Technical Note

Analysis of Forward Model, Data Type, and Prior Information in Probabilistic Inversion of Crosshole GPR Data

Hui Qin [1] , Zhengzheng Wang [1] , Yu Tang [2,*] and Tiesuo Geng [1]

1 State Key Laboratory of Coastal and Offshore Engineering, Dalian University of Technology, Dalian 116024, China; hqin@dlut.edu.cn (H.Q.); wangzhengzheng@dlut.edu.cn (Z.W.); gengts@dlut.edu.cn (T.G.)
2 School of Hydraulic Engineering, Dalian University of Technology, Dalian 116024, China
* Correspondence: ytang@dlut.edu.cn

Abstract: The crosshole ground penetrating radar (GPR) is a widely used tool to map subsurface properties, and inversion methods are used to derive electrical parameters from crosshole GPR data. In this paper, a probabilistic inversion algorithm that uses Markov chain Monte Carlo (MCMC) simulations within the Bayesian framework is implemented to infer the posterior distribution of the relative permittivity of the subsurface medium. Close attention is paid to the critical elements of this method, including the forward model, data type and prior information, and their influence on the inversion results are investigated. First, a uniform prior distribution is used to reflect the lack of prior knowledge of model parameters, and inversions are performed using the straight-ray model with first-arrival traveltime data, the finite-difference time-domain (FDTD) model with first-arrival traveltime data, and the FDTD model with waveform data, respectively. The cases using first-arrival traveltime data require an unreasonable number of model evaluations to converge, yet are not able to recover the real relative permittivity field. In contrast, the inversion using the FDTD model with waveform data successfully infers the correct model parameters. Then, the smooth constraint of model parameters is employed as the prior distribution. The inversion results demonstrate that the prior information barely affects the inversion results using the FDTD model with waveform data, but significantly improves the inversion results using first-arrival traveltime data by decreasing the computing time and reducing uncertainties of the posterior distribution of model parameters.

Keywords: crosshole ground penetrating radar (GPR); probabilistic inversion; Markov chain Monte Carlo (MCMC); prior; forward model

1. Introduction

The crosshole ground penetrating radar (GPR) method has been widely applied for geological, environmental, and engineering investigations to characterize subsurface properties [1–6]. This technique uses high-frequency electromagnetic (EM) waves and transmission measurements through two boreholes to retrieve the spatial distribution of electrical properties, including mainly the dielectric permittivity and electrical conductivity, which determine the geometrical and physical features of the subsurface medium [7].

The inversion of crosshole GPR data involves estimating the dielectric parameters from the information provided by crosshole measurements [8]. In order to search for model parameters, deterministic inversion algorithms are extensively used to find an optimal solution that best fits the observed data [9–11]. In contrast, probabilistic inversion methods are capable of providing an ensemble of solutions that are statistically acceptable and quantifying parameter uncertainties [12–17]. Probabilistic inversion treats parameters and data as random variables, and Bayes theorem is often implemented to derive the posterior distribution of model parameters from the prior distribution and measured data [18]. The prior distribution reflects the knowledge of model parameters before observations, and an informative prior helps to reduce inversion uncertainties and improve the parameter

searching efficiency [19,20]. The likelihood function measures how well the model fits data by performing forward simulations and computing the distance between the simulated and measured data. According to the forward assumptions, the ray-based forward kernels use the first-arrival traveltime or first-cycle amplitude data, while the wave-based simulators employ the waveform data. The ray-based forward models are usually computational efficient, yet utilize only a small portion of the measured data that may bias the inversion result [21]. Instead, the wave-based forward models (e.g., the finite-difference time-domain (FDTD) simulator of Maxwell's equations) take advantage of the full-waveform data for a better characterization of model parameters at the price of considerable computational resources [22–24].

In our previous work, a Bayesian inversion method was developed to infer the relative permittivity values of underground structures from crosshole GPR data [25]. This method used the FDTD forward simulator with waveform data and achieved better accuracy than the traditional ray tomography. Yet one main concern of this approach is that it requires thousands to millions of forward model evaluations for the parameters to converge, leading to dozens of hours of inversion time, which limits the practicability of this method. In a later study, a straight-ray forward model was adopted and the computational efficiency was significantly improved [26]. In probabilistic inversion, the prior information is also an important aspect which is capable of reducing ambiguity by imposing constraints on the model parameters [27,28], yet improper prior distribution may distort the inversion result.

Based on the previous research, this paper focuses on the key elements of probabilistic inversion of crosshole GPR data, and systematically investigates the impact of the forward model, data type, and prior information on the inversion results. The remainder of this paper is structured as follows. First, the formulation of the probabilistic inversion method is briefly introduced. Next, the inversion results using the straight-ray model with first-arrival traveltime data, FDTD model with first-arrival traveltime data, and FDTD model with waveform data are investigated. Then, the influence of the prior information on the inversion results are analyzed, and followed by a discussion of the strengths and weaknesses of this method. Finally, this paper is concluded with a summary of the main findings.

2. Formulation of Probabilistic Inversion

The inversion of crosshole GPR data aims at estimating model parameters through measured data. In a Bayesian sense, the inverse problem can be formulated as

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{m})p(\mathbf{d}|\mathbf{m})}{p(\mathbf{d})} \propto p(\mathbf{m})L(\mathbf{m}|\mathbf{d}), \qquad (1)$$

where $p(\mathbf{m})$ and $p(\mathbf{m}|\mathbf{d})$ are the prior and posterior probability density of model parameters $\mathbf{m}$, respectively, $L(\mathbf{m}|\mathbf{d}) \equiv p(\mathbf{d}|\mathbf{m})$ is the likelihood function, and $p(\mathbf{d})$ is the probability density of data $\mathbf{d}$ that serves as a normalizing constant. In this study, the model parameters refer to the relative permittivity (ε_r) values of the subsurface structure, while the data are the first-arrival traveltimes or waveforms obtained by crosshole GPR measurements. The Bayesian approach treats all information as random variables and infers the posterior distribution of model parameters from the combination of prior knowledge and measured data.

2.1. Prior

The prior distribution summarizes all the model parameter information before collecting crosshole GPR data. In this work, a uniform distribution was used because of the lack of knowledge about the model parameters. In practice, the changes in the spatial distribution of underground media occur gradually. Hence, a smooth constraint of the model

parameters with the following form is feasible to be included in the prior information to confine the model structure [27].

$$p(\mathbf{m}) = \frac{1}{(2\pi\lambda^2)^{R_x+R_z}} \exp\left(\frac{1}{2\lambda^2}\left(\mathbf{m}^{\mathrm{T}}\left(\mathbf{W}_x^{\mathrm{T}}\mathbf{W}_x + \mathbf{W}_z^{\mathrm{T}}\mathbf{W}_z\right)\mathbf{m}\right)\right), \tag{2}$$

where $\mathbf{W}_x$ and $\mathbf{W}_z$ are the difference operators in x and z directions with rank R_x and R_z, respectively, and λ is the standard deviation of the model gradient.

2.2. Likelihood

The likelihood function describes the distance between the simulated and measured crosshole GPR data. Assuming that the measurement errors are independent and identically distributed with a normal distribution of zero mean and standard deviation of σ, the likelihood function can be written as

$$L(\mathbf{m}|\mathbf{d}) = \prod_{i=1}^{n} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{1}{2}\left(\frac{f_i(\mathbf{m}) - \mathbf{d}_i}{\sigma}\right)^2\right), \tag{3}$$

where $f_i(\mathbf{m})$ and $\mathbf{d}_i$ are the i-th simulated and measured crosshole GPR data, respectively. n denotes the total number of data. To prevent the value of Equation (3) from approaching zero when using a large number of data, it is feasible the following logarithmic form

$$l(\mathbf{m}|\mathbf{d}) = -\frac{1}{2}n\log(2\pi) - \frac{1}{2}n\log(\sigma^2) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{f_i(\mathbf{m}) - \mathbf{d}_i}{\sigma}\right)^2. \tag{4}$$

2.3. Forward Model

Evaluation of Equation (4) requires a forward model $f(\mathbf{m})$ to generate simulated data. The crosshole GPR uses EM waves for subsurface characterization, and the well-known Maxwell's equations govern this process. In two-dimensional space, the Maxwell's equations can be simplified in transverse electric (TE) mode

$$\begin{aligned} \frac{\partial E_x}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial z} - \sigma_e E_x\right) \\ \frac{\partial E_z}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_y}{\partial x} - \sigma_e E_z\right), \\ \frac{\partial H_y}{\partial t} &= \frac{1}{\mu}\left(\frac{\partial E_z}{\partial x} - \frac{\partial E_x}{\partial z}\right) \end{aligned} \tag{5}$$

where E_x and E_z are the electric fields in the x and z directions, and H_y signifies the magnetic field in the y-direction. ε, σ_e and μ are the dielectric permittivity, electric conductivity, and magnetic permeability of the subsurface medium, respectively.

As analytical solutions of Equation (5) are always tricky, numerical methods are usually used to resolve the electromagnetic fields. In this work, the FDTD algorithm is implemented for numerical solutions [29]. This approach uses central difference to discretize the Maxwell's equations in both the space and time domains, and is capable of simulating the electromagnetic field at each time step for a given time duration with high precision. Therefore, the first-arrival traveltimes, first-cycle amplitudes, and full-waveforms, which are used for the crosshole GPR inversion, can be obtained with an FDTD solver. However, within the Bayesian inversion framework, thousands to millions of forward model evaluations are needed for the model parameters to converge to their target distribution, and this CPU-intensive forward kernel makes the inversion a time-consuming task.

To improve the computational efficiency of the probabilistic inversion method, a straight-ray based forward approach is also considered [30]. This ray-based approach assumes that the propagation of the EM wave from the source to receiver is along a straight raypath l, and computes the first-arrival traveltime t by

$$t_i = \int_{l_i} s(l_i)dl_i, \tag{6}$$

where i denotes the i-th measurement, and s signifies the slowness of the EM wave along the raypath. s is the reciprocal of velocity v derived by $v = c/\sqrt{\varepsilon_r}$, in which c is the EM wave velocity in vacuum. Equation (6) simplifies the EM wave propagation into a straight ray and runs much faster than the FDTD model. However, with the straight-ray model, only the first-arrival traveltime data can be simulated [31].

2.4. Posterior

MCMC simulation is adopted to derive the posterior distribution of model parameters. This method uses a Markov chain that generates a trial move from the current state, $\mathbf{m}^{(i)}$ to a new position, $\mathbf{m}^*$, and decides whether to accept this jump using the Metropolis ratio [32]

$$p_a\left(\mathbf{m}^{(i)}, \mathbf{m}^*\right) = \min\left(1, \frac{p(\mathbf{m}^*)}{p(\mathbf{m}^{(i)})}\right), \tag{7}$$

where $p_a(\cdot)$ is the acceptance probability. In this paper, the DREAM$_{(ZS)}$ package [33], which is an adaptive MCMC algorithm designed for high-dimensional problems, was used to explore the posterior distribution of model parameters [34].

3. Reference Model and Data

In this section, a synthetic model is created for simulating crosshole GPR detection of a defect in an underground concrete structure. As shown in Figure 1a, a 1.0 m $\times$ 1.0 m model was built with $\varepsilon_r = 9$ to simulate the concrete structure (background), and a 0.2 m $\times$ 0.2 m higher relative permittivity ($\varepsilon_r = 12$) area was created as the defect (target). Crosshole GPR measurements were simulated using 11 transmitters on the left side and 26 receivers on the right side of the model.

In order to obtain crosshole GPR data, the gprMax software [35,36], which is an FDTD solver of the Maxwell's equations, was used to perform numerical simulations. The model depicted in Figure 1a is discretized, with a grid size of 0.02 m $\times$ 0.02 m, leading to a total number of $50 \times 50 = 2500$ unknown relative permittivity values, subject to estimate. The FDTD simulations use the Ricker wavelet with a center frequency of 500 MHz as the source, and collect a 20 ns GPR waveform for each transmitter–receiver pair.
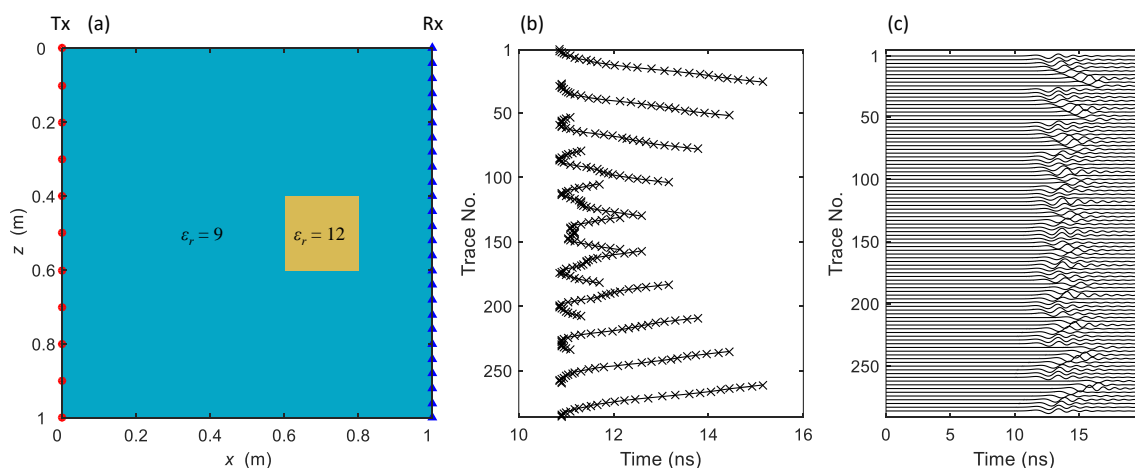

Figure 1. (**a**) Reference relative permittivity (ε_r) model, (**b**) synthetic first-arrival traveltime data, and (**c**) waveform data. The red dots and blue triangles denote the positions of crosshole ground penetrating radar (GPR) transmitters and receivers, respectively.

At each transmitter position, GPR waveforms were collected at all locations of the 26 receivers. A total of $11 \times 26 = 286$ GPR traces were obtained as waveform data, and the first breaks of the GPR traces were extracted as the first-arrival traveltime data. Figure 1b,c display the synthetic traveltime and waveform data, respectively. Then, these data were perturbed with white noise and served as the measurement data. For the first-arrival

traveltime data, a zero mean and standard deviation of 3% of the mean value (0.24 ns) Gaussian noise was used as the measurement error. For the waveform data, the signal-to-noise ratio (SNR) was used to control the noise level, and a Gaussian random noise with SNR = 20 was generated to simulate the measurement error, which was equivalent to a standard deviation of 2.06.

4. Effect of Forward Model and Data Type

In this section, the impact of the forward model and data type on the inversion results is investigated. The straight-ray model computes only first-arrival traveltime data, whereas the FDTD model simulates the EM fields, from which both first-arrival traveltime and waveform data can be obtained. Therefore, three cases are discussed in this section, which are the straight-ray model with first-arrival traveltime data, the FDTD model with first-arrival traveltime data, and the FDTD model with waveform data.

The forward model and discretized grid size uses the same values of the reference model shown in Figure 1a, thus a total number of 2500 relative permittivity values need to be estimated. The discrete cosine transform (DCT) strategy was adopted for dimensionality reduction [37,38], and 256 DCT-coefficients were used to represent the full parameter space. The simulated data were computed using source and receiver positions identical to those used in the synthetic measurements, which produced $11 \times 26 = 286$ first-arrival traveltimes for the straight-ray model, and 286 GPR traces for the FDTD model. For each GPR trace, the waveform from 10 to 20 ns, which contains most of the information, was used as the simulated waveform data. Each waveform consists of 212 data points, and a total of $286 \times 212 = 60,632$ data points constitute the waveform data for the FDTD model.

MCMC simulation with the $DREAM_{(ZS)}$ algorithm was deployed to sample the posterior distribution of model parameters. A uniform distribution of model parameters was used in this section as non-informative prior, and Equation (4) was adopted as the likelihood function. The setup of the probabilistic inversions with different forward models and data types is summarized in Table 1. The inversion results are discussed in the following subsections.

Table 1. Bayesian inversion parameters using uniform prior.

Case No.	Forward Model	Data Type	No. Data	No. Parameters	Prior Distribution	Prior Range
1	Ray Model	Traveltime Data	286	256	Uniform	0–40
2	FDTD Model	Traveltime Data	286	256	Uniform	0–40
3	FDTD Model	Waveform Data	60,632	256	Uniform	0–40

4.1. Straight-Ray Model with Traveltime Data

First, the straight-ray forward model with first-arrival traveltime data was used to infer the relative permittivity field. After 1,440,000 model evaluations, all parameters converged to their target distribution, and the last 50% of the parameters after convergence were considered as posterior samples. Figure 2 displays the histograms of the prior (blue bars) and posterior (red bars) distribution of three randomly selected parameters. Before inversion, the parameters were uniformly distributed in the prior range. After inversion, the parameters moved toward the true values marked with black crosses. However, the mean values were estimated with large errors, and the posterior parameters were distributed in a wide range with large variance.

Figure 3a–c present three randomly selected posterior realizations. They show significant differences in model structure, and none of them resembles the true relative permittivity model. Figure 3d,e illustrate the maximum a-posteriori (MAP) and posterior mean realizations, respectively, which fail to reconstruct the correct model structure again. Figure 3f shows a large variance of posterior model parameters, which means the inverted parameters have large uncertainties. Therefore, with the straight-ray forward model and

first-arrival traveltime data, the model parameters converge to a broad range, and the posterior models do not recover the relative permittivity field correctly.
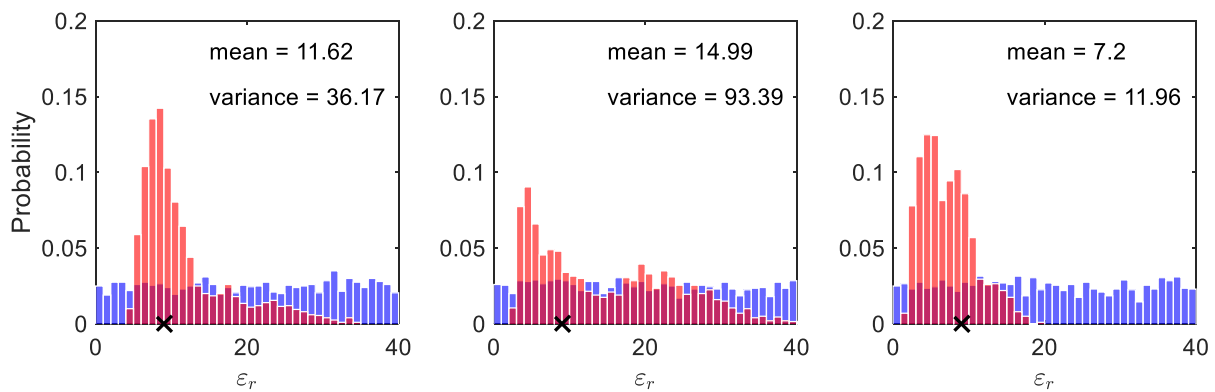

Figure 2. Histograms of the prior (blue bars) and posterior (red bars) distribution of three randomly selected parameters derived from Bayesian inversion using the straight-ray forward model with first-arrival traveltime data. The true values of the parameters are denoted with black crosses.
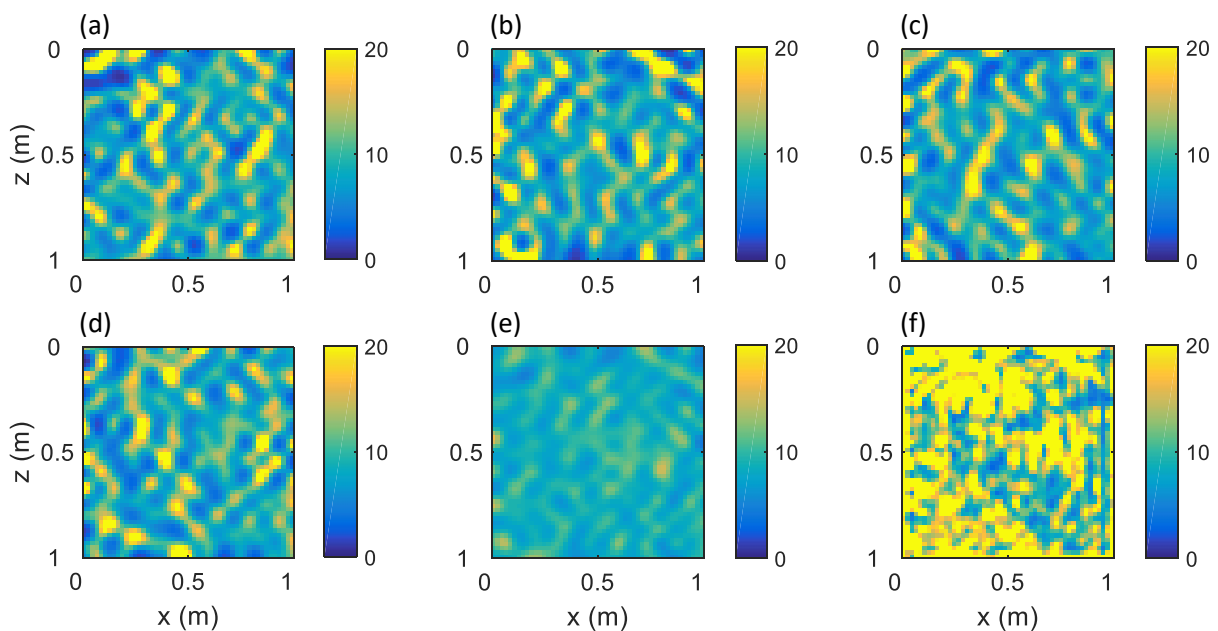

Figure 3. Posterior solutions using the straight-ray model with first-arrival traveltime data: (**a**–**c**) three randomly selected realizations from the posterior distribution, (**d**) the maximum a-posteriori (MAP) density solution, (**e**) the posterior mean realization, and (**f**) variance of the posterior solutions.

Figure 4 shows the data fit of the inverted models. It can be seen that the posterior simulated data are in good agreement with the measured data, and the root mean squared error (RMSE) values of the posterior models are distributed around the measurement error (0.24 ns), which means the posterior models fit the data very well. However, the posterior realizations differ significantly from the actual model (see Figure 3), indicating that a large number of different model parameters exists that explain the measured data very well. Therefore, the information provided by the first-arrival traveltime data is insufficient to retrieve the correct relative permittivity field.
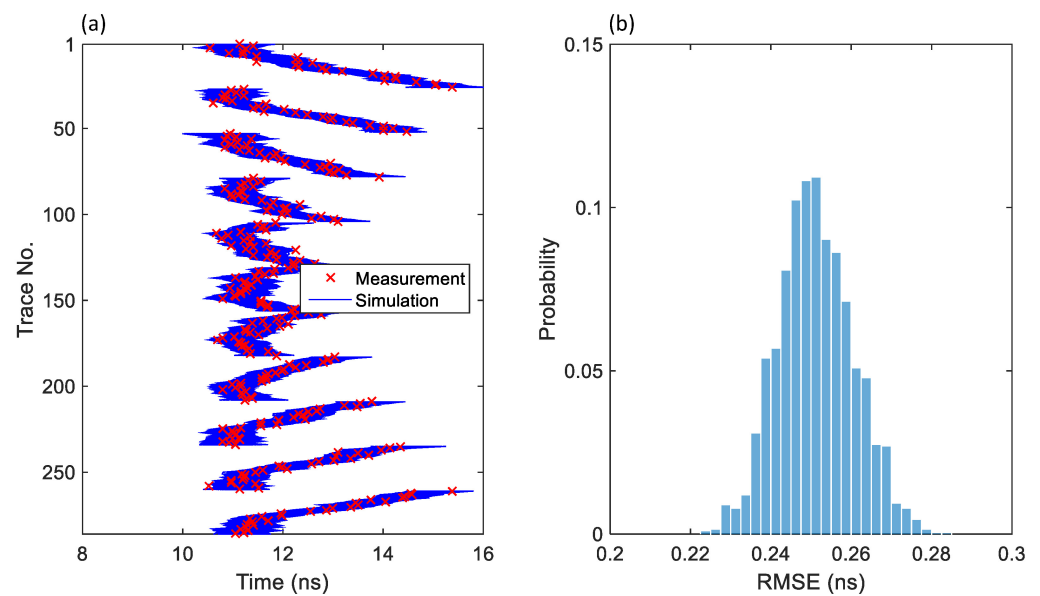
Figure 4. Fitting of the measurement data by the posterior models derived from Bayesian inversion using the straight-ray forward model: (**a**) measured and simulated first-arrival traveltimes, and (**b**) histogram of the root mean squared error (RMSE) values of the posterior solutions.

4.2. FDTD Model with Traveltime Data

In the second experiment, the FDTD forward model with first-arrival traveltime data was evaluated. A total of 1,638,000 iterations were required for all parameters to converge to their posterior distribution. Figure 5 summarizes all posterior samples and illustrates the prior and posterior distribution of three randomly selected parameters. After Bayesian inversion, the model parameters were sampled around their actual values. Compared with the results using the straight-ray forward model, the posterior model parameters derived from the FDTD model converged to a smaller range with reduced variances.
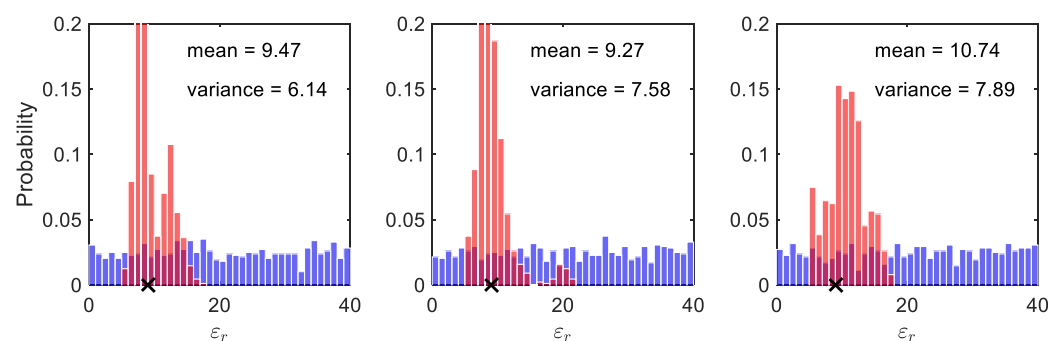

Figure 5. Histograms of the prior (blue bars) and posterior (red bars) distribution of three randomly selected parameters derived from Bayesian inversion using the finite-difference time-domain (FDTD) forward model with first-arrival traveltime data. The true values of the parameters are denoted with black crosses.

The posterior reconstructed relative permittivity models are displayed in Figure 6. Similar to the realizations obtained from the inversion using the straight-ray forward model, the FDTD model with first-arrival traveltime data also fails to resolve the reference model correctly. Although the FDTD forward model reduces parameter uncertainties, it is not sufficient to produce relative permittivity models that resemble the true one.

The simulated first-arrival traveltimes of the posterior parameters using the FDTD forward model are shown in Figure 7a. Unlike the case using the straight-ray forward model, the posterior simulated data using the FDTD model are distributed in a much

smaller range close to the measured data, and the RMSE values of the posterior models are slightly higher than the measurement error (see Figure 7b). In this case, the posterior models using the FDTD forward simulations with first-arrival traveltime data explain the observed data very well with smaller uncertainties, yet they still failed to converge to the actual relative permittivity field.
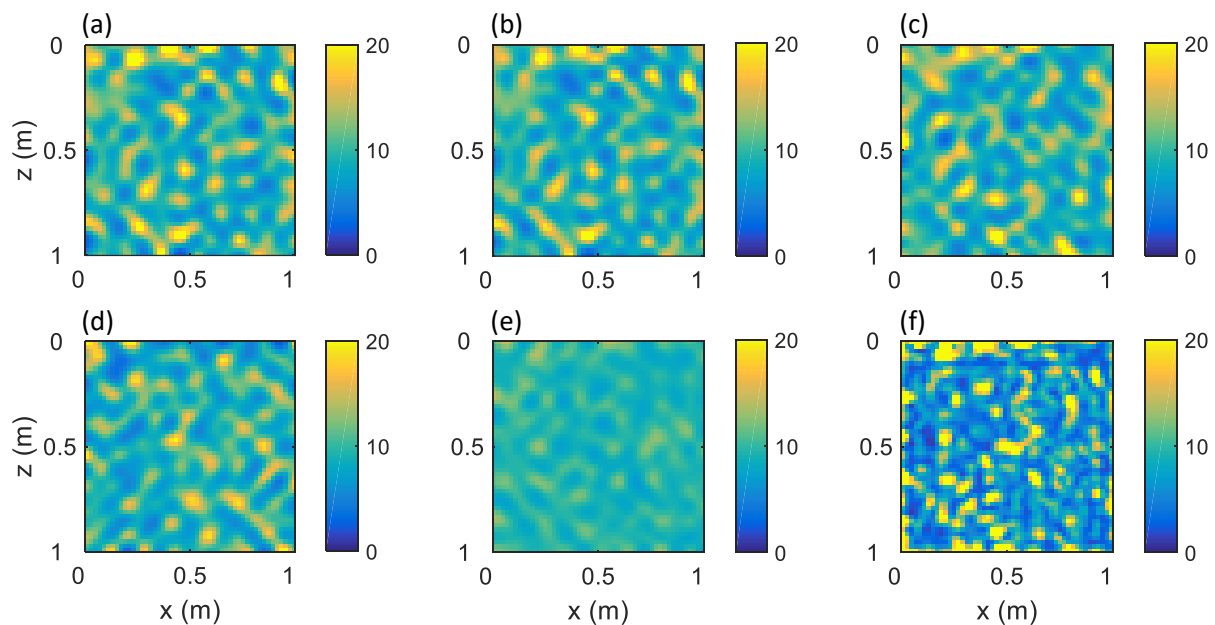

Figure 6. Posterior solutions using the FDTD model with first-arrival traveltime data: (**a**–**c**) three randomly selected realizations from the posterior distribution, (**d**) the maximum a-posteriori (MAP) density solution, (**e**) the posterior mean realization, and (**f**) variance of the posterior solutions.
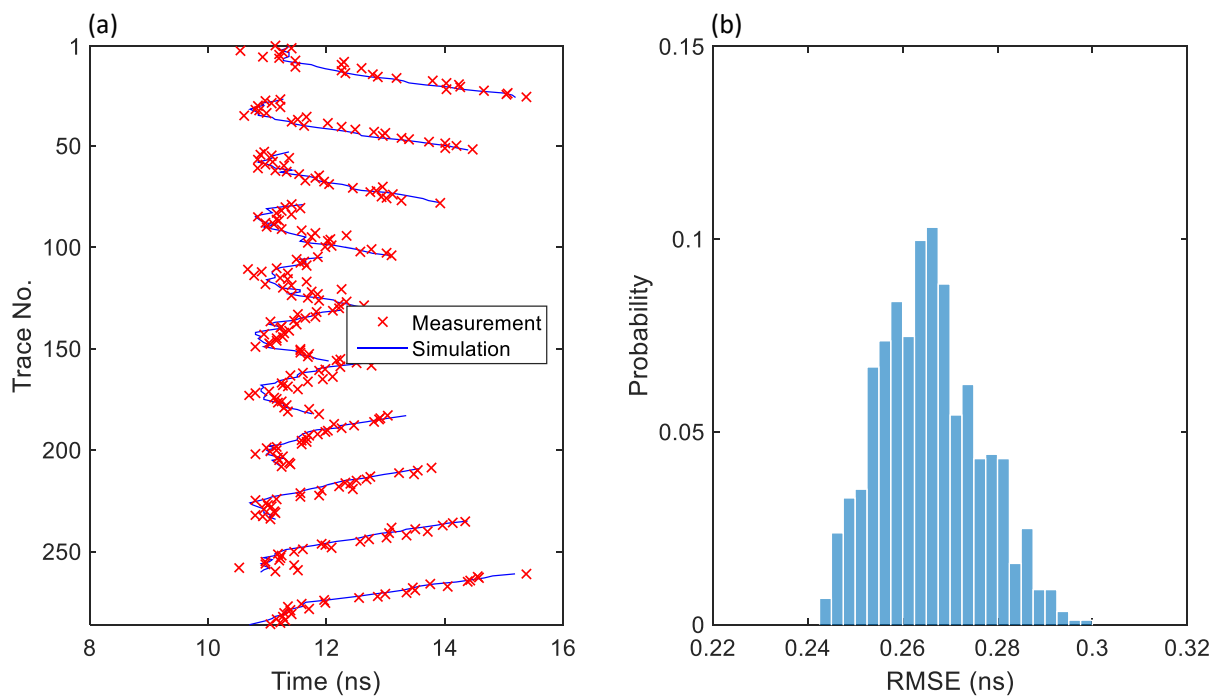

Figure 7. Fitting of the measurement data by the posterior models derived from Bayesian inversion using the FDTD forward model: (**a**) measured and simulated first-arrival traveltimes, and (**b**) histogram of the RMSE values of the posterior solutions.

4.3. FDTD Model with Waveform Data

Next, the use of the FDTD forward model with waveform data to infer the relative permittivity field was investigated. All the parameters converged to their target distribution after 156,000 model evaluations, much less than the cases using the first-arrival traveltime data. The histograms of the prior and posterior distribution of three randomly selected parameters are presented in Figure 8, in which the posterior parameters are distributed in a very narrow range and focus on the true values.
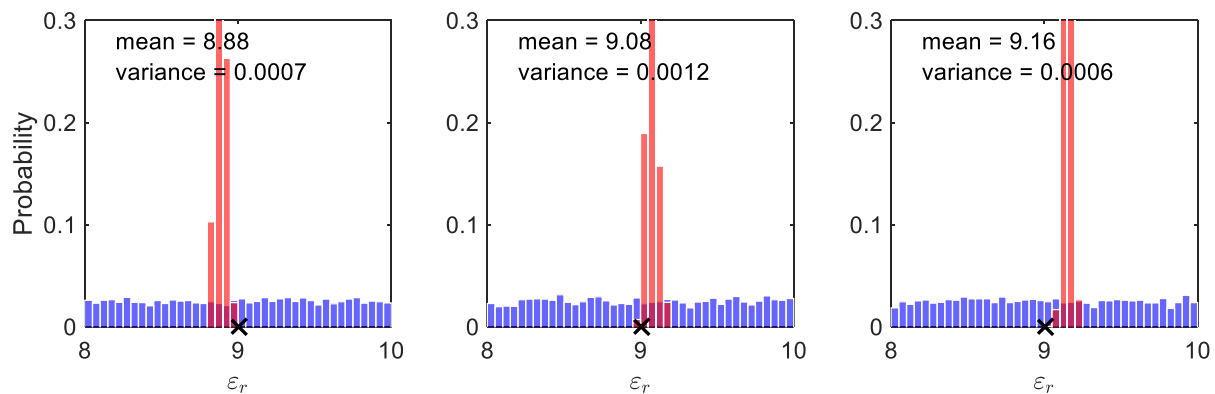
Figure 8. Histograms of the prior (blue bars) and posterior (red bars) distribution of three randomly selected parameters derived from Bayesian inversion using the FDTD forward model with waveform data. The true values of the parameters are denoted with black crosses.

The posterior realizations shown in Figure 9 successfully reconstruct the proper relative permittivity field. The MAP and posterior mean models present similar patterns and are visually identical to the randomly selected posterior realizations. Small variances are observed, indicating that the waveform data provide sufficient information for the inversion algorithm to retrieve the correct model parameters.
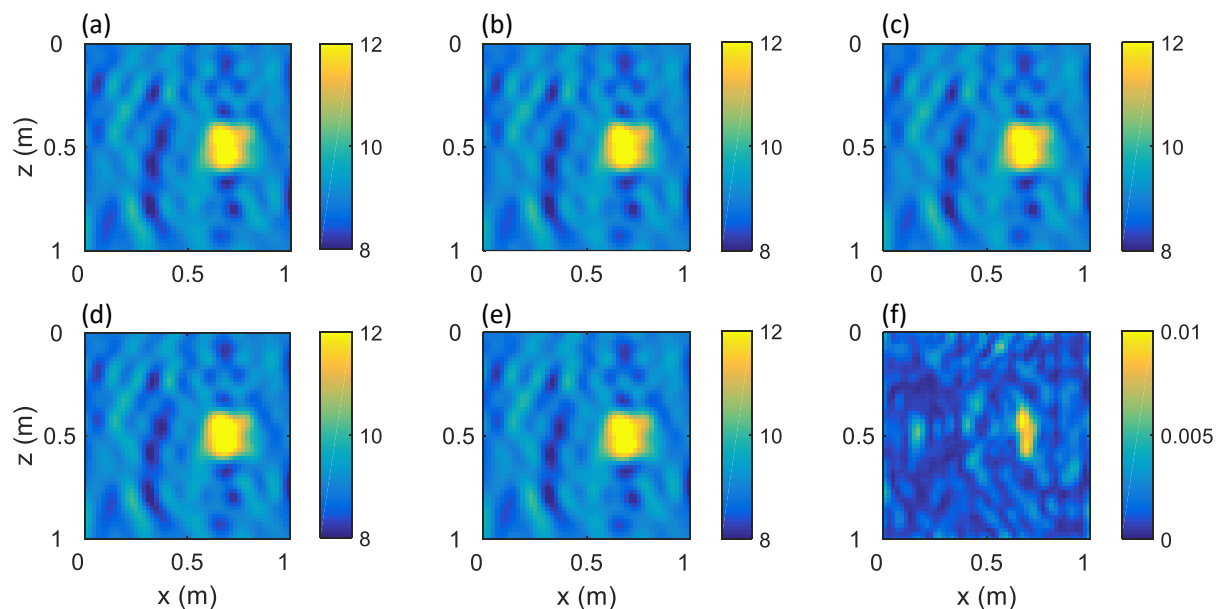
Figure 9. Posterior solutions using the FDTD model with waveform data: (**a**–**c**) three randomly selected realizations from the posterior distribution, (**d**) the maximum a-posteriori (MAP) density solution, (**e**) the posterior mean realization, and (**f**) variance of the posterior solutions.

Figure 10a presents the simulated waveforms of the posterior parameters using the transmitter position of $z = 0.5$ m and all receiver positions from 0 to 1 m. The posterior

models exhibit an excellent fit to the observed waveforms. Figure 10b displays the distribution of the RMSE values of the posterior solutions, which is close to the real value of the measurement error. Therefore, the FDTD forward model with waveform data is capable of characterizing the relative permittivity field.
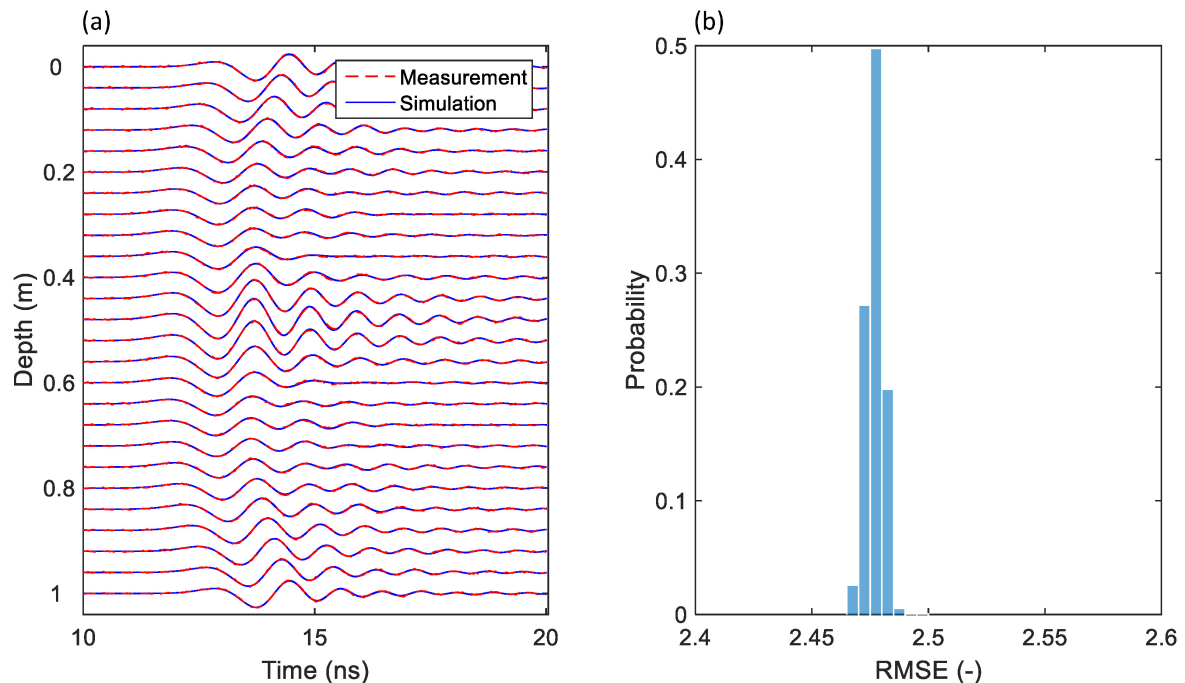

Figure 10. Fitting of the measurement data by the posterior models derived from Bayesian inversion using the FDTD forward model: (**a**) measured and simulated waveforms, and (**b**) histogram of the RMSE values of the posterior solutions.

The inversion results of the three cases using uniform prior distribution are summarized as follows. With first-arrival traveltime data, neither the straight-ray model nor the FDTD model reconstructs the relative permittivity field correctly. When using the waveform data, the model parameters converge to the real values, and the model structure is recovered correctly. This is because there are 2500 unknown model parameters, yet the first-arrival traveltime data contain only 286 data points, far less than the number of unknowns. Without additional knowledge about model parameters, a large number of permittivity models exist that fit the data well. Therefore, the information provided by the first-arrival traveltime data is not enough to lead the model parameters to their correct values. On the contrary, the waveform data, which contain 60,632 data points, are much more informative to infer the model parameters correctly.

5. Effect of Prior Information

In this section, the informative prior defined in Equation (2) is considered, and the impact of prior information on the inversion results is discussed. Three inversion cases including the straight-ray model with first-arrival traveltime data, the FDTD model with first-arrival traveltime data, and the FDTD model with waveform data are analyzed. Except for the prior distribution, the same inversion setup was used as that in the previous section. Table 2 summarizes the inversion parameters of the three cases.

When the prior information was accounted for, it cost 144,000, 236,800, and 156,000 model evaluations for the three cases to converge to their posterior distribution. For the cases using the first-arrival traveltime data, much fewer iterations were needed compared with their flat-prior counterparts. However, when the waveform data were used, the prior distribution had little influence on the convergence rate. Figure 11 illustrates the distribution of a randomly selected parameter for each case. It is obvious that for all three cases, the parameter converges to a small range and is distributed around the actual value.

The prior information reduces the parameter uncertainty substantially for the cases using the first-arrival traveltime data, yet has little impact on the parameter distribution using waveform data.

Table 2. Bayesian inversion parameters using smooth constraint as prior information.

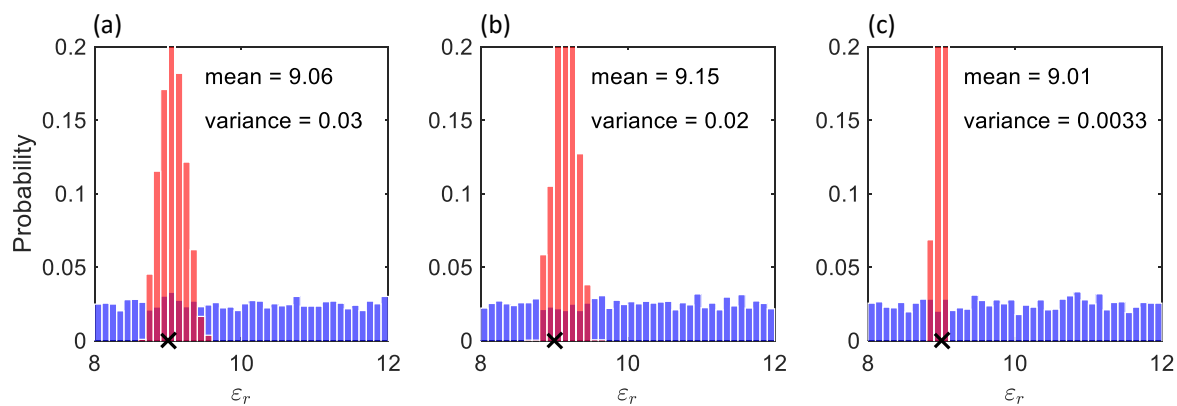| Case No.	Forward Model	Data Type	No. Data	No. Parameters	Prior Distribution	Prior Range
4	Ray Model	Traveltime Data	286	256	Smooth Constraint	0–40
5	FDTD Model	Traveltime Data	286	256	Smooth Constraint	0–40
6	FDTD Model	Waveform Data	60,632	256	Smooth Constraint	0–40

Figure 11. Histograms of the prior (blue bars) and posterior (red bars) distribution of three randomly selected parameters derived from Bayesian inversion with smooth model constraint using (**a**) the straight-ray forward model with first-arrival traveltime data, (**b**) FDTD model with first-arrival traveltime data, and (**c**) FDTD model with waveform data. The true values of the parameters are denoted with black crosses.

Figure 12 displays the posterior realizations using the straight-ray model with first-arrival traveltime data, the FDTD model with first-arrival traveltime data, and the FDTD model with waveform data, respectively. For the two cases using the first-arrival traveltime data, the use of prior information improves inversion results significantly so that the higher permittivity area can be identified in the MAP and posterior mean realizations clearly. The MAP models exhibit better spatial resolution, yet suffers from more noise. The posterior mean models, on the other hand, present a smoother pattern. The posterior model variance also becomes much smaller compared with the results using the uniform prior distribution. For the case using the FDTD model with waveform data, the result is little affected by the prior information.

In Figure 13, the posterior simulated data of the three cases are in good agreement with the measured data. The use of prior information refines the convergence of parameters to the correct values, but does not affect the fitting capacity. The RMSE values of the three cases are close to the correct values and distributed in a small range compared with the uniform prior cases.

Taken together, Table 3 summarizes the performance of the above numerical experiments involving different forward models, data types, and prior distribution. The number of iterations and corresponding inversion time measure the computational efficiencies, the RMSE values of the posterior mean simulations evaluate how well the simulated models fit the measured data, and the peak signal-to-noise ratios (PSNR) quantify the similarity between the posterior mean models and the real relative permittivity model [39]. With uniform prior, the inversions using first-arrival traveltime data (Cases 1 and 2) require more than one million iterations for the parameters to converge, yet the realizations do not resemble the true relative permittivity field with low PSNR values of 17.22 and 19.30, respectively. In contrast, the inversion using the FDTD model with waveform data cost only 120,000 model evaluations (13 times less than those of Case 2) and achieved a better in-

version accuracy, with a PSNR value of 28.95. The results indicate that, without additional knowledge, the first-arrival traveltime data are unable to provide sufficient information to infer the model parameters. Although the simulated models nicely fit the measurement data with RMSE values close to the measurement error (0.24 ns), the inverted model structures differ significantly from the true model. When the smooth constraint was taken into consideration as prior information, the inversion results using first-arrival traveltime data (Cases 4 and 5) improved considerably by reducing the number of iterations to 144,000 and 236,800, and increasing the PSNR values to 26.83 and 26.59, respectively. Yet, the inversion results using the FDTD model with waveform data (Cases 3 and 6) were barely affected by the prior information, indicating that the waveform data contain sufficient information to infer the model parameters. The choice of forward model also influenced the computational efficiency greatly. By comparing Cases 4 and 5, the straight-ray model inversion achieved almost the same inversion accuracy (PSNR = 26.83) as that of the FDTD model (PSNR = 26.59), but runs more than 120 times faster than the FDTD model inversion.
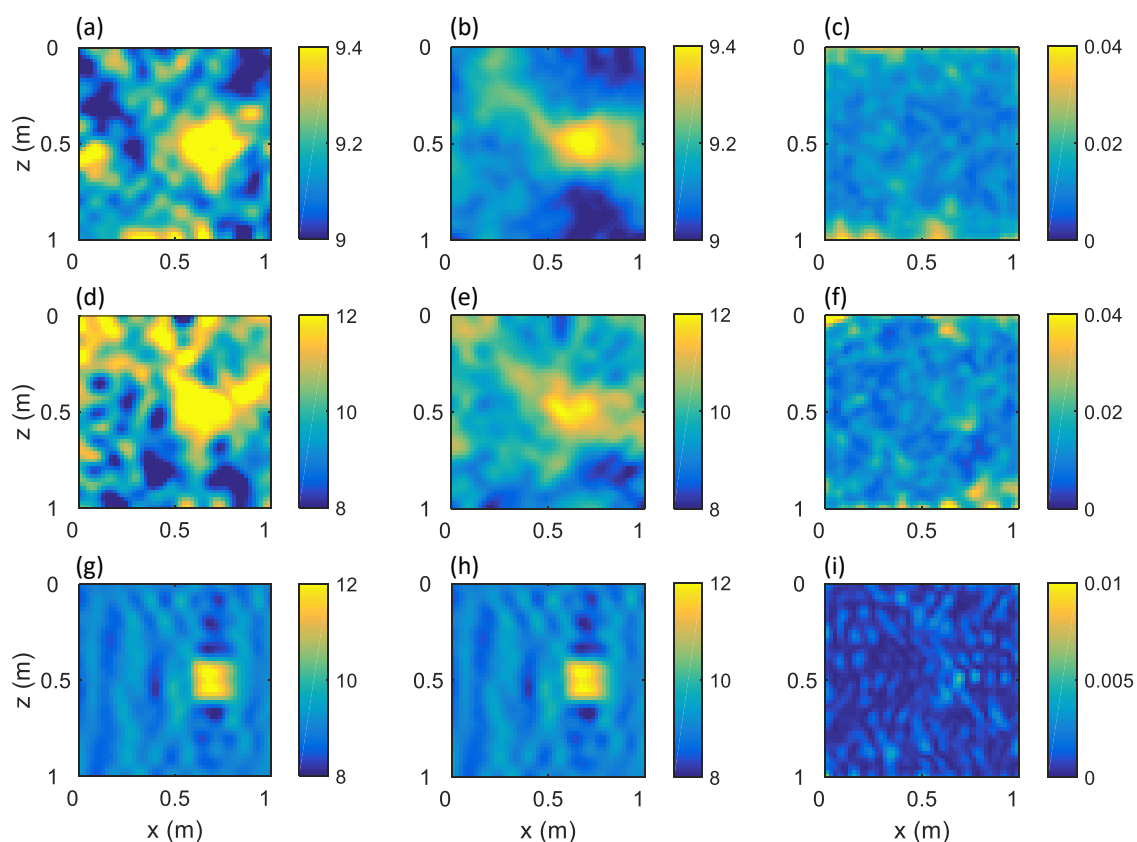

Figure 12. Posterior solutions of the Bayesian inversion with smooth model constraint using (**a**–**c**) the straight-ray model with first-arrival traveltime data, (**d**–**f**) FDTD model with first-arrival traveltime data, and (**g**–**i**) FDTD model with waveform data. The left and middle columns display the maximum a-posteriori (MAP) density and posterior mean solutions, respectively, whereas the right column presents the variance of the posterior solutions.

Table 3. Summary of the Bayesian inversion results.

| Case No.	Forward Model and Data Type	Prior Distribution	Correctly Converged	No. Iterations	Inversion Time (h)	RMSE	PSNR
1	Ray Model + Traveltime Data	Uniform	No	1,440,000	9.52	0.2513	17.22
2	FDTD Model + Traveltime Data	Uniform	No	1,638,000	767.38	0.2656	19.30
3	FDTD Model + Waveform Data	Uniform	Yes	120,000	54.17	2.4770	28.95
4	Ray Model + Traveltime Data	Smooth Constraint	Yes	144,000	0.84	0.2504	26.83
5	FDTD Model + Traveltime Data	Smooth Constraint	Yes	236,800	102.44	0.2634	26.59
6	FDTD Model + Waveform Data	Smooth Constraint	Yes	156,000	61.85	2.5037	29.84
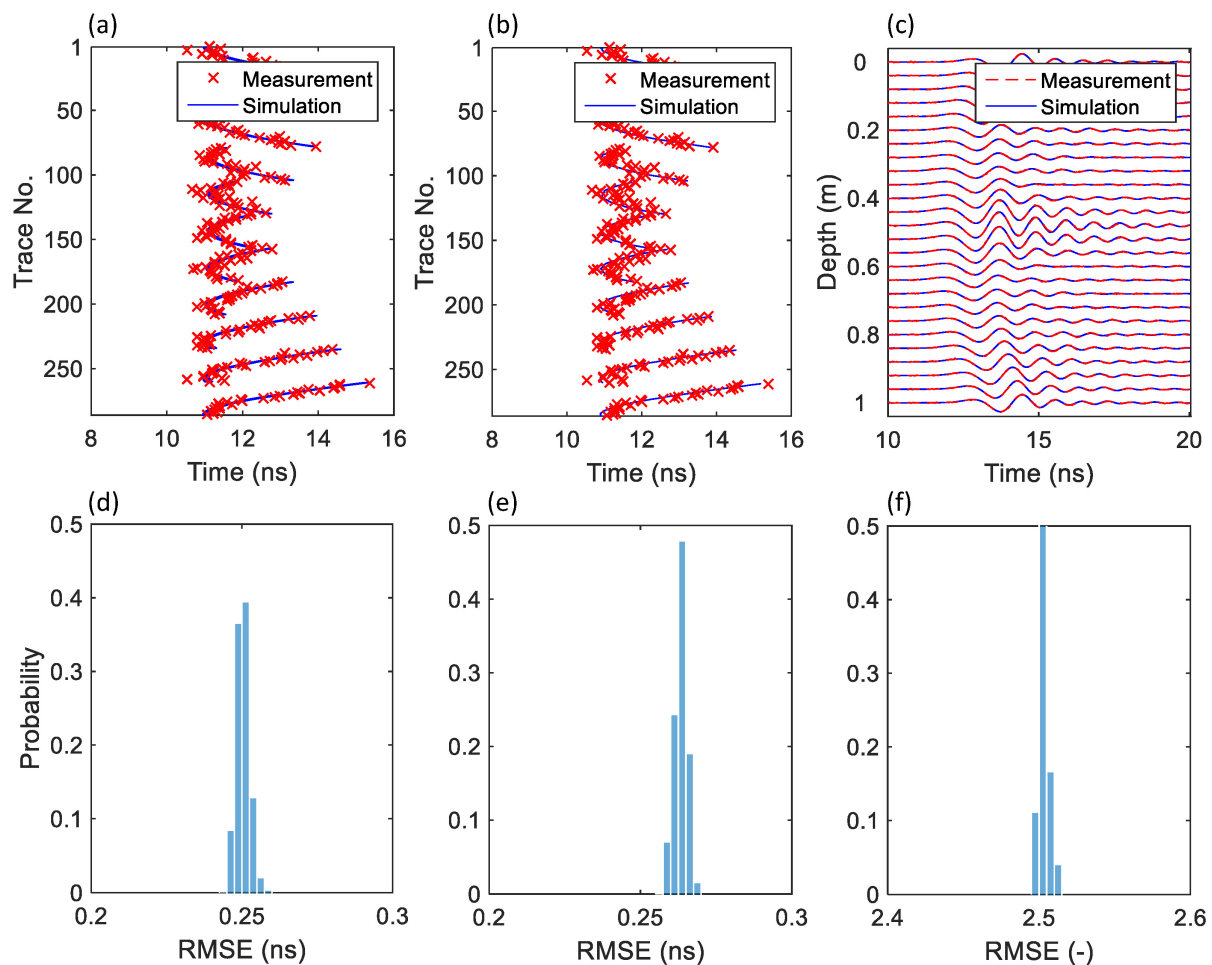
Figure 13. Fitting of the measurement data by the posterior models derived from Bayesian inversion with smooth model constraint using (**a**) the straight-ray model with first-arrival traveltime data, (**b**) FDTD model with first-arrival traveltime data, and (**c**) FDTD model with waveform data, respectively. (**d**–**f**) Histograms of the RMSE values of the posterior solutions in correspondence with the three cases.

6. Discussion

The probabilistic inversion with Bayesian inference is able to explicitly treat different sources of errors, providing a set of statistically acceptable solutions [25]. Compared with deterministic inversion methods that offer a single realization fitting the observed data, the probabilistic inversion results are delivered in the form of a posterior distribution of model parameters that quantifies the uncertainties of the estimation. The probabilistic inversion method summarizes the information of measured data and prior knowledge, and explores the posterior distribution of parameters using global optimization approaches (e.g., the DREAM$_{(ZS)}$ algorithm in this work). This process usually requires a large number of iterations (thousands to millions, depending on the number of parameters) to converge, and the results are affected by the prior distribution and data fitting. Therefore, the choice of forward model, data type, and prior distribution should be carefully considered. Although the inversion using the FDTD model with waveform data has the best accuracy, the low computational efficiency caused by the CPU-intensive FDTD forward solver is unacceptable for practical use. The straight-ray forward model, which is much more computationally appealing, reduces the inversion time remarkably, yet it uses only a small portion of the information of GPR data and leads to biased posterior distribution. An informative prior (smooth model constraint) was therefore used to confine the model structure, thus deriving an improved inversion result. The use of the smooth constraint reduces the parameter ambiguity, yet it sacrifices the inversion accuracy by forcing the parameters

to vary gradually. It turns out that there is a trade-off between computational efficiency and inversion accuracy. For our cases, the inversion using the straight-ray forward model and first-arrival traveltime data with smooth model constraint achieves very high computational efficiency and acceptable inversion accuracy. Note that when the inversion problem involves larger parameter differences, greater target sizes or multiple objects, the applicability of this approach should be further investigated.

7. Conclusions

In this paper, the key components of the probabilistic inversion for the relative permittivity parameters with crosshole GPR data were investigated. The impact of the forward model, data type, and prior distribution on the inversion efficiency and accuracy was analyzed. The inversion results on a synthetic example demonstrate that, (1) the waveform data are much more informative than the first-arrival traveltime data, and are able to infer the model parameters accurately without additional information; (2) prior knowledge in the form of model constraint is necessary for the inversions using the first-arrival traveltime data to reconstruct the relative permittivity field correctly; (3) the straight-ray forward model is much more CPU-friendly than the FDTD model (more than 120 times faster), and better facilitates the probabilistic inversion, which requires a large number of model evaluations. Future work should involve analysis of the applicability to more complex model structures and real-world cases.

Author Contributions: Conceptualization, methodology, writing—original draft preparation, H.Q.; software, data curation, Z.W.; formal analysis, validation, Y.T.; writing—review and editing, T.G. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data used to support the findings of this study are included within the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhou, H.; Sato, M. Subsurface cavity imaging by crosshole borehole radar measurements. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 335–341. [CrossRef]
2. Liu, S.; Liu, X.; Meng, X.; Fu, L.; Lu, Q.; Deng, L. Application of time-domain full waveform inversion to cross-hole radar data measured at Xiuyan jade mine, China. *Sensors* **2018**, *18*, 3114. [CrossRef] [PubMed]
3. Scholer, M.; Irving, J.; Looms, M.C.; Nielsen, L.; Holliger, K. Bayesian Markov-chain-Monte-Carlo inversion of time-lapse crosshole GPR data to characterize the vadose zone at the Arrenaes site, Denmark. *Vadose Zone J.* **2012**, *11*, 1–19. [CrossRef]
4. Qin, H.; Xie, X. Design and test of an improved dipole antenna for detecting enclosure structure defects by cross-hole gpr. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 108–114. [CrossRef]
5. Qin, H.; Vrugt, J.A.; Xie, X.; Zhou, Y. Improved characterization of underground structure defects from two-stage Bayesian inversion using crosshole GPR data. *Autom. Constr.* **2018**, *95*, 233–244. [CrossRef]
6. Qin, H.; Xie, X.; Tang, Y.; Wang, Z. Detection of diaphragm wall defects using crosshole GPR. In Proceedings of the IEEE 17th International Conference on Ground Penetrating Radar, Rapperswil, Switzerland, 18–21 June 2018; pp. 1–4. [CrossRef]
7. Irving, J.D.; Knight, R.J. Numerical simulation of antenna transmission and reception for crosshole ground-penetrating radar. *Geophysics* **2006**, *71*, K37–K45. [CrossRef]
8. Tarantola, A.; Valette, B. Inverse problems = quest for information. *J. Geophys.* **1982**, *50*, 159–170.
9. Dines, K.A.; Lytle, R.J. Computerized geophysical tomography. *Proc. IEEE* **1979**, *67*, 1065–1073. [CrossRef]
10. Greenhalgh, S.A.; Bing, Z.; Green, A. Solutions, algorithms and inter-relations for local minimization search geophysical inversion. *J. Geophys. Eng.* **2006**, *3*, 101–113. [CrossRef]
11. Tronicke, J.; Tweeton, D.R.; Dietrich, P.; Appel, E. Improved crosshole radar tomography by using direct and reflected arrival times. *J. Appl. Geophys.* **2001**, *47*, 97–105. [CrossRef]

12. Bikowski, J.; Huisman, J.A.; Vrugt, J.A.; Vereecken, H.; van der Kruk, J. Integrated analysis of waveguide dispersed GPR pulses using deterministic and Bayesian inversion methods. *Near Surf. Geophys.* **2012**, *10*, 641–652. [CrossRef]

13. Laloy, E.; Linde, N.; Vrugt, J.A. Mass conservative three-dimensional water tracer distribution from Markov chain Monte Carlo inversion of time-lapse ground-penetrating radar data. *Water Resour. Res.* **2012**, *48*, 1–15. [CrossRef]

14. Laloy, E.; Linde, N.; Jacques, D.; Vrugt, J.A. Probabilistic inference of multi-Gaussian fields from indirect hydrological data using circulant embedding and dimensionality reduction. *Water Resour. Res.* **2015**, *51*, 4224–4243. [CrossRef]

15. Hunziker, J.; Laloy, E.; Linde, N. Bayesian full-waveform tomography with application to crosshole ground penetrating radar data. *Geophys. J. Int.* **2019**, *218*, 913–931. [CrossRef]

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

17. Köpke, C.; Irving, J.; Roubinet, D. Stochastic inversion for soil hydraulic parameters in the presence of model error: An example involving ground-penetrating radar monitoring of infiltration. *J. Hydrol.* **2019**, *569*, 829–843. [CrossRef]

18. Dafflon, B.; Barrash, W. Three-dimensional stochastic estimation of porosity distribution: Benefits of using ground-penetrating radar velocity tomograms in simulated-annealing-based or Bayesian sequential simulation approaches. *Water Resour. Res.* **2012**, *48*, 1–13. [CrossRef]

19. Moghadas, D.; Vrugt, J.A. The influence of geostatistical prior modeling on the solution of DCT-based Bayesian inversion: A case study from Chicken Creek catchment. *Remote Sens.* **2019**, *11*, 1549. [CrossRef]

20. Hansen, T.M.; Cordua, K.S.; Mosegaard, K. Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling. *Comput. Geosci.* **2012**, *16*, 593–611. [CrossRef]

21. Balkaya, C.; Akcıg, Z.; Gokturkler, G. A comparison of two travel-time tomography schemes for crosshole radar data: Eikonal-equation-based inversion versus ray-based inversion. *J. Environ. Eng. Geophys.* **2010**, *15*, 203–218. [CrossRef]

22. Ernst, J.R.; Maurer, H.; Green, A.G.; Holliger, K. Full-waveform inversion of crosshole radar data based on 2-D finite-difference time-domain solutions of Maxwell's equations. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 2807–2828. [CrossRef]

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

24. Klotzsche, A.; Vereecken, H.; van der Kruk, J. Review of crosshole ground-penetrating radar full-waveform inversion of experimental data: Recent developments, challenges, and pitfalls. *Geophysics* **2019**, *84*, H13–H28. [CrossRef]

25. Qin, H.; Xie, X.; Vrugt, J.A.; Zeng, K.; Hong, G. Underground structure defect detection and reconstruction using crosshole GPR and Bayesian waveform inversion. *Autom. Constr.* **2016**, *68*, 156–169. [CrossRef]

26. Qin, H.; Xie, X.; Tang, Y. Evaluation of a straight-ray forward model for Bayesian inversion of crosshole ground penetrating radar data. *Electronics* **2019**, *8*, 630. [CrossRef]

27. Rosas-Carbajal, M.; Linde, N.; Kalscheuer, T.; Vrugt, J.A. Two-dimensional probabilistic inversion of plane-wave electromagnetic data: Methodology, model constraints and joint inversion with electrical resistivity data. *Geophys. J. Int.* **2014**, *193*, 1508–1524. [CrossRef]

28. Lochbühler, T.; Vrugt, J.A.; Sadegh, M.; Linde, N. Summary statistics from training images as prior information in probabilistic inversion. *Geophys. J. Int.* **2015**, *201*, 157–171. [CrossRef]

29. Yee, K.S. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Trans. Antennas Propag.* **1966**, *AP-14*, 302–307. [CrossRef]

30. Giroux, B.; Gloaguen, E.; Chouteau, M. bh_tomo—A Matlab borehole georadar 2D tomography package. *Comput. Geosci.* **2007**, *33*, 126–137. [CrossRef]

31. Belina, F.A.; Ernst, J.R.; Holliger, K. Inversion of crosshole seismic data in heterogeneous environments: Comparison of waveform and ray-based approaches. *J. Appl. Geophys.* **2009**, *68*, 85–94. [CrossRef]

32. Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092. [CrossRef]

33. Vrugt, J.A. Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environ. Model. Softw.* **2015**, *75*, 273–316. [CrossRef]

34. ter Braak, C.J.F.; Vrugt, J.A. Differential evolution Markov chain with snooker updater and fewer chains. *Stat. Comput.* **2008**, *18*, 435–446. [CrossRef]

35. Giannopoulos, A. Modelling ground penetrating radar by GprMax. *Constr. Build. Mater.* **2005**, *19*, 755–762. [CrossRef]

36. Warren, C.; Giannopoulos, A.; Giannakis, I. gprMax: Open source software to simulate electromagnetic wave propagation for ground penetrating radar. *Comput. Phys. Commun.* **2016**, *209*, 163–170. [CrossRef]

37. Jafarpour, B.; Goyal, V.K.; McLaughlin, D.B.; Freeman, W.T. Transform-domain sparsity regularization for inverse problems in geosciences. *Geophysics* **2009**, *74*, R69–R83. [CrossRef]

38. Moghadas, D. Probabilistic inversion of multiconfiguration electromagnetic induction data using dimensionality reduction technique: A numerical study. *Vadose Zone J.* **2019**, *18*, 1–16. [CrossRef]

39. Huynh-Thu, Q.; Ghanbari, M. Scope of validity of PSNR in image/video quality assessment. *Electron. Lett.* **2008**, *44*, 800–801. [CrossRef]