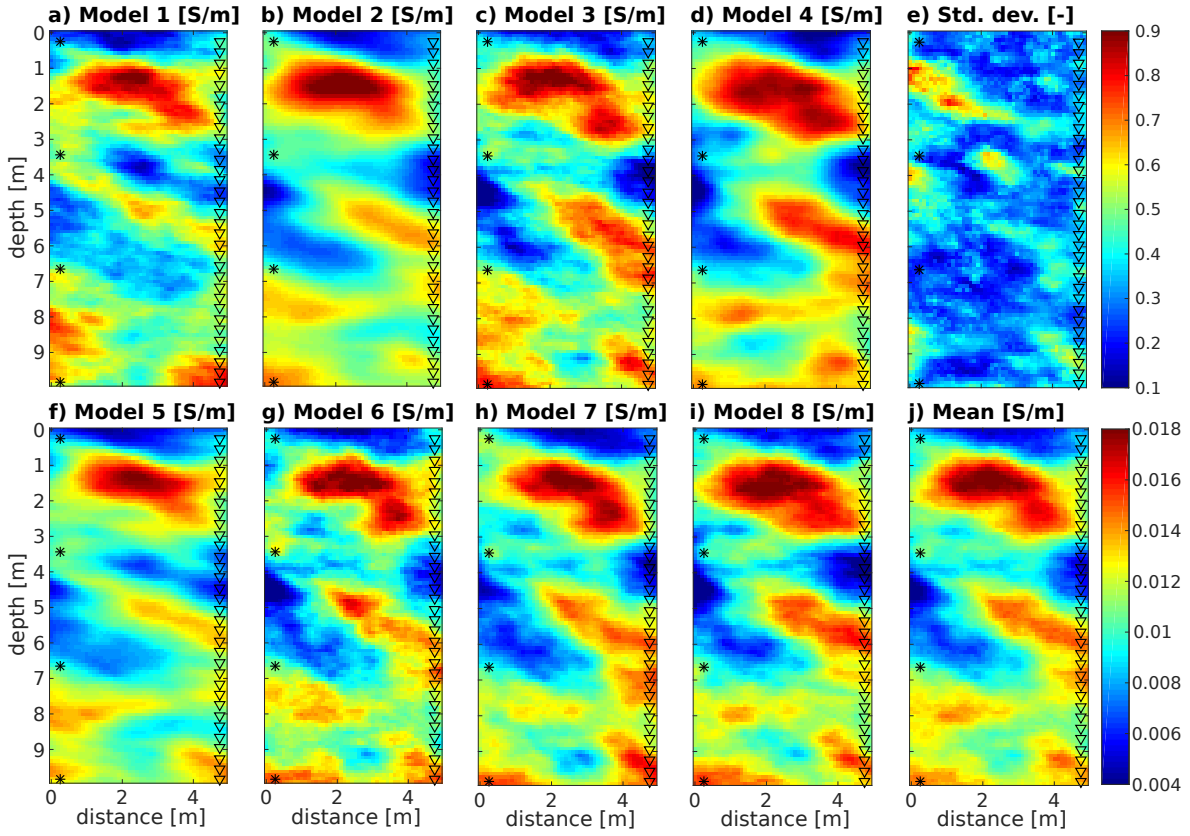

Figure 7. (a-d) and (f-i) Electric conductivity realizations of the posterior distribution corresponding to the relative permittivity model realizations shown in Figure 6. (e) The standard deviation of each pixel of the subsurface structure based on 256 relative conductivity posterior realizations normalized by the overall standard deviation of the true conductivity structure (Figure 1b). (j) The mean of the same 256 relative conductivity posterior realizations. Asterisks and triangles represent sources and receivers, respectively. The colorbar of the mean conductivity structure is also valid for the individual posterior realizations.

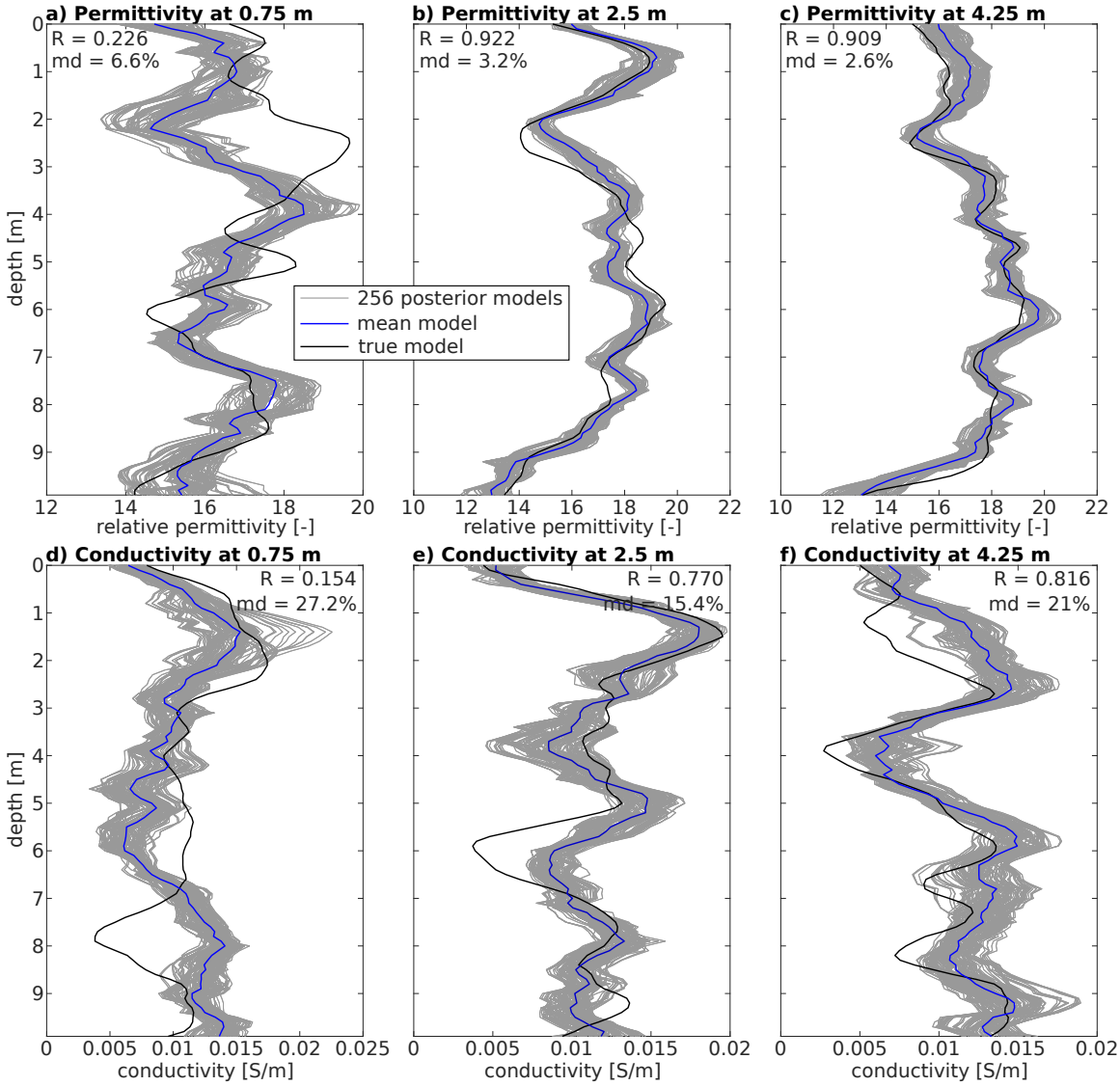

Figure 8. (a,b,c) Relative permittivity logs and (d,e,f) electric conductivity logs in virtual boreholes at a horizontal offset of (a,d) 0.75 m, (b,e) 2.5 m and (c,f) 4.25 m. The offsets correspond to 0.5 m from the source borehole, in the middle between the two boreholes and 0.5 from the receiver borehole, respectively. The gray logs correspond to the 256 posterior realizations used to compute the mean and standard deviation plotted in Figures 6e, 6j, 7e and 7j. The blue logs correspond to the mean and the black logs to the true subsurface structure at those offsets. The correlation coefficient R and the mean deviation md between the mean subsurface structure and the true subsurface structure at that offset are given in each plot. The latter is given relative to the mean permittivity or conductivity of the true subsurface structure.

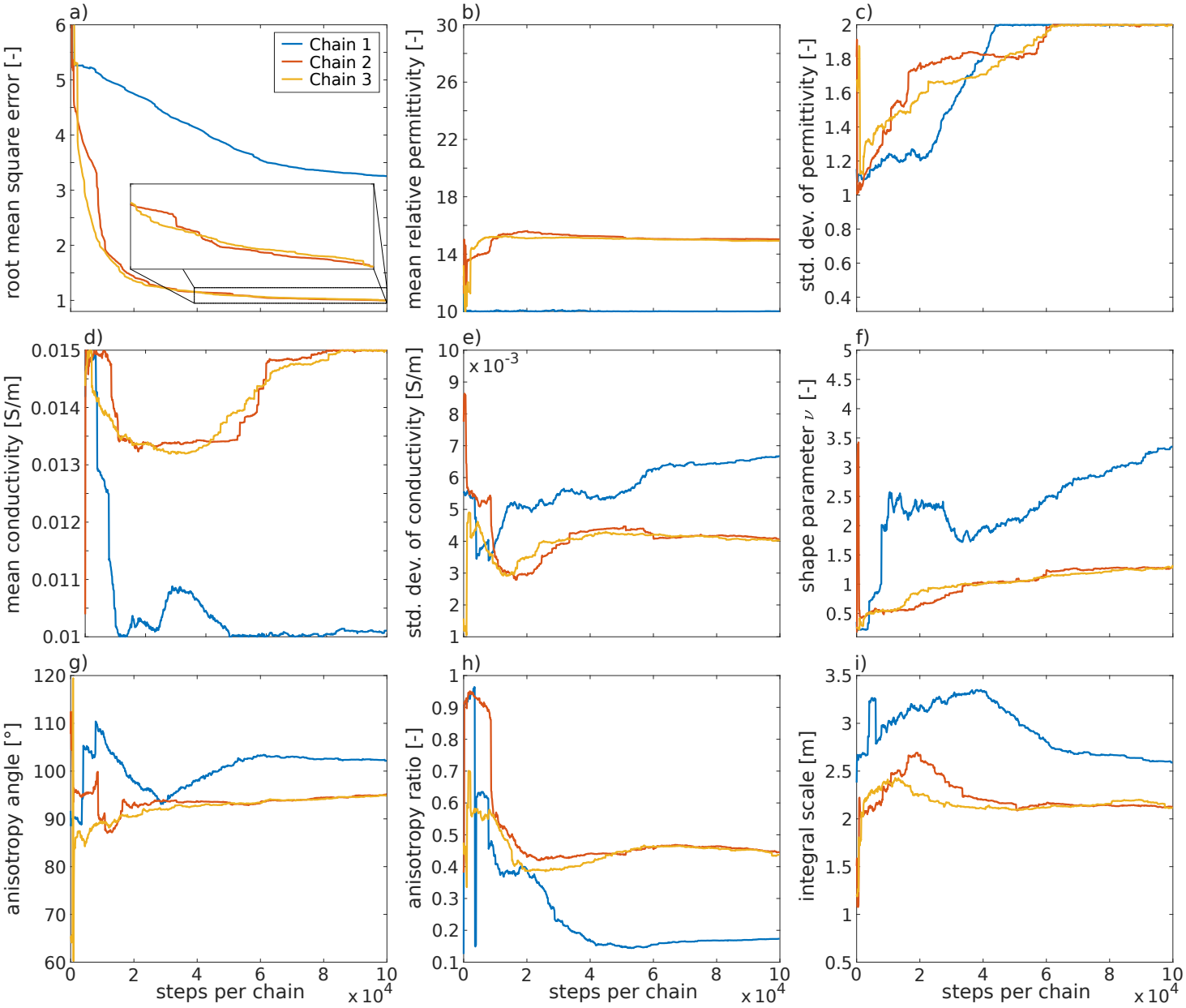

Figure 9. The inversion of the field-data example was run once using three Markov chains. For these three chains are shown: (a) the weighted root mean square error, (b) the mean of the relative permittivity, (c) the standard deviation of the relative permittivity, (d) the mean of the conductivity, (e) the standard deviation of the conductivity, (f) the shape parameter ν, (g) the anisotropy angle, (h) the anisotropy ratio and (i) the integral scale.

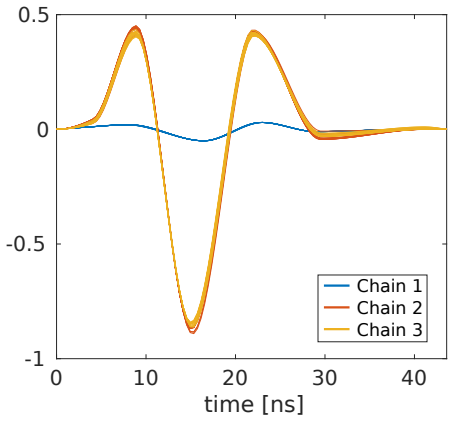

Figure 10. For each chain, 16 evenly-spaced wavelet estimates drawn from the posterior distribution are plotted.

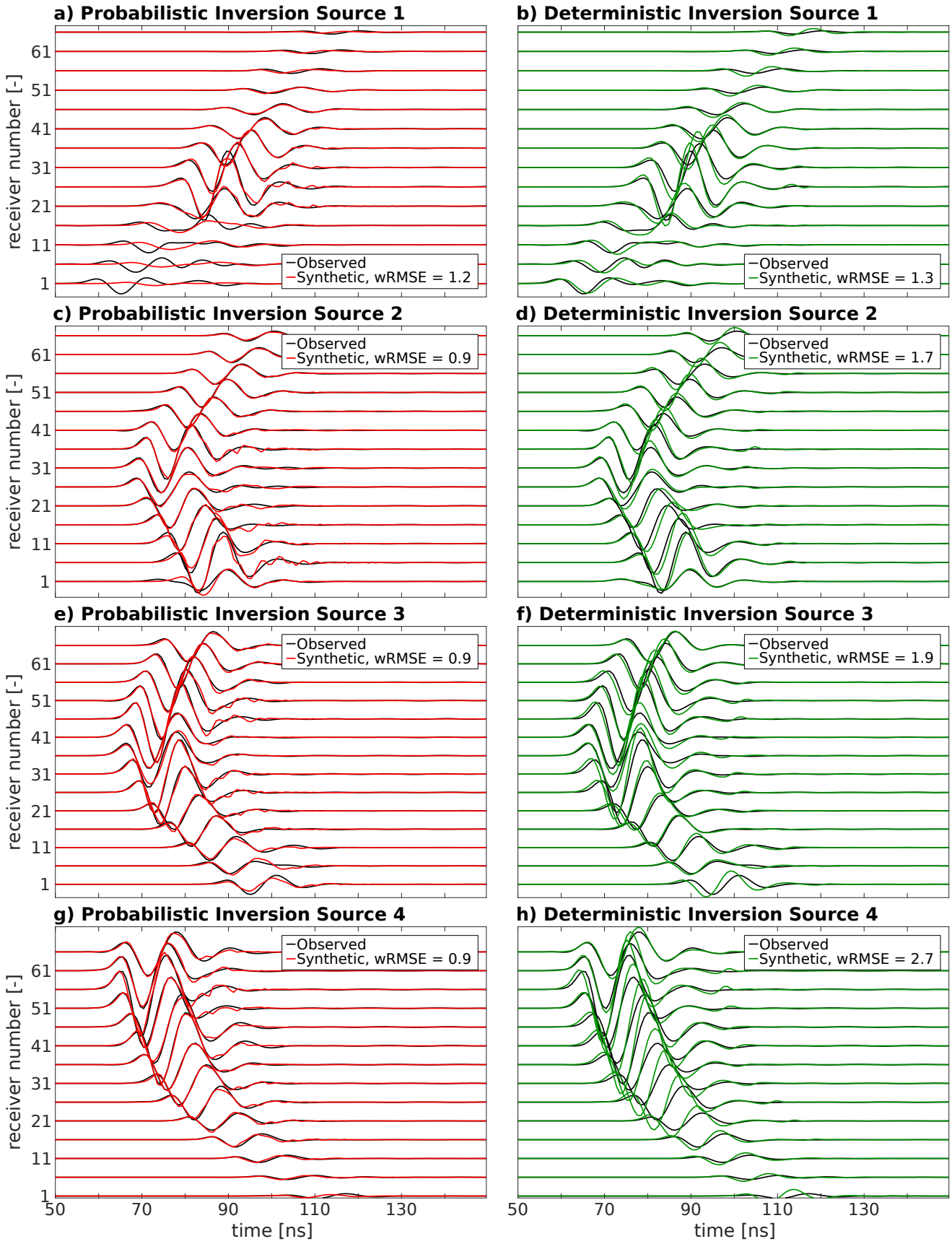

Figure 11. Comparison between observed data (black lines), simulated data based on the last model of the third Markov chain of the Bayesian inversion (red lines) and simulated data based on the model obtained from the deterministic inversion (green lines). All four source locations used in the Bayesian inversion are displayed, where source 1 is the topmost source in the borehole and source 4 the bottommost source in the same borehole. For clarity, traces are only shown for every fifth receiver. The legend gives for each source the weighted RMSE between the observed and the simulated data.

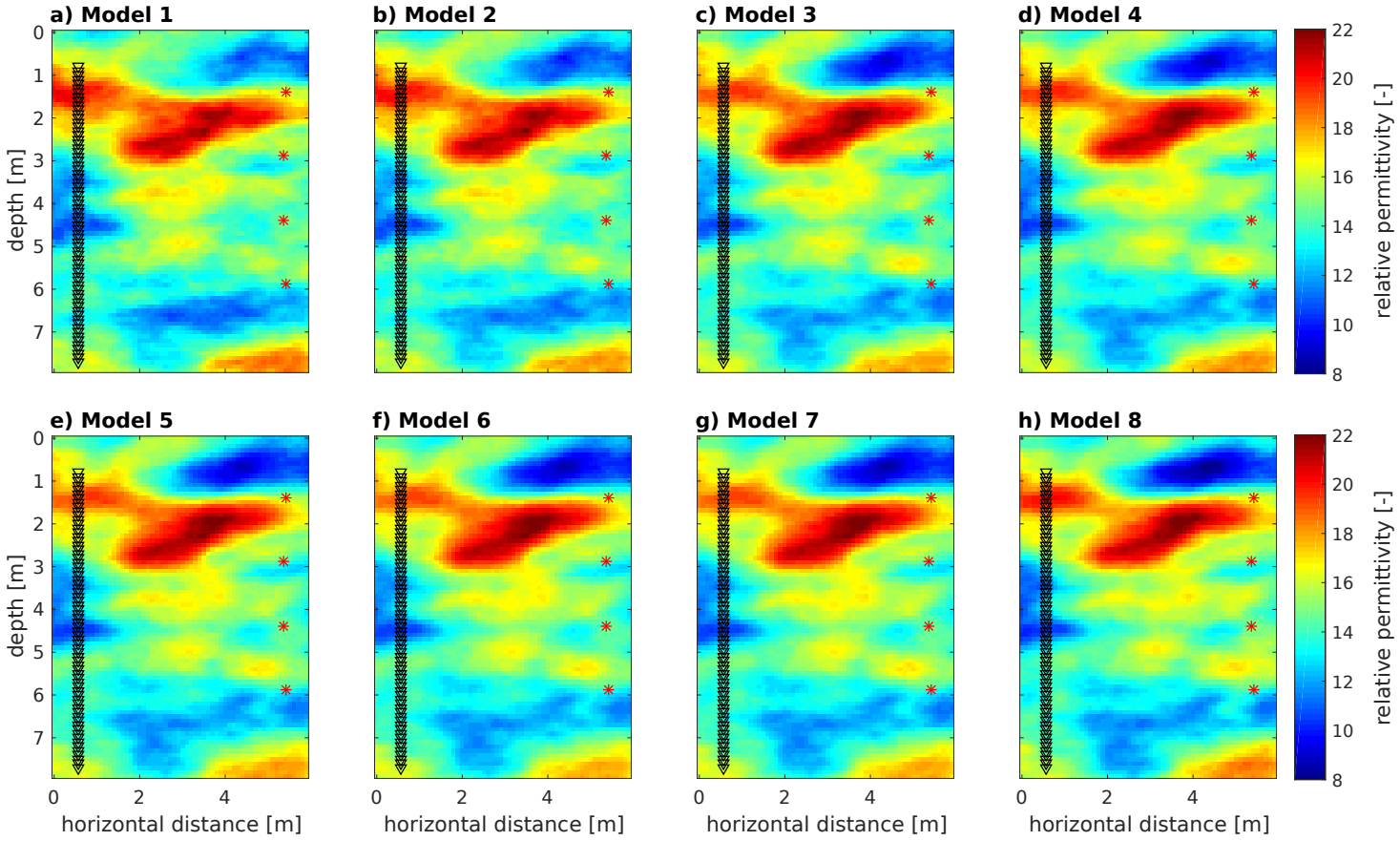

Figure 12. Randomly chosen relative permittivity realizations of the posterior distribution (the last 50% of the Markov chains). Red asterisks and black triangles represent sources and receivers, respectively.

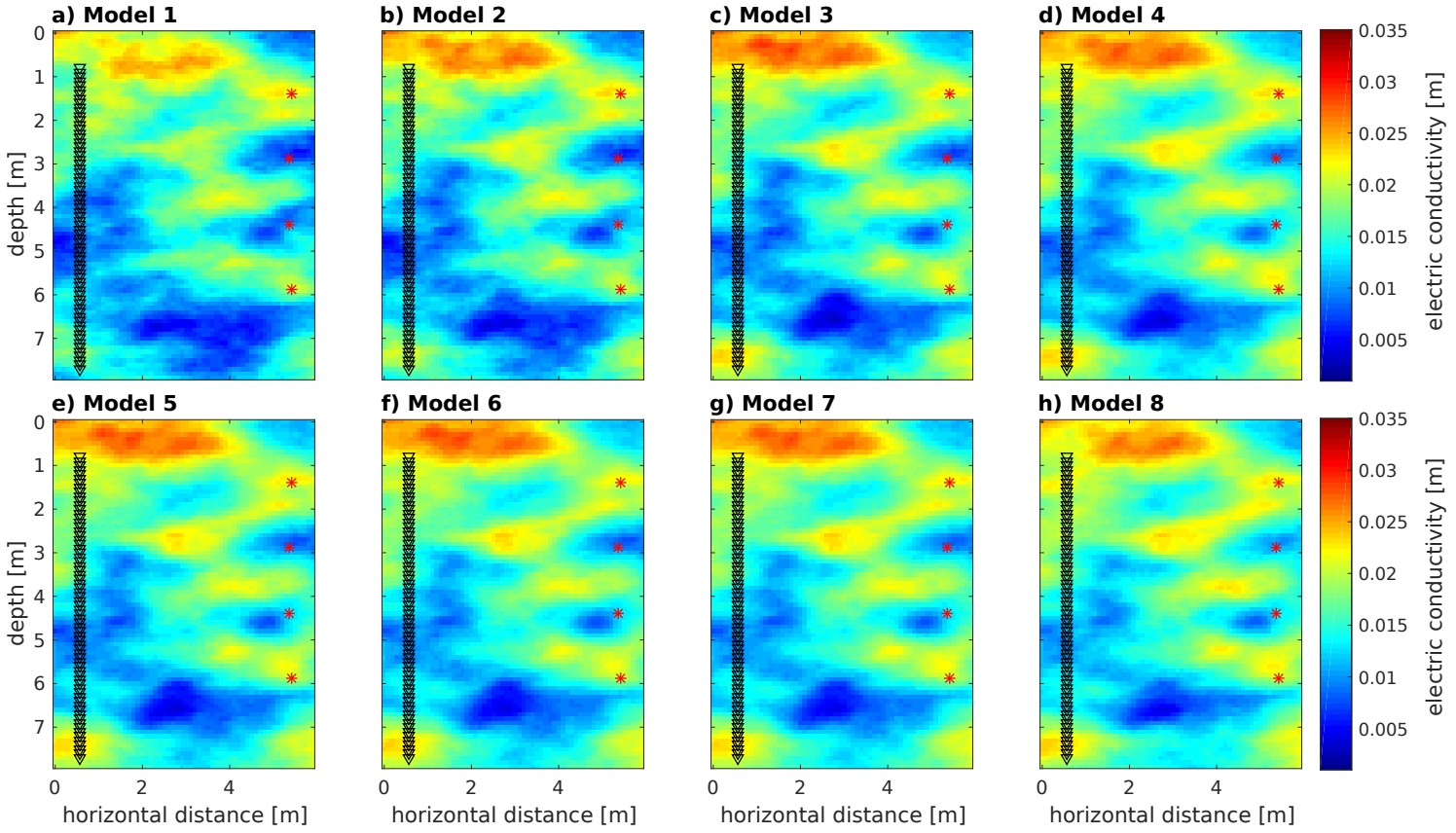

Figure 13. Electric conductivity structures corresponding to the relative permittivity model realizations shown in Figure 12. Red asterisks and black triangles represent sources and receivers, respectively.

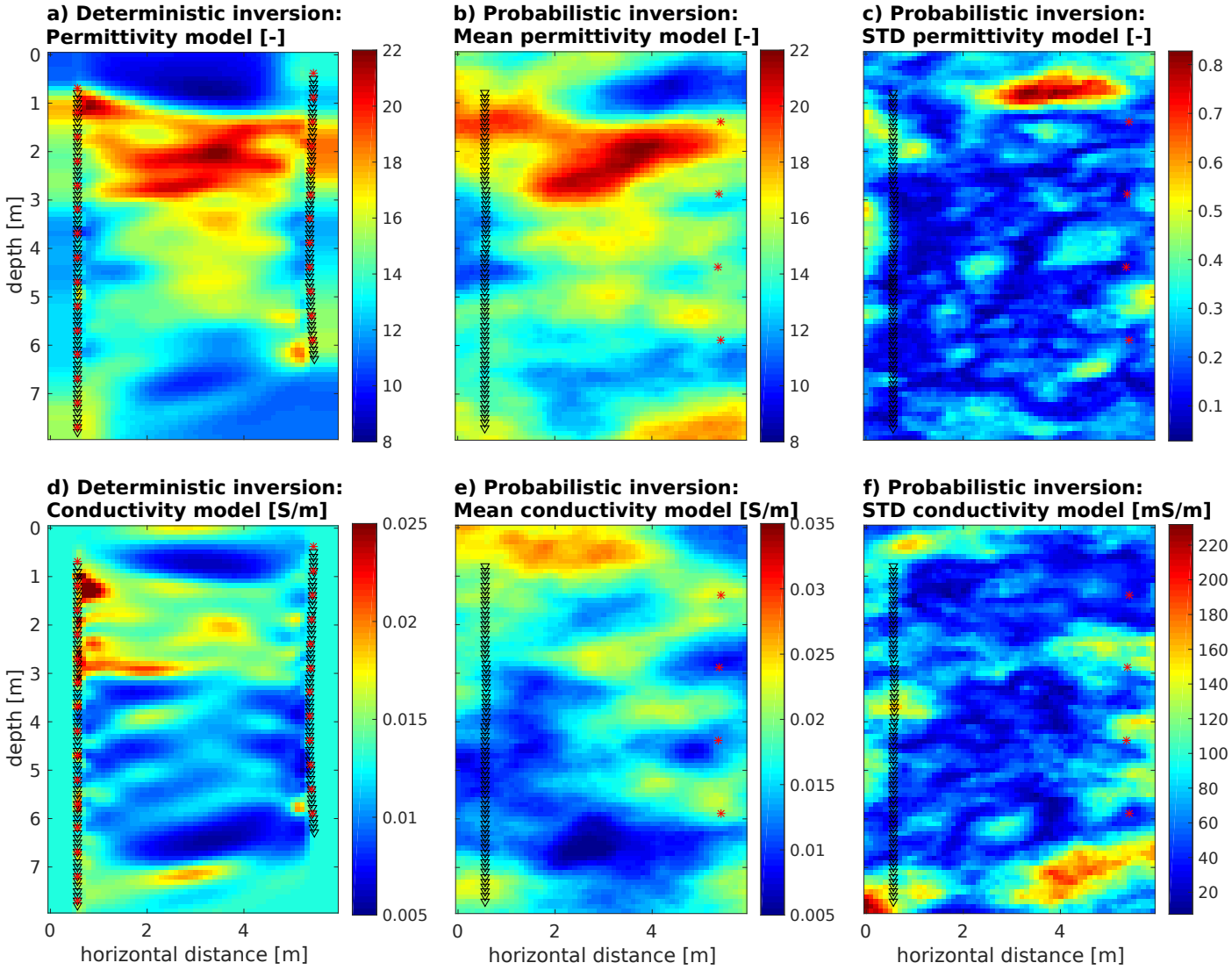

Figure 14. Comparison of deterministic and probabilistic inversion results. (a) and (d): Permittivity and conductivity model obtained from the deterministic inversion; (b) and (e): Mean permittivity and mean conductivity model obtained from the probabilistic inversion; (c) and (f): Standard deviation of the permittivity and conductivity models obtained from the probabilistic inversion. The mean and the standard deviation for the probabilistic inversion results are calculated based on 32 posterior realizations. Red asterisks and black triangles represent sources and receivers, respectively, used for the different inversion approaches.

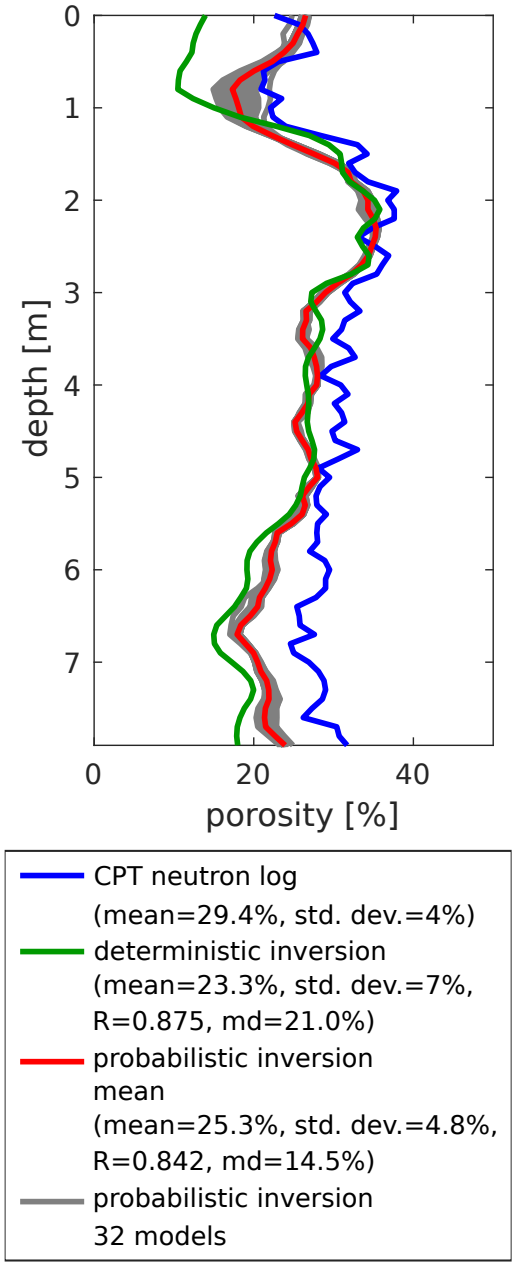

Figure 15. Comparison of the deterministic inversion results (green), the probabilistic inversion results (red and gray) and a borehole log (blue) for the porosity. The borehole is located in the middle between the two boreholes containing the GPR antennas. The permittivity model obtained from the GPR data has been converted to porosity using the CRIM equation as described by Gueting et al. (2017). The correlation coefficient R and the mean deviation md between the GPR inversion results and the borehole log is given in brackets in the legend.

LIST OF TABLES

Table 1. Unknowns in the FWI. Those marked with an asterisk (*) are unknowns of the geostatistical model.

Quantity	number of unknowns
Dimension reduction variables relative electric permittivity	250
Dimension reduction variables electric conductivity	250
*mean and standard deviation of relative electric permittivity	2
*mean and standard deviation of electric conductivity	2
*anisotropy angle and ratio	2
*integral scale	1
*shape parameter	1
Standard deviation of the data error σ_e	1
Wavelet	13
Total	522

Table 2. The geostatistical model parameters used to generate the subsurface structure depicted in Figure 1 and the corresponding prior PDF from which values are drawn in the inversion process.

Parameter	Value	Unit	Lower bound	Upper bound	distribution
Mean of the relative permittivity	17	-	10	30	uniform
Standard deviation of the relative permittivity	1.41	-	0.32	2	uniform
Mean of the electric conductivity	0.01	S/m	0.005	0.02	uniform
Standard deviation of the electric conductivity	0.0032	S/m	10^{-3}	10^{-2}	uniform
Anisotropy angle	69	$^{\circ}$	60	120	uniform
Anisotropy ratio	0.5	-	0.1	1	uniform
Integral scale of the major axis	2	m	0.5	3.5	uniform
Shape parameter	2.1	-	0.1	5	log-uniform