WATER RESOURCES RESEARCH, VOL. 45, W12404, doi:10.1029/2008WR007471, 2009

Click Here for Full Article

Estimation of the lateral correlation structure of subsurface water content from surface-based ground-penetrating radar reflection images

James Irving,[1] Rosemary Knight,[2] and Klaus Holliger[1]

Received 23 September 2008; revised 1 July 2009; accepted 8 July 2009; published 4 December 2009.

[1] Over the past decade, significant interest has been expressed in relating the spatial statistics of surface-based reflection ground-penetrating radar (GPR) data to those of the imaged subsurface volume. A primary motivation for this work is that changes in the radar wave velocity, which largely control the character of the observed data, are expected to be related to corresponding changes in subsurface water content. Although previous work has indeed indicated that the spatial statistics of GPR images are linked to those of the water content distribution of the probed region, a viable method for quantitatively analyzing the GPR data and solving the corresponding inverse problem has not yet been presented. Here we address this issue by first deriving a relationship between the 2-D autocorrelation of a water content distribution and that of the corresponding GPR reflection image. We then show how a Bayesian inversion strategy based on Markov chain Monte Carlo sampling can be used to estimate the posterior distribution of subsurface correlation model parameters that are consistent with the GPR data. Our results indicate that if the underlying assumptions are valid and we possess adequate prior knowledge regarding the water content distribution, in particular its vertical variability, this methodology allows not only for the reliable recovery of lateral correlation model parameters but also for estimates of parameter uncertainties. In the case where prior knowledge regarding the vertical variability of water content is not available, the results show that the methodology still reliably recovers the aspect ratio of the heterogeneity.

Citation: Irving, J., R. Knight, and K. Holliger (2009), Estimation of the lateral correlation structure of subsurface water content from surface-based ground-penetrating radar reflection images, *Water Resour. Res.*, *45*, W12404, doi:10.1029/2008WR007471.

1. Introduction

[2] A critical problem in hydrological studies is the accurate description of subsurface heterogeneity at a scale relevant to the prediction of contaminant fate and transport. To this end, geophysical methods offer the potential to bridge a gap in terms of resolution and coverage between traditional hydrological characterization techniques such as borehole log and core analyses and pumping and tracer tests [e.g., *Rubin and Hubbard*, 2005]. Of particular interest in the recent past has been the idea of relating the lateral correlation statistics of surface-based reflection ground-penetrating radar (GPR) data to those of the water content distribution of the probed region [e.g., *Rea and Knight*, 1998; *Dafflon et al.*, 2006; *Knight et al.*, 2007]. The primary reasons for this interest are that (1) changes in water content are widely recognized as the predominant cause of changes in the radar wave velocity, and thus of GPR reflections, in the subsurface; (2) in electrically resistive media, surface-based GPR reflection surveys provide a quick and noninvasive means of imaging the shallow subsurface with unsurpassed resolution; and (3) geostatistical information on subsurface heterogeneity is particularly useful for many hydrological applications [e.g., *Gelhar*, 1993]. In the saturated zone, information regarding the lateral correlation structure of water content is equivalent to information regarding lateral porosity variability, which is extremely difficult to assess from borehole data alone yet is critically important in hydrological studies because of its inherent ties with storage and permeability. In the vadose zone, the spatial distribution of water content is a complicated function of soil properties, boundary and initial conditions, and distance from the water table. Nevertheless, being able to estimate key aspects of the stochastic structure of water content in unsaturated regions could provide important information regarding the spatial distribution of soil textures, which tend to critically affect the unsaturated hydraulic conductivity and thus vadose zone contaminant transport processes [e.g., *Hillel*, 2007].

[3] A number of studies have explored the relationship between the lateral correlation structure of seismic and radar velocity fields and those of the corresponding seismic and GPR reflection data [e.g., *Gibson*, 1991; *Hurich*, 1996; *Pullammanappallil et al.*, 1997; *Line et al.*, 1998; *Rea and Knight*, 1998; *Bean et al.*, 1999; *Knight et al.*, 2004; *Oldenborger et al.*, 2004; *Dafflon et al.*, 2006; *Carpentier and Roy-Chowdhury*, 2007; *Knight et al.*, 2007]. Although the seismic and GPR communities have to date worked

[1]Institute of Geophysics, University of Lausanne, Lausanne, Switzerland.
[2]Geophysics Department, Stanford University, Stanford, California, USA.

Copyright 2009 by the American Geophysical Union.
0043-1397/09/2008WR007471$09.00

largely independently of one another, it is important to note that the results obtained are mutually transferrable because of the strong mathematical analogies that exist between seismic and electromagnetic wave propagation [e.g., *Carcione and Cavallini*, 1995; *Belina et al.*, 2009]. In the seismic literature, the current consensus appears to be that, although there clearly exists a relationship between the lateral correlation properties of seismic data and those of the underlying velocity distribution for weakly scattering media, the nature of this relationship is largely unclear. Work by *Holliger et al.* [1994] and *Pullammanappallil et al.* [1997] initially suggested that the average lateral correlation structures of both of these fields should be equivalent. However, the more recent results of *Bean et al.* [1999] and *Carpentier and Roy-Chowdhury* [2007] have pointed out the fundamental dependence of the lateral correlation properties of a seismic image on bandwidth and on the vertical derivative operator that acts to create reflection coefficients from an impedance field, respectively. Clearly, such details concerning the physics of the experiment must be considered in any strategy to estimate the correlation structure of subsurface velocity from seismic reflection data.

[4] Within the GPR community, similar research efforts have led to corresponding results and conclusions. *Rea and Knight* [1998] and *Dafflon et al.* [2006], for example, saw very good agreement between the lateral geostatistics of a cliff face photograph and those of a GPR image collected along the top of the cliff. In both cases, the rather weak assumption was made that the gray scale tones in the photograph were representative of sediment grain size, with the further assumption that finer-grained sediments would be associated with higher water contents. *Knight et al.* [2007] also observed similarities between the horizontal correlation statistics of GPR data and neutron probe measurements. However, *Knight et al.* [2004] and *Oldenborger et al.* [2004] both noticed that the lateral correlation structure of a GPR image is significantly affected by its vertical resolution, which is controlled by the GPR antenna frequency. Again, this suggests that the elementary physics of the underlying wave propagation phenomena must be accounted for in any correlation analysis procedure. What is critically needed to further research in both the seismic and GPR domains, therefore, is (1) a consistent methodological framework linking the lateral correlation properties of a reflection image to those of the subsurface properties and (2) an appropriate corresponding inversion strategy.

[5] Here, we address the above issue and present a new approach for estimating the parameters describing the lateral variability of water content from surface-based GPR reflection data. Our approach accounts for the essential details of wave propagation phenomena and their effect on the correlation structure of a GPR image, something which has been missing in previous related efforts. We begin by deriving a relatively simple relationship between the 2-D spatial autocorrelation of a water content distribution and that of the corresponding GPR image. Next, we pose the problem of estimating the water content correlation model parameters from the GPR data in the context of Bayesian geophysical inverse theory and describe how Markov chain Monte Carlo (MCMC) sampling can be effectively used in this regard. Finally, we apply our methodology to a realistic synthetic example as well as to field GPR data collected at two different frequencies, and we assess and discuss the limitations and potential of the technique.

2. Methodology

[6] To relate the 2-D spatial autocorrelation of a GPR reflection image to that of the underlying water content distribution, we begin with the assumption that the recorded GPR data, after suitable processing, can be approximately modeled by what is known as a primary reflectivity section (PRS) [e.g., *Gibson*, 1991; *Pullammanappallil et al.*, 1997; *Bean et al.*, 1999]. That is, we assume that the processed and migrated/imaged GPR data can be treated as the convolution product of a reflection coefficient distribution, which is approximately obtained by vertical differentiation of the underlying radar wave velocity distribution, with a source wavelet function. In the depth domain, this is expressed as follows:

$$\begin{aligned} d(x,z) &= r(x,z) * w(z) \\ &\approx \frac{\partial}{\partial z} v(x,z) * w(z), \end{aligned} \tag{1}$$

where the asterisk denotes the convolution operator, x and z are the spatial coordinates, $d(x,z)$ represents the processed and depth-migrated GPR amplitude data, $r(x,z)$ is the reflection coefficient distribution, $v(x,z)$ is the velocity distribution, and $w(z)$ is the source wavelet. The PRS is composed of a series of 1-D, vertical incidence, primaries-only GPR traces and can be regarded as the ideal GPR image. The model is based on the assumptions that (1) single scattering prevails, and thus the effects of multiple scattering have either been removed by processing or are sufficiently benign to be ignored, an assumption that is indeed inherent to most seismic and GPR processing, imaging, and interpretation strategies; (2) dispersion in the GPR data is minor or has been corrected such that an approximately constant wavelet shape can be assumed; and (3) the GPR data have been properly migrated. Under these conditions, which are widely regarded as quite realistic, the PRS model is able to capture the essential features of a migrated reflection GPR data set.

[7] One limitation of the PRS model in the context of this study is its inability to account for the inherently limited lateral resolution of real GPR data. It is well known that horizontal resolution in an unmigrated GPR or seismic reflection section is limited by the Fresnel zone, which essentially describes the area on a subsurface reflector that contributes to the recorded data [e.g., *Berkhout*, 1984]. The radius of the Fresnel zone increases with depth and wavelength, and hence, unmigrated reflection sections have a horizontal resolution which worsens with increasing depth and decreasing frequency. The process of migration acts to collapse the size of the Fresnel zone to a uniform theoretical value on the order of the dominant wavelength and thus to improve the horizontal resolution everywhere in a reflection image [e.g., *Stolt and Benson*, 1986]. For accurate estimation of the horizontal correlation statistics of water content under a wide range of scenarios, however, it is necessary to take into account this lateral resolution limit of a migrated

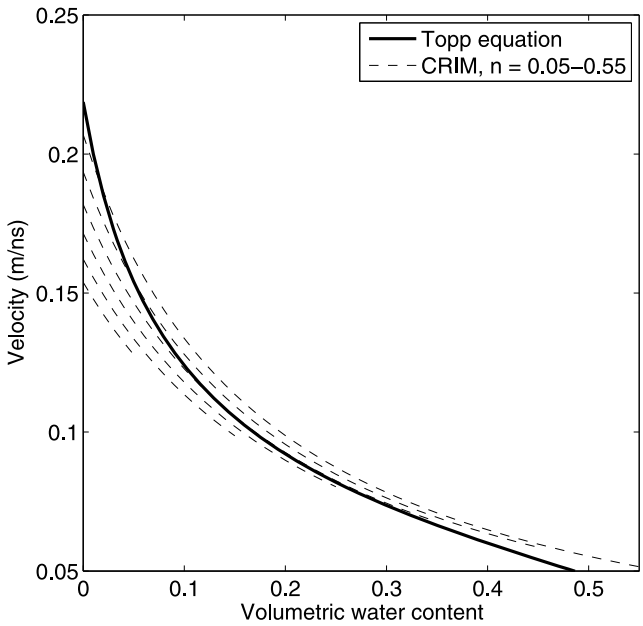

Figure 1. Electromagnetic wave velocity versus water content relationship predicted by the Topp equation and CRIM model with porosity values between 0.05 and 0.55 and a relative dielectric permittivity of $\varepsilon_s = 4$ for the dry matrix.

GPR image. We do this by modifying equation (1) to include a horizontal resolution filter, $h(x)$, as follows:

$$d(x,z) \approx \frac{\partial}{\partial z} v(x,z) * w(z) * h(x). \tag{2}$$

We have found that a simple Gaussian low-pass filter, whose width is dependent on the dominant signal wavelength, is a very effective choice for $h(x)$. This filter is of the form

$$h(x) = \exp\left(-\frac{x^2}{2c^2}\right), \tag{3}$$

where c determines the filter width and is set such that the distance between the two points where the Gaussian reaches 1% of its maximum amplitude is equal to the dominant wavelength. Extensive tests that we have performed under a range of realistic conditions and frequencies confirm that GPR data modeled using equation (2) closely match their corresponding finite difference modeled counterparts. Most importantly, equation (2) captures very effectively the overall statistical character of a GPR image and is thus well suited for use in the correlation analysis methodology presented here.

[8] In many cases, the relationship between radar wave velocity and water content at a given site, considering the vadose and saturated zones separately, can be adequately described as linear. In Figure 1, we plot velocity versus water content using the commonly used empirical relation by *Topp et al.* [1980] and the complex refractive index method (CRIM) model of *Roth et al.* [1990] for a variety of matrix porosities. Although the relationship is clearly nonlinear over the entire range of water content values shown, we see that over restricted ranges characteristic of many field scenarios, a linear relationship is not a bad approximation. This is especially the case for ranges of higher water content typically encountered in the saturated zone but is also valid enough in many unsaturated settings for the geostatistical work considered here. Considering a linear relationship between radar velocity and water content with constant of proportionality k, equation (2) can be rewritten as follows:

$$d(x,z) \approx \frac{\partial}{\partial z}[k\ \theta(x,z)] * w(z) * h(x), \tag{4}$$

where $\theta(x,z)$ is water content. In other words, the GPR response to the subsurface according to our modified PRS model can be approximately formulated in terms of the spatial structure of vertical changes in subsurface water content convolved with the radar pulse and horizontal resolution filter.

[9] Noting that the derivative operator in equation (4) will yield the same result when applied to the mean-removed water content distribution, and also that this operator can be treated as a filter whose position in the equation can be shifted to act on the wavelet, we obtain

$$d(x,z) \approx \tilde{\theta}(x,z) * f(z) * h(x), \tag{5}$$

where $\tilde{\theta}(x,z)$ is the mean-removed water content distribution and $f(z)$ is the differentiated wavelet into which we absorb the proportionality constant

$$f(z) = k\frac{\partial}{\partial z} w(z). \tag{6}$$

Transforming equation (5) into the frequency domain and taking the squared magnitude of both sides, we arrive at the following relationship between the power spectra of all quantities:

$$|D(k_x,k_z)|^2 \approx \left|\tilde{\theta}(k_x,k_z)\right|^2\ |F(k_z)|^2 |H(k_x)|^2. \tag{7}$$

Now taking the inverse Fourier transform and making use of the Wiener-Khintchine theorem linking the power spectrum and autocorrelation functions, we arrive at the final result:

$$R_{dd}(x,z) \approx R_{\tilde{\theta}\tilde{\theta}}(x,z) * R_{ff}(z) * R_{hh}(x). \tag{8}$$

[10] Equation (8) states that the 2-D spatial autocorrelation of a GPR image, $R_{dd}(x,z)$, is approximately equivalent to the 2-D autocorrelation of the mean-removed water content distribution, $R_{\tilde{\theta}\tilde{\theta}}(x,z)$, filtered vertically by the autocorrelation of the differentiated source wavelet, $R_{ff}(z)$, and horizontally by the autocorrelation of lateral resolution filter, $R_{hh}(x)$. What this means is that, with an estimate of the wavelet or its spectrum, and thus the autocorrelation of the differentiated wavelet, we can estimate the correlation properties of the mean-removed water content distribution from those of the GPR image. It must be emphasized that although the convolution with $R_{ff}(z)$ in equation (8) operates only in the vertical direction, this vertical convolution has a

pronounced effect on the lateral correlation structure of the image, independent of the horizontal convolution with $R_{hh}(x)$. What this means is that both the vertical and lateral correlation properties will, in general, be significantly different between a water content distribution and the corresponding GPR image. In our view this is a key aspect that has so far not been properly accounted for in previous efforts to estimate the lateral correlation statistics of water content from GPR data [e.g., *Rea and Knight*, 1998; *Dafflon et al.*, 2006; *Knight et al.*, 2007].

[11] To use equation (8) for the estimation of water content correlation parameters from a GPR image, we employ a Bayesian inversion strategy with MCMC sampling from the posterior distribution of model parameters [e.g., *Mosegaard and Tarantola*, 1995]. The inversion procedure consists of the following steps.

[12] 1. Draw a proposed set of model parameters, **m**, which parameterize the 2-D autocorrelation function of the mean-removed water content distribution, from a prescribed prior probability distribution for these parameters. We assume in our analysis that the autocorrelation of water content follows some simple parametric form such that relatively few parameters are required for its description. This greatly improves the estimation problem and makes the MCMC inversion approach computationally tractable.

[13] 2. For the set of parameters drawn in step 1, compute the 2-D autocorrelation of water content, $R_{\tilde{\theta}\tilde{\theta}}(\mathbf{m})$, using the assumed parametric equation. Then calculate the predicted autocorrelation of the corresponding GPR image, $R_{dd}^{pred}(\mathbf{m})$, using equation (8) and our best knowledge of the source wavelet characteristics.

[14] 3. Calculate the likelihood of the currently proposed set of model parameters by comparing the predicted autocorrelation with that calculated from the observed GPR data, R_{dd}^{obs}. We assume that the errors in R_{dd}^{obs} are independent and Gaussian, with variances given by $\sigma_{d_i}^2$, where i varies over the data for different spatial lag vectors. In this case, the likelihood function takes the following form [*Mosegaard and Tarantola*, 1995]:

$$L(\mathbf{m}) = \exp\left[-\sum_{i=1}^{N} \frac{\left(R_{dd_i}^{pred}(\mathbf{m}) - R_{dd_i}^{obs}\right)^2}{\sigma_{d_i}^2}\right]. \tag{9}$$

[15] 4. If the likelihood of the currently proposed set of model parameters is greater than or equal to that of the previously accepted set, or if this is the first iteration of the procedure, accept **m** and return to step 1 for another iteration.

[16] 5. If the likelihood of the currently proposed set of model parameters is less than that of the previously accepted set, do not necessarily reject **m**, but instead subject it to a random decision rule whereby the probability of its acceptance is equal to the ratio of the two likelihoods, $L_{\text{old}}/L_{\text{new}}$. Then return to step 1 for another iteration.

[17] It can be shown that after a sufficient number of iterations known as the "burn-in" period, the above procedure will converge to produce samples from the Bayesian posterior distribution of model parameters. This distribution takes into account both the available prior information and observed autocorrelation data and is essentially an update of the prior distribution into a refined state of knowledge about the system. Once a sufficient number of posterior samples have been generated, they can then be analyzed to determine the most likely water content correlation model parameters that gave rise to the recorded GPR reflection data along with estimates of the associated uncertainties. In addition, the suite of realizations can be examined for interrelationships existing between the model parameters. The above procedure is schematically illustrated and conceptually summarized by the flowchart shown in Figure 2.

3. Synthetic Data Example

[18] We now present a realistic synthetic example of how the inversion strategy described in section 2 can be used to estimate the correlation parameters of subsurface water content that are consistent with reflection GPR data. Figure 3a shows the stochastic water content distribution that we consider for this purpose. This distribution was generated using an anisotropic exponential autocorrelation model whose parametric form is given by

$$R(x,z) = \exp\left[-\sqrt{\left(\frac{x}{a_x}\right)^2 + \left(\frac{z}{a_z}\right)^2}\right], \tag{10}$$

where the model parameters a_x and a_z are the horizontal and vertical correlation lengths, respectively. We assume with equation (10) that the axis of anisotropy for the heterogeneity is aligned with the x and z axes. To obtain the realization in Figure 3a, we used a spectral simulation technique which involved assigning a uniformly random phase spectrum to the amplitude spectrum of the above autocorrelation model and taking the inverse Fourier transform [e.g., *Goff et al.*, 1994].

[19] The mean and standard deviation of the water content distribution in Figure 3a are 0.25 and 0.015, respectively. The minimum and maximum values are 0.19 and 0.31. This range is representative of a saturated zone scenario. For simplicity, we did not include in our water content model an unsaturated region near the Earth's surface. Note that this in no way limits the applicability of our findings, as in practice only the GPR data from below the water table would be selected for analysis. Also note that although our synthetic example deals with a saturated zone situation, the inversion methodology could be equally well applied to data collected in the vadose zone when the distribution of water content can be adequately described by a stochastic correlation model. Indeed, the field data presented in section 4 were collected in an unsaturated setting.

[20] To simulate a GPR experiment over the water content distribution in Figure 3a, we first needed to obtain values for subsurface electrical properties required by the finite difference time domain (FDTD) solution of Maxwell's equations in 2-D Cartesian coordinates that was used to perform the modeling [*Irving and Knight*, 2006]. Values for the dielectric permittivity, ε, were obtained from the water content model using the CRIM equation, assuming a dielectric permittivity for the dry matrix of $\varepsilon_s = 4\varepsilon_0$, with ε_0 denoting the permittivity of free space. As most surficial geological materials are nonmagnetic, the magnetic permeability, μ, was set to its value in free space, μ_0, throughout

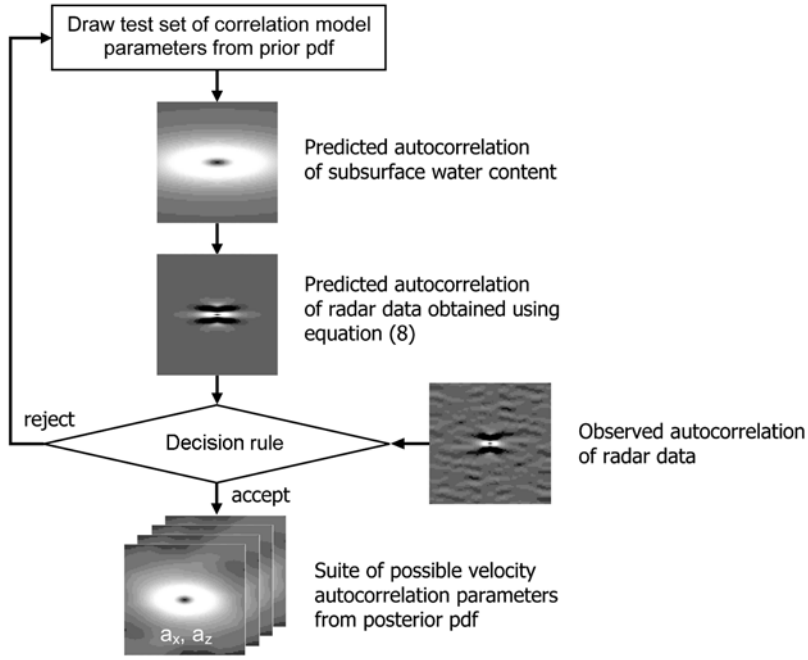

Figure 2. Flowchart illustrating our Bayesian-MCMC approach for estimating the parameters describing the correlation model of subsurface water content from reflection GPR data.

the simulation region. For the electrical conductivity, σ, we assumed a uniform value of 1 mS/m. This is reasonable for the sand-and-gravel-type materials amenable to GPR wave propagation, as the effects of conductivity variations with water content predicted by Archie's law [*Archie*, 1942] would be minimal in this case. Figure 3b shows the radar wave velocity field corresponding to Figure 3a, which was obtained from the electrical properties using the low-loss approximation

$$v \approx \frac{1}{\sqrt{\mu\varepsilon}}. \tag{11}$$

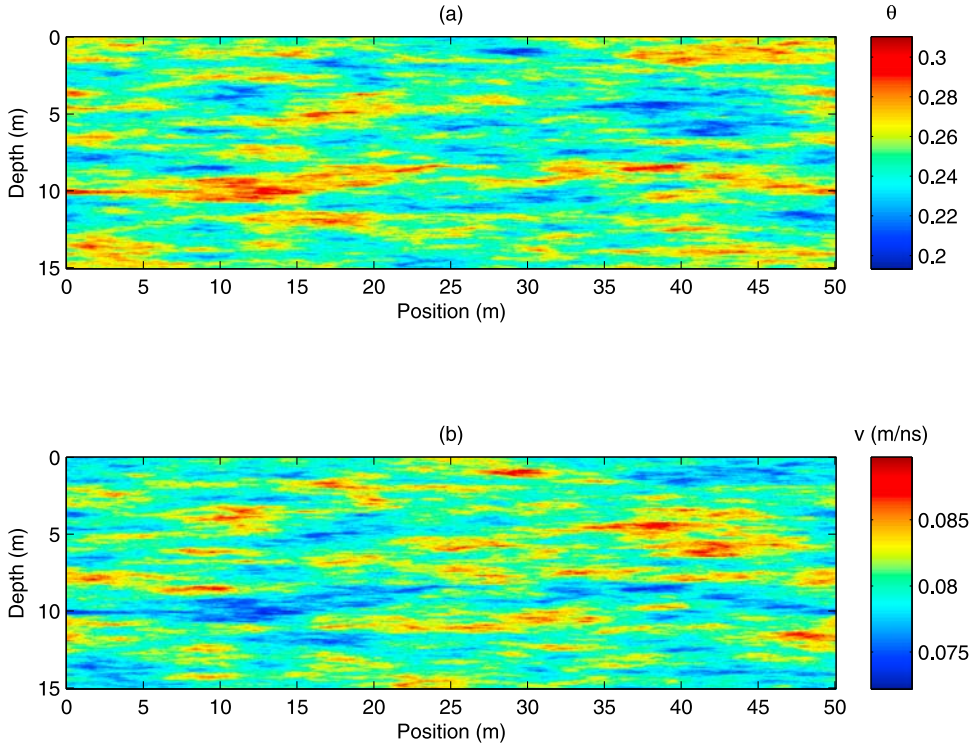

Figure 3. (a) Water content distribution generated using an anisotropic, exponential autocorrelation model. (b) Corresponding radar wave velocity distribution obtained using the CRIM model.

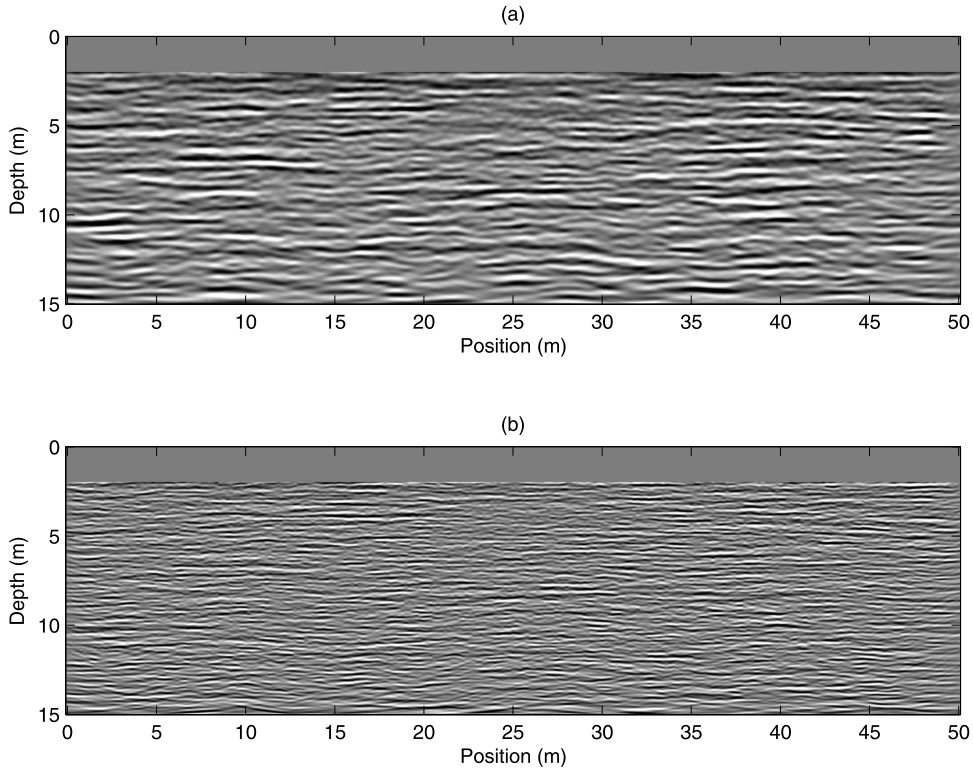

Figure 4. (a) Zero-offset 50 MHz GPR reflection image obtained by the FDTD modeling over the velocity distribution shown in Figure 3b followed by gaining and migration. (b) Corresponding 100 MHz GPR reflection image.

[21] GPR reflection data were simulated over the environment in Figure 3 at two different frequencies. Zero-offset traces were recorded at lateral increments of 0.2 m along the Earth's surface. The source current function that was input into the modeling grid was a differentiated Gaussian pulse having a dominant frequency of 50 and 100 MHz for the first and second data sets, respectively. After the numerical simulations were complete, the data were first gained in time using a smoothly varying function inversely proportional to the average trace envelope. A smooth gain such as this preserves the relative amplitudes of reflections along a trace but corrects for the effects of geometrical spreading and attenuation; however, we have found that our correlation analysis methodology is relatively insensitive to the type of gain used. Next, the GPR data were depth migrated in the frequency–wave number domain using the constant velocity algorithm of *Stolt* [1978] and a mean subsurface velocity value of 0.08 m/ns. For the 2-D synthetic data considered here, this type of migration is perfectly acceptable as the radiation characteristics of the line source are uniform with angle and the velocity variations for imaging purposes are minimal. In the case of field GPR data, which we consider in section 4, there is a long and successful history of employing standard seismic processing and migration algorithms such as this one [e.g., *Annan*, 2005]. Although other GPR-specific and potentially more accurate imaging methodologies are available, we believe that the seismic algorithms are sufficient for our purposes because most practical approaches to amplitude scaling, including ours, largely obliterate any traces of the directivity of the transmitter and receiver antennas in the GPR data. As a final step, band-limited Gaussian random noise with a standard deviation of 5% of the RMS amplitude of the data was added to simulate more realistic conditions.

[22] Figures 4a and 4b show the processed and migrated 50 and 100 MHz GPR images, respectively. The top parts of the data were muted to suppress waves traveling directly between the antennas. This muted zone was not considered for analysis. Notice the similarities between the two images in terms of the location and nature of the dominant reflecting interfaces. At the same time, however, because of the difference in frequency content of the GPR pulses used, Figure 4b appears to have a distinctly longer horizontal correlation length than Figure 4a. Any reasonable inversion strategy to determine the underlying subsurface correlation structure must account for the effective changes in lateral correlation of the GPR image produced by the predominantly vertical interaction with the radar pulse.

[23] Figures 5a and 5b show the calculated 2-D autocorrelations corresponding to the original water content model and radar wave velocity distribution in Figures 3a and 3b, respectively. As prescribed, these autocorrelations follow an exponential parametric model. Best fitting correlation lengths to the autocorrelation data in Figure 5a were determined using a nonlinear least squares algorithm to be $a_z = 0.35$ m and $a_x = 3.53$ m. For the data in Figure 5b, they were determined to be $a_z = 0.34$ m and $a_x = 3.48$ m. The correlation model parameters for the water content and velocity distributions are thus approximately equal. This results because of the roughly linear relationship between

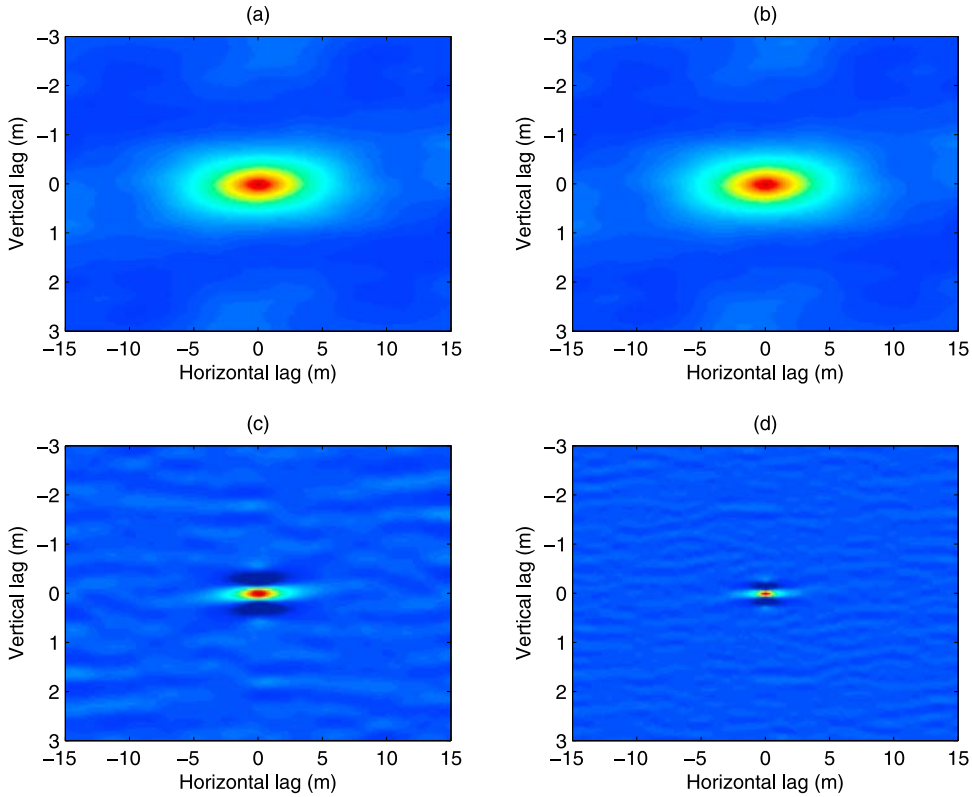

Figure 5. (a) Autocorrelation of the mean-removed water content distribution from Figure 3a. Best fitting exponential model parameters are a_z = 0.35 m and a_x = 3.53 m. (b) Autocorrelation of the mean-removed velocity distribution from Figure 3b. Best fitting model parameters are a_z = 0.34 m and a_x = 3.48 m. (c) Autocorrelation of the 50 MHz GPR image from Figure 4a. (d) Autocorrelation of the 100 MHz GPR image from Figure 4b.

these two variables over the range of water content values encountered in Figure 3a (see also Figure 1).

[24] In Figures 5c and 5d, we see the calculated 2-D autocorrelations corresponding to the 50 and 100 MHz radar images in Figures 4a and 4b, respectively. Notice the distinct difference between these autocorrelations and those in Figures 5a and 5b. In both cases, we have a visibly much shorter correlation length in the horizontal direction, and the vertical correlation properties of the radar images are largely controlled by the spectral content of the GPR wavelet rather than the vertical structure of the underlying water content heterogeneity. This is an excellent example of how, in going from water content to the corresponding GPR reflection image, we may encounter significant changes in both the vertical and horizontal correlation properties. Between the 50 and 100 MHz data autocorrelations, we also have a significant difference in the autocorrelation behavior; the 50 MHz data show a distinctly longer correlation length in the vertical and horizontal directions. Again, any reasonable method for estimating the lateral correlation structure of water content from GPR data must account for all of these effects by adequately capturing the physics of the GPR experiment.

[25] The goal of our numerical example is to estimate, from the autocorrelation data in Figures 5c and 5d, the parameters describing the autocorrelation in Figure 5a. Assuming that the data can be adequately explained by our model in equation (8), the GPR image and water content autocorrelations will be related through convolution with the autocorrelations of the differentiated source pulse and horizontal resolution filter. To estimate $R_{ff}(z)$, we require, at a minimum, an estimate of the spectral content of $f(z)$. Here we determined $R_{ff}(z)$ from an estimate of $w(z)$ obtained from the ground wave pulse traveling directly between the GPR antennas at a far horizontal offset, where it was possible to separate this pulse from the one arriving directly through the air. Although we believe that it is reasonable to have knowledge of $R_{ff}(z)$, obtained either in this manner or through some other method, uncertainty about the details of the source wavelet could be accounted for in our inversion procedure through the use of additional model parameters.

[26] Also required for the Bayesian-MCMC inversion are prior probability distributions for the autocorrelation model parameters, in our case the vertical and horizontal correlation lengths a_z and a_x. For this example, we first assumed relatively good knowledge of a_z but very little knowledge of a_x. We also assumed to know the parametric form of the autocorrelation model. We think that this is reasonable because information regarding the autocorrelation model and vertical structure could be estimated from sparse borehole data or perhaps from some form of deconvolution applied to the recorded data [e.g., *Poppeliers and Levander*, 2004; *Poppeliers*, 2007]. Moreover, we have found that knowledge regarding the vertical variability is crucial in some cases for reliable estimation of the lateral correlation

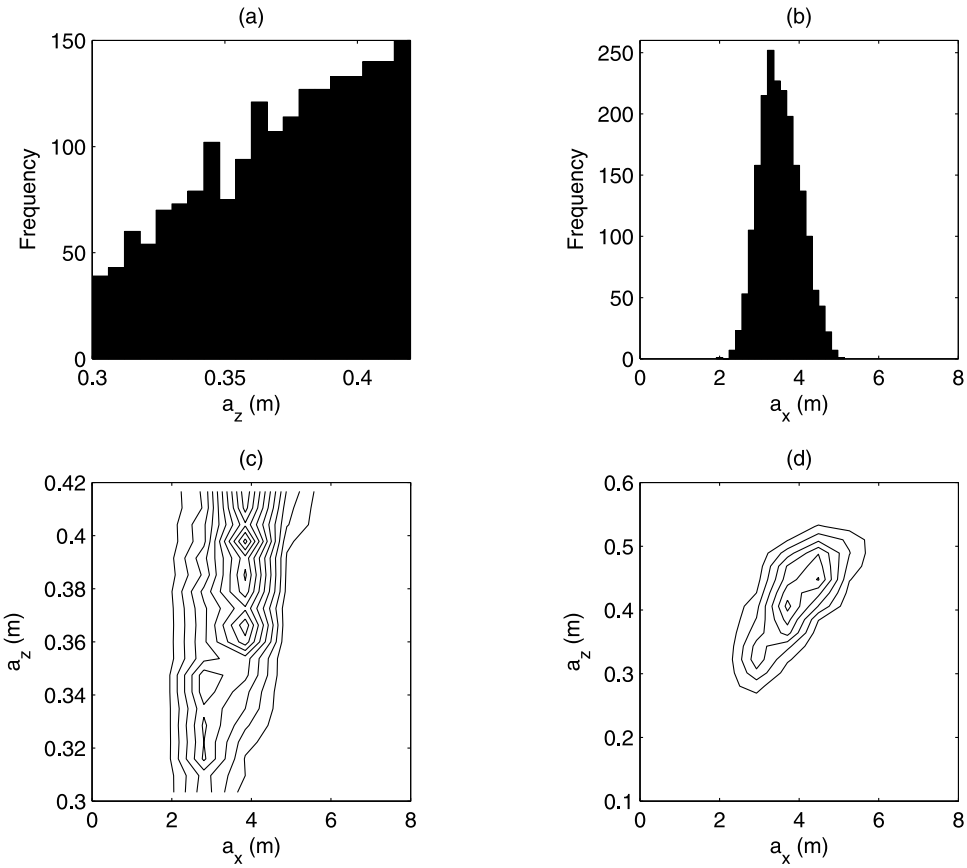

Figure 6. Results of MCMC analysis of the synthetic 50 MHz GPR image. (a and b) Marginal posterior histograms for a_z and a_x, respectively, obtained using uniform prior ranges of a_z = [0.30, 0.42] m and a_x = [0.2, 20.0] m. (c) Corresponding contoured joint posterior histogram. (d) Contoured joint posterior histogram for the case where the prior range for a_z was extended to [0.1, 2.0] m.

properties because in the common situation where the vertical correlation length is less than the dominant propagating wavelengths, the vertical structure tends to be largely obliterated through averaging by the source wavelet. That is, the vertical autocorrelation of the GPR image often shows limited sensitivity to the vertical correlation structure of water content, yet the horizontal correlation structure of the data is strongly dependent on the vertical resolution. For a_z, we first assumed a uniform prior probability distribution with lower and upper bounds of 0.30 and 0.42 m. For a_x, we also assumed a uniform prior distribution but with much broader bounds between 0.2 and 20 m. With these values, the Bayesian-MCMC algorithm was run separately on both the 50 and 100 MHz data sets until 2000 realizations from the posterior probability distribution for a_z and a_x had been generated. On a dated 2.4 GHz desktop computer with 1 GB of RAM, this took approximately 1 h for each data set. Past the burn-in period, all posterior samples were used in the subsequent analysis. This is because the proposed sets of model parameters in our two-parameter MCMC code were drawn independently from the prior distribution.

[27] Figures 6a–6c show the results of our inversion for a_z and a_x based on the autocorrelation of the 50 MHz GPR data in Figure 5c. Figures 6a and 6b show the marginal histograms computed for these parameters from the suite of posterior realizations, whereas Figure 6c shows a contour plot of the joint 2-D histogram. The first thing to notice from these results is that, although quite specific prior information was provided in the inversion regarding the vertical correlation length of water content, there is very little refinement of this information in the posterior distribution. In fact, it is clear from Figures 6a and 6c that a wide range of a_z values is likely to be valid within the considered interval, with preference actually being given to a_z values near the upper limit of the uniform range instead of the true value of 0.35 m. As mentioned, we believe that this is related to the vertical averaging effects of the source wavelet, which result in the vertical autocorrelation of the GPR data being relatively insensitive to the underlying vertical stochastic structure of water content. For the horizontal correlation length, on the other hand, Figures 6b and 6c demonstrate that the 50 MHz GPR data allow for a significant reduction of uncertainty regarding this parameter compared to the broad uniform prior distribution. The mean value of a_x from the suite of posterior realizations is 3.60 m with a standard deviation of 0.50 m, compared to the true value of 3.53 m. In Figure 6d, we explore the effect on the stochastic inversion when detailed information regarding a_z is not available. Here, we show the contoured joint histogram for the case where the MCMC procedure was run with a broad uniform prior for a_z between 0.1 and 2.0 m. Clearly,

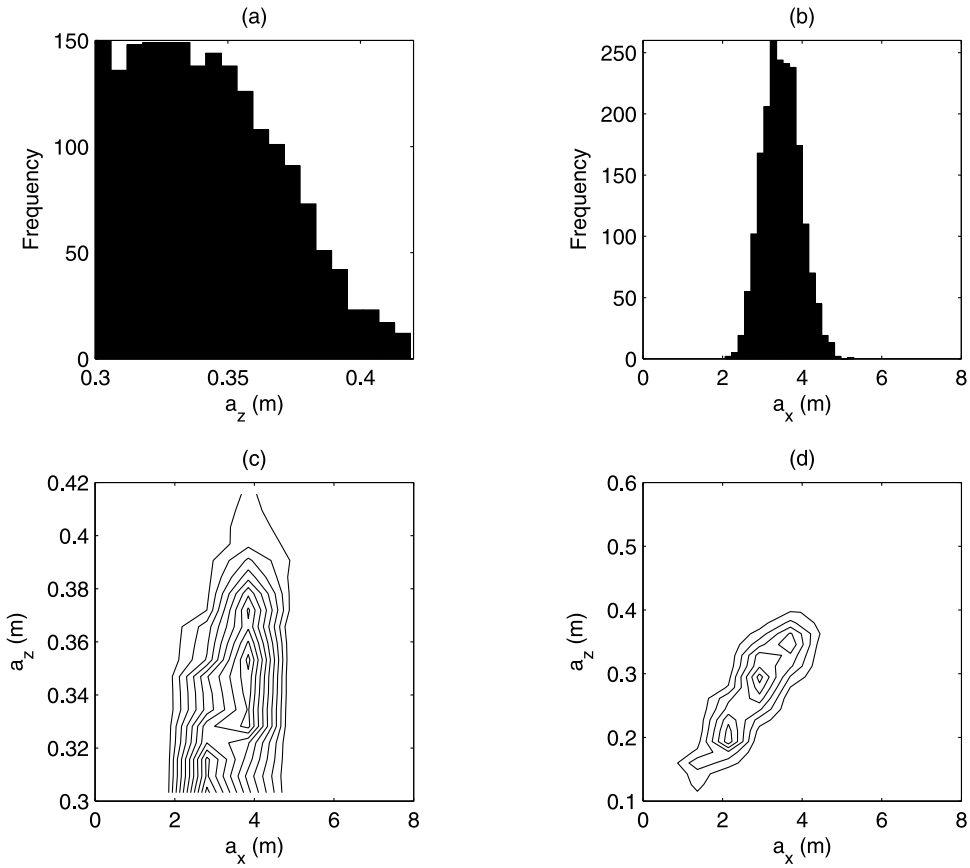

Figure 7. Results of MCMC analysis of the synthetic 100 MHz GPR image. (a and b) Marginal posterior histograms for a_z and a_x, respectively, obtained using uniform prior ranges of a_z = [0.30, 0.42] m and a_x = [0.2, 20.0] m. (c) Corresponding contoured joint posterior histogram. (d) Contoured joint posterior histogram for the case where the prior range for a_z was extended to [0.1, 2.0] m.

a variety of a_z-a_x combinations can explain very well the recorded GPR data, which suggests that without reliable knowledge of a_z, it may not be possible to accurately recover the horizontal correlation length. The mean value of a_x from the posterior realizations in this case is 3.86 m, with a standard deviation of 0.72 m. Although this is not a bad estimate of a_x, we see that without proper bounds on a_z, the posterior realizations are biased and have a larger standard deviation than in the case where the vertical structure is well constrained. Nevertheless, Figure 6d shows that there is a trade-off between a_z and a_x, with the most acceptable solutions occurring along a ridge-type feature having slope approximately equal to 10, which corresponds to the aspect ratio of the underlying water content heterogeneity.

[28] In Figure 7, we see the results of our inversion for a_z and a_x based on the autocorrelation of the 100 MHz GPR data in Figure 5d. Similar to the 50 MHz case, Figures 7a and 7b show the marginal histograms computed for a_z and a_x from the suite of posterior realizations, whereas Figures 7c and 7d show contour plots of the joint 2-D histogram of these parameters assuming the same narrow and broad prior ranges for a_z, respectively. Again, we see in Figure 7 that there is relatively little refinement in the information regarding a_z through the MCMC inversion, with values this time near the lower range of the prior uniform distribution being preferred. However, as in the case with the 50 MHz data, there is a significant reduction of uncertainty regarding a_x, with the mean value of the posterior realizations equal to 3.48 m with a standard deviation of 0.46 m. In the case where accurate information regarding a_z is not available, Figure 7d illustrates again that there appears to be a trade-off between a_z and a_x, with the range of most likely parameter combinations falling along a ridge-like feature having a slope approximately equal to 10. Here the mean value of a_x from the posterior realizations is 2.80 m with a standard deviation of 0.74 m. Again, the estimate is biased and more uncertain with inadequate knowledge of a_z, although it is still not a completely unreasonable value for the horizontal correlation length.

4. Field Data Example

[29] We now show the application of our inversion methodology to field GPR data acquired at a gravel quarry in northern Switzerland in an unsaturated glaciofluvial environment. The sediments at the site are unconcolidated and consist of structurally complex, gravel- and sand-dominated braided stream deposits. The local geology and its expression in radar images have been previously described by *Huggenberger* [1993] and *Beres et al.* [1999]. Common-offset reflection GPR data were collected at nominal antenna

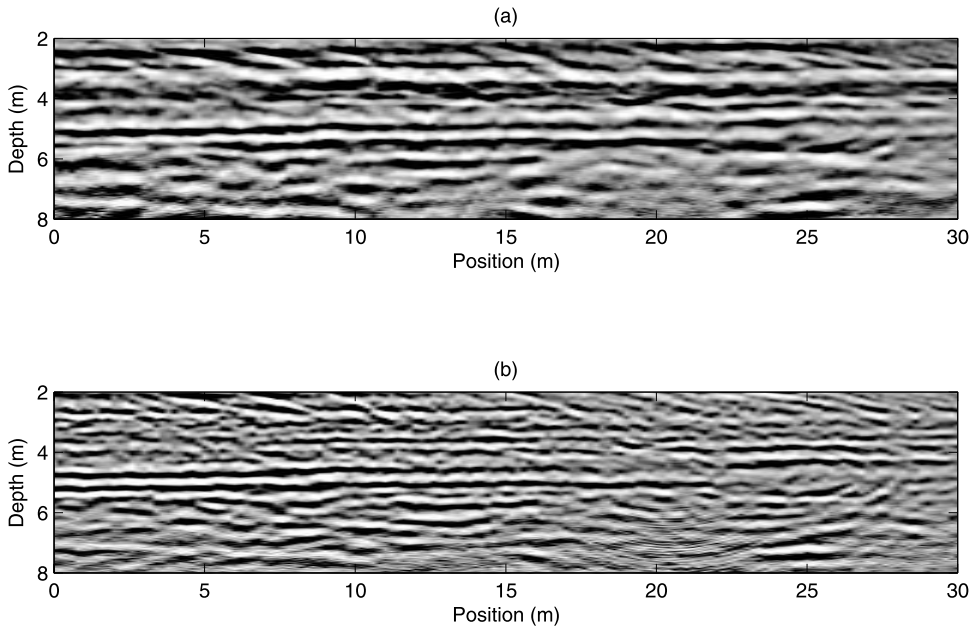

Figure 8. (a) Zero-offset 100 MHz field GPR data after gaining, migration, and cropping. (b) Corresponding 200 MHz GPR image.

center frequencies of 100 and 200 MHz using a commercial GPR system along an approximately 75 m long profile. To acquire the data, a time-sampling interval of 0.2 ns and trace spacing of 0.2 m were employed, and 64 vertical stacks were taken at each point to improve the signal-to-noise ratio. Common midpoint (CMP) data were also collected at two locations along the GPR survey line, such that an estimate of the radar wave velocity and its vertical variability could be obtained for migration of the reflection data. In the CMP analysis, which consisted of examining the hyperbolic curvature of reflections as a function of antenna separation, the velocity was found to be remarkably uniform with depth and approximately equal to 0.11 m/ns. As in our synthetic example, we also used the CMP data to determine the direct pulse traveling through the ground between the radar antennas, which provided an estimate of the radar wavelet, and thus $R_{ff}(z)$, for the inversion procedure.

[30] Figures 8a and 8b show the 100 and 200 MHz field GPR data after gaining, migration, and cropping of the images to select a suitable region for the statistical analysis. Only the last 30 m of the 75 m long survey line was considered as the rest of the collected data did not reach a suitable depth of penetration. One of the CMP data sets was located approximately at the center of this 30 m profile. Depths between 2 and 8 m were considered in the correlation analysis in order to avoid the direct arrivals at the top of the section and the regions where noise was clearly dominant in the lower part of the section. Again, the constant-velocity migration algorithm by *Stolt* [1978] was employed to image the data since the velocity was found to be approximately uniform. As mentioned previously, we feel that this last step is justified.

[31] Notice again the similarities between the two GPR images shown in Figure 8 in terms of their response to dominant reflecting interfaces in the subsurface. However, as in the case of our synthetic example, we see that the higher-frequency data appear to have a shorter lateral correlation length. This is because reflectors which effectively "line up" when imaged at 100 MHz become horizontally discontinuous when imaged at 200 MHz, again confirming the dependence on vertical resolution of the horizontal correlation structure of the GPR data [*Knight et al.*, 2004]. This effective decrease in correlation length of the data with increasing frequency is better seen in Figure 9, which shows the calculated 2-D autocorrelations of the 100 and 200 MHz data along with horizontal and vertical slices through the centers of these autocorrelations.

[32] Clearly, our goal with the field data example is to estimate the lateral correlation structure of water content and then to validate the results in some manner. However, in virtually all practical situations, obtaining the "true" spatial structure of the subsurface water content distribution for such a validation, especially in the lateral direction, is not possible. In the absence of such conclusive information, we focus here on inverting independently the 100 and 200 MHz field data. We then assess the validity of the estimated correlation model parameters through the degree of consistency between the two different frequencies. Unfortunately, no concrete information regarding the most suitable parametric correlation model and range for a_z were available at this particular site, the latter of which we saw in the synthetic example to be important for the accurate determination of a_x. Such information might be obtained at other field sites from the analysis of borehole porosity log data, if they are available. For our inversion, we again assumed that the water content correlation structure can be adequately captured by an exponential model as given by equation (10), which is arguably the most widely used geostatistical model. We also assumed a_z to lie within a uniform prior interval between 0.10 and 0.80 m, which we believe is broad enough to capture all realistic possibilities for this parameter at the field site on the basis of geological considerations. For the

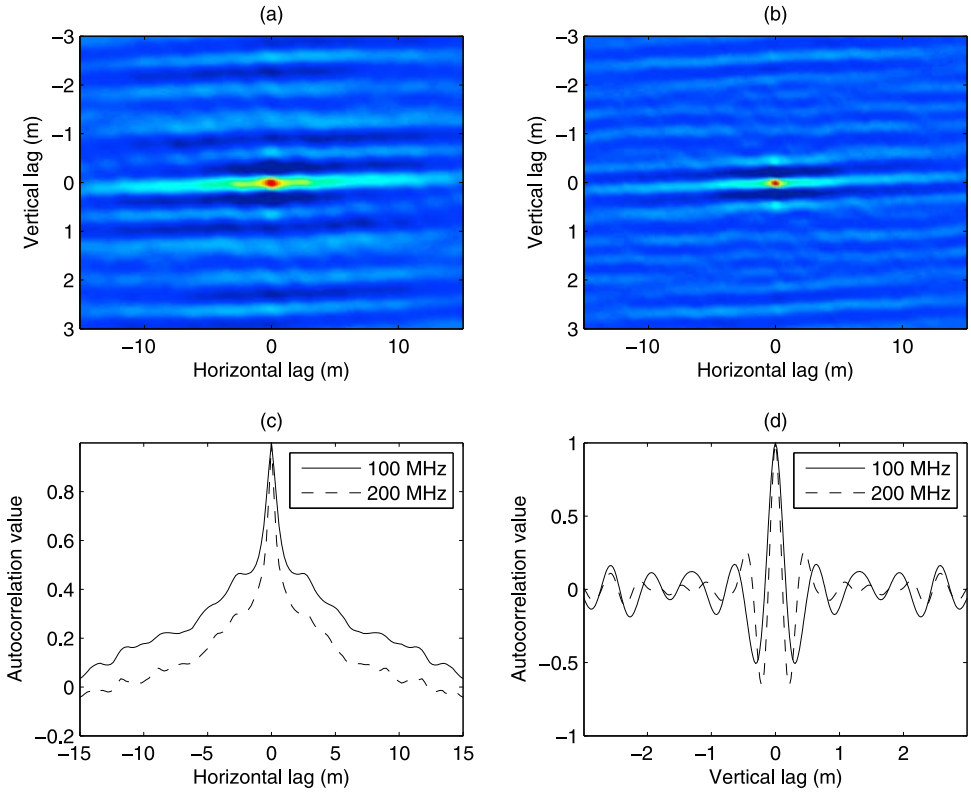

Figure 9. (a) Autocorrelation of the 100 MHz image from Figure 8a. (b) Autocorrelation of the 200 MHz image from Figure 8b. (c) Horizontal slices through the centers of Figures 9a and 9b. (d) Vertical slices through the centers of Figures 9a and 9b.

bounds on a_x, we assumed a very broad uniform prior distribution between 0.2 and 20 m. Like in the synthetic case, the MCMC inversion was run separately on the two autocorrelations in Figures 9a and 9b until 2000 samples from the posterior distribution of correlation model parameters were generated.

[33] Figures 10 and 11 show the histograms obtained from the posterior samples for the inversion of the 100 and 200 MHz data, respectively. In both Figures 10 and 11, we see a slight refinement in our knowledge of a_z through the inversion, but, most importantly, we see a consistent and significant refinement in our knowledge regarding a_x. The posterior histograms in Figures 10a and 11a show that only a_z values between 0.1 and 0.6 m can be considered as acceptable given the GPR data, even though the input prior range for this parameter extended to 0.8 m. In the case of the 200 MHz data, a_z values around 0.35 m are preferred. For a_x, the inversion of both data sets can be seen to result in posterior histograms having a distinctly normal appearance. The posterior mean and standard deviation for the 100 MHz results in Figure 10b are 4.17 and 1.60 m, respectively, whereas those for the 200 MHz results in Figure 11b are 4.70 and 1.58 m. The similarity of the posterior realizations obtained from the two inversions, within the bounds of uncertainty provided, indicates the internal consistency of our analysis and suggests the likely realism of the inferred parameter values. The larger mean value estimated from the 200 MHz data likely results from our lack of prior information regarding a_z and thus slightly biased results in both cases. Finally, from the joint 2-D histograms presented in Figures 10c and 11c, we see again that a range of acceptable a_z-a_x combinations is able to fit the recorded 100 and 200 MHz GPR data, with the slope of the acceptable solution domain being approximately equal to 13. This is consistent with published values for structural aspect ratios that are typically observed in comparable alluvial environments [e.g., *Gelhar*, 1993].

5. Discussion

[34] We have presented a novel methodology for the estimation of the lateral correlation structure of subsurface water content from surface-based reflection GPR data. Compared to previous related efforts in this domain, our methodology represents a significant step forward in that it is based on a theoretical foundation that captures critical aspects of the underlying wave propagation phenomena. Furthermore, the inversion strategy is posed within a versatile Bayesian-MCMC framework that allows for estimates of parameter uncertainties and the incorporation of additional unknowns. In applying our estimation procedure to both synthetic and field GPR images using multiple antenna frequencies, consistent information regarding the horizontal correlation length of water content was obtained. In the case where detailed information regarding the vertical correlation structure of water content was available, the estimates of the horizontal correlation length were very accurate. In the absence of such information, our technique proved to provide a rather robust estimate of the aspect ratio of the heterogeneity, which in its own right serves as valuable geostatistical information.

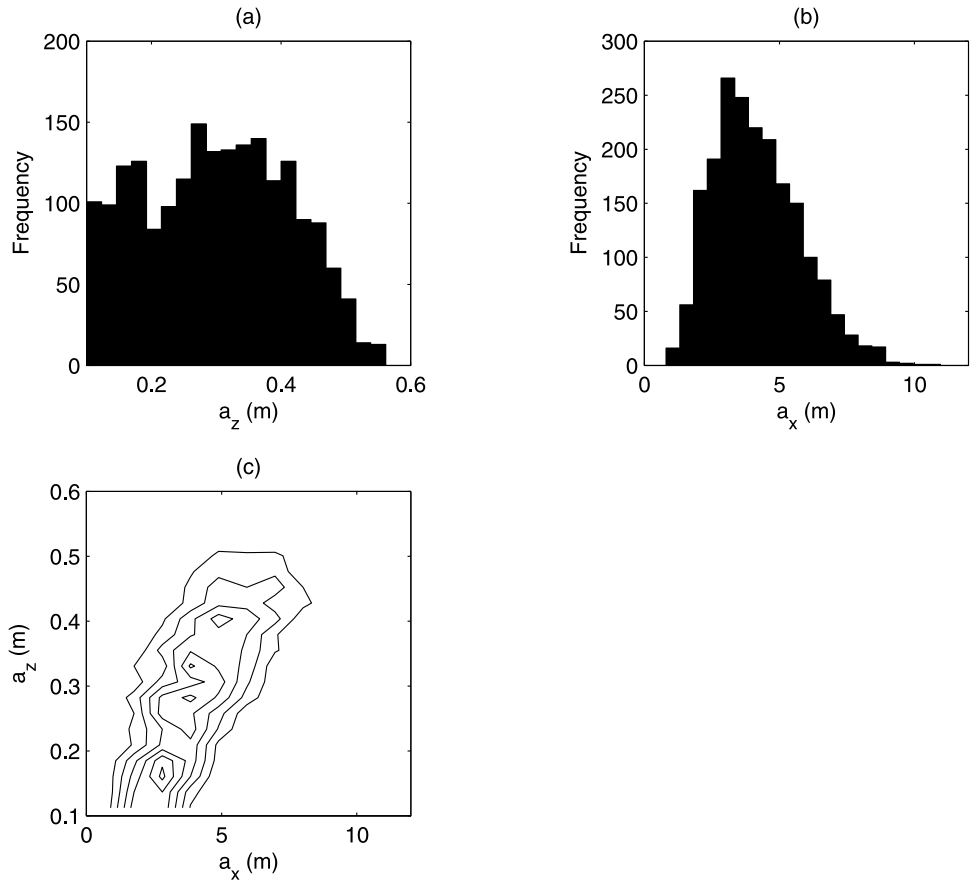

Figure 10. Results of MCMC analysis of the field 100 MHz GPR image. (a and b) Marginal posterior histograms for a_z and a_x, respectively, obtained using uniform prior ranges of a_z = [0.1, 0.8] m and a_x = [0.2, 20.0] m. (c) Corresponding contoured joint posterior histogram.

[35] In our formulation of the inverse problem and in our synthetic and field examples, a number of assumptions were made that warrant some additional discussion. First, we have mentioned that the PRS model, which forms the basis of our method, assumes that we are within the regime of weak, single scattering in the subsurface. We believe this to be reasonable for most GPR surveys, both above and below the water table, on the basis of expected variations in soil water content in both regions. Nevertheless, future work will include the investigation of whether our methodology can be adapted or modified for use in stronger scattering environments. In fact, with our technique, only the 2-D spatial autocorrelation of the GPR image is used to estimate the parameters describing the water content correlation model. That is, only the information contained in the 2-D Fourier power spectrum of the data is employed, and the phase information is discarded. Further accuracy and general applicability to stronger scattering environments might be obtained by attempting to link the full Fourier spectrum of the GPR data to the correlation properties of the underlying water content distribution.

[36] Another assumption inherent to the PRS model is that of wavelet stationarity. By working with this model in the depth domain in equation (1), we implicitly assume an approximately constant background subsurface velocity such that changes in wavelet shape in the migrated GPR image are negligible. Although strong background velocity trends can occur with GPR, in our experience this is quite rare except when crossing the water table. Furthermore, problems related to the assumption of wavelet stationarity could likely be overcome by applying our inversion procedure to regions of a migrated image where wavelet shape is visibly similar. Regarding changes in the GPR wavelet resulting from frequency-dependent material properties in the subsurface, such effects, if moderate enough to allow for successful GPR imaging of the subsurface region, could likely be corrected prior to analysis using, for example, the technique described by *Irving and Knight* [2003].

[37] In our synthetic and field examples, we assumed that knowledge was available regarding the parametric model for the autocorrelation of subsurface water content, in our case an exponential model described by a_x and a_z. We also found that prior information regarding the range of a_z was very beneficial for the reliable recovery of a_x. In the case where borehole data are available, we believe these assumptions to be reasonable. Otherwise, it may be possible to include the correlation model itself as an unknown in the MCMC inversion procedure, such that models other than the exponential one can be considered with regard to their ability to fit the GPR data. Indeed, preliminary results indicate that GPR data contain pertinent information not only on the lateral correlation structure but also on the parametric form of the autocorrelation model. More problematic in the absence of borehole data is having knowledge

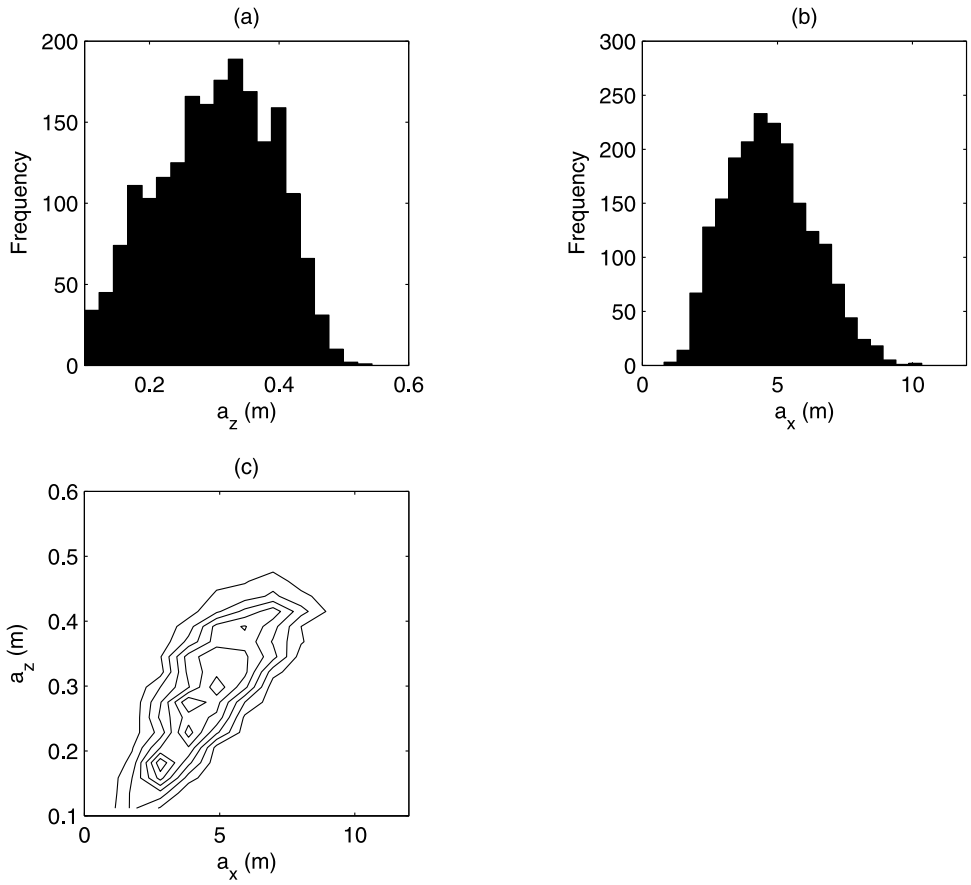

Figure 11. Results of MCMC analysis of the field 200 MHz GPR image. (a and b) Marginal posterior histograms for a_z and a_x, respectively, obtained using uniform prior ranges of a_z = [0.1, 0.8] m and a_x = [0.2, 20.0] m. (c) Corresponding contoured joint posterior histogram.

of the vertical correlation length of water content. The results in Figures 6, 7, 10, and 11 suggest that because of a trade-off between a_z and a_x, at least in some cases where knowledge of a_z is unavailable, it may only be possible to recover the aspect ratio of the subsurface variability or a very broad and possibly biased distribution for a_x. Nevertheless, as stated above, such information is useful in its own right.

[38] **Acknowledgments.** We would like to thank Marie Scholer for helpful discussions regarding horizontal resolution and the successful application of a Gaussian horizontal resolution filter. We also thank Associate Editor Lee Slater, John Bradford, and three anonymous reviewers for comments that helped to significantly improve this manuscript.

References

Annan, A. P. (2005), GPR methods for hydrogeological studies, in *Hydrogeophysics*, edited by Y. Rubin and S. S. Hubbard, pp. 185–213, Springer, New York.

Archie, G. E. (1942), The electrical resistivity log as an aid in determining some reservoir characteristics, *Trans. Am. Inst. Min. Metall. Pet. Eng.*, *146*, 54–62.

Bean, C. J., D. Marsan, and F. Martini (1999), Statistical measures of crustal heterogeneity from reflection seismic data: The role of seismic bandwidth, *Geophys. Res. Lett.*, *26*, 3241–3244, doi:10.1029/1999GL005400.

Belina, F. A., J. R. Ernst, and K. Holliger (2009), Inversion of crosshole seismic data in heterogeneous environments: Comparison of waveform and ray-based approaches, *J. Appl. Geophys.*, *68*, 85–94, doi:10.1016/j.jappgeo.2008.10.012.

Beres, M., P. Huggenberger, A. G. Green, and H. Horstmeyer (1999), Using two- and three-dimensional georadar methods to characterize glaciofluvial architecture, *Sediment. Geol.*, *129*, 1–24, doi:10.1016/S0037-0738(99)00053-6.

Berkhout, A. (1984), *Seismic Resolution: A Quantitative Analysis of Resolving Power of Acoustical Echo Techniques*, 228 pp., Geophys. Press, London.

Carcione, J. M., and F. Cavallini (1995), On the acoustic-electromagnetic analogy, *Wave Motion*, *21*, 149–162, doi:10.1016/0165-2125(94)00047-9.

Carpentier, S., and K. Roy-Chowdhury (2007), Underestimation of scale lengths in stochastic fields and their seismic response: A quantification exercise, *Geophys. J. Int.*, *169*, 547–562, doi:10.1111/j.1365-246X.2007.03333.x.

Dafflon, B., J. Tronicke, and K. Holliger (2006), Inferring the lateral subsurface correlation structure from georadar data: Methodological background and experimental evidence, in *Geostatistics for Environmental Applications*, edited by P. Renard, H. Demougeot-Renard, and R. Froidevaux, pp. 467–478, Springer, New York.

Gelhar, L. W. (1993), *Stochastic Subsurface Hydrology*, 480 pp., Prentice-Hall, Englewood Cliffs, N. J.

Gibson, B. (1991), Analysis of lateral coherency in wide-angle seismic images of heterogeneous targets, *J. Geophys. Res.*, *96*, 10,261–10,273, doi:10.1029/91JB00340.

Goff, J. A., K. Holliger, and A. Levander (1994), Modal fields: A new method for characterization of random seismic velocity heterogeneity, *Geophys. Res. Lett.*, *21*, 493–496, doi:10.1029/94GL00311.

Hillel, D. (2007), *Soil in the Environment: Crucible of Terrestrial Life*, 307 pp., Elsevier, Burlington, Mass.

Holliger, K., R. Carbonell, A. Levander, and R. Hobbs (1994), Some attributes of wavefields scattered from Ivrea-type lower crust, *Tectonophysics*, *232*, 267–279, doi:10.1016/0040-1951(94)90089-2.

Huggenberger, P. (1993), Radar facies: Recognition of facies patterns and heterogeneities within Pleistocene Rhine gravels, NE, Switzerland, in

Braided Rivers, edited by J. L. Best and C. S. Bristow, *Geol. Soc. Spec. Publ.*, *75*, 163–176.

Hurich, C. A. (1996), Statistical description of seismic reflection wavefields: A step towards quantitative interpretation of deep seismic reflection profiles, *Geophys. J. Int.*, *125*, 719–728, doi:10.1111/j.1365-246X.1996.tb06019.x.

Irving, J., and R. Knight (2003), Removal of wavelet dispersion from ground-penetrating radar data, *Geophysics*, *68*, 960–970, doi:10.1190/1.1581068.

Irving, J., and R. Knight (2006), Numerical modeling of ground-penetrating radar in 2-D using MATLAB, *Comput. Geosci.*, *32*, 1247–1258, doi:10.1016/j.cageo.2005.11.006.

Knight, R., P. Tercier, and J. Irving (2004), The effect of vertical measurement resolution on the correlation structure of a ground penetrating radar reflection image, *Geophys. Res. Lett.*, *31*, L21607, doi:10.1029/2004GL021112.

Knight, R., J. Irving, P. Tercier, G. Freeman, C. Murray, and M. Rockhold (2007), A comparison of the use of radar images and neutron probe data to determine the horizontal correlation length of water content, in *Subsurface Hydrology: Data Integration for Properties and Processes*, *Geophys. Monogr. Ser.*, vol. 171, edited by D. W. Hyndman, F. D. Day-Lewis, and K. Singha, pp. 31–44, AGU, Washington, D. C.

Line, C. E. R., R. W. Hobbs, and D. B. Snyder (1998), Estimates of upper-crustal heterogeneity in the Baltic Shield from seismic scattering and borehole logs, *Tectonophysics*, *286*, 171–183, doi:10.1016/S0040-1951(97)00263-1.

Mosegaard, K., and A. Tarantola (1995), Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res.*, *100*, 12,431–12,447, doi:10.1029/94JB03097.

Oldenborger, G. A., M. D. Knoll, and W. Barrash (2004), Effects of signal processing and antenna frequency on the geostatistical structure of ground-penetrating radar data, *J. Eng. Environ. Geophys.*, *9*, 201–212.

Poppeliers, C. (2007), Estimating vertical stochastic scale parameters from seismic reflection data: Deconvolution with non-white reflectivity, *Geophys. J. Int.*, *168*, 769–778, doi:10.1111/j.1365-246X.2006.03239.x.

Poppeliers, C., and A. Levander (2004), Estimation of vertical stochastic scale parameters in the Earth's crystalline crust from seismic reflection data, *Geophys. Res. Lett.*, *31*, L13607, doi:10.1029/2004GL019538.

Pullammanappallil, S., A. Levander, and S. P. Larkin (1997), Estimation of crustal stochastic parameters from seismic exploration data, *J. Geophys. Res.*, *102*, 15,269–15,286, doi:10.1029/97JB01144.

Rea, J., and R. Knight (1998), Geostatistical analysis of ground-penetrating radar data: A means of describing spatial variation in the subsurface, *Water Resour. Res.*, *34*, 329–339, doi:10.1029/97WR03070.

Roth, K., R. Schulin, H. Flühler, and W. Attinger (1990), Calibration of time domain reflectometry for water content measurement using a composite dielectric approach, *Water Resour. Res.*, *26*, 2267–2273, doi:10.1029/WR026i010p02267.

Rubin, Y., and S. Hubbard (2005), *Hydrogeophysics*, 523 pp., Springer, Dordrecht, Netherlands.

Stolt, R. H. (1978), Migration by Fourier transform, *Geophysics*, *43*, 23–48, doi:10.1190/1.1440826.

Stolt, R. H., and A. Benson (1986), *Seismic Migration: Theory and Practice*, 382 pp., Geophys. Press, London.

Topp, G. C., J. L. Davis, and A. P. Annan (1980), Electromagnetic determination of soil water content: Measurements in coaxial transmission lines, *Water Resour. Res.*, *16*, 574–582, doi:10.1029/WR016i003p00574.

K. Holliger and J. Irving, Institute of Geophysics, University of Lausanne, CH-1015 Lausanne, Switzerland. (klaus.holliger@unil.ch)

R. Knight, Geophysics Department, Stanford University, Stanford, CA 94305-2215, USA.