Stochastic inversion for soil hydraulic parameters in the presence of model error: an example involving ground-penetrating radar monitoring of infiltration

Corinna Köpke[1], James Irving[1†], and Delphine Roubinet[2]

[1]Institute of Earth Sciences, University of Lausanne, Switzerland
[2]Géosciences Montpellier, UMR 5243 CNRS, University of Montpellier, France
[†]Corresponding author: james.irving@unil.ch

Revised manuscript for *Journal of Hydrology*

November 11, 2018

Abstract

Proxy forward solvers are commonly used in Bayesian solutions to inverse problems in hydrology and geophysics in order to make sampling of the posterior distribution, for example using Markov-chain-Monte-Carlo (MCMC) methods, computationally tractable. However, use of these solvers introduces model error into the problem, which can lead to strongly biased and overconfident parameter estimates if left uncorrected. Focusing on the specific example of estimating unsaturated hydraulic parameters in a layered soil from time-lapse ground-penetrating radar data acquired during a synthetic infiltration experiment, we show how principal component analysis, conducted on a suite of stochastic model-error realizations, can for some problems be used to build a sparse orthogonal basis for the model error arising from known forward solver approximations and/or simplifications. Projection of the residual onto this basis during MCMC permits identification and removal of the model error before calculation of the likelihood. Our results indicate that, when combined with an informal likelihood metric based on the expected behaviour of the ℓ_2-norm of the residual, this methodology can yield posterior parameter estimates exhibiting a marked reduction in bias and overconfidence when compared to those obtained with no model-error correction, at reasonable computational cost.

1 Introduction

Stochastic parameter estimation and inversion have become increasingly popular in hydrology and geophysics over the past decade. In particular, it is now computationally feasible and common to solve many inverse problems in these domains in a Bayesian manner, whereby prior knowledge about the subsurface parameters of interest is combined with measured data to yield a posterior probability distribution. The latter is typically sampled using Markov-chain-Monte-Carlo (MCMC) methods (*Linde et al.*, 2017). Notable advantages of the Bayesian-MCMC approach are that (i) it is highly flexible and can incorporate any information that can be expressed as a probability density into the inverse problem; (ii) it provides a natural framework for data integration; and (iii) it has the potential to provide more accurate parameter uncertainty estimates than traditional methods based on linearization. This does, however, come at the cost of being highly computationally expensive. Indeed, many thousands if not millions of MCMC iterations, each requiring a numerical solution of the forward problem, are typically required to obtain a sufficient number of posterior samples for use in subsequent probabilistic forecasting and risk analysis (e.g., *Ruggeri et al.*, 2015).

A critical component of framing an inverse problem in a Bayesian context is proper characterization of the expected statistical nature of the residual. This is, in order to formulate the likelihood, we must have detailed knowledge about the statistical distribution of the difference between the measured data and those calculated through the numerical solution of the forward problem on the "true" set of subsurface model parameters. In arguably most cases, the residual is attributed solely to data-measurement errors and described as multi-Gaussian, usually with independent and identically distributed elements (e.g., *Bodin and Sambridge*, 2009; *Gallagher et al.*, 2009; *Irving and Singha*, 2010; *Linde and Vrugt*, 2012; *Scholer et al.*, 2013; *Vrugt et al.*, 2008). This is despite the fact that, in order to improve the computational tractability of the Bayesian-MCMC approach, approximate versions of the forward solver, for example using coarsened discretizations and/or simplifications of the underlying physics, are typically employed (e.g., *Christen and Fox*, 2005; *Cui et al.*, 2011; *Efendiev et al.*, 2008; *Hinnell et al.*, 2010; *Ray et al.*, 2015; *Scholer et al.*, 2013). The use of such computationally efficient "proxy solvers" leads to model error, which if left uncorrected has the potential to overwhelm the effects of data measurement uncertainties and lead to strongly biased and overconfident posterior distributions (*Brynjarsdóttir and O'Hagan*, 2014).

In recent years, a number of studies in the hydrological and geophysical literature have attempted to address the issue of model error in Bayesian inversions, with the aim of making more effective use of proxy solvers when dealing with computationally expensive forward operators. In general, the approaches that have been presented can be divided into two categories. In the first category, researchers have focused on the overall or "global" statistical characterization of the model error, with the goal of using this information to develop more appropriate parametric likelihood functions that better reflect the true nature of the residual. This has generally been accomplished through the analysis of stochastic model-error realizations, which are generated by running the full and approximate forward solvers for randomly drawn sets of model parameters. Typically, multi-Gaussian statistics for the model error are assumed, meaning that means and covariances estimated from the realizations can be incorporated into a Gaussian likelihood (e.g., *Arridge et al.*, 2006; *Hansen et al.*, 2014; *Kaipio and Somersalo*, 2007; *Lehikoinen et al.*, 2010; *Stephen*, 2007), but other parametric likelihood functions have also been considered (e.g., *Del Giudice et al.*, 2013; *Schoups and Vrugt*, 2010; *Smith et al.*, 2010; *Smith et al.*, 2015). Accounting for model error in this manner has been shown to lead to broadened posterior distributions and a reduction in parameter bias. One key issue, however, concerns the validity of the assumption that the errors can be adequately described by a given parametric distribution. In many inverse problems in geophysics and hydrology, for example, model errors will exhibit complex statistics and correlations that arise from the typically high dimension of the data and/or model-parameter spaces in these problems, combined with the non-linearity of the forward operators involved. Indeed, there has been much increased interest in "likelihood-free" inference methods such as approximate Bayesian computation (ABC) (e.g., *Vrugt and Sadegh*, 2013) and generalized likelihood uncertainty estimation (GLUE) (e.g., *Beven and Binley*, 1992), to a large extent because of this issue.

In the second category of developed approaches for addressing model error, researchers have focused on building a parameter-dependent or "local" error model in order to describe the discrepancy between the full and approximate forward solvers. As with the approaches mentioned above, this is constructed based on computed realizations of the model error for different parameter sets. However, in this case the results are used to effectively correct the output of the approximate solver rather than to develop a more appropriate Bayesian likelihood function. Construction of the error model can be done in a number of different ways. This includes simple nearest-neighbour or linear interpolation between model-error realizations (e.g., *Cui et al.*, 2011; *O'Sullivan and Christie*, 2006), representing the discrepancy as a Gaussian process conditioned to the points in the parameter space where the model error is known (e.g., *Kennedy and O'Hagan*, 2001; *Xu and Valocchi*, 2015), or using statistical regression approaches (e.g., *Doherty and Christensen*, 2011; *Josset et al.*, 2015). In all of this work, the implicit assumption is that the full and approximate model-response surfaces are regular enough such that the model error for a set of parameter values where it is unknown can be effectively predicted through some kind of interpolation between the existing realizations. While this may be the case for some inverse problems, difficulties can arise in the presence of strongly non-linear forward solvers and/or large numbers of model parameters. That is, it may not be possible to sufficiently sample the model-parameter space with model-error realizations such that interpolation between these realizations will provide a reliable model-error estimate at some new location.

Recently, *Köpke et al.* (2018) presented a new approach to account for the model error arising from the use of proxy forward solvers in Bayesian-MCMC inversions, whereby information about the error is gathered during the inversion procedure through occasional runs of the approximate and full solvers together, the results of which are stored in a dictionary. In contrast to the existing methods mentioned above, the approach of *Köpke et al.* (2018) focuses on the projection-based identification of the model-error component of the residual through the construction of a local, parameter-dependent, orthogonal model-error basis, rather than on attempting to fit the overall model-error statistics to a prescribed statistical distribution or develop an interpolation-based error model. The model error estimated by projecting onto the basis is then subtracted from the residual before computing the likelihood of a proposed set of model parameters in MCMC. Application of the approach of *Köpke et al.* (2018) to a high-dimensional spatially distributed tomographic example was found to yield parameter estimates exhibiting a notable reduction in bias compared to those obtained when the model error was ignored. The presented method does, however, still require occasional runs of the full forward solver along the Markov chain as MCMC iterations progress, which can be computationally costly depending on the problem at hand.

In this paper, we build on the work of *Köpke et al.* (2018) and show that, for some inverse problems, it may be possible to derive a suitable global basis for the model error over the entire parameter space through the application of principal component analysis to a large number of stochastic model-error realizations. These realizations can be conveniently computed in parallel prior to MCMC, and they must be properly organized in the data space before analysis to maximize similarity in their spatial characteristics. We begin in Section 2 with an overview of Bayesian-MCMC inversion, followed by a detailed description of our developed approach. This is followed in Section 3 with application to synthetic data corresponding to a vadose-zone inverse problem that has been the subject of much investigation in previous work, which is the estimation of unsaturated soil hydraulic parameters from time-lapse zero-offset-profile (ZOP) ground-penetrating radar (GPR) data acquired during infiltration. We compare the results obtained with our methodology to those obtained when no model-error correction is applied, and conclude in Section 4 with an overall assessment of the method with regard to its advantages and limitations.

2 Methodological background

2.1 Bayesian-MCMC inversion

Consider to begin the forward problem linking a set of M subsurface model parameters of interest $\mathbf{m}_{true} \in \mathbb{R}^M$ to a set of N measured or observed data $\mathbf{d}_{obs} \in \mathbb{R}^N$:

$$\mathbf{d}_{obs} = F(\mathbf{m}_{true}) + \mathbf{e}_d, \tag{1}$$

where forward operator $F : \mathbb{R}^M \to \mathbb{R}^N$ contains the physics and geometry of the measurements and $\mathbf{e}_d$ is vector of data measurement errors. The goal of the corresponding inverse problem is to estimate $\mathbf{m}_{true}$ given $\mathbf{d}_{obs}$, which requires knowledge of F and in most cases some prior information about the model parameters. From a probabilistic point of view, this can be formulated using Bayes' theorem, whereby an initial prior state of information for the model parameters $\rho(\mathbf{m})$ is updated to a more refined posterior state of knowledge $\sigma(\mathbf{m})$ based on the available data (e.g., *Tarantola*, 2005). That is,

$$\sigma(\mathbf{m}) = k\, L(\mathbf{m})\, \rho(\mathbf{m}), \tag{2}$$

where normalization constant k ensures that the posterior probability distribution integrates to unity, and likelihood $L(\mathbf{m})$ expresses the conditional probability of model parameter set $\mathbf{m}$ given the observed data $\mathbf{d}_{obs}$. Assuming that (i) the underlying physics are completely known and considered in the inverse problem; and (ii) the data measurement errors are independent and identically normally distributed having mean zero and standard deviation s_d, $L(\mathbf{m})$ takes on the simple multi-Gaussian form

$$L(\mathbf{m}) = \frac{1}{(2\pi s_d^2)^{N/2}} \exp\left[-\frac{||\mathbf{r}(\mathbf{m})||^2}{2s_d^2}\right], \tag{3}$$

where $||\cdot||$ denotes the ℓ^2-norm and $\mathbf{r}(\mathbf{m})$ is the residual or difference between the observed data and those predicted for some model parameter set $\mathbf{m}$ using F. The latter quantity is given by

$$\begin{aligned} \mathbf{r}(\mathbf{m}) &= \mathbf{d}_{pred} - \mathbf{d}_{obs} \\ &= \underbrace{F(\mathbf{m}) - [F(\mathbf{m}_{true})}_{\substack{\text{parameter-error} \\ \text{component}}} + \mathbf{e}_d], \end{aligned} \tag{4}$$

where $\mathbf{d}_{pred}$ denotes the predicted data. We see in equation (3) that $L(\mathbf{m})$ will be maximized when the ℓ^2-norm of $\mathbf{r}(\mathbf{m})$ is minimized, which corresponds to the case where $\mathbf{m} = \mathbf{m}_{true}$ and the parameter-error component as defined in equation (4) is zero. The spread of the likelihood distribution about the maximum value is controlled by the data measurement error standard deviation s_d along with the number of data N, with larger errors and lesser amounts of data yielding broader likelihoods.

Equations (2) through (4) together provide a means of calculating the posterior probability of a set of model parameters $\mathbf{m}$. This is commonly used within MCMC sampling algorithms to quantify posterior uncertainty, thereby solving the inverse problem, since it is not generally possible to perform the multi-dimensional integrations needed to obtain the statistical moments of $\sigma(\mathbf{m})$. In this regard, a basic Metropolis-Hastings algorithm (*Metropolis et al.*, 1953; *Hastings*, 1970) that is guaranteed (after burn-in) to generate a Markov chain of samples $\{\mathbf{m}_1, ..., \mathbf{m}_k\}$ from the Bayesian posterior distribution proceeds as follows:

1. Draw the first model in the Markov chain $\mathbf{m}_1$ from the Bayesian prior distribution $\rho(\mathbf{m})$. Set $i = 1$.

2. Draw a perturbed model-parameter set $\mathbf{m}'$ from the proposal distribution $Q(\mathbf{m}'|\mathbf{m}_i)$, whose width around $\mathbf{m}_i$ is chosen so as to provide a balance between efficiently moving through the parameter space and generating proposals that have a reasonable probability of being accepted.

3. Calculate the probability of accepting $\mathbf{m}'$ as the next model in the Markov chain using

$$P_{acc} = \min\left[1, \frac{\sigma(\mathbf{m}')\,Q(\mathbf{m}_i|\mathbf{m}')}{\sigma(\mathbf{m}_i)\,Q(\mathbf{m}'|\mathbf{m}_i)}\right]. \tag{5}$$

4. Draw a random number $x \in U(0, 1)$. If $x \leq P_{acc}$, then set $\mathbf{m}_{i+1} = \mathbf{m}'$. Otherwise set $\mathbf{m}_{i+1} = \mathbf{m}_i$.

5. Set $i = i + 1$ and go to Step 2.

Note that, in the case where the proposal distribution is symmetric (i.e., $Q(\mathbf{m}'|\mathbf{m}_i) = Q(\mathbf{m}_i|\mathbf{m}')$), the above algorithm reduces to the original MCMC sampler of *Metropolis et al.* (1953) where the acceptance probability is given by $P_{acc} = \min[1, \sigma(\mathbf{m}')/\sigma(\mathbf{m}_i)]$. We consider the latter sampler for the example inversions presented in Section 3.

2.2 Accounting for model error

Likelihood equation (3) is perfectly theoretically valid for the case where the only contribution to the difference between the observed and predicted data, when considering the correct set of model parameters $\mathbf{m}_{true}$, is a set of Gaussian data-measurement errors having standard deviation s_d. However, as mentioned previously, approximate forward solvers are typically used in hydrological and geophysical problems to improve the computational efficiency of the Bayesian-MCMC procedure, meaning that the residual more realistically takes the form

$$
\begin{aligned}
\mathbf{r}(\mathbf{m}) &= \hat{F}(\mathbf{m}) - [F(\mathbf{m}_{true}) + \mathbf{e}_d] \\
&= \underbrace{\hat{F}(\mathbf{m}) - F(\mathbf{m})}_{\substack{\text{model-error}\\ \text{component}}} + \underbrace{F(\mathbf{m}) - [F(\mathbf{m}_{true})}_{\substack{\text{parameter-error}\\ \text{component}}} + \mathbf{e}_d],
\end{aligned}
\tag{6}
$$

where $\hat{F}$ is the approximate forward operator. The presence of an additional model-error term in equation (6) as compared with equation (4), which is commonly of large magnitude, strongly correlated, and/or highly non-Gaussian (*Kaipio and Somersalo*, 2007; *Schoups and Vrugt*, 2010; *Smith et al.*, 2010), makes use of likelihood expression (3) inappropriate. In particular, it means that (i) the residual will not necessarily be minimized when $\mathbf{m} = \mathbf{m}_{true}$, implying posterior parameter bias; and (ii) feasible model parameter sets may have an extremely low likelihood when considering realistic levels of data error. Although simple inflation of s_d can be used to broaden the Gaussian likelihood and reduce the latter issue, it cannot address the former and be viewed as an effective solution for reliable posterior uncertainty quantification.

In order to address the model-error issue, we build on the work of *Köpke et al.* (2018) in this paper and focus on learning about the nature of the model error through stochastic simulation such that it may be identified and removed from the residual during MCMC. The overall idea is that, for some problems, a representative set of stochastic model-error realizations, computed prior to MCMC for random model parameter sets using the full and approximate forward solvers, can be used to construct an orthonormal basis for the model error. Projection of the residual onto this basis in each MCMC iteration is used to isolate the model-error component, which is the subtracted from $\mathbf{r}(\mathbf{m})$ before calculating the likelihood. Note that, whereas *Köpke et al.* (2018), used a dictionary-based K-nearest-neighbour (KNN) approach to construct a different local model-error basis for each proposed set of model parameters in MCMC, with runs of the full forward solver being required periodically along the entire Markov chain, we focus here on the development of a global basis (i.e., over the entire model parameter space) before posterior sampling begins. Although not appropriate for all problems, this methodology has the advantage that all expensive forward solver computations can be run in a simple parallel manner outside of the MCMC iterations. The corresponding set of model-error realizations can also be directly reused in any subsequent inversions. Our approach proceeds as follows:

1. Generate k random sets of model parameters $\{\mathbf{m}_1, ..., \mathbf{m}_k\}$ from the Bayesian prior distribution $\rho(\mathbf{m})$.

2. Compute the corresponding set of stochastic model-error realizations $\{\mathbf{E}_1, ..., \mathbf{E}_k\}$, where $\mathbf{E}_i = \hat{F}(\mathbf{m}_i) - F(\mathbf{m}_i)$.

3. If necessary, organize the information in each realization to improve coherency for subsequent analysis (see Section 3.3).

4. Perform principal component analysis (PCA) on the model-error realizations $\{\mathbf{E}_1, ..., \mathbf{E}_k\}$ in order to obtain a sparse orthonormal basis $\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_b]$ for the model error. The number of basis vectors b should be chosen to be the minimum required to capture a high percentage of the variance of the realizations, typically around 98%. In this way, the basis will be able to capture the model-error behaviour, but will have minimal ability to represent contributions to the residual that do not resemble model error such as data measurement uncertainties.

5. For each set of model parameters $\mathbf{m}'$ tested within MCMC, calculate the best approximation of the residual $\mathbf{r}(\mathbf{m}') = \hat{F}(\mathbf{m}') - \mathbf{d}_{obs}$ using the model-error basis, obtained in a least-squares sense using $\mathbf{B}\,\mathbf{B}^T\mathbf{r}(\mathbf{m}')$, and remove this result from the residual. This yields the remainder
$$\mathbf{R}(\mathbf{m}') = \mathbf{r}(\mathbf{m}') - \mathbf{B}\,\mathbf{B}^T\mathbf{r}(\mathbf{m}') \tag{7}$$

6. Use $\mathbf{R}(\mathbf{m}')$ to determine $L(\mathbf{m}')$ within MCMC (see Section 2.3).

It is important to note that the success of the modified MCMC approach described above, in terms of providing refined and unbiased posterior parameter estimates using an approximate forward operator, hinges on our ability to effectively separate the model-error component of equation (6) from (i) data-measurement errors, and (ii) parameter-related errors. The implicit assumption in our work is that these two other sources of error lie orthogonal to the elements of $\mathbf{B}$, such that projection of the residual onto the basis will preserve only the model-error component. With regard to (i), we have found that this is a reasonable expectation as the limited number of model-error basis vectors, which tend to possess a high degree of spatial correlation, are generally not capable of representing random data-measurement errors through a linear combination (*Köpke et al.*, 2018). With respect to (ii), although there is no guarantee that the basis cannot represent at least part of the parameter-error term through a linear combination, our experience has been that the model and parameter-related errors typically possess significantly different statistical characteristics meaning that the latter tend to be quite effectively attenuated through projection onto $\mathbf{B}$. If this is not the case and a particular incorrect model-parameter set tested within MCMC happens to yield a parameter-error component that resembles what was observed in the model-error realizations, this error will be removed and the parameter set will have a reasonably high chance of being accepted (*Köpke et al.*, 2018). This latter point is discussed in further detail in Section 3.5.

2.3 Likelihood evaluation

Ideally, the remainder $\mathbf{R}(\mathbf{m})$ in equation (7) should represent the residual in (6) with the model-error component perfectly removed, meaning that it should be identical to equation (4) and thus suitable for inclusion into Gaussian expression (3) to evaluate the likelihood. In reality, however, small but correlated and non-Gaussian errors in the approximation of the model-error component of the residual, related to our inability to perfectly separate model error from data measurement and parameter uncertainty using the sparse basis $\mathbf{B}$, mean that $\mathbf{R}(\mathbf{m}_{true})$ will deviate somewhat from multi-Gaussian and use of equation (3) can be problematic. Indeed, the strong ranking of models provided by a Gaussian likelihood function is well understood to pose difficulties for Bayesian inference when the underlying statistical assumptions regarding the residual are violated, in the sense that sets of model parameters that are perfectly acceptable may be mapped to extremely low likelihoods (e.g., *Beven and Binley*, 1992; *O'Sullivan and Christie*, 2006). To address this issue, we evaluate the likelihood in this work using a statistically informal but more practical metric based on the expected univariate distribution of the ℓ_2-norm of the remainder in equation (7), as opposed to the expected multivariate distribution of the vector $\mathbf{R}(\mathbf{m})$. Specifically, assuming for lack of better information that the elements of $\mathbf{R}(\mathbf{m})$ are uncorrelated and normally distributed having mean zero and standard deviation s_R, it can be shown that the ℓ_2-norm $||\mathbf{R}(\mathbf{m})||$ will follow a scaled chi distribution (*Forbes et al.*, 2010), leading to the following equation:

$$L(\mathbf{m}) = \frac{2^{1-\frac{N}{2}}}{\Gamma\left(\frac{N}{2}, 0\right)} \; s_R^{-N} \, ||\mathbf{R}(\mathbf{m})||^{N-1} \exp\left[-\frac{||\mathbf{R}(\mathbf{m})||^2}{2s_R^2}\right], \tag{8}$$

where $\Gamma(\cdot, \cdot)$ is the incomplete gamma function.

For the typical case where s_R is unknown and can only be bounded between lower and upper values s_{R_1} and s_{R_2}, respectively, equation (8) can be integrated over s_R yielding

$$L(\mathbf{m}) \propto \Gamma\left(\frac{N-1}{2}, \frac{||\mathbf{R}(\mathbf{m})||^2}{2s_R^2}\right)\Bigg|_{s_{R_1}}^{s_{R_2}}. \tag{9}$$

Figure 1 shows $L(\mathbf{m})$ calculated using equation (9) as a function of $||\mathbf{R}(\mathbf{m})||$ for $N = 1000$ with $s_{R_1} = 0.1$ and $s_{R_2} = 0.2$. We see that there is a range for the ℓ_2-norm of $\mathbf{R}(\mathbf{m})$ over which the likelihood is approximately constant, outside of which it falls off rapidly to near-zero values. In other words, sets of model parameters for which $||\mathbf{R}(\mathbf{m})||$ is consistent with the ℓ_2-norm of a vector of normally distributed values with $s_R \in [0.1, 0.2]$ are considered to be approximately equally likely, whereas those that do not fit this criterion are given almost zero likelihood. Equation (9) is much less sensitive to small changes in $\mathbf{R}(\mathbf{m})$ compared with equation (3), and represents a significantly more relaxed and inherently conservative constraint than that provided by a formal Gaussian likelihood function. Indeed, the use of such informal likelihood measures within stochastic inverse methods has gained widespread acceptance in hydrology and other domains in recent years (e.g., *Beven and Freer*, 2001; *Beven and Binley*, 2014; *Blasone et al.*, 2008; *Nott et al.*, 2012; *Sadegh and Vrugt*, 2013a,b; *Wilkinson*, 2013), as researchers have realized the shortcomings of placing too much importance on the detailed statistical properties of the residual for many real-world problems. Equation (9) can in fact be considered as a slight variation of the generalized likelihood uncertainty estimation (GLUE) approach, originally proposed by *Beven and Binley* (1992), where the distinction between "behavioural" and "non-behavioural" models is quantified by using the ℓ_2-norm of the data misfit. Models whose remainder norm falls within the bounds indicated in Figure 1 will have a high chance of being accepted in MCMC, whereas those falling significantly outside these bounds will tend to be rejected.

[Figure 1 about here.]

3 Example: GPR monitoring of infiltration

We now apply the model-error methodology presented in Section 2 to a synthetic example involving GPR monitoring of an infiltration experiment. Zero-offset-profile (ZOP) GPR data, acquired between two boreholes over the course of the experiment, provide estimates of horizontally averaged soil water content as a function of depth and time (e.g., *Annan*, 2006). Together with a numerical model for the infiltration process, the latter results are then used to estimate unsaturated soil hydraulic properties in a layered subsurface. This particular problem has been the focus of much previous research in the field of hydrogeophysics (e.g., *Binley and Beven*, 2003; *Cassiani and Binley*, 2005; *Looms et al.*, 2008; *Rucker and Ferré*, 2004; *Rucker*, 2011), and was most recently investigated within the context of Bayesian-MCMC inversion by *Scholer et al.* (2011, 2012, 2013). Here, we consider the model errors arising from a simplifying assumption common to all past work, which is that water movement occurs in a purely vertical direction through the subsurface.

3.1 Governing equations and model simplifications

The general movement of water through unsaturated soils is described by Richards' equation (*Richards*, 1931), given by

$$\frac{\partial \theta(h)}{\partial t} = \nabla \cdot [K(h) \nabla h] + \frac{\partial K(h)}{\partial z}, \tag{10}$$

where θ is the volumetric water content, K is the unsaturated hydraulic conductivity, h is pressure head, t is time, and z is elevation. The relationships $\theta(h)$ and $K(h)$ for different soils are commonly described using the van Genuchten - Mualem (VGM) model (*Mualem*, 1976; *van Genuchten*, 1980). With this model, the soil water retention, expressed in terms of effective saturation S_e, is given by

$$S_e(h) = \frac{\theta(h) - \theta_r}{\theta_s - \theta_r} = \begin{cases} (1 + |\alpha h|^n)^{-m} & \text{, for } h \leq 0 \\ 1 & \text{, for } h > 0 \end{cases} \tag{11}$$

where θ_r and θ_s are the residual and saturated water contents, respectively, and α, m, and n are empirical shape factors with $m = 1 - 1/n$. The unsaturated hydraulic conductivity is described by

$$K(h) = K_s S_e(h)^{1/2} \left[1 - (1 - S_e(h)^{1/m})^m\right]^2, \tag{12}$$

where K_s is the hydraulic conductivity value at full saturation. A total of five parameters (K_s, θ_r, α, n, and θ_s) therefore describe a soil's hydraulic properties using the VGM model.

Equations (10) through (12) provide a link between a set of subsurface VGM parameters and the corresponding spatiotemporal distribution of water content in response to infiltration. That is, knowing the distribution of soil VGM parameters along with the boundary and initial conditions of the infiltration experiment, we can calculate the evolution of water content in the subsurface. This forward link provides the basis for inverting for the soil hydraulic properties given a set of dynamic GPR-derived water-content measurements. However, in the context of stochastic inversion, repeated solution of a fully 3D unsaturated flow model based on (10) can be extremely computationally demanding. As a result, previous work in this domain has typically assumed that flow occurs only in the vertical direction (e.g., *Binley and Beven*, 2003; *Cassiani and Binley*, 2005; *Looms et al.*, 2008; *Scholer et al.*, 2012), such that the following 1D version of Richards' equation can be utilized in the inversion procedure:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left[K(h)\left(\frac{\partial h}{\partial z} + 1\right)\right]. \tag{13}$$

The vertical flow assumption may hold in layered subsurface environments under natural loading conditions (e.g., *Binley and Beven*, 2003), but it will be clearly violated during infiltration experiments where the area over which loading occurs is spatially restricted and loading rates are significantly higher. This will be particularly the case where there exist large contrasts in subsurface hydraulic properties (e.g., *Looms et al.*, 2008; *Rucker*, 2011). As a result, the 1D flow assumption represents a significant source of model error. Although such errors and their potential for posterior parameter bias have been acknowledged in previous research efforts (*Scholer et al.*, 2013), they have never before been examined and accounted for in the inversion procedure.

3.2 Infiltration experiment and data

Figure 2 shows the overall setup considered for our synthetic infiltration experiment. Infiltration at a rate of 2 cm/h is applied to a circular region on the Earth's surface having a diameter of 3 m. The infiltration is carried out for a period of 11.6 d, during which GPR-derived estimates of horizontally averaged water content are considered to be available every 2.8 h. The water-content measurements are considered between boreholes 2-m apart and 8-m deep, with a depth sampling interval of 0.1 m. The subsurface consists of two layers whose VGM parameters are given in Table 1, which texturally describes a sandy soil underlain by a less permeable silt loam. The boundary between the layers is located at 3-m depth.

[Figure 2 about here.]

[Table 1 about here.]

To determine the spatiotemporal distribution of water content corresponding to the experimental setup described above, we used the code VS2D (*Lappala et al.*, 1987) to solve the general 3D Richards' equation (10) under the assumption of rotational symmetry about the vertical axis, meaning that the model domain was parameterized in terms of radius (r) and depth (z), with $r = 0$ corresponding to the center of the infiltration region. A specified-flux boundary condition was imposed at the Earth's surface ($z = 0$ m) with no-flow conditions assumed outside of the infiltration region ($r > 1.5$ m). No-flow conditions were also assumed along the outside of the model domain, the latter of which was set at $r = 4$ m. At the bottom of the domain ($z = 10$ m), a fixed-pressure-head value of $h = -0.5$ m was specified in order to simulate the presence of the water table at 10.5-m depth. The initial distribution of soil water content prior to running the infiltration experiment was obtained using a 1D steady-state infiltration code based on the work of *Rockhold et al.* (1997) assuming a constant infiltration rate of 0.036 cm/h.

Figure 3 shows snapshots of the modeled subsurface water-content distribution over the course of the infiltration experiment for times $t = 0, 1, 2, 5, 7,$ and 11 d. We see that, once the experiment begins, the infiltration front moves approximately vertically through the sandy soil layer, making its way to the boundary with the silt loam in just under 2 d. From this point onwards, although the front continues to move downwards, a strong lateral component to the flow is observed because of the lesser permeability of the lower layer. Indeed, as the water is not able to infiltrate as quickly into the silt loam, it begins to build up at the interface between the two soils and spread horizontally. Such behaviour cannot be captured using a 1D flow model based on equation (13), which is discussed in further detail below.

[Figure 3 about here.]

We next simulated the GPR-derived water-content measurements acquired during the infiltration experiment, which again represent the data to be inverted for the VGM parameters in each soil layer. To this end, every 2.8 h, the horizontal average of the water-content field between the boreholes was calculated from the VS2D results using a depth discretization interval of 0.1 m. This yielded 81 measurements in depth across 101 GPR acquisition times, to which zero-mean Gaussian random noise with a standard deviation of 0.01 (roughly 5%) was added to simulate the effects of measurement error. It is important to note that, for the sake of simplicity in this example, we did not explicitly model the propagation of GPR energy between the transmitter and receiver antennas in the two boreholes based on the VS2D results, but rather assumed that the ZOP GPR experiment provided a measure of the horizontal average of soil water content as a function of depth. Although, in doing this, we admittedly neglect several aspects of the physics that would be encountered in a field setting such as critical refractions of GPR energy and frequency-dependent resolution limitations (e.g., *Rossi et al.*, 2012; *Rucker and Ferré*, 2004), these aspects were not considered essential for this numerical study into the effects of model error arising from the 1D flow assumption.

Figure 4a shows the simulated GPR-derived water-content data, organized into a matrix with depth on the vertical axis and measurement time on the horizontal axis. To gain insight into the importance of model error for this example, Figure 4b shows the corresponding water content calculated as a function of depth and time assuming purely vertical flow, such that the 1D Richards' equation (13) could be applied. The results obtained using the 3D and 1D models for the same set of VGM parameters and boundary conditions are clearly and significantly different, most notably with respect to: (i) the speed at which the infiltration front travels through the lower layer, which is greater using the 1D model; and (ii) the evolution of water content in the upper layer after the infiltration front reaches the soil interface, in that the upper layer is seen to "fill up" to full saturation in the 1D case rather than pool and spread laterally. Figure 4c shows the difference between and 1D and 3D simulation results, equal to the sum of the model error and Gaussian measurement uncertainties. Here we see that there are parts of the data space where the magnitude of the error is almost 50%, and that the model error exhibits a high degree of correlation. All of this means that using a 1D flow model to stochastically invert the data in Figure 4a, without accounting for model error, will result in a strong bias in the estimated VGM parameters and unreliable posterior statistics (see Section 3.4). Finally, Figure 4d shows the error image from Figure 4c with the results reorganized such that they are plotted relative to the arrival time of the infiltration front as a function of depth observed in the data (Figure 4a). The importance of this data arrangement step is explained in the following section.

[Figure 4 about here.]

3.3 Model-error realizations and analysis

The first step in our approach to dealing with a known source model error in this paper involves generation of a set of stochastic model-error realizations corresponding to parameter sets randomly drawn from the Bayesian prior distribution. Again, this is done so that we can learn about the overall characteristics of the model error, with the goal of using this information to identify the model-error component of the residual during MCMC. Table 2 shows the lower and upper bounds of the uniform prior distributions that were assumed for the different VGM parameters in our synthetic study. Note that these distributions are rather broad and encompass a wide range of soil types (e.g., *Carsel and Parrish*, 1988), and that the same priors were assumed for each soil layer. In this way, relatively little information about the hydraulic properties is provided to the inversion procedure and we rely strongly upon the data to resolve them. Also note that the prior bounds for K_s are specified in terms of its logarithm, which is consistent with previous work and reflects the wide range of natural variability of this parameter (e.g., *Scholer et al.*, 2012, 2013).

[Table 2 about here.]

Each model-error realization was generated by: (i) drawing a random set of VGM parameters for each soil layer from the prior distributions in Table 2; (ii) computing the corresponding GPR-derived water-content data as a function of depth and time based on the general 3D Richards' equation (10); (iii) computing the GPR-derived water-content data under the assumption of purely vertical flow using the 1D Richards' equation (13); and (iv) calculating the difference between the 3D and 1D simulation results. It is important to reiterate that this part of our model-error approach is easily parallelized in the sense that different model-error realizations can be computed on different processors of a cluster, thereby greatly reducing the time needed to run the relatively large number of expensive 3D unsaturated flow simulations required. In this regard, runs of the 3D solver for our example took approximately 100 s on a standard desktop computer, whereas runs of the 1D solver were over 60 times faster at 1.5 s.

Figure 5 shows an example of 18 model-error realizations, each of which has been plotted relative to the arrival time of the infiltration front observed in the 3D simulation results, as was done for Figure 4d. This latter step, whereby the realizations are effectively "aligned" on the curve representing the 3D infiltration-front arrival in depth, is important for this problem because, without it, the realizations would be highly dissimilar in the data space and not amenable to any kind of global analysis. In contrast, after alignment, the model-error realizations are seen to take on a similar form which is described by: (i) a triangular region below 3-m depth that results from the difference in the speed of propagation of the infiltration front in the lower layer between the 3D and 1D simulations; and (ii) another triangular region above 3-m depth that results when the upper layer "fills up" in the 1D simulation for cases where the lower layer is less permeable. Although the widths and amplitudes of these triangular regions are significantly different across the various realizations in Figure 5, the images, due to their similarity in form, are generally well suited to PCA analysis with the aim of generating a compact orthonormal basis for the model error. At the same time, however, it is important to note that the strong variations between the realizations in Figure 5 in terms of width and amplitude mean that the model error is not well described using a simple parametric distribution, and thus not amenable to the global statistical approaches for model error mentioned earlier. Indeed, detailed analysis of the 6500 model-error realizations indicates that the model-error values are highly non-Gaussian-distributed with complex correlation patterns in the data space.

[Figure 5 about here.]

To construct the model-error basis, a total of 6500 realizations were analyzed using PCA, the results of which showed that only the first 50 principal components (out of 6561) were necessary to capture 98% of the variance of the input. Note that the number of principal components required to capture this percent of variance tends to increase with the number of model-error realizations considered, as smaller sets of realizations will generally exhibit a lesser range of variability that can be represented by a smaller basis (Figure 6). Our choice of 6500 realizations represents a point after which this trend stabilizes and the addition of further realizations does not require more principal components to capture 98% of the variance. Figure 7 shows the first 15, and last 3, vectors in the orthonormal model-error basis, ordered with respect to their decreasing contribution to the total variance and plotted as images in the data space. As expected, we see a gradual increase in the spatial frequency content of each vector as its index increases, with the first few vectors tending to capture the overall large-scale trends seen in the realizations in Figure 5 and the higher-order basis elements being necessary to resolve the finer details. Again, under the assumptions of orthogonality stated in Section 2.2, projection of the residual onto this basis during MCMC should adequately identify the model-error component, which can then be removed prior to computing the likelihood.

[Figure 6 about here.]

[Figure 7 about here.]

3.4 Stochastic inversion results

We now present the results of three different Bayesian-MCMC inversions to estimate the "true" VGM parameters in Table 1 from the GPR-derived water-content data in Figure 4a, all of which are based on use of a simplified 1D flow model. We begin by presenting the results of inverting using a "standard" Gaussian likelihood given by equation (3), where no correction for model error is considered and the standard deviation of the data errors s_d is artificially inflated in order to compensate for the additional error source. This is followed by inverting using the informal ℓ_2-norm-based likelihood measure given by equation (9), again with no correction for model error, such that the results obtained using this measure and using the Gaussian likelihood can be directly compared. Finally, we show the posterior results obtained for the case where the ℓ_2-norm-based likelihood is combined with the correction for model error described in Section 2.2. For each inversion, a uniform MCMC proposal distribution $Q(\mathbf{m}'|\mathbf{m}_i)$, centered on the current state of the Markov chain and whose width was chosen to provide a model acceptance rate of approximately 30% (*Gelman et al.*, 1996), was employed. A total of 800,000 MCMC iterations were run in each case, from which the first 10,000 samples were discarded as burn-in. These latter values were deemed appropriate based on visual inspection of each model parameter, along with its mean and variance, as a function of iteration (e.g., *Hassan et al.*, 2009).

3.4.1 Gaussian likelihood, no model-error correction

Figure 8 shows the marginal posterior histograms obtained for the VGM parameters in each soil layer for the case where the data in Figure 4a were inverted using a standard Gaussian likelihood function. The error standard deviation in equation (3) in this case was arbitrarily set to $s_d = 0.2$, which is 20 times the level of the random noise added to the data, in order to compensate for the model-error contribution to the residual and counteract the strong ranking of models provided by a Gaussian likelihood when the true residual statistics do not agree precisely with those that are assumed. Without such error inflation, the use of equation (3) would result in a highly peaked posterior distribution that could only be sampled with an extremely narrow proposal distribution and unreasonably large number of MCMC iterations. Indeed, *Brynjarsdóttir and O'Hagan* (2014) point out that, when model errors are present and not accounted for in Bayesian inference, the posterior tends to become narrowly focused around the wrong set of model parameters, and this only gets worse as more data are considered. Error inflation permits, at the very least, for the biased parameter set(s) to be identified at the expense of the posterior parameter uncertainties being arbitrary.

[Figure 8 about here.]

We see in Figure 8 that, because model error is present but has not been accounted for in the inversion procedure, the posterior VGM-parameter histograms are consistently focused on the wrong values. That is, there exists a set of incorrect parameter values whose predicted data, obtained using a 1D flow model, are a better match to the observed data than the true parameters in Table 1. The most significant bias in parameters occurs for the saturated hydraulic conductivity in both layers and the saturated water content of the upper layer, where the true values are seen to fall outside of the limits of the posterior distributions despite that fact that the error inflation imposed in this example is significant. As infiltration occurs significantly more rapidly in a 1D simulation than in 3D for the same set of model parameters (Figure 4), an inversion based on the 1D flow model will tend to select lower values for K_s in both layers in order to best match the observed, 3D-generated data. Further, the 1D flow model predicts a greater accumulation of water at the interface between the two layers, which can be reduced by selecting lower values for θ_s in the upper layer.

To gain insight into how such model-error-related biases translate into quantities relevant to flow and transport, Figure 9 shows the water retention and unsaturated hydraulic conductivity functions for the two soil layers corresponding to (i) the posterior VGM-parameter sets (color); (ii) the prior parameter ranges (grayscale); and (iii) the true parameter set in Table 1 (blue curve). Here we observe that the true curves often fall either at the limits of the posterior ranges or outside of them, meaning that the posterior parameter sets do not well reflect the soil hydraulic behaviour. Clearly, the model errors arising from the 1D vertical flow assumption cannot be neglected if we wish to have reliable predictions of flow and transport through this system.

[Figure 9 about here.]

3.4.2 L2-norm likelihood, no model-error correction

Figure 10 shows the marginal posterior histograms obtained for the case where the data in Figure 4a were inverted using the informal L_2-norm-based likelihood measure developed in Section 2.3. Again, the advantage of using this measure is that the likelihood is determined based on the expected behaviour of a summary measure of the residual (i.e., its L_2-norm), rather than on the residual vector itself, thereby avoiding an overly strong preference for model parameter sets whose corresponding residual statistics fit exactly the assumed Gaussian model. As in the Gaussian likelihood case, no attempt was made to remove the effects of model error in this inversion. To account for the increased residual energy due to model error and allow for effective MCMC sampling, a modest amount of error inflation was made by setting the residual standard deviation in equation (9) to lie between $s_{R_1} = 0.01$ and $s_{R_2} = 0.04$.

[Figure 10 about here.]

We observe in Figure 10 that, as was the case with the standard Gaussian likelihood, a strong bias exists in the posterior results because of the model error coming from the 1D flow assumption. Indeed, the marginal posterior histograms look similar to those in Figure 8, with the true parameter values for K_s and θ_s often falling far outside of the limits of the posterior distributions. In terms of the water retention and unsaturated hydraulic conductivity functions, Figure 11 shows results that are almost identical to those in Figure 9. Note, however, that because of the use of the informal likelihood measure, the results presented here were obtained with significantly less error inflation than in the Gaussian likelihood case. That is, in using the L_2-norm-based likelihood, we greatly increase the probability of acceptance of model parameter sets whose residual norm fits our expectations, but whose residual vector may deviate slightly from Gaussian.

[Figure 11 about here.]

3.4.3 L2-norm likelihood, correction for model error

Finally, Figure 12 shows the marginal posterior histograms obtained for the case where the water-content data in Figure 4a were inverted using our informal L_2-norm-based likelihood measure combined with the proposed correction for model error described in Section 2.2. Only a small amount of error inflation was done in this inversion by setting $s_{R_1} = 0.01$ and $s_{R_2} = 0.015$ in order to account for the fact that, even with the correction, it is unlikely that the model-error component of the residual will be perfectly removed. As a result, the energy in the remainder should be slightly larger than the level of noise added to the data.

[Figure 12 about here.]

We see in Figure 12 that, as a result identifying and subtracting the model-error component of the residual before evaluation of the likelihood in MCMC, the posterior VGM-parameter histograms are no longer biased, with the true parameter values falling in most cases near the middle of the posterior ranges. With regard to the corresponding hydraulic behaviour, we observe in Figure 13 that the true water retention and unsaturated hydraulic conductivity functions now lie well within the extent of the posterior curves. It is important to point out that, despite the fact that a minimal amount of error inflation was done for this inversion compared to the Gaussian- and informal-likelihood inversions, the posterior distributions are broader, most notably for parameters K_s and θ_s. This results from the fact that (i) a significant amount of residual energy is removed with our PCA-based correction before calculating the likelihood; and (ii) there exist some incorrect model-parameter sets whose corresponding parameter-error component of the residual will resemble (and will thus be identified as) model error, leading to the parameter sets being accepted in the MCMC inversion procedure. This latter important point is discussed in further detail in the following section.

[Figure 13 about here.]

3.5 Discussion

It is clear from the previous results that, in the context of the considered example problem, our proposed correction for model error offers an effective means of overcoming the posterior parameter bias related to use of a simplified forward model, thereby providing more accurate and useful uncertainty estimates. We now attempt to gain further insight into the reason why, with this correction, particular sets of incorrect model parameters may be accepted in the

MCMC procedure, which contributes to the broadening of the obtained posterior distributions. Figure 14 presents the results of an analysis of three different parameter sets, the first row corresponding to the true subsurface VGM parameters (Table 1) and the last two rows corresponding to random "test" sets of VGM parameters drawn from the prior distribution (Table 3). In the columns of the figure we show (i) the predicted GPR-derived water-content data assuming 1D vertical flow; (ii) the residual obtained by subtracting the "observed" data in Figure 4a and expressing the results relative to the arrival time of the infiltration front; (iii) the projection of this residual onto the model-error basis, which represents our estimate of the model-error component of the residual; and (iv) the corresponding remainder, obtained by subtracting the projection from the residual.

[Figure 14 about here.]

[Table 3 about here.]

We see in Figure 14 that, when the true set of VGM parameters is considered and thus when the parameter-error component of the residual is zero (see equation (6)), projection of the residual onto the model-error basis correctly identifies the model-error component, which after subtraction leaves a low-amplitude remainder that is mostly comprised of Gaussian data-measurement uncertainties. As the L_2-norm of the remainder determines the likelihood, the true parameter set stands a high chance of being accepted in MCMC. For the first (incorrect) set of test model parameters, we observe that the corresponding residual, which now is comprised of non-zero model- and parameter-error components, closely resembles the model-error realizations presented in Figure 5. Ideally, projection of this residual onto the PCA-derived basis would isolate only the model-error component. However, in this case the entire residual is identified as model error, which again results in a low-amplitude remainder and a correspondingly high probability of acceptance. In other words, when the sum of the model- and parameter-error components of the residual tends to look similar to the stochastic model-error realizations, both of these components will be subtracted in our correction procedure, leading to a high likelihood of an incorrect parameter set. Finally, for the last set of test model parameters, we observe the intended functioning of the algorithm; the projection of the residual onto the model-error basis correctly identifies the model-error component, but leaves the parameter-error component which then forms part of the remainder. The high amplitudes observed in the remainder yield a low probability of the parameter set being accepted.

The fact that certain sets of incorrect model parameters, whose residuals under the 1D flow assumption appear similar to the stochastic model-error realizations, are given a high likelihood in our modified inversion procedure may be initially disconcerting. However, it must be emphasized that, in any situation where parameter-related errors cannot be distinguished from model errors, the corresponding model-parameter set cannot be rejected as a possibility. Indeed, in this regard, our proposed algorithm should be viewed as a conservative stochastic inversion approach in the presence of model error; if a residual appears to resemble model error based on the generated model-error realizations, then the corresponding parameter set should not be excluded from the Bayesian posterior distributions. The strong advantage of our approach compared to not accounting for model error is that bias is strongly reduced and the true parameter set becomes well represented by these distributions.

4 Conclusions

Building on the recent work of *Köpke et al.* (2018), we have presented in this paper a methodology for accounting for model errors in Bayesian-MCMC inversions that is geared towards the common case where such errors arise from the use of an computationally efficient simplified forward model in place of a more accurate but computationally burdensome numerical solution. Our approach is based on the analysis of a suite stochastic model-error realizations, created before the MCMC iterations by running the simplified and full forward solvers together for randomly drawn model-parameter sets from the prior distribution, which leads to the development of an orthonormal basis for the model error. Under the assumption that the model errors for the considered problem can be well described by this basis and that the model-error component of the residual lies orthogonal to the parameter-error and data-measurement-error components, projection of the residual onto the basis identifies the model error, which is then subtracted from the residual before evaluating the likelihood.

We saw through the considered example problem that application of our model-error correction, combined with an informal likelihood measure based on the expected behaviour of the L_2-norm, leads to a strong reduction in bias and notably better characterization of posterior uncertainties. This comes at the cost of needing to perform a number of full forward model runs (in our case a few thousand) to generate the model-error basis prior to MCMC. Note again, however, that these full numerical simulations can be conducted in parallel. Our approach represents a remarkable computational savings when compared to MCMC based entirely on the full forward solver, in which hundreds of thousands of expensive model evaluations, conducted in series, would be necessary. For the specific example presented in this paper involving 800,000 Metropolis iterations, use of the fully 3D Richards' equation solver within MCMC would require over 900 days on a standard desktop computer. In contrast, running our algorithm based on the 1D forward solver with model-error correction took less that two weeks.

A critical assumption in our proposed approach is that of orthogonality between the model- and parameter-error components of the residual. As much as our experience until now suggests that this will be approximately true in many cases, hence explaining the success of our method, it cannot be proven and we observed that some incorrect model-parameter sets may produce a residual that looks like model error. In the latter cases, the model- and parameter-error components of the residual cannot be distinguished by projecting onto the basis and the parameter sets will stand a good chance of being accepted. In our view, this is not a concern as it simply means that the posterior parameter distributions will be broadened to include such parameters; i.e., our approach will conservatively include the parameters as possibilities. However, if this behaviour is undesirable, a two-stage MCMC algorithm could be proposed in which our approach would be used in a first accept/reject phase to effectively filter out unreasonable parameter sets from being tested with the full numerical solution, albeit at greatly increased computational cost.

It must be emphasized that the approach described herein is only intended for known sources of model error, for which random realizations of the error can be generated and used to help identify the model-error component of the residual. Although this will often be the major source of bias for inverse problems in geophysics and hydrology, there are situations where even our best forward solution will not provide a good enough description of the physical process involved. In the case of such unknown or unspecified model errors, our methodology can still be expected to effectively deal with the model errors for which it was intended, thereby providing more reliable posterior uncertainty estimates. Another related issue is the fact that, for many problems, particularly those of high dimension with spatially distributed parameters, the nature of the model errors may change significantly over the model parameter space and it may not be possible to effectively describe them using a single global basis. In this case, the work of *Köpke et al.* (2018) shows that a KNN dictionary-based approach to model-error identification, whereby the basis is constructed locally at each MCMC iteration, can be a highly effective means of obtaining reliable posterior parameter distributions when using an approximate forward solver. It is also likely that the computational efficiency of the approach of *Köpke et al.* (2018) can be further improved by using parallel computation within the MCMC procedure to generate local model-error realizations simultaneously. Finally, future work should investigate whether the approach proposed in this paper might be adapted for use with gradient-based MCMC methods employing an adjoint solver based on the simplified forward model.

Acknowledgments

This work was supported by a grant to J. Irving from the Swiss National Science Foundation, grant number 200021-140864. The authors also wish to thank Richard Healey from the U.S. Geological Survey for support with the VS2D software during the initial stages of this research project.

References

Annan, A. (2006), GPR methods for hydrogeological studies, in *Hydrogeophysics*, edited by Y. Rubin and S. S. Hubbard, pp. 185–213, Springer.

Arridge, S., J. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and

M. Vauhkonen (2006), Approximation errors and model reduction with an application in optical diffusion tomography, *Inverse Problems*, *22*(1), 175.

Beven, K., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrological Process*, *6*, 279–298.

Beven, K., and A. Binley (2014), GLUE: 20 years on, *Hydrological Processes*, *28*(24), 5897–5918.

Beven, K., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *Journal of Hydrology*, *249*(1), 11–29.

Binley, A., and K. Beven (2003), Vadose zone flow model uncertainty as conditioned on geophysical data, *Ground Water*, *41*, 119–127.

Blasone, R.-S., J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski (2008), Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling, *Advances in Water Resources*, *31*(4), 630–648.

Bodin, T., and M. Sambridge (2009), Seismic tomography with the reversible jump algorithm, *Geophysical Journal International*, *178*(3), 1411–1436.

Brynjarsdóttir, J., and A. O'Hagan (2014), Learning about physical parameters: The importance of model discrepancy, *Inverse Problems*, *30*(11), 114,007.

Carsel, R. F., and R. S. Parrish (1988), Developing joint probability distributions of soil water retention characteristics, *Water Resources Research*, *24*, 755–769, doi: 10.1029/WR024i005p00755.

Cassiani, G., and A. Binley (2005), Modeling unsaturated flow in a layered formation under quasi-steady state conditions using geophysical data constraints, *Advances in Water Resources*, *28*(5), 467–477.

Christen, J. A., and C. Fox (2005), Markov chain Monte Carlo using an approximation, *Journal of Computational and Graphical Statistics*, *14*(4), 795–810.

Cui, T., C. Fox, and M. J. O'Sullivan (2011), Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm, *Water Resources Research*, *47,W10521*, doi:10.1029/2010WR010352.

Del Giudice, D., M. Honti, A. Scheidegger, C. Albert, P. Reichert, and J. Rieckermann (2013), Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, *Hydrology and Earth System Sciences*, *17*, 4209–4225, doi:10.5194/hess-17-4209-2013.

Doherty, J., and S. Christensen (2011), Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, *Water Resources Research*, *47*, W12,534, doi:10.1029/2011WR010763.

Efendiev, Y., A. Datta-Gupta, X. Ma, and B. Mallick (2008), Modified Markov Chain Monte Carlo method for dynamic data integration using streamline approach, *Mathematical Geosciences*, *40*(2), 213–232.

Forbes, C., M. Evans, N. Hastings, and B. Peacock (2010), *Statistical Distributions*, Wiley Series in Probability and Statistics - Applied Probability and Statistics Section Series, John Wiley & Sons.

Gallagher, K., K. Charvin, S. Nielsen, M. Sambridge, and J. Stephenson (2009), Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *Marine and Petroleum Geology*, *26*(4), 525–535.

Gelman, A., G. O. Roberts, W. R. Gilks, et al. (1996), Efficient Metropolis jumping rules, *Bayesian statistics*, *5*(599-608), 42.

van Genuchten, M. T. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soil, *Soil Science Society of America Journal*, *44*, 892–898.

Hansen, T. M., K. S. Cordua, B. H. Jacobsen, and K. Moosegaard (2014), Accounting for imperfect forward modeling in geophysical inverse problems - exemplified for cross hole tomography, *Geophysics*, *79*, H1–H21, doi:10.1190/geo2013-0215.1.

Hassan, A. E., H. M. Bekhit, and J. B. Chapman (2009), Using Markov Chain Monte Carlo to quantify parameter uncertainty and its effect on predictions of a groundwater flow model, *Environmental Modelling & Software*, *24*(6), 749–763.

Hastings, W. K. (1970), Monte Carlo sampling methods using Markov chains and their applications, *Biometrika*, *57*(1), 97–109.

Hinnell, A. C., T. P. A. Ferré, J. A. Vrugt, J. A. Huisman, S. Moysey, J. Rings, and M. B. Kowalsky (2010), Improved extraction of hydrologic information from geophysical data through coupled hydrogeophysical inversion, *Water Resources Research*, *46*, W00D40, doi: 10.1029/2008WR007060.

Irving, J., and K. Singha (2010), Stochastic inversion of tracer test and electrical geophysical data to estimate hydraulic conductivities, *Water Resources Research*, *46*(11).

Josset, L., D. Ginsbourger, and I. Lunati (2015), Functional error modeling for uncertainty quantification in hydrogeology, *Water Resources Research*, *51*, 10501068, doi: 10.1002/2014WR016028.

Kaipio, J., and E. Somersalo (2007), Statistical inverse problems: Discretization, model reduction and inverse crimes, *Journal of Computational and Applied Mathematics*, *198*, 493–504, doi:10.1016/j.cam.2005.09.027.

Kennedy, M. C., and A. O'Hagan (2001), Bayesian calibration of computer models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *63*(3), 425–464.

Köpke, C., J. Irving, and A. H. Elsheikh (2018), Accounting for model error in Bayesian solutions to hydrogeophysical inverse problems using a local basis approach, *Advances in Water Resources*, *116*, 195–207.

Lappala, E. G., R. W. Healy, E. P. Weeks, et al. (1987), *Documentation of computer program VS2D to solve the equations of fluid flow in variably saturated porous media*, Department of the Interior, US Geological Survey.

Lehikoinen, A., J. M. J. Huttunen, S. Finsterle, M. B. Kowalsky, and J. P. Kaipio (2010), Dynamic inversion for hydrological process monitoring with electrical resistance tomography under model uncertainties, *Water Resources Research*, *46,W04513*, doi: 10.1029/2009WR008470.

Linde, N., and J. A. Vrugt (2012), Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion, *Vadose Zone Journal*, *12*, 238–248, doi: 10.2136/vzj2012.0101.

Linde, N., D. Ginsbourger, J. Irving, F. Nobile, and A. Doucet (2017), On uncertainty quantification in hydrogeology and hydrogeophysics, *Advances in Water Resources*, *110*, 166–181.

Looms, M. C., A. Binley, K. H. Jensen, and L. Nielsen (2008), Identifying unsaturated hydraulic parameters using an integrated data fusion approach on cross-borehole geophysical data, *Vadose Zone Journal*, *7*, 238–248, doi:10.2136/vzj2007.0087.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, *21*(6), 1087–1092.

Mualem, Y. (1976), A new model for predicting the hydraulic conductivity of unsaturated porous media, *Water Resources Research*, *12*, 513–522, doi:10.1029/WR012i003p00513.

Nott, D. J., L. Marshall, and J. Brown (2012), Generalized likelihood uncertainty estimation (GLUE) and approximate Bayesian computation: What's the connection?, *Water Resources Research*, *48*(12).

O'Sullivan, A., and M. Christie (2006), Simulation error models for improved reservoir prediction, *Reliability Engineering & System Safety*, *91*(10), 1382–1389.

Ray, J., Z. Hou, M. Huang, K. Sargsyan, and L. Swiler (2015), Bayesian calibration of the community land model using surrogates, *Journal of Uncertainty Quantification*, *3*, 199–233, doi:10.1137/140957998.

Richards, L. A. (1931), Capillary conduction of liquids through porous mediums, *Physics*, *1*(5), 318–333.

Rockhold, M. L., C. S. Simmons, and M. J. Fayer (1997), An analytical solution technique for one-dimensional, steady vertical water flow in layered soils, *Water Resources Research*, *33*(4), 897–902.

Rossi, M., G. Cassiani, and A. Binley (2012), A stochastic analysis of cross-hole ground-penetrating radar zero-offset profiles for subsurface characterization, *Vadose Zone Journal*, *11*(4).

Rucker, D. F. (2011), Inverse upscaling of hydraulic parameters during constant flux infiltration using borehole radar, *Advances in Water Resources*, *34*(2), 215–226.

Rucker, D. F., and T. Ferré (2004), Parameter estimation for soil hydraulic properties using zero-offset borehole radar, *Soil Science Society of America Journal*, *68*(5), 1560–1567.

Ruggeri, P., J. Irving, and K. Holliger (2015), Systematic evaluation of sequential geostatistical resampling within MCMC for posterior sampling of near-surface geophysical inverse problems, *Geophysical Journal International*, *202*(2), 961–975.

Sadegh, M., and J. Vrugt (2013a), Approximate Bayesian Computation in hydrologic modeling: equifinality of formal and informal approaches, *Hydrology & Earth System Sciences Discussions*, *10*(4).

Sadegh, M., and J. Vrugt (2013b), Bridging the gap between GLUE and formal statistical approaches: Approximate Bayesian computation, *Hydrology and Earth System Sciences*, *17*(12).

Scholer, M., J. Irving, A. Binley, and K. Holliger (2011), Estimating vadose zone hydraulic properties using ground penetrating radar: The impact of prior information, *Water Resources Research*, *47*(10).

Scholer, M., J. Irving, M. C. Looms, L. Nielsen, and K. Holliger (2012), Bayesian Markov-

Chain-Monte-Carlo inversion of time-lapse crosshole GPR data to characterize the vadose zone at the Arrenaes site, Denmark, *Vadose Zone Journal*, *11*(4).

Scholer, M., J. Irving, M. Looms, L. Nielsen, and K. Holliger (2013), Examining the information content of time-lapse crosshole GPR data collected under different infiltration conditions to estimate unsaturated soil hydraulic properties, *Advances in Water Resources*, *54*, 38–56.

Schoups, G., and J. A. Vrugt (2010), A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, *Water Resources Research*, *46,W10531*, doi:10.1029/2009WR008933.

Smith, T., A. Sharma, L. Marshall, R. Mehrotra, and S. Sisson (2010), Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, *Water Resources Research*, *46,W12551*, doi:10.1029/2010WR009514.

Smith, T., L. Marshall, and A. Sharma (2015), Modeling residual hydrologic errors with Bayesian inference, *Journal of Hydrology*, *528*, 29–37.

Stephen, K. (2007), Scale and process dependent model errors in seismic history matching, *Oil & Gas Science and Technology – Revue de l'IFP*, *62*(2), 123–135.

Tarantola, A. (2005), *Inverse problem theory and methods for model parameter estimation*, SIAM.

Vrugt, J. A., and M. Sadegh (2013), Towards diagnostic model calibration and evaluation: Approximate Bayesian computation, *Water Resources Research*, *49*, 4335–4345, doi:10.1002/wrcr.20354.

Vrugt, J. A., C. J. Ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resources Research*, *44*(12).

Wilkinson, R. D. (2013), Approximate Bayesian computation (ABC) gives exact results under the assumption of model error, *Statistical applications in genetics and molecular biology*, *12*(2), 129–141.

Xu, T., and A. J. Valocchi (2015), A Bayesian approach to improved calibration and prediction of groundwater models with structural error, *Water Resources Research*, *51*(11), 9290–9311.

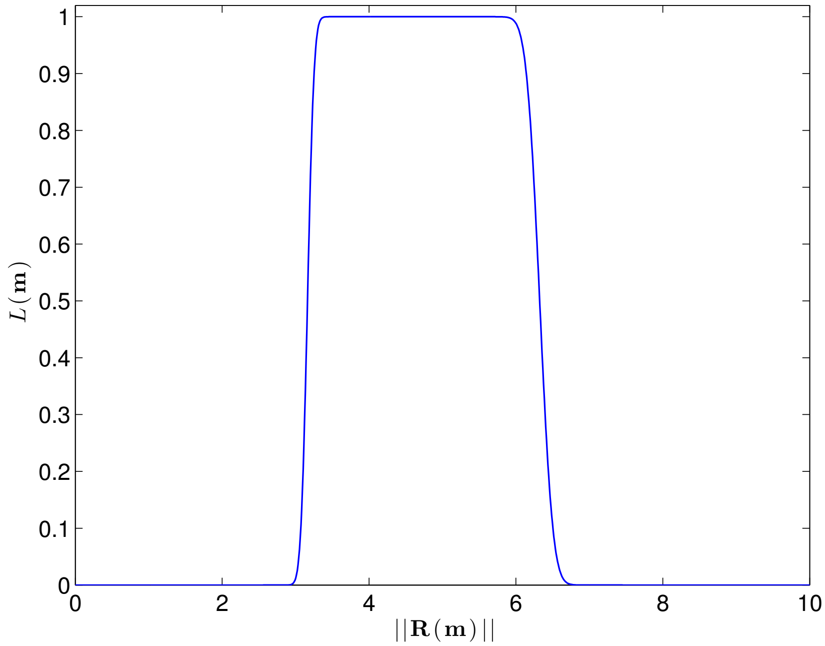

Figure 1: Normalized ℓ_2-norm likelihood given by equation (9) as a function of $||\mathbf{R}(\mathbf{m})||$ for $N = 1000$ with $s_{R_1} = 0.1$ and $s_{R_2} = 0.2$.

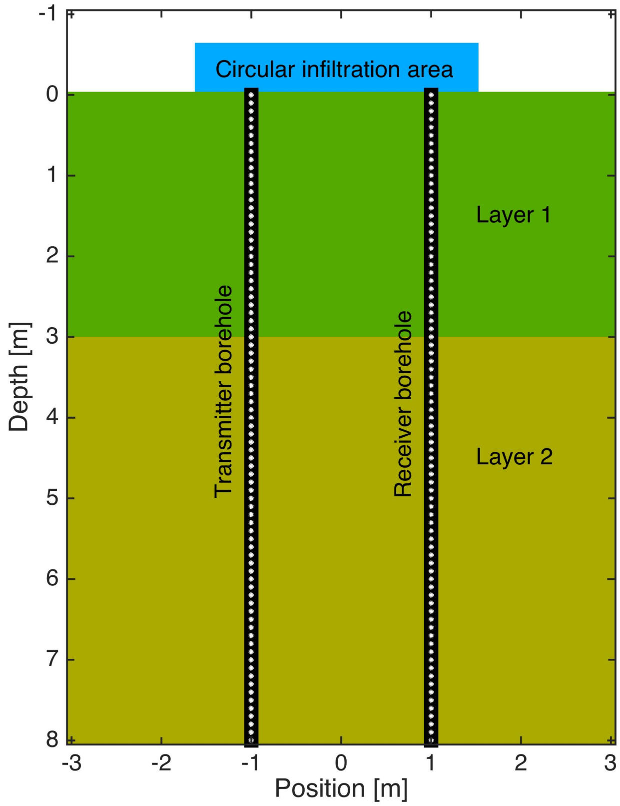

Figure 2: Setup for the synthetic infiltration experiment considered in this study. The white dots represent transmitter and receiver antenna positions for the ZOP GPR measurements.

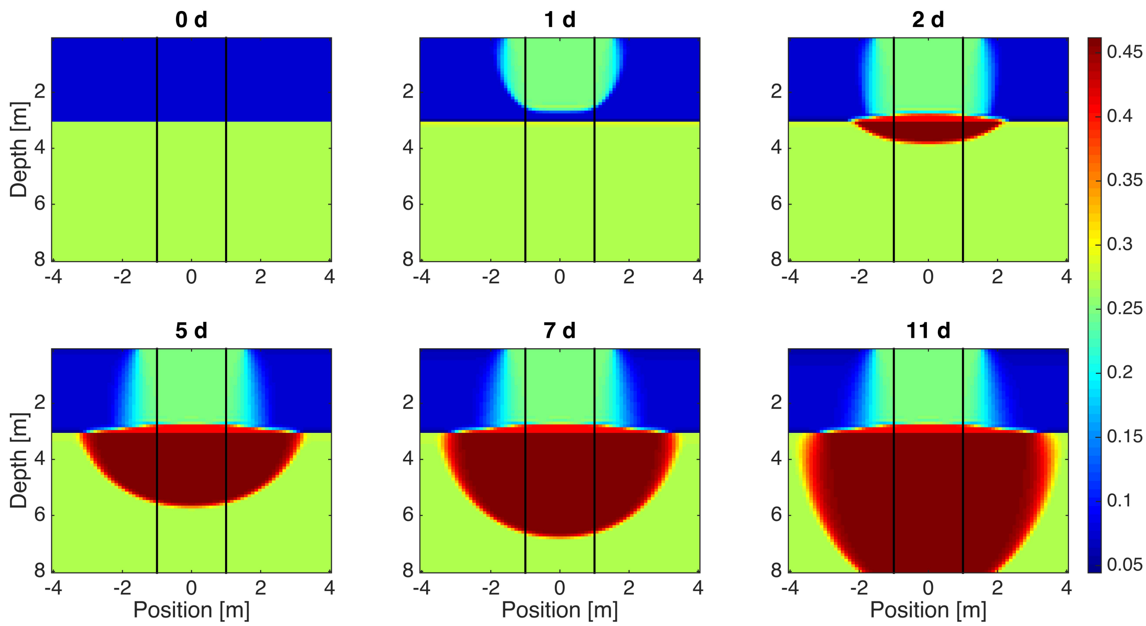

Figure 3: Spatial distribution of water content in the subsurface at various times throughout the infiltration experiment. The GPR boreholes are shown for reference.

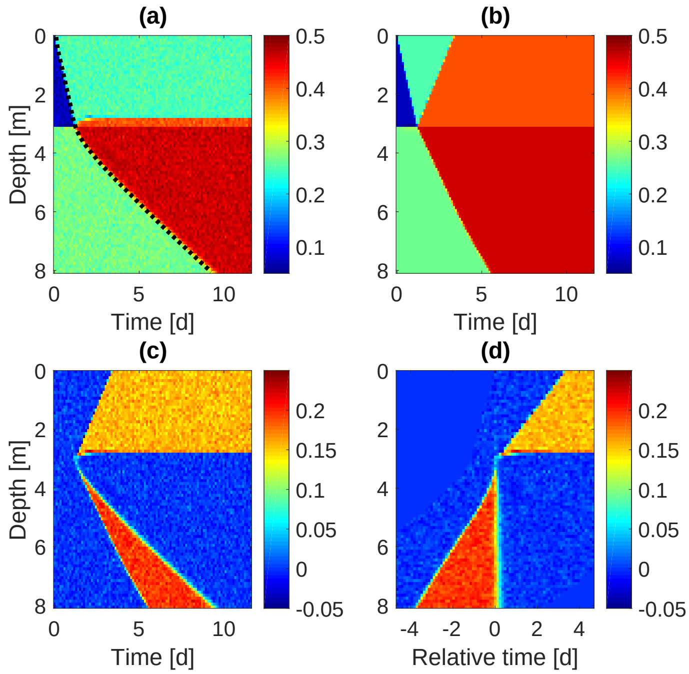

Figure 4: (a) Average soil water content between the boreholes as a function of depth and measurement time, computed using a 3D infiltration model with the addition of Gaussian measurement noise, representing the synthetic data to be inverted for the VGM parameters in Table 1. The arrival time of the infiltration front as a function of depth is indicated with a black dashed line. (b) Corresponding water-content distribution obtained assuming purely vertical (1D) flow. (c) Difference (b)-(a), which is equal to the sum of the model error and measurement uncertainties. (d) Error image from (c) expressed relative to the arrival time of the infiltration front in (a).

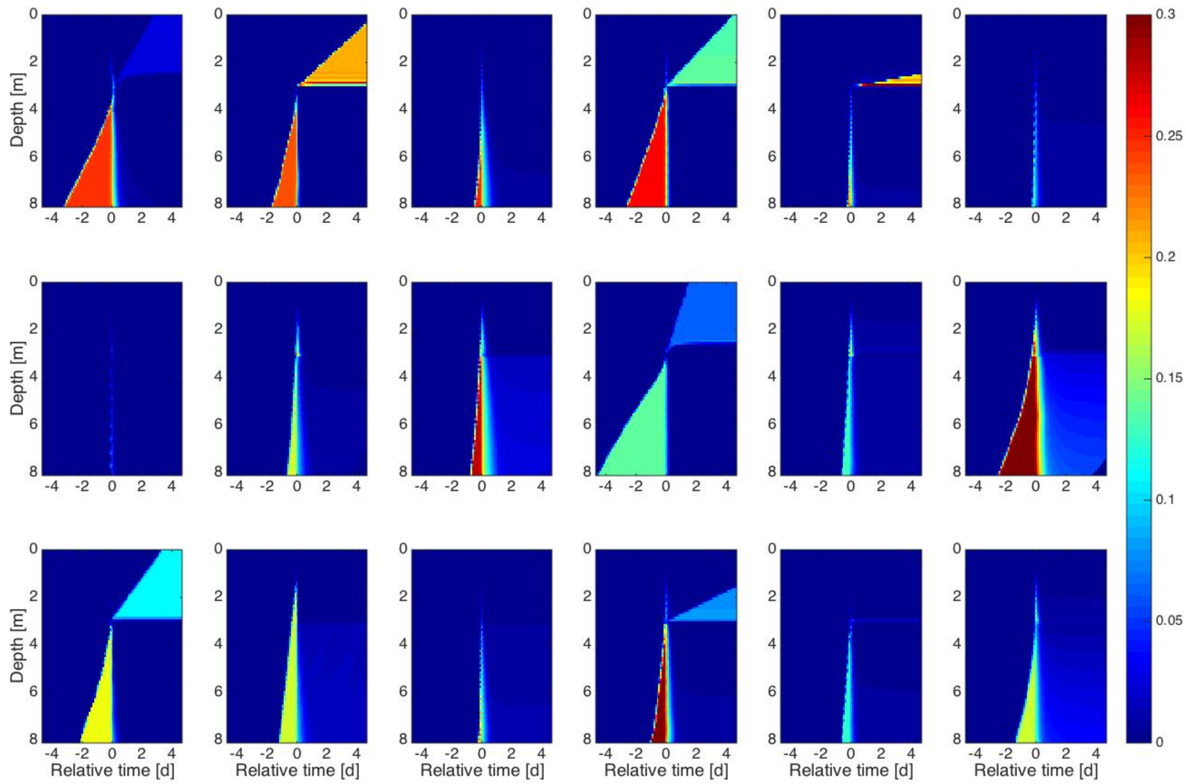

Figure 5: Example stochastic realizations of the model error corresponding to random sets of VGM parameters drawn from the prior distributions in Table 2. For greater coherency between the images, each has been expressed relative to the arrival time of the infiltration front in depth as observed in the 3D flow simulation. A total of 6500 realizations were generated to construct the model-error basis.

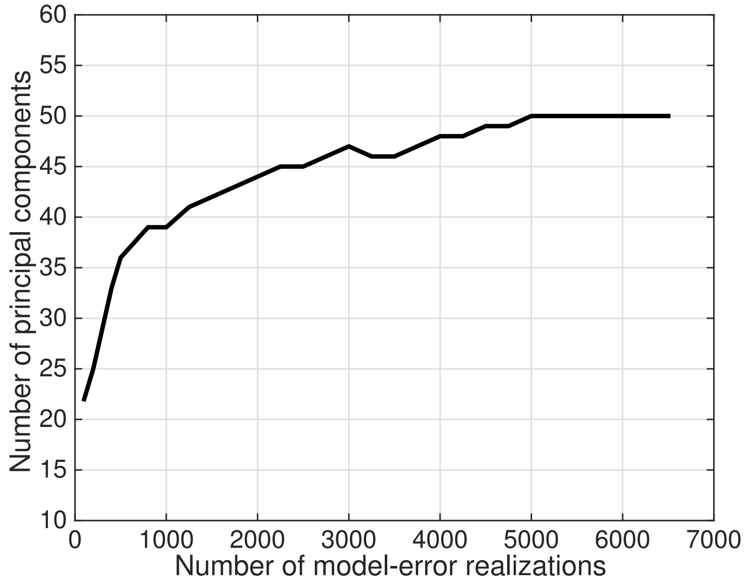

Figure 6: Number of principal components needed to capture 98% of the variance of the model-error realizations as a function of the number of realizations considered.

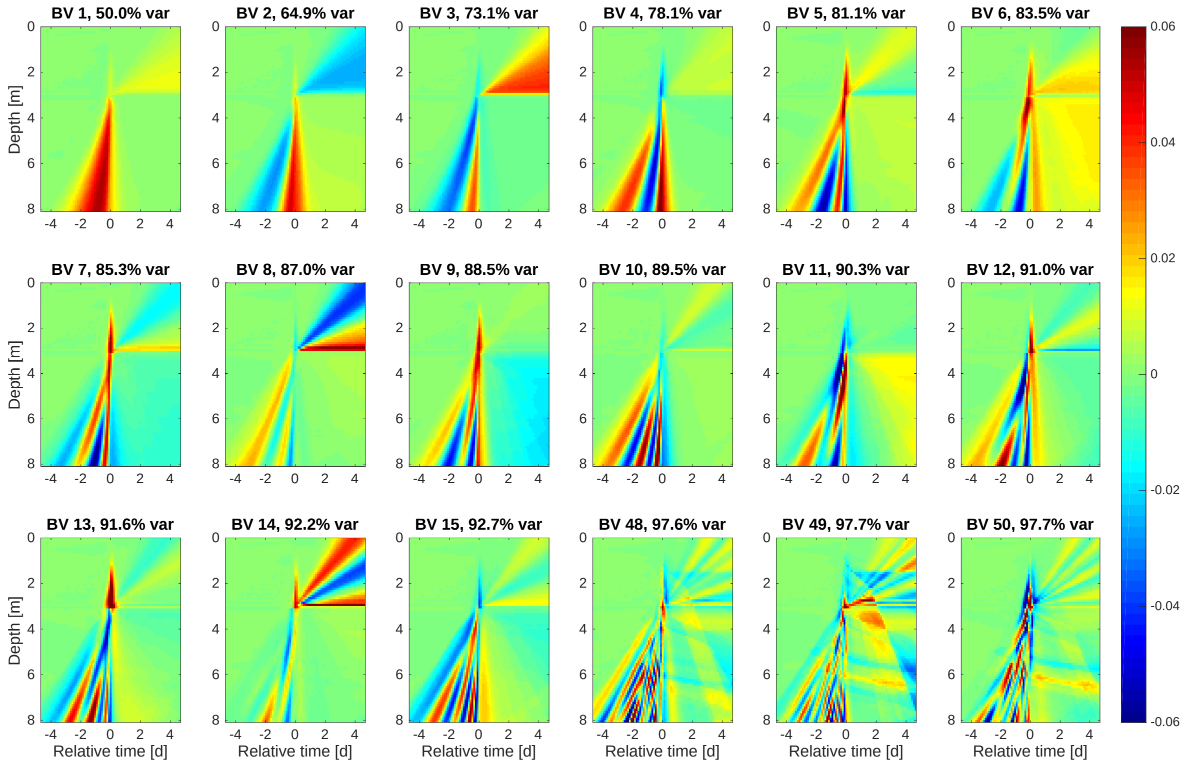

Figure 7: The first 15, and the last 3, of 50 model-error basis vectors, arranged in decreasing order with respect to their contribution of the total variance. The vectors were obtained by performing PCA on the set of 6500 stochastic model-error realizations. Each vector is plotted as an image with the cumulative contribution to the variance noted in the title.

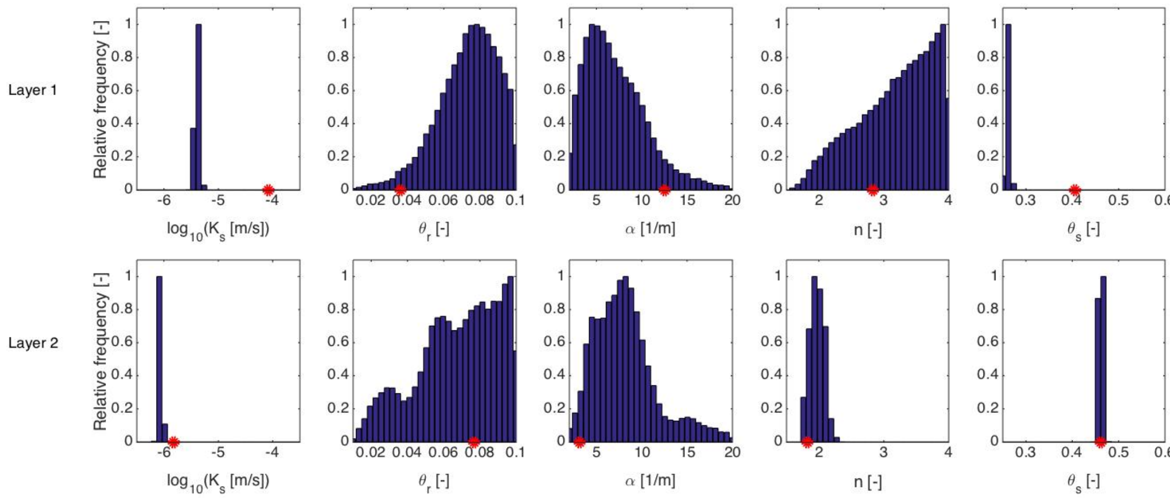

Figure 8: Marginal posterior histograms for the VGM parameters in each soil layer, obtained through MCMC sampling using an inflated Gaussian likelihood function for the residuals with no model-error correction. The red dots indicate the true parameter values. The limits of the horizontal axis on each plot represent the prior uniform parameter bounds.

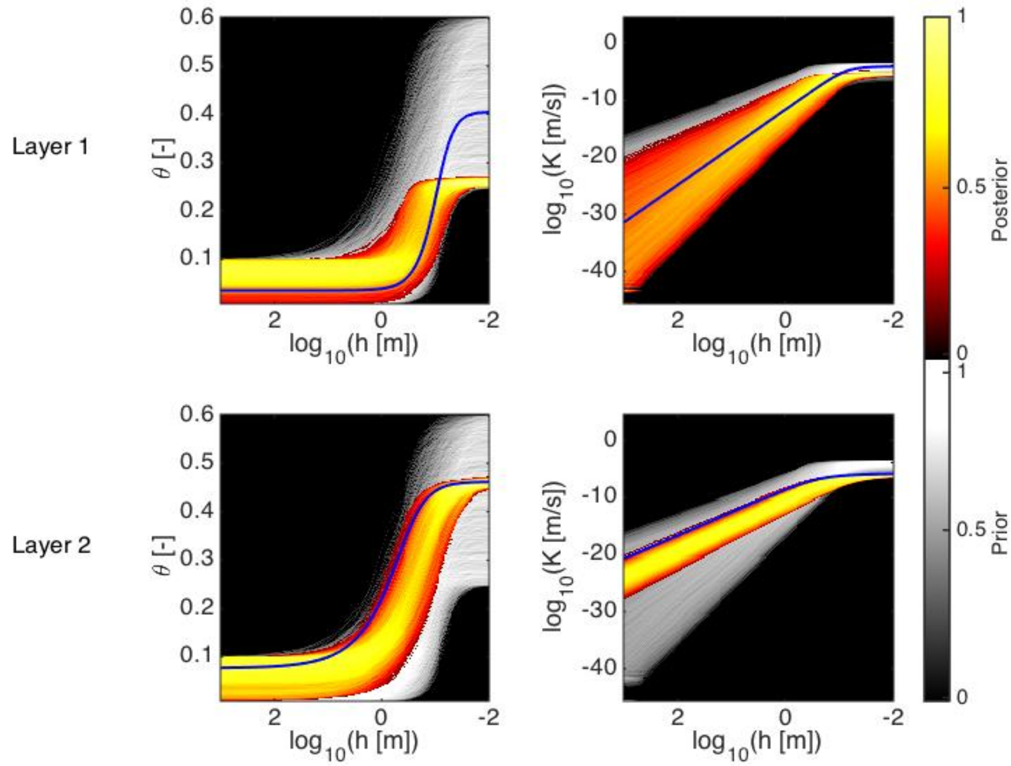

Figure 9: Water retention (left) and unsaturated hydraulic conductivity (right) functions for each soil layer corresponding to the prior distribution (gray; Table 2) and the posterior distribution obtained using an inflated Gaussian likelihood function for the residuals with no model-error correction (color; Figure 8). The blue lines represent the curves corresponding to the true parameter set in Table 1. The prior and posterior results are expressed in terms of curve densities.

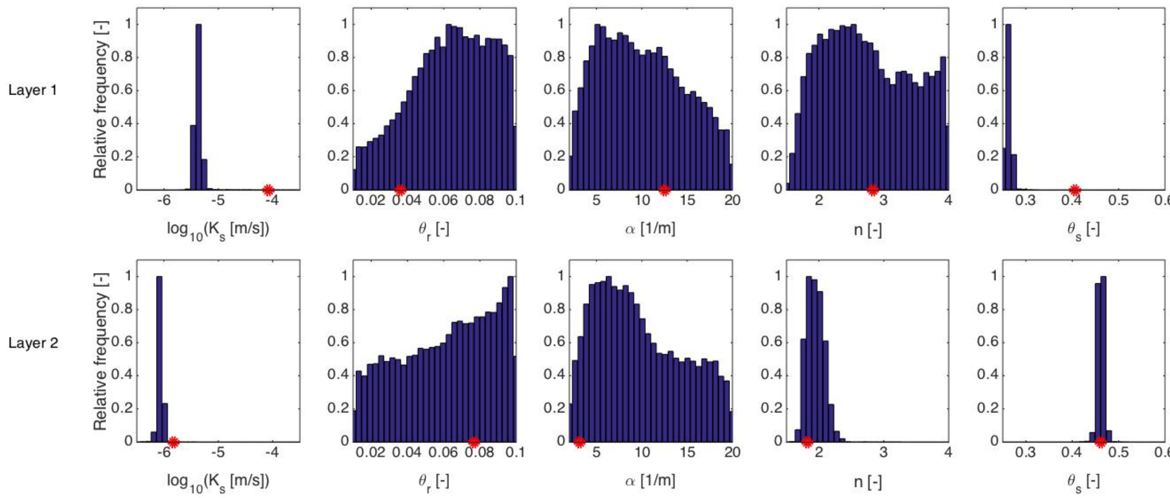

Figure 10: Marginal posterior histograms for the VGM parameters in each soil layer, obtained through MCMC sampling using an L_2-norm-based likelihood measure for the residuals with no model-error correction (see text for details). The red dots indicate the true parameter values. The limits of the horizontal axis on each plot represent the prior uniform parameter bounds.

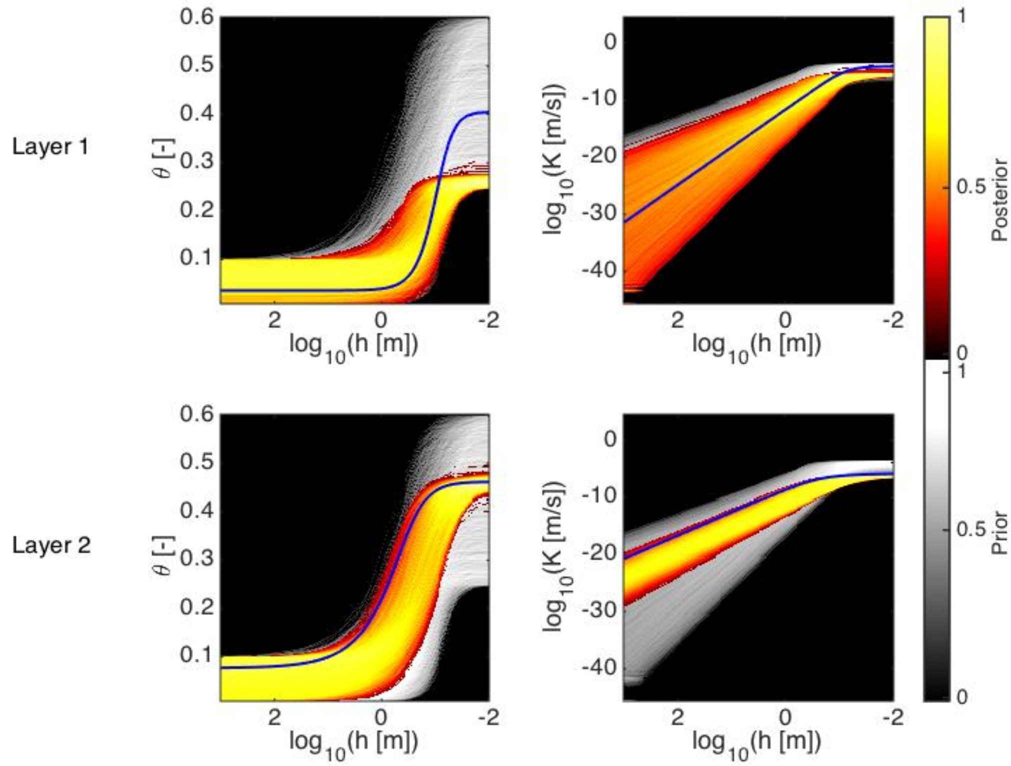

Figure 11: Water retention (left) and unsaturated hydraulic conductivity (right) functions for each soil layer corresponding to the prior distribution (gray; Table 2) and the posterior distribution obtained using an L_2-norm-based likelihood measure for the residuals with no model-error correction (color; Figure 10). The blue lines represent the curves corresponding to the true parameter set in Table 1. The prior and posterior results are expressed in terms of curve densities.

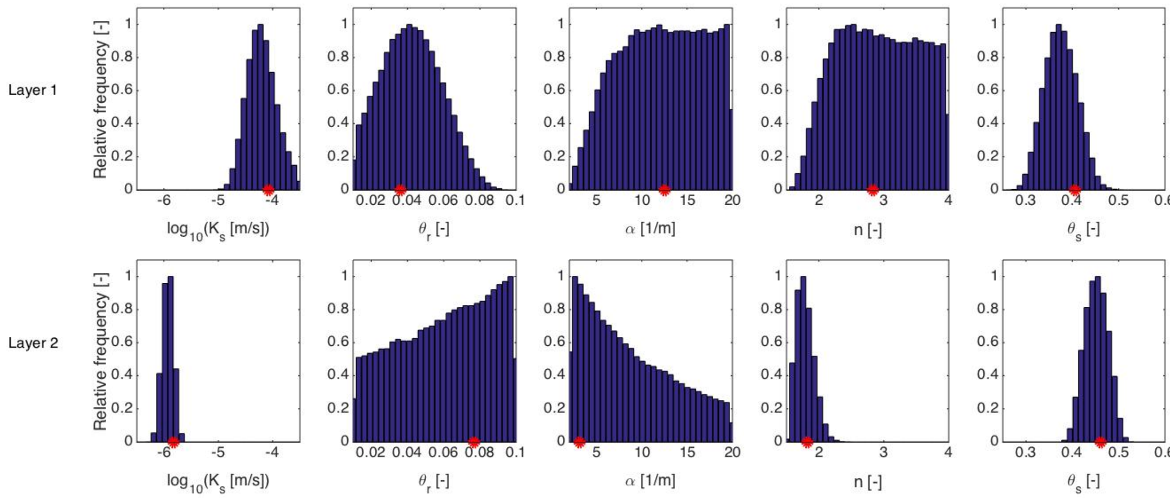

Figure 12: Marginal posterior histograms for the VGM parameters in each soil layer, obtained through MCMC sampling using an L_2-norm-based likelihood measure for the residuals after correcting for model error (see text for details). The red dots indicate the true parameter values. The limits of the horizontal axis on each plot represent the prior uniform parameter bounds.

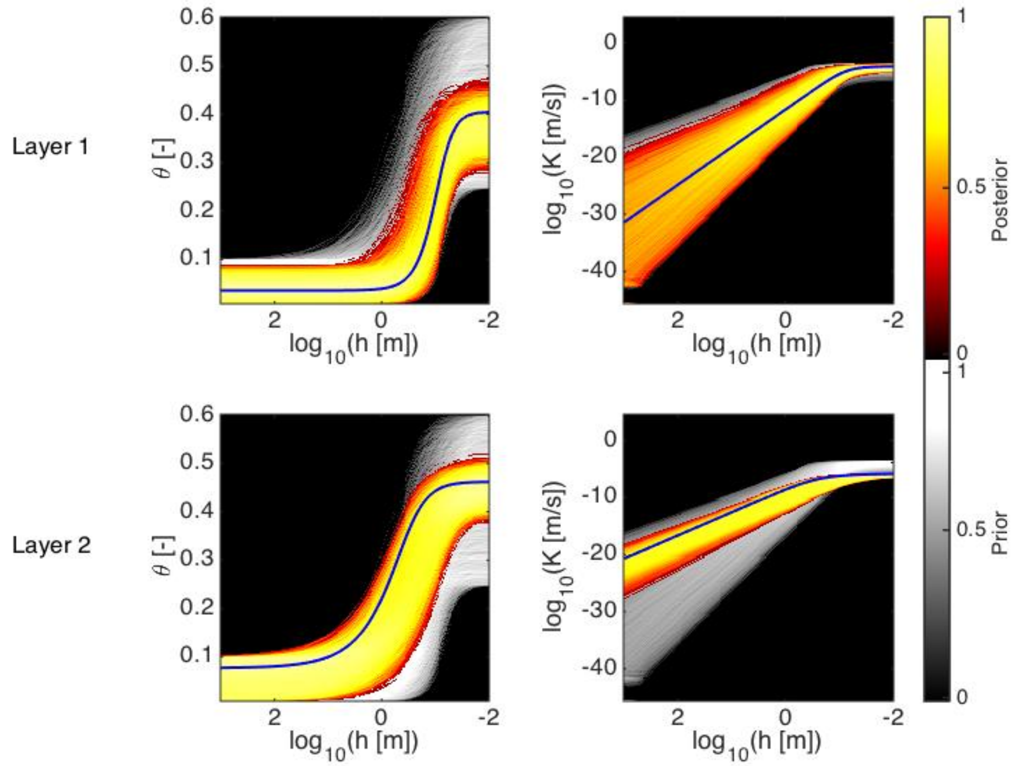

Figure 13: Water retention (left) and unsaturated hydraulic conductivity (right) functions for each soil layer corresponding to the prior distribution (gray; Table 2) and the posterior distribution obtained using an L_2-norm-based likelihood measure for the residuals after correcting for model error (color; Figure 12). The blue lines represent the curves corresponding to the true parameter set in Table 1. The prior and posterior results are expressed in terms of curve densities.

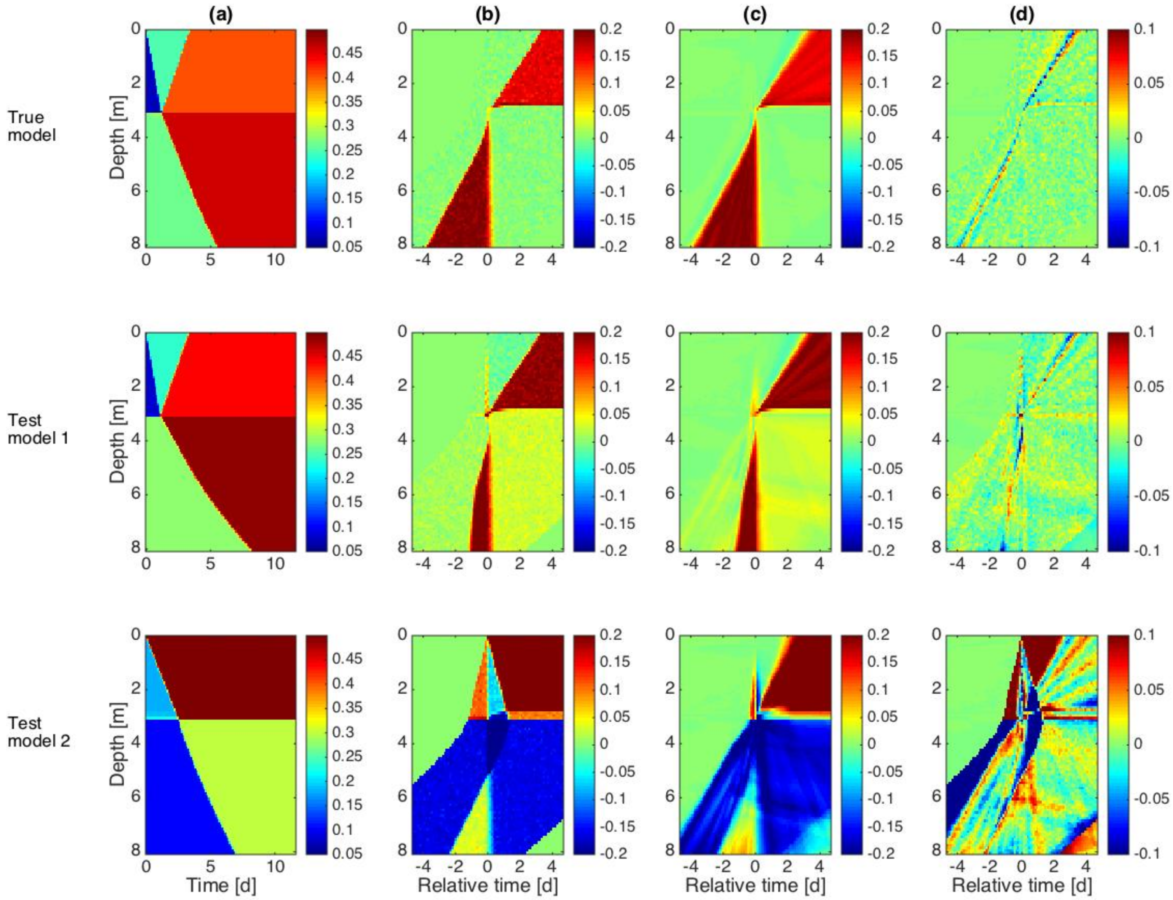

Figure 14: For the true set of VGM parameters (Table 1) and two sets of incorrect "test" model parameters (Table 3): (a) Predicted water content assuming 1D vertical flow; (b) Residual obtained by subtracting the synthetic data in Figure 4a from the results in (a) and expressing relative to the arrival time of the infiltration front observed in the data; (c) Projection of the residual in (b) onto the model-error basis; (d) Corresponding remainder (b)-(c).

	$\log_{10}(K_s$ [m/s])	θ_r [-]	α [1/m]	n [-]	θ_s [-]
Layer 1	-4.074	0.036	12.552	2.830	0.405
Layer 2	-5.826	0.077	3.165	1.819	0.462

Table 1: VGM parameters considered for the 2-layer synthetic example.

	$\log_{10}(K_s$ [m/s])	θ_r [-]	α [1/m]	n [-]	θ_s [-]
Lower bound	-6.500	0.010	2.000	1.500	0.250
Upper bound	-3.500	0.100	20.000	4.000	0.600

Table 2: Lower and upper bounds of the uniform Bayesian prior distributions assumed for the VGM parameters in each layer.

		$\log_{10}(K_s$ [m/s])	θ_r [-]	α [1/m]	n [-]	θ_s [-]
Test model 1	Layer 1	-3.950	0.050	5.600	3.500	0.440
	Layer 2	-6.030	0.060	4.100	1.900	0.490
Test model 2	Layer 1	-5.300	0.080	7.000	2.600	0.500
	Layer 2	-5.700	0.030	10.000	3.000	0.300

Table 3: VGM parameters corresponding to the two test models considered in Figure 14.