

Integrated analysis of waveguide dispersed GPR pulses using deterministic and Bayesian inversion methods

Jutta Bikowski^{1*}, Johan A. Huisman¹, Jasper A. Vrugt², Harry Vereecken¹ and Jan van der Kruk¹

¹Agrosphere, Forschungszentrum Juelich, Germany

²Dept. of Civil and Environmental Engineering, University of California, Irvine

ABSTRACT

Ground-penetrating radar (GPR) data affected by waveguide dispersion are not straightforward to analyse. Therefore, waveguide dispersed common midpoint measurements are typically interpreted using so-called dispersion curves, which describe the phase velocity as a function of frequency. These dispersion curves are typically evaluated with deterministic optimization algorithms that derive the dielectric properties of the subsurface as well as the location and depth of the respective layers. However, these methods do not provide estimates of the uncertainty of the inferred subsurface properties. Here, we applied a formal Bayesian inversion methodology using the recently developed Differential Evolution Adaptive Metropolis DREAM_(ZS) algorithm. This Markov Chain Monte Carlo simulation method rapidly estimates the (non-linear) parameter uncertainty and helps treat the measurement error explicitly. We found that the frequency range used in the inversion has an important influence on the posterior parameter estimates, essentially because parameter sensitivity varies with measurement frequency. Moreover, we established that the measurement error associated with the dispersion curve is frequency dependent and that the estimated model parameters become severely biased if this frequency-dependent nature of the measurement error is not properly accounted for. We estimated these frequency-dependent measurement errors together with the model parameters using the DREAM_(ZS) algorithm. The posterior distribution of the model parameters derived in this way compared well with inversion results for a reduced frequency bandwidth. This more subjective method is an alternative to reduce the bias introduced by this frequency-dependent measurement error. Altogether, our inversion procedure provides an integrated and objective methodology for the analysis of dispersive GPR data and appropriately treats the measurement error and parameter uncertainty.

INTRODUCTION

Ground-penetrating radar (GPR) is a geophysical technique that uses pulsed electromagnetic waves to explore the subsurface. The transmitted waves will be partly reflected and partly transmitted when contrasts in dielectric permittivity associated with subsurface structures occur. The propagation velocity of the GPR waves depends on the dielectric permittivity, which in turn can be related to soil moisture content and soil porosity amongst other factors (e.g., van Overmeeren *et al.* 1997; Huisman *et al.* 2001; Galagedara *et al.* 2003, Huisman *et al.* 2003a, Moyssey 2004, Bradford 2008, Westermann *et al.* 2010; Haarder *et al.* 2011; Rhim 2011; Steelman and Endres 2011).

For on-ground GPR, propagation velocity and therewith, the dielectric permittivity can be determined when GPR measurements are made with multiple antenna offsets, for example using a common midpoint (CMP) measurement where the antenna

separation is increased while keeping the same midpoint. In such a CMP measurement, reflected GPR waves can be identified by their hyperbolic shape, which can be used to estimate the depth of the reflecting layer and the average propagation velocity of the layer above the reflecting boundary (Greaves *et al.* 1996; van Overmeeren *et al.* 1997; Dannowski and Yaramanci 1999; Endres *et al.* 2000; Bohidar and Hermance 2002; Garambois *et al.* 2002; Grote *et al.* 2003; Lunt *et al.* 2005; Turesson 2006; Gerhards *et al.* 2008). Another wave that has been used for soil water content determination is the direct subsurface ground wave, which is the direct transmission from the transmitter to the receiver antenna through the top part of the soil. This ground wave can be recognized in a CMP measurement by its traveltime curve that shows a linear increase in arrival time with antenna separation. The ground wave propagation velocity can be determined from the slope of the traveltime curve. The ground wave velocity has been widely used to measure the spatio-temporal development of soil water content variability (e.g., Galagedara *et*

*j.bikowski@fz-juelich.de

al. 2003; Huisman *et al.* 2003b; Weihermüller *et al.* 2007).

Although the ground wave is promising for soil water content measurements, difficulties arise when the subsurface is highly heterogeneous with distinct layers or gradients in soil moisture content that introduce thin horizontal layers with a strong contrast in dielectric permittivity. If the depth of these layers is comparable to or smaller than the wavelength of the GPR signal, they can act as a waveguide in which the electromagnetic waves are trapped. This leads to positive interference related to total reflection of the trapped wave at the boundaries of the layer. Field situations where such waveguides have been reported include a thin ice sheet floating on water (Arcone 1984; van der Kruk *et al.* 2007), an organic-rich sandy silt layer overlying a gravel unit (Arcone *et al.* 2003; van der Kruk *et al.* 2006), a mountain slope with a soil cover (Strobbia and Cassiani 2007) and thawing of a frozen soil layer (van der Kruk *et al.* 2009; Steelman *et al.* 2010). In the presence of such waveguides, CMP measurements are difficult to interpret because the arrival time and the first cycle amplitude of the ground wave cannot be identified due to interfering waves and dispersion.

To enable interpretation of GPR data with waveguide dispersion, van der Kruk *et al.* (2006) presented a deterministic inversion algorithm to simultaneously estimate the thickness and permittivity of the dispersive waveguide and the permittivity of the material below the waveguide. This work was inspired by inversion algorithms for Rayleigh and Love waves commonly observed in multi-offset seismic data and were extended to higher order modes by van der Kruk (2006, 2007). More recent extensions include the inversion for multiple layers acting as waveguides (van der Kruk *et al.* 2010).

Inversion of GPR data affected by waveguide dispersion requires a forward model that accurately describes the dispersive characteristics of GPR data for a given subsurface structure described by a set of model parameters. Optimization methods are then used to seek a set of model parameters that minimizes the discrepancy between simulated and measured GPR data. In general, numerical modelling and inversion methods for GPR data have greatly improved in the last decade, which obviously enhances the inversion quality, as well as the range of GPR data that can be inverted. Yet, traditional GPR inversion algorithms are deterministic and estimate only a single ‘best’ set of model parameters without consideration of parameter uncertainty (e.g., Pettinelli *et al.* 2007; Steelman and Endres 2010; Wollschläger *et al.* 2010). Therefore, it is not yet well established how errors in GPR measurements and models propagate through the processing and inversion of dispersive GPR data. For non-dispersive data, reported confidence intervals of wave velocity and hence, implicitly of dielectric permittivity, vary widely depending on the field settings and methods used (e.g., Jacob and Hermance 2004). Typical errors of GPR measurements introduced during data acquisition include inaccuracies in offset, timing, antenna orientation and other antenna effects (Slob 2010). An additional error source that is more complicated to address is the structural

error in the model, which is often introduced by simplifying assumptions regarding the modelling of GPR and the representation of the subsurface.

Clearly, it is desirable to simultaneously estimate the ‘best’ model parameters and their associated uncertainty. Bayesian inversion algorithms based on Markov Chain Monte Carlo (MCMC) simulation methods are particularly well suited for this task. Such methods are not new in the field of geophysical inversion (Mosegaard and Tarantola 1995; Sambridge and Mosegaard 2002) but the ever increasing computational power and development of advanced MCMC simulation schemes has resulted in their increased use in recent years, especially in the field of hydrogeophysics (e.g., Strobbia and Cassiani 2007; Irving and Singha 2010; Hinnell *et al.* 2010; Huisman *et al.* 2010).

The aim of this study is to present an integrated analysis of parameter uncertainty associated with the inversion of synthetic and experimental GPR data with waveguide dispersion using MCMC simulation methods. In order to do so, we first describe deterministic and MCMC inversion methods for dispersive GPR data. Then, we use synthetic dispersive GPR data to illustrate the effects of frequency bandwidth and measurement noise on the

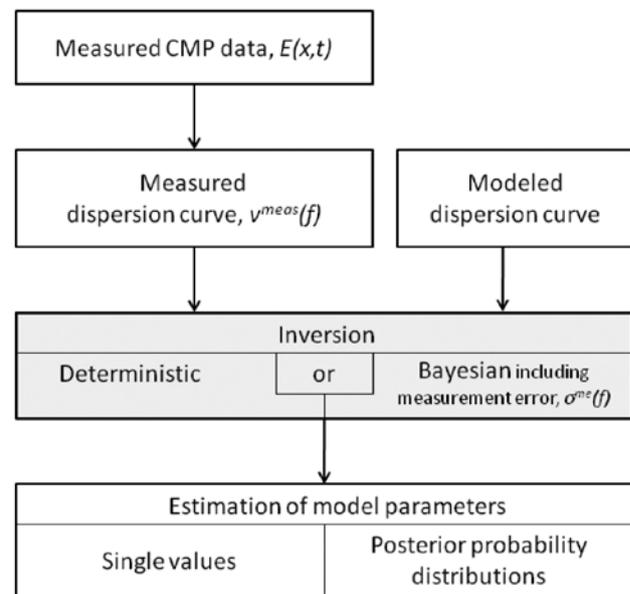


FIGURE 1

Schematic outline of the analysis approach. The measured CMP data are transformed into a dispersion curve that is affected by the measurement error. The forward model provides the modelled dispersion curve given the model parameters describing the waveguide. Either a deterministic or Bayesian inversion method is applied. Only the Bayesian inversion approach can be used to simultaneously estimate model parameters and the measurement error associated with the dispersion curve. The result of the deterministic method is a single value for each model parameter, whereas the Bayesian inversion with MCMC simulation (DREAM_(ZS)) returns the posterior probability distribution of the model parameters, which provides the most likely model parameters and their uncertainty.

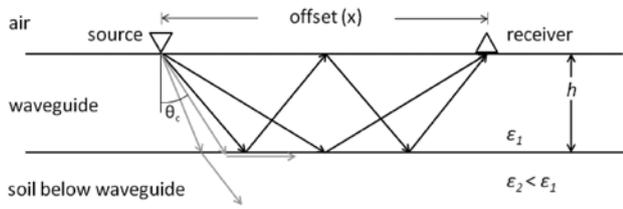


FIGURE 2

Single-layer subsurface model acting as a waveguide. Due to total reflection beyond the critical angle θ_c , multiple reflections occur that are trapped within the waveguide layer. The model parameters that influence the dispersion characteristics of the electromagnetic waves are the thickness h , the relative permittivity of the waveguide ϵ_1 and the relative permittivity of the material below the waveguide, ϵ_2 .

final inversion results. Next, we analyse actual dispersive GPR data and demonstrate how model parameters and measurement error can be jointly estimated from the dispersive GPR data. Finally, the last section of this paper summarizes our main results, states directions of further research and provides some conclusions.

METHODOLOGY

Dispersion curve

The general flow of our deterministic and Bayesian inversion strategies is summarized in Fig. 1. The CMP measurements, $E(x, t)$, with x being the antenna offset and t time, are processed to obtain a so-called dispersion curve (Park *et al.* 1998; van der Kruk 2006). More precisely, $E(x, t)$ is Fourier transformed into the frequency domain, $\hat{E}(x, f)$. Then, the phase-velocity spectrum is calculated:

$$D(v, f) = \left| \sum_x \frac{\hat{E}(x, f)}{|\hat{E}(x, f)|} \exp\left(i \frac{2\pi f}{v} x\right) \right|, \quad (1)$$

where v is the phase velocity, f denotes frequency and i signifies the square root of -1. In the last step, the dispersion curve, $v^{\text{meas}}(f)$, is obtained by selecting the phase velocity with maximum amplitude for each frequency, $v^{\text{meas}}(f) = \max_v D(v, f)$. This $v^{\text{meas}}(f)$ constitutes the measured data in the following deterministic and Bayesian inversion strategies.

The forward model used to simulate dispersion curves is based on the modal theory (Budden 1961; van der Kruk 2006) and does not rely on the use of a source wavelet. It assumes that the waveguide consists in a single high permittivity layer on top of material with lower permittivity (Fig. 2). The model parameters of the single layer model are the dielectric permittivity and thickness of the waveguide (ϵ_1 and h) and the permittivity of the material below the waveguide (ϵ_2). The highest phase velocity of a single-layer guided wave is given by $c_0 / \sqrt{\epsilon_2}$, where c_0 denotes the speed of light (van der Kruk 2006; van der Kruk *et al.* 2006). It is obtained for low frequencies because the dispersion curve decreases monotonically with increasing frequency. A first rough estimate of ϵ_2 can thus be calculated from the highest phase

TABLE 1

True parameter values and their prior ranges used for the synthetic and experimental data.

parameter	true parameter	min	max
ϵ_1	20	10	30
ϵ_2	10	3	20
h	0.25	0.05	1

velocity in the measured dispersion curve. This also suggests that an adequate representation of low frequencies is required to obtain accurate estimates of ϵ_2 . For high frequencies, the dispersion curve asymptotically approaches $c_0 / \sqrt{\epsilon_1}$, which suggests that an adequate representation of high frequencies is required for accurate estimates of ϵ_1 . However, the direct estimation of ϵ_1 and ϵ_2 using the simple procedure outlined above can be difficult because of the low signal-to-noise ratio in the dispersion curve for low and high frequencies.

DETERMINISTIC INVERSION

Two inversion algorithms are used in this study: the deterministic inversion algorithm of van der Kruk *et al.* (2006) and the Differential Evolution Adaptive Metropolis, DREAM_(z) algorithm of Vrugt *et al.* (2009). The deterministic inversion approach of van der Kruk (2006) aims to find the model parameters ($m = \epsilon_1, \epsilon_2, h$) that minimize the mean of the absolute difference between modelled and measured dispersion curves. In contrast to van der Kruk *et al.* (2006), we describe the misfit, C , between the measured dispersion curve, $v^{\text{meas}}(f)$ and modelled dispersion curve, $v^{\text{mod}}(f, m)$, using a mean squared difference,

$$C(\mathbf{m}) = \frac{1}{N} \sum_{k=1}^N \left(v^{\text{mod}}(f_k, \mathbf{m}) - v^{\text{meas}}(f_k) \right)^2, \quad (2)$$

where N denotes the number of frequencies used. The deterministic inversion was carried out with a global/local optimization approach using a feasible parameter space defined by the lower and upper bounds specified in Table 1. This space is first sampled using a regular grid and then each grid point serves as a starting value for a local search using the Simplex method (Lagarias *et al.* 1998). The final model parameter combination with the smallest value of C is assumed to represent the best possible subsurface model describing the dispersive GPR data. Application of this inversion algorithm to synthetic and measured dispersive GPR data showed robust and reliable results (van der Kruk 2006; van der Kruk *et al.* 2006, 2007, 2010; Steelman *et al.* 2010).

BAYESIAN INVERSION USING MCMC SIMULATION

If the inversion is cast in a Bayesian framework, the misfit is not described by a single value as in the deterministic method but by the likelihood, which is the probability of observing the data given the model parameters. There are different formulations of likelihood functions (e.g., Schoups and Vrugt 2010). When it is assumed that the residuals between modelled and measured dis-

person curves are normally distributed and mutually independent, the likelihood is given by:

$$p(v^{meas} | \mathbf{m}, \sigma) = \prod_{k=1}^N \left(\frac{1}{\sqrt{2\pi(\sigma)^2}} \exp \left[-\frac{1}{2} \left(\frac{v^{mod}(f_k, \mathbf{m}) - v^{meas}(f_k)}{\sigma} \right)^2 \right] \right) \quad (3)$$

It is important to note that the likelihood not only depends on the model parameters but also on the standard deviation of the residuals, σ .

Bayesian inference allows us to combine information from measurements with prior information about the system. This is achieved through Bayes' theorem:

$$p(\mathbf{m}, \sigma | v^{meas}) \propto p(v^{meas} | \mathbf{m}, \sigma) p(\mathbf{m}) p(\sigma) \quad (4)$$

where $p(\mathbf{m})$ and $p(\sigma)$ denote the prior probability of σ and the model parameters and $p(\mathbf{m}, \sigma | v^{meas})$ represents the posterior probability, which is defined as the probability of the model parameters and σ after considering available measurement data. In most geophysical applications, relatively little prior knowledge is available about the values of the model parameters. In such cases a uniform prior distribution suffices with upper and lower bounds of each individual parameter (e.g., Table 1). All model parameters are assumed to have equal probability within these bounds.

The posterior probability, $p(\mathbf{m}, \sigma | v^{meas})$ distribution, can be determined analytically for some cases but is typically intractable for non-linear models. Instead, MCMC algorithms have

found widespread use to estimate $p(\mathbf{m}, \sigma | v^{meas})$. Unfortunately, standard MCMC algorithms are generally inefficient and even very simple problems typically require many thousands of model evaluations to converge to the posterior probability distribution. This work capitalizes on recent developments in MCMC simulation and uses the DREAM_(ZS) algorithm (Vrugt *et al.* 2009) that runs multiple Markov chains (random walk trajectories) in parallel. A key step in MCMC simulation is the selection of an appropriate proposal distribution, which is used to generate new candidate points in each Markov chain. Whereas standard MCMC algorithms require extensive tuning of the proposal distribution for a particular problem, DREAM_(ZS) automatically determines the orientation and scale of the proposal distribution during sampling and has been shown to work well for a range of problems (Vrugt *et al.* 2009). The only information to be specified by the user is the feasible parameter space and the likelihood function. A detailed description of the algorithm and proof of convergence can be found in Vrugt *et al.* (2009).

The choice of an appropriate value for σ in equation (2) is of particular interest herein. In the likelihood function, σ should reflect the standard deviation of the residuals between the measured and modelled dispersion curve. An appropriate value of σ is unknown *a priori* because it represents an integrated value over several error sources, such as measurement error and model structural errors. In practice, the value of σ should preferably be estimated directly from the measurements themselves. Yet, this is

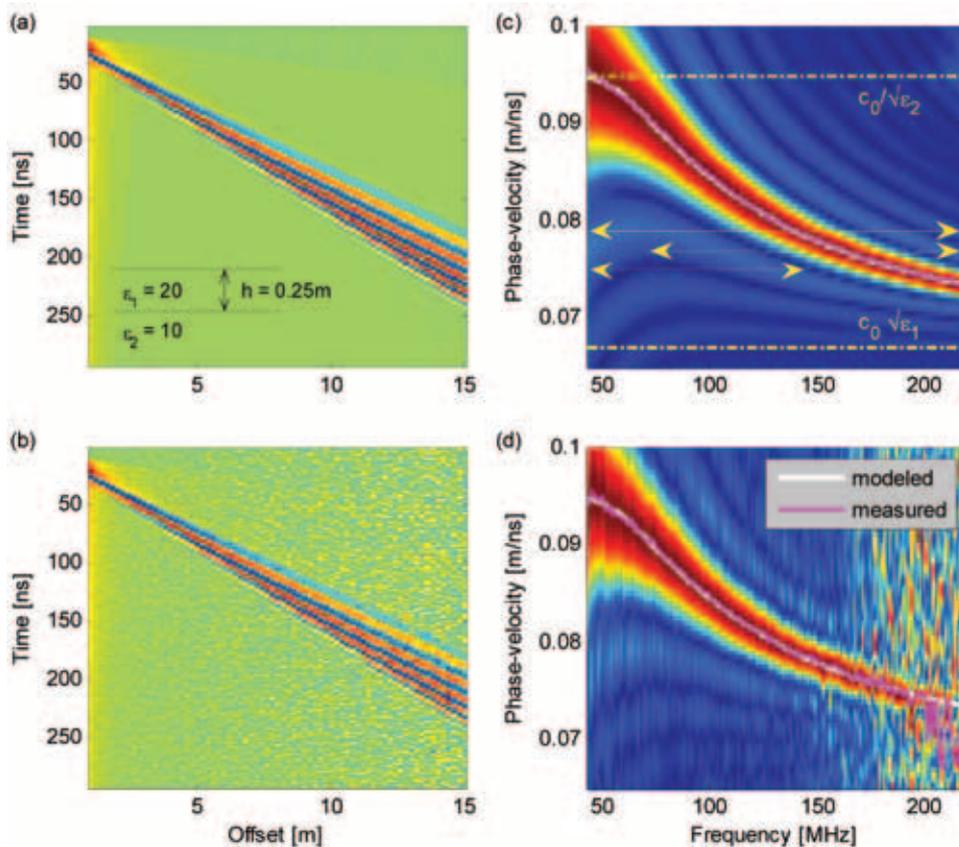


FIGURE 3 Numerically created CMP data (a) without and (b) with noise and (c, d) the corresponding frequency-normalized phase-velocity spectra. Red indicates the largest amplitude and blue the smallest. The modelled dispersion curve calculated directly from the known model parameters is plotted in the phase velocity spectra in white, whereas the measured dispersion curve determined from the maximum in the phase-velocity spectrum is plotted in magenta. The theoretical upper and lower bounds of the dispersion curve are indicated with dashed-dotted yellow lines and the different frequency ranges used in the inversion are indicated with yellow arrows.

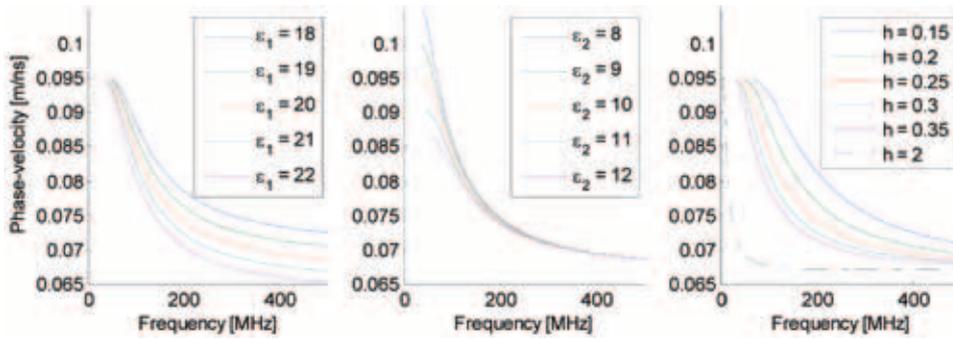


FIGURE 4

Modelled dispersion curves with a varying model parameter. In each plot, one parameter is varied to investigate its influence on the dispersion curve, while the others are fixed. The fixed parameter values are $\epsilon_1 = 20$, $\epsilon_2 = 10$ and thickness = 0.25.

typically difficult, if not impossible. As a consequence, the terminology ‘measurement error’ might not be fully appropriate for the definition of σ . Nevertheless we will use this terminology in the remainder of the paper for lack of a better name.

The appropriate choice of σ is required for an accurate estimation of the posterior distribution. If the value of σ is too large, the posterior distribution will be too wide and the model parameter uncertainty derived from the posterior distribution will be too large. A conservative choice for σ might significantly underestimate the actual parameter uncertainty. A common approach to avoid an explicit definition of σ is to integrate out σ from equation (4) (Vrugt and Bouten 2002; Huisman *et al.* 2004; Kavetski *et al.* 2006; Scharnagl *et al.* 2011). To do this, we assume a Jeffreys prior for σ , $p(\sigma) \propto 1/\sigma$. When using this approach, it is assumed that the standard deviation of the residuals between measured and modelled dispersion curves (i.e., the root mean square error, RMSE) is an appropriate estimate of the absolute frequency-independent measurement error. In the case of the dispersion curve, the nature of the measurement error is not yet well established and therefore, we allow σ to be frequency dependent, $\sigma(f)$ by assigning specific values to $\sigma(f)$ *a priori* (synthetic data) or by treating $\sigma(f)$ as unknown parameters to be estimated within the MCMC framework.

We use the DREAM_(ZS) algorithm with prior ranges of the parameters specified in Table 1. These bounds are consistent with the deterministic inversion approach and appropriate for the synthetic and real-world dispersion curves considered herein. Note that practical experience suggests that larger bounds do not really affect the performance of both methods – only the efficiency with which the optimal values are found. In each MCMC trial, convergence of DREAM_(ZS) to the limiting posterior distribution was monitored using the diagnostic of Gelman and Rubin (1992). After convergence, the last 5000 parameter sets created with DREAM_(ZS) were used to represent the posterior parameter distribution.

SYNTHETIC DATA

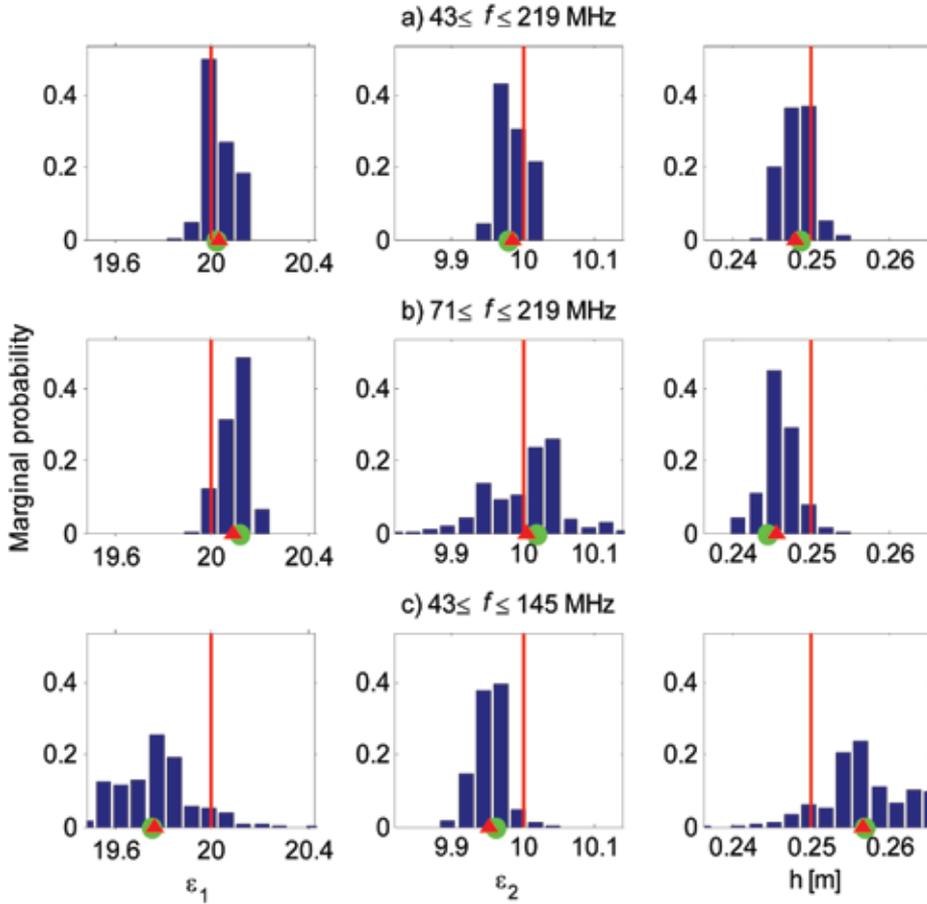
We simulated CMP data corresponding to a single layer model (Fig. 2) using a numerical solution of an exact forward model for a horizontally layered medium (van der Kruk *et al.* 2006) and a Gaussian source wavelet with a centre frequency of 100 MHz. The resulting dispersive CMP data for a waveguide with a thickness of

0.25 m and a relative permittivity of $\epsilon_1 = 20$ overlying material with $\epsilon_2 = 10$ are shown in Fig. 3(a). The shingling events indicate different phase and group velocities, which are characteristic for waveguide dispersion. To investigate parameter uncertainty, we corrupted the radargram by adding a normally distributed random error term to the simulated CMP data (Fig. 3b). This random noise had a mean of zero and the standard deviation was set to 1% of the maximum amplitude of the simulated CMP data. The visual impression of a higher noise level for larger offsets is due to the trace normalization in the visualization of the CMP data and the lower signal strength for large offsets. The phase-velocity spectra, $D(v,f)$, of the noise-free data and the noisy data as well as measured and modelled dispersion curve are presented in Fig. 3(c,d). The measured dispersion curve is determined by finding the phase velocity with the largest magnitude for each frequency. The modelled dispersion curve is obtained by the modal theory using the prescribed model parameters.

Each of the three model parameters has a distinct influence on the dispersion curve, which is illustrated in Fig. 4. The permittivity of the waveguide, ϵ_1 , determines the high-frequency asymptote, whereas ϵ_2 determines the highest phase velocity observed for the low frequencies. The thickness of the waveguide mostly determines the shape of the dispersion curve.

INVERSION RESULTS FOR SYNTHETIC GPR DATA

To investigate how the use of different frequency ranges influences the model parameter estimates, we selected three frequency ranges (43–219 MHz, 71–219 MHz and 43–145 MHz, Fig. 3c). The marginal posterior probability distributions for ϵ_1 , ϵ_2 and h and the results of the deterministic inversion (green points) are shown in Fig. 5. The marginal posterior probability distribution is obtained by projecting (‘marginalizing’) the posterior distribution on a single model parameter. The mean of the marginal posterior probability distributions (red triangle) and the true model parameters (red line) are included as well. Compared to the marginal posterior probability distributions for the complete frequency range (top panel of Fig. 5), a decreasing frequency range resulted in an increasing uncertainty in model parameter estimates as expressed by the width of the marginal posterior probability distribution. Removing the low frequencies (middle panel of Fig. 5) resulted in a significant increase in the parameter uncertainty for ϵ_2 . In comparison, the parameter

**FIGURE 5**

Histograms of the marginal posterior distributions of the model parameters obtained from the inversion of the dispersion curve shown in Fig. 3(c) with frequency ranges; a) $43 < f < 219$ MHz, b) $71 < f < 219$ MHz and c) $43 < f < 145$ MHz (see also arrows in Fig. 3c). These histograms are created using the last 5000 samples generated with $DREAM_{(ZS)}$. The green circles represent the results of the deterministic inversion, the red triangles denote the mean of the marginal posterior distributions and the red lines indicate the true parameter values used to generate the synthetic data.

uncertainty for ε_1 and h increased only slightly, although a small shift of the mean was observed. The effect of excluding the high frequencies is illustrated in the bottom panel of Fig. 5. The uncertainty in the model parameters ε_1 and h is now larger, whereas the uncertainty in ε_2 has only slightly increased. These results clearly show that ε_1 and h are more sensitive to high frequencies than ε_2 . In addition, the close agreement of the mean of the marginal posterior probability distributions and the results for the deterministic inversion inspire confidence that both algorithms have successfully located the global minimum.

Next, we investigated the effect of noise in GPR data on the model parameter estimates. First, we considered noise in the radargram and its influence on the dispersion curve. A first indication of this effect can be obtained from a comparison of the frequency spectra averaged over all offsets as shown in Fig. 6(a). The ratio of the noise-free average frequency spectrum (blue line) and the noise (green line) determines the signal-to-noise ratio (SNR). The variation of the SNR with frequency is evident in Fig. 6(a), which implies that noise mostly influences low and high frequencies. This was already apparent in Fig. 3(d), where the measured dispersion curve deviates more from the modelled dispersion curve for high and low frequencies. In contrast, the measured dispersion curves obtained from the noise-free CMP data (Fig. 3c) correspond better with the modelled dispersion

curve for all frequencies. These results suggest that the measurement error of the dispersion curve is not constant but instead depends explicitly on the frequency. Higher data errors are found for low and high frequencies.

As a first step in our inversion, we nevertheless assume that the measurement error of the dispersion curve is independent of frequency and equal to the RMSE of the best fit to the measured dispersion curve obtained from noisy CMP data (Fig. 3d) for the frequency range from 34–219 MHz. This resulted in a measurement error of $\sigma_1 = 0.0013 \text{ m(ns)}^{-1}$. Figure 3(d) shows that most of this error is associated with the highest frequencies because there the misfit between the measured and modelled dispersion curves is largest. Therefore, we considered a reduced frequency range to exclude the error-prone frequency range. The reduced frequency range from 45 MHz to 201 MHz resulted in a significantly lower estimate of the measurement error ($\sigma_2 = 0.0003 \text{ m(ns)}^{-1}$). Finally, we also considered frequency dependent measurement error ($\sigma_3(f)$) in a third scenario. To assign values to $\sigma_3(f)$, we created 50 different realizations of noisy CMP data and determined the corresponding dispersion curves that are shown in Fig. 6(b). The standard deviation of these 50 different dispersion curves for each frequency served as an estimate of $\sigma_3(f)$ and is plotted in Fig. 6(c) together with the two frequency-independent measurement errors.

The inversion results for the three different specifications of σ are presented in Fig. 7. For the wide frequency range (34–219 MHz) and the frequency-independent measurement error σ_1 , the posterior distributions do not contain the true values for ε_1 and h (top panel of Fig. 7). Clearly, too much confidence is placed on the uncertain ends of the dispersion curves during inversion, which results in biased parameter estimates. When the most uncertain part of the dispersion curve is not considered and the corresponding measurement error σ_2 is used, the means of the marginal posterior probability distributions are much closer to the true values used to generate the synthetic data (middle panel of Fig. 7b). Moreover, the uncertainty of the estimated model parameters is smaller compared to the use of the full frequency bandwidth, illustrating that a smaller measurement error reduces parameter uncertainty. The deterministic inversion results show behaviour similar to that of the mean of the posterior probability distribution for these two measurement error definitions.

To obtain accurate and precise model parameter estimates with a frequency-independent measurement error, it was necessary to discard the uncertain parts of the dispersion curve.

Alternatively, the frequency-dependent measurement error $\sigma_3(f)$ and the full frequency range can be used. Although the uncertain parts of the dispersion curve are now considered in the inversion, they receive less weight because of their relatively high $\sigma_3(f)$ values. The inversion results with this frequency-dependent measurement error are shown in the bottom panel of Fig. 7. We indeed obtain marginal posterior distributions that encapsulate the true parameter values. This demonstrates that a better description of the measurement error results in more realistic parameter estimates. The results for the deterministic inversion are not available in this case because the deterministic inversion was not extended to include the frequency-dependent measurement error.

EXPERIMENTAL DATA

The measured GPR data were obtained on a terrace of braided river sediments in New Zealand (van der Kruk *et al.* 2006). A trench 80 m away from the site of the GPR measurements revealed a 0.13–0.38 m thick layer of rich organic sandy silt on top of a gravel layer. The CMP data were recorded with a pulseEKKO 100A system and 100 MHz antennas. The time range

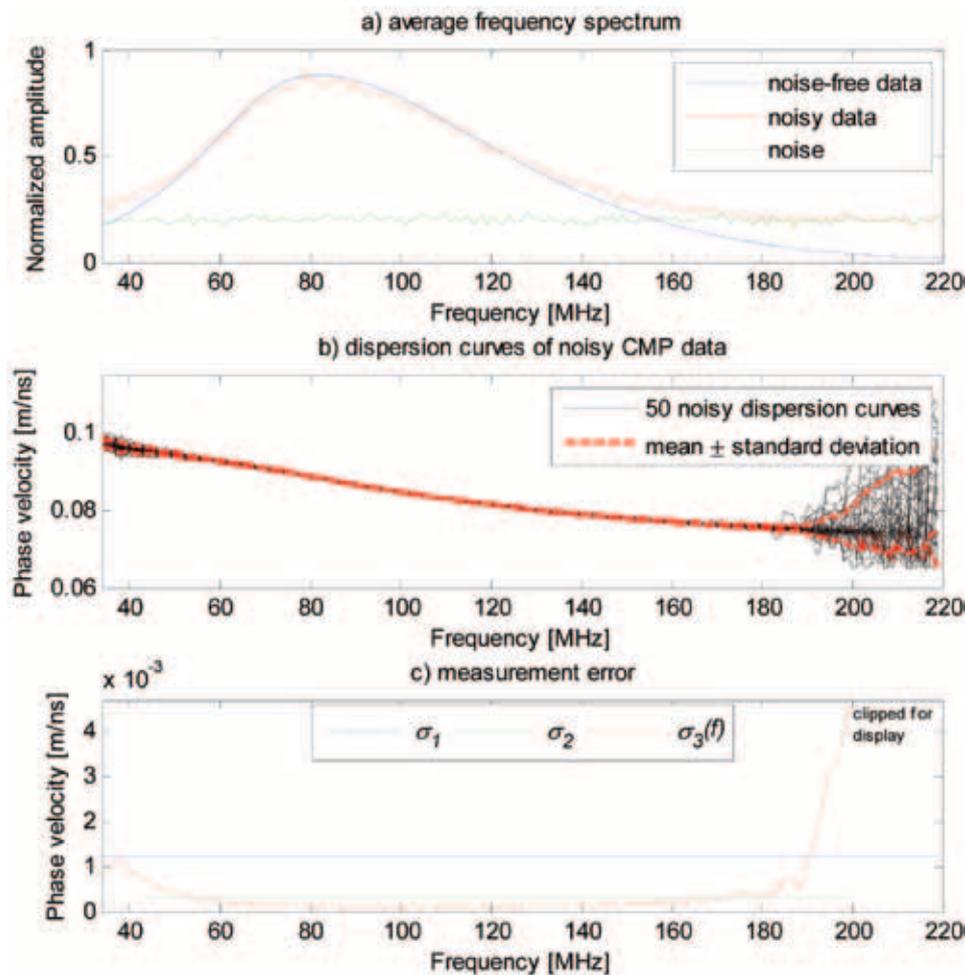
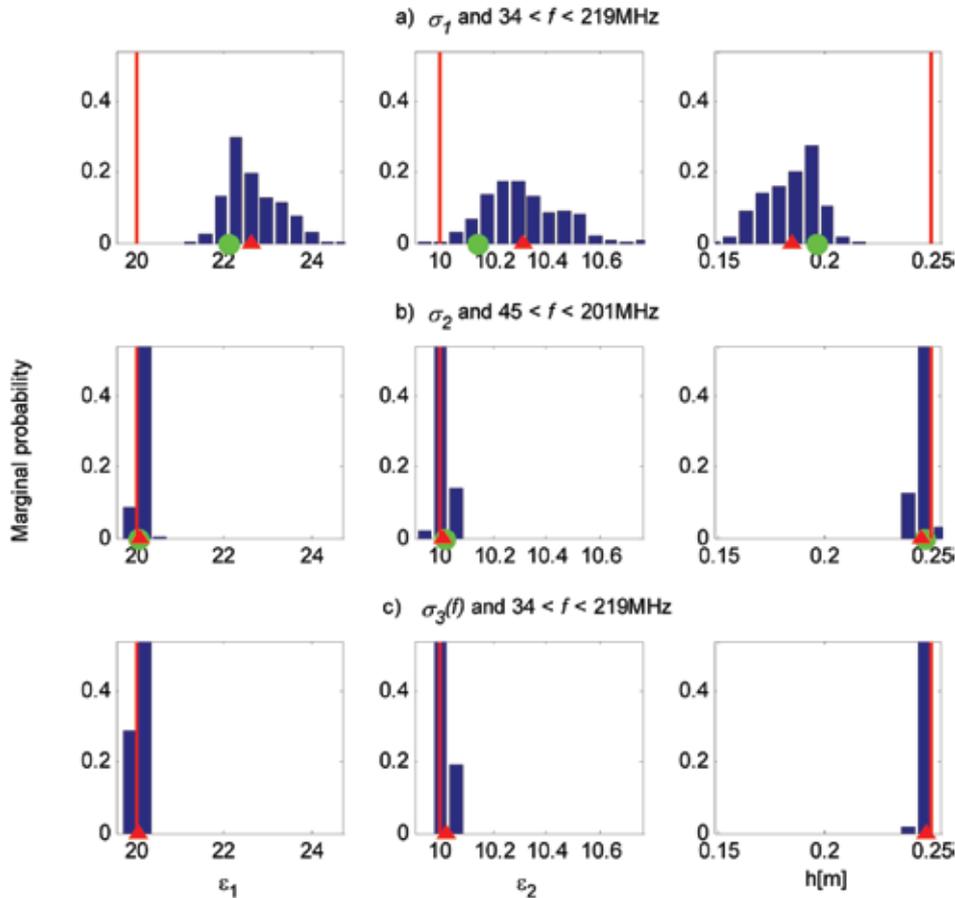


FIGURE 6

(a) Average frequency spectra of the noise-free data (Fig. 3a), the noisy synthetic data (Fig. 3b) and the noise. (b) Dispersion curves of different realizations of noisy CMP data (black lines). The red lines indicate the standard deviation added and subtracted from the mean of all dispersion curves (68% of all the dispersion curves lay between the red lines). (c) Three different strategies to define the measurement error: σ_1 (blue line) is the frequency-independent measurement error, which is determined from the RMSE of the best fit to the data, σ_2 (green line) is obtained by limiting the frequency range of the dispersion curve to $45 < f < 201$ MHz and $\sigma_3(f)$ (red line) is determined from the standard deviation of the 50 different dispersion curves shown in Fig. 6(b).

**FIGURE 7**

Histograms of the marginal posterior parameter distributions obtained from the inversion of the dispersion curve shown in Fig. 3(d) using the three different definitions of measurement error shown in Fig. 6(c): a) σ_1 and $34 < f < 219$ MHz, b) σ_2 and $45 < f < 201$ and c) $\sigma_3(f)$ and $34 < f < 219$ MHz. The red triangles indicate the mean of marginal posterior distributions, the green dots represent the results of the deterministic inversion and the red lines indicate the true model parameters.

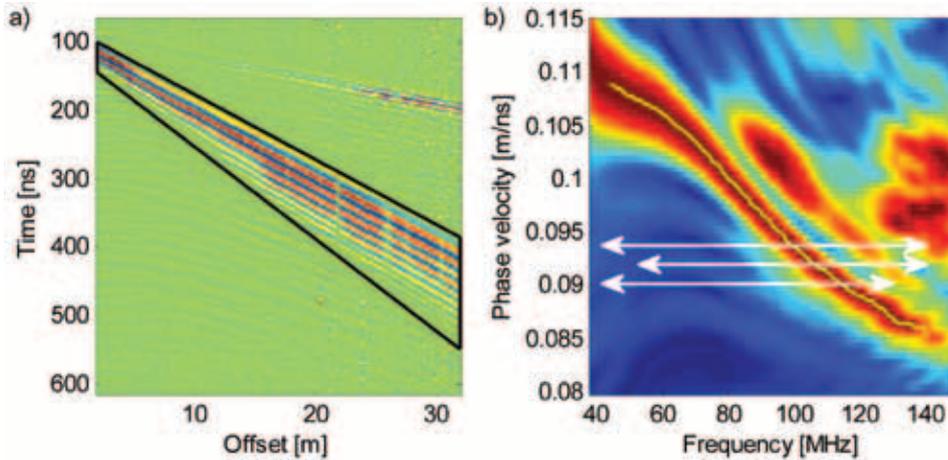
was 600 ns with a sample discretization of 0.5 ns and the spatial sampling resolution was 0.2 m. The trace-normalized CMP data are shown in Fig. 8(a). For the inversion, the air wave and the reflected waves were muted. Therefore, the phase-velocity spectrum shown in Fig. 8(b) was determined from the region enclosed by the black lines in Fig. 8(a). The measured dispersion curve obtained from the maximum amplitude of this spectrum is depicted by the white line.

INVERSION RESULTS FOR EXPERIMENTAL GPR DATA

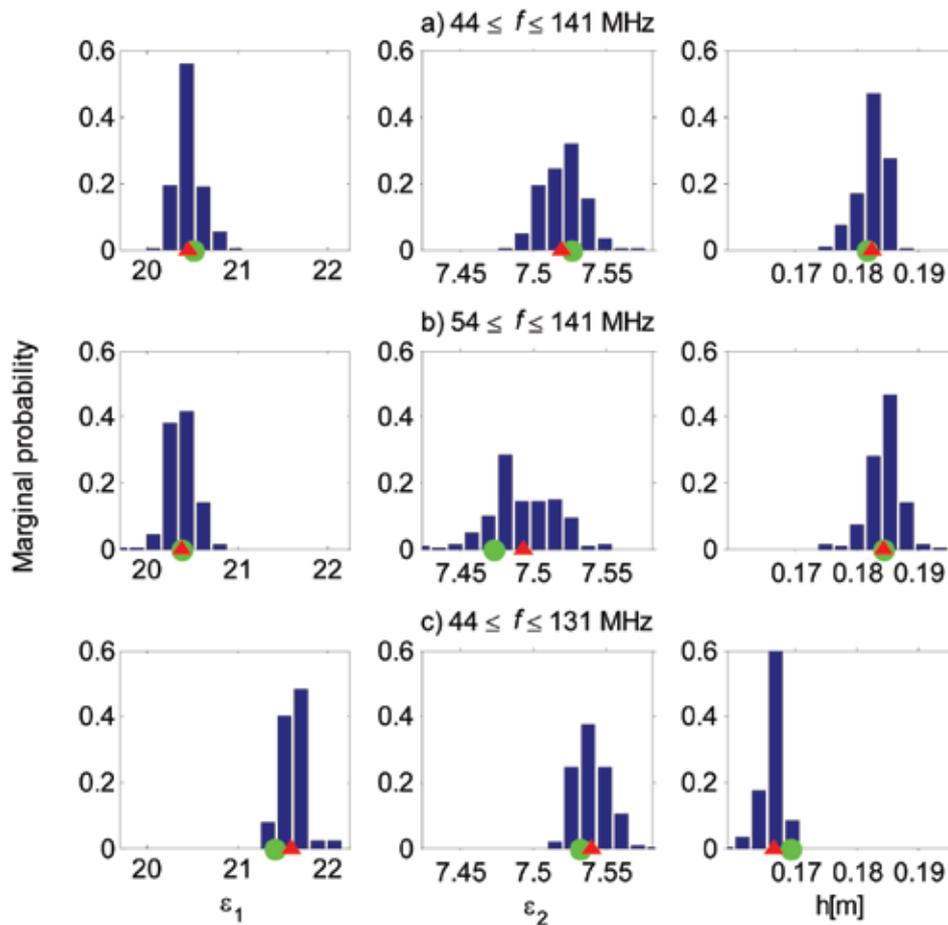
To investigate the sensitivity of the estimated model parameters to the use of different frequency bandwidths, we selected three different frequency ranges: a) the frequency range approximately used in van der Kruk (2006), 44–141 MHz, b) a reduction of the low frequencies, 54–141 MHz and c) a reduction of the high frequencies, 44–131 MHz. These frequency ranges are highlighted in Fig. 8(b) with arrows. The marginal posterior distributions for these different frequency ranges and the deterministic inversion results are displayed in Fig. 9. As with the synthetic data, we observe that the marginal posterior distributions of ε_1 and h remain similar when the low frequencies are removed (top and middle panel of Fig. 9), whereas the uncertainty for ε_2 is nearly doubled and the mean of the marginal posterior distribu-

tion changed slightly. Again, this confirms that ε_2 is mostly sensitive to lower frequencies. The inversion results that do not consider the highest frequencies of the measured dispersion curve are shown in the bottom panel of Fig. 9. The marginal posterior distribution of ε_2 was similar to the results for the full frequency bandwidth. However, significant differences were observed in the mean of the marginal posterior distribution of ε_1 and h , which changed from $\varepsilon_1^{mean} = 20.4$ to $\varepsilon_1^{mean} = 21.6$ and from $h^{mean} = 0.182$ m to $h^{mean} = 0.166$ m, respectively. Again, this confirms the sensitivity of ε_1 and h to high frequencies.

Figure 9 shows that exclusion of high frequencies in the inversion resulted in marginal posterior parameter distributions of ε_1 and h , which do not overlap with their counterparts derived using the entire frequency bandwidth. The analysis of the synthetic data indicated that this might be related to an inappropriate definition of the measurement error, σ . Indeed, it was already shown that the value of σ is frequency dependent but this was not considered in our analysis of the measured GPR data thus far. Unfortunately, it is not straightforward to obtain a reliable estimate of this frequency dependent σ of the measured dispersion curve. Recently, several studies using MCMC simulation have included σ as an additional parameter to be estimated (e.g., Vrugt *et al.* 2009). To test the usefulness of this approach for our data, we first assumed a single, frequency independent σ and esti-


FIGURE 8

a) Trace-normalized measured CMP data, b) corresponding phase-velocity spectrum using the data enclosed by the black lines. The yellow line indicates the selected dispersion curve and the different frequency ranges used in the inversions are indicated with white arrows.


FIGURE 9

Histograms of the marginal posterior parameter distributions obtained from the inversion of the dispersion curve of Fig. 8(b) and three different frequency ranges: a) 44–141 MHz, b) 54–141 MHz (low frequencies not considered) and c) 44–131 MHz (high frequencies not considered). The measurement error is assumed to be frequency independent. The red triangles denote the mean of the marginal posterior distributions and the green dots illustrate the results of the deterministic waveguide inversion.

mated it along with the model parameters using the DREAM_(zs) algorithm. The median value of the marginal posterior distribution of σ corresponded very well with the RMSE of the best model fit for the entire frequency bandwidth (black arrow, Fig. 10) and the 95th percentile posterior uncertainty of σ (blue area, Fig. 10) nicely encapsulated this RMSE value. Consequently, the marginal posterior parameter distributions (Fig. 11, top panel) were similar to those derived previously using the full frequency bandwidth (Fig. 9, top panel).

It is rather encouraging to conclude that the measured GPR dispersion curve contains sufficient information to warrant inference of the measurement error. Yet, our approach has considered σ to be frequency-independent; an assumption that was already shown to be unrealistic when using the full frequency bandwidth in the inversion. We therefore proceeded with another DREAM_(zs) run in which the measurement error was assumed to be frequency dependent, $\alpha(f)$. Besides the three model parameters, four additional parameters were introduced that specify the measurement

error at four different frequencies equally distributed along the frequency axis (Fig. 10). Cubic hermite interpolation between these four points was subsequently used to estimate the measurement error at the remaining frequencies. About 70 000 model runs were needed with DREAM_(zS) to converge to a limiting distribution. This is significantly more than the 20 000 model runs originally required to estimate the model parameters only.

In Fig. 10, the median and corresponding 95th percentile uncertainty range of $\sigma(f)$ are plotted. The shape of $\sigma(f)$ illustrates

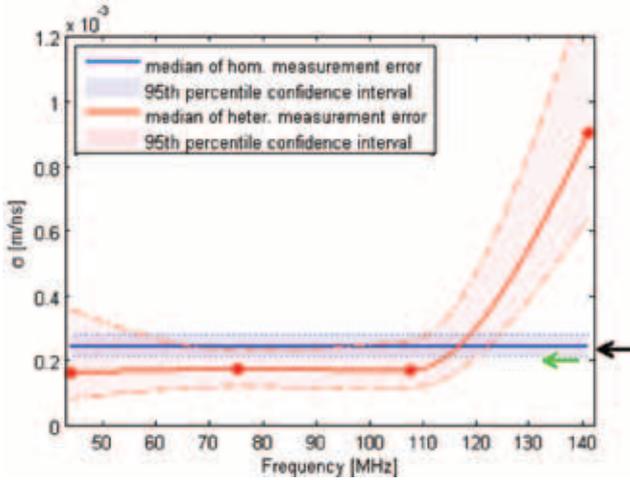


FIGURE 10
The estimated frequency-independent and frequency-dependent measurement error of the dispersion curve in blue and red, respectively. The solid line is the median and the coloured band illustrates the 95% confidence interval of the estimated measurement error. The red dots represent the median of the estimated interpolation points and the black and green arrows indicate the RMSE of the best fit of the results shown in Fig. 9(a) and Fig. 9(c), respectively.

that the measured dispersion curve is most reliable in the frequency range between 70–110 MHz. Above 120 MHz, the measurement error of the dispersion curve is significantly larger implying that this part of the dispersion curve receives less weight in the inversion. Consequently, the marginal posterior parameter distributions (bottom panel, Fig. 11) were similar to those derived with a reduced frequency bandwidth (bottom panel, Fig. 9). The $\sigma(f)$ curves shown in Fig. 10 are very similar to those derived for the synthetic GPR data, which are plotted in Fig. 6. This attests to the ability of our inversion procedure to return a reasonable estimate of the frequency-dependent measurement error of the dispersion curve.

We argue that the posterior parameter distributions presented in Fig. 11 (bottom panel) best summarize the actual subsurface properties considered herein based on the results of our synthetic study. We demonstrated that the model parameters can only be correctly retrieved using two approaches. First, a meaningful *a priori* reduction of the frequency bandwidth can be applied before inversion of the dispersion curve. However, this approach is rather subjective and in practice it remains difficult to pinpoint an appropriate frequency bandwidth to invert the dispersion curve. Moreover, the synthetic and experimental case study illustrated a strong sensitivity of the model parameter estimates to high and low frequencies. A reduction of the frequency bandwidth could therefore lead to a loss of valuable information. Alternatively, the full frequency bandwidth together with a proper description of the frequency-dependent measurement error can be used in the Bayesian inversion. The MCMC inversion approach introduced simultaneously estimates the model parameters and the frequency-dependent measurement error and their underlying uncertainty in an objective manner. Therefore, the MCMC inversion approach removes the need for subjective decisions on the frequency bandwidth and provides an attractive and accurate approach to analyse dispersive GPR data.

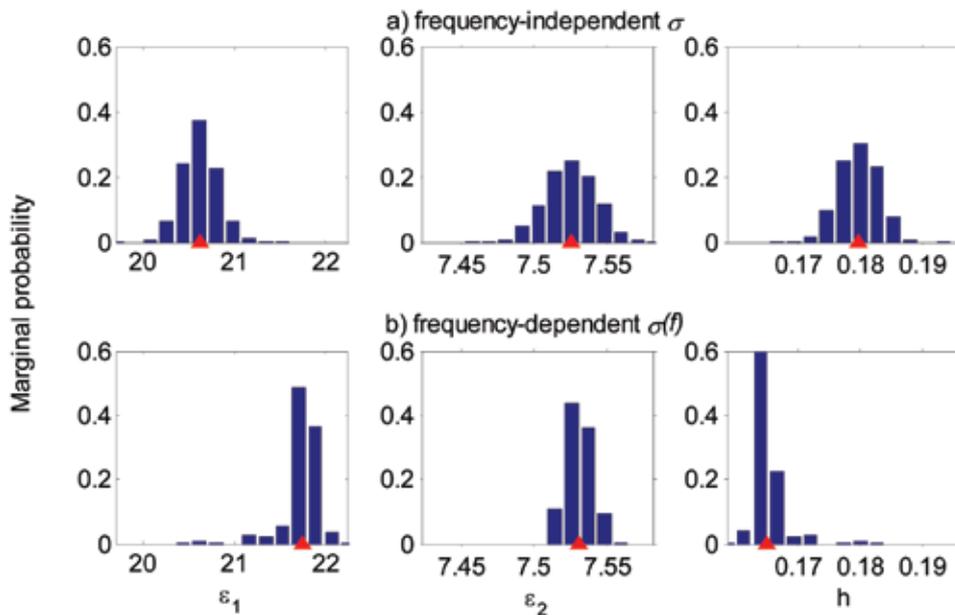


FIGURE 11
Parameter posterior distributions when measurement error σ and model parameters are estimated simultaneously. In the top panel, σ is independent of frequency and in the bottom panel a frequency dependent measurement error is assumed. The red triangles denote the mean of the marginal posterior distributions.

SUMMARY AND CONCLUSIONS

We applied a deterministic and a Bayesian inversion method to synthetic and experimental on-ground GPR data with waveguide dispersion assuming a single layer model of the subsurface. The Bayesian inversion used the recently developed DREAM_(ZS) adaptive MCMC method, which is especially designed to rapidly explore the posterior parameter distribution. Unlike deterministic inversion methods, MCMC simulation with DREAM_(ZS) provides robust estimates of parameter uncertainty, which is of crucial importance for model predictive uncertainty assessment and sensitivity studies. Overall, the estimated 'best' parameters derived from synthetic and measured dispersion curves depended strongly on the frequency bandwidth used in the inversion. More precisely, the relative permittivity of the material below the waveguide was sensitive to the low frequencies of the dispersion curve, whereas the relative permittivity and the thickness of the waveguide were sensitive to high frequencies. Detailed analysis of the synthetic data showed that the measurement error associated with the dispersion curve was frequency dependent. In particular, the extreme ends of the dispersion curve were more uncertain. When such frequency-dependent measurement errors were not properly handled during the inversion, the resulting model parameters were biased.

One possible way to remove this bias is to exclude the low- and high-frequency parts of the dispersion curve during the inversion. Although this procedure led to plausible results for the deterministic and Bayesian inversion, the choice of an appropriate frequency bandwidth is quite arbitrary. A more satisfactory approach was presented and consisted in estimating the measurement error properties simultaneously with the model parameters. This resulted in plausible estimates of all three model parameters that compared well with inversion results for a reduced frequency bandwidth.

Future research with Bayesian inversion is warranted. One possible application area constitutes combined TE-TM and multi-layer inversion problems. Another area of potential application could include models with continuously changing parameters for the monitoring of wetting and thawing fronts. In addition, the performance of different likelihood functions or a generalized likelihood function should be investigated.

Altogether, the Bayesian inversion methodology presented provides an objective and integrated framework to successfully retrieve the posterior distribution of model parameters and the measurement error. This approach circumvents the need to make subjective decisions on which frequency bandwidth to use in the inversion.

ACKNOWLEDGEMENTS

We thank Steve Arcone and one anonymous reviewer for their detailed comments that helped to further improve the paper. The DREAM_(ZS) algorithm used in this study was provided by the third author (jasper@uci.edu) and is available upon request.

REFERENCES

- Arcone S.A. 1984. Field Observations of Electromagnetic Pulse-Propagation in Dielectric Slabs. *Geophysics* **49**, 1763–1773.
- Arcone S.A., Peapples P.R. and Liu L.B. 2003. Propagation of a ground-penetrating radar (GPR) pulse in a thin-surface waveguide. *Geophysics* **68**, 1922–1933.
- Bohidar R.N. and Hermance J.F. 2002. The GPR refraction method. *Geophysics* **67**, 1474–1485.
- Bradford J.H. 2008. Measuring water content heterogeneity using multi-fold GPR with reflection tomography. *Vadose Zone Journal* **7**, 184–193.
- Budden K.G. 1961. *The Wave-Guide Mode Theory of Wave Propagation*. Prentice-Hall Inc., London
- Dannowski G. and Yaramanci U. 1999. Estimation of water content and porosity using combined radar and geoelectrical measurements. *European Journal of Environmental and Engineering Geophysics* **4**, 71–85.
- Endres A.L., Clement W.P. and Rudolph D.L. 2000. Ground penetrating radar imaging of an aquifer during a pumping test. *Ground Water* **38**, 566–576.
- Galagedara L.W., Parkin G.W. and Redman J.D. 2003. An analysis of the ground-penetrating radar direct ground wave method for soil water content measurement. *Hydrological Processes* **17**, 3615–3628.
- Garambois S., Senechal P. and Perroud H. 2002. On the use of combined geophysical methods to assess water content and water conductivity of near-surface formations. *Journal of Hydrology* **259**, 32–48.
- Gelman A. and Rubin D.B. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7**, 457–472.
- Gerhards H., Wollschlager U., Yu Q., Schiwiek P., Pan X. and Roth K. 2008. Continuous and simultaneous measurement of reflector depth and average soil-water content with multichannel ground-penetrating radar. *Geophysics* **73**, J15–J23.
- Greaves R.J., Lesmes D.P., Lee J.M. and Toksoz M.N. 1996. Velocity variations and water content estimated from multi-offset, ground-penetrating radar. *Geophysics* **61**, 683–695.
- Grote K., Hubbard S. and Rubin Y. 2003. Field-scale estimation of volumetric water content using ground-penetrating radar ground wave techniques. *Water Resources Research* **39**, 1321.
- Haarder E.B., Looms M.C., Jensen K.H. and Nielsen L. 2011. Visualizing Unsaturated Flow Phenomena Using High-Resolution Reflection Ground Penetrating Radar. *Vadose Zone Journal* **10**, 84–97.
- Hinnell A.C., Ferre T.P.A., Vrugt J.A., Huisman J.A., Moysey S., Rings J. and Kowalsky M.B. 2010. Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion. *Water Resources Research* **46**, W00D40.
- Huisman, J.A., Bouten, W. and Vrugt, J.A. 2004. Accuracy of frequency domain analysis scenarios for the determination of complex dielectric permittivity. *Water Resources Research* **40**, W02401
- Huisman J.A., Hubbard S.S., Redman J.D. and Annan A.P. 2003a. Measuring Soil Water Content with Ground Penetrating Radar: A Review. *Vadose Zone Journal* **2**, 476–491.
- Huisman J.A., Rings J., Vrugt J.A., Sorg J. and Vereecken H. 2010. Hydraulic properties of a model dike from coupled Bayesian and multi-criteria hydrogeophysical inversion. *Journal of Hydrology* **380**, 62–73.
- Huisman J.A., Snepvangers J.J.J.C., Bouten W. and Heuvelink G.B.M. 2003b. Monitoring temporal development of spatial soil water content variation: Comparison of ground-penetrating radar and time domain reflectometry. *Vadose Zone Journal* **2**, 519–529.
- Huisman J.A., Sperl C., Bouten W. and Verstraten J.M. 2001. Soil water content measurements at different scales: Accuracy of time domain reflectometry and ground-penetrating radar. *Journal of Hydrology* **245**, 48–58.

- Irving J. and Singha K. 2010. Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities. *Water Resources Research* **46**, W11514.
- Jacob R.W. and Hermance J.F. 2004. Assessing the precision of GPR velocity and vertical two-way travel time estimates. *Journal of Environmental and Engineering Geophysics* **9**, 143–153.
- Kavetski D., Kuczera G. and Franks S.W. 2006. Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research* **42**, W03407.
- van der Kruk J. 2006. Properties of surface waveguides derived from inversion of fundamental and higher mode dispersive GPR data. *IEEE Transactions on Geoscience and Remote Sensing* **44**, 2908–2915.
- van der Kruk J., Arcone S.A. and Liu L. 2007. Fundamental and higher mode inversion of dispersed GPR waves propagating in an ice layer. *IEEE Transactions on Geoscience and Remote Sensing* **45**, 2483–2491.
- van der Kruk J., Jacob R.W. and Vereecken H. 2010. Properties of precipitation-induced multilayer surface waveguides derived from inversion of dispersive TE and TM GPR data. *Geophysics* **75**, WA263–WA273.
- van der Kruk J., Steelman C. M., Endres A. L. and Vereecken H. 2009. Dispersion inversion of electromagnetic pulse propagation within freezing and thawing soil waveguides. *Geophysical Research Letters* **36**, L18503–L18507.
- van der Kruk J., Streich R. and Green A.G. 2006. Properties of surface waveguides derived from separate and joint inversion of dispersive TE and TM GPR data. *Geophysics* **71**, K19–K29.
- Lagarias J.C., Reeds J.A., Wright M.H. and Wright P.E. 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *Siam Journal on Optimization* **9**, 112–147.
- Lunt I.A., Hubbard S.S. and Rubin Y. 2005. Soil moisture content estimation using ground-penetrating radar reflection data. *Journal of Hydrology* **307**, 254–269.
- Mosegaard K. and Tarantola A. 1995. Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research* **100**, 12431–12448.
- Moysey S. and Knight R. 2004. Modeling the field-scale relationship between dielectric constant and water content in heterogeneous systems. *Water Resources Research* **40**, 10.
- van Overmeeren R.A., Sariowan S.V. and Gehrels J.C. 1997. Ground penetrating radar for determining volumetric soil water content; Results of comparative measurements at two test sites. *Journal of Hydrology* **197**, 316–338.
- Park C.B., Miller R.D. and Xia J.X. 1998. Imaging dispersion curves of surface waves on multi-channel record. In: *Imaging dispersion curves of surface waves on multi-channel record*, pp. 1377–1380.
- Pettinelli E., Vannaroni G., Di Pasquo B., Mattei E., Di Matteo A., De Santis A. and Annan P.A. 2007. Correlation between near-surface electromagnetic soil parameters and early-time GPR signals: An experimental study. *Geophysics* **72**, A25–A28.
- Rhim H.C. 2011. Measurements of dielectric constants of soil to develop a landslide prediction system. *Smart Structures and Systems* **7**, 319–328.
- Sambridge M. and Mosegaard K. 2002. Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics* **40**, 1009.
- Scharnagl B., Vrugt J.A., Vereecken H. and Herbst M. 2011. Inverse modelling of in situ soil water dynamics: Investigating the effect of different prior distributions of the soil hydraulic parameters. *Hydrology and Earth System Sciences* **15**, 3043–3059.
- Schoups G. and Vrugt J.A. 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research* **46**, 17.
- Slob E. 2010. Uncertainty in ground penetrating radar models. *Proceedings of 13th International Conference on Ground Penetrating Radar (GPR) 2010*, doi: 10.1109/ICGPR.2010.5550154
- Steelman C.M. and Endres A.L. 2010. An examination of direct ground wave soil moisture monitoring over an annual cycle of soil conditions. *Water Resources Research* **46**, 16.
- Steelman C.M. and Endres A.L. 2011. Comparison of Petrophysical Relationships for Soil Moisture Estimation using GPR Ground Waves. *Vadose Zone Journal* **10**, 270–285.
- Steelman C.M., Endres A.L. and van der Kruk J. 2010. Field Observations of Shallow Freeze and Thaw processes using High-Frequency Ground-Penetrating Radar. *Hydrological Processes* **24**, pp. 2022–2033.
- Strobbia C. and Cassiani G. 2007. Multilayer ground-penetrating radar guided waves in shallow soil layers for estimating soil water content. *Geophysics* **72**, J17–J29.
- Turesson A. 2006. Water content and porosity estimated from ground-penetrating radar and resistivity. *Journal of Applied Geophysics* **58**, 99–111.
- Vrugt J.A. and Bouten W. 2002. Validity of first-order approximations to describe parameter uncertainty in soil hydrologic models. *Soil Science Society of America Journal* **66**, 1740–1751.
- Vrugt J.A., ter Braak C.J.F., Clark M.P., Hyman J.M. and Robinson B. 2008. Treatment of input uncertainty in hydrologic modeling: Doing hydrology backwards with Markov chain Monte Carlo simulation. *Water Resources Research* **44**, W00B09.
- Vrugt J.A., ter Braak C.J.F., Diks C.G.H., Robinson B.A., Hyman J.M. and Higdon D. 2009. Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. *International Journal of Nonlinear Sciences and Numerical Simulations* **10**, 273–290.
- Vrugt J.A., ter Braak C.J.F., Gupta H.V. and Robinson B.A. 2009. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? *Stochastic Environmental Research and Risk Assessment* **23**, 1011–1026.
- Westermann S., Wollschläger U. and Boike J. 2010. Monitoring of active layer dynamics at a permafrost site on Svalbard using multi-channel ground-penetrating radar. *Cryosphere* **4**, 475–487.
- Wollschläger U., Gerhards H., Yu Q. and Roth K. 2010. Multi-channel ground-penetrating radar to explore spatial variations in thaw depth and moisture content in the active layer of a permafrost site. *Cryosphere* **4**, 269–283.