Accounting for model error in Bayesian solutions to hydrogeophysical inverse problems using a local basis approach

Corinna Köpke[a,*], James Irving[a], Ahmed H. Elsheikh[b]

[a] *University of Lausanne, Institute of Earth Sciences, 1015 Lausanne, Switzerland*
[b] *Heriot-Watt University, Institute of Petroleum Engineering, Edinburgh EH14 4AS, United Kingdom*

Abstract

Bayesian solutions to geophysical and hydrological inverse problems are dependent upon a forward model linking subsurface physical properties to measured data, which is typically assumed to be perfectly known in the inversion procedure. However, to make the stochastic solution of the inverse problem computationally tractable using methods such as Markov-chain-Monte-Carlo (MCMC), fast approximations of the forward model are commonly employed. This gives rise to model error, which has the potential to significantly bias posterior statistics if not properly accounted for. Here, we present a new methodology for addressing model error in Bayesian solutions to hydrogeophysical inverse problems that is geared towards the common case where the error cannot be (i) effectively characterized through some parametric statistical distribution; and (ii) estimated by interpolating between a small number of computed model-error realizations. To this end, we focus on identification and removal of the model-error component of the residual during MCMC using a projection-based approach, whereby the orthogonal basis employed for the projection is derived in each iteration from the K-nearest neighboring entries in a model-error dictionary. The latter is constructed during the inversion and grows at a specified rate

*Corresponding author. Tel.: +41 21 692 44 09
Email addresses: `corinna.koepke@unil.ch` (Corinna Köpke), `james.irving@unil.ch` (James Irving), `A.Elsheikh@hw.ac.uk` (Ahmed H. Elsheikh)

as the iterations proceed. We demonstrate the performance of our technique on the inversion of synthetic crosshole ground-penetrating radar travel-time data considering three different subsurface parameterizations of varying complexity. Synthetic data are generated using the eikonal equation, whereas a straight-ray forward model is assumed for their inversion. In each case, our developed approach enables us to remove posterior bias and obtain a more realistic characterization of uncertainty.

Keywords: Model error, Bayesian inference, MCMC, Proxy model

1. Introduction

Bayesian inversion of hydrological and geophysical data using Markov-chain-Monte-Carlo (MCMC) methods has become increasingly popular over the past decade. Key advantages of this approach are that: (i) it allows for more comprehensive quantification of posterior parameter uncertainty when compared to traditional linearized uncertainty estimates; (ii) it is extremely flexible in the sense that any information that can be expressed probabilistically (e.g., model prior information, data measurement errors) can be incorporated into the inverse problem; and (iii) it provides a natural framework within which to perform data integration. The Bayesian-MCMC approach does, however, have the notable disadvantage of being limited by its high computational cost, which results from the typically large numbers of model parameters in geophysical and hydrological problems combined with the need for small model perturbations along the Markov chain in order to ensure reasonable rates of proposal acceptance. That is, millions of forward model runs are commonly required to obtain meaningful posterior statistics, which is computationally prohibitive for many real-world applications (e.g., [1]).

A variety of techniques exist for reducing the computational load of Bayesian-MCMC inversions. Recent algorithmic developments for MCMC methods, which take advantage of parallel architectures and incorporate chain history and posterior gradient information into the proposal distribution, have been shown to significantly improve computational efficiency past the standard Metropolis-Hastings approach (e.g., [2, 3, 4, 5, 6, 7]). Model reduction, through the use of basis functions that exploit the spatial correlation naturally present in subsurface properties (e.g., [8, 9, 10, 11]), can also be performed to reduce the dimensionality, and thus the numerical complexity, of the inverse problem. Yet another means of reducing the computational load of Bayesian-MCMC inversions, and arguably the most intuitive and commonly employed approach, is to use a fast approximation of the forward solver in place of the slower "full" numerical solution. This can be accomplished via simplification of the physics of the problem (e.g., [12, 13]), reduction of the numerical accuracy of the solution by coarsening the model discretization (e.g., [14, 15]), or the construction of response-surface proxies based on, for example, polynomial chaos expansion, artificial neural networks, or Gaussian processes (e.g., [16, 17, 18, 19]). While the use of approximate forward solvers in this manner can be highly effective, it can lead to strongly biased and overconfident posterior statistics if the discrepancies between the approximate and detailed solutions are not taken into account [20]. Indeed, such "model errors" have the potential to overwhelm the effects of data measurement uncertainties and may have a controlling influence on posterior inference. Despite this fact, the issue of model error has been largely ignored in the vast majority of geophysical and hydrological studies to date where Bayesian-MCMC methods have been employed.

In recent years, a number of techniques have appeared in the scientific and engineering literature to address the model error problem, thus allowing for more effective use of approximate forward solvers in Bayesian stochastic inversions. One popular avenue of research focuses on the overall or "global" statistical characterization of these errors, whereby a small number of stochastic model-error realizations, generated by running the approximate and detailed forward solvers on random parameter sets drawn from the prior distribution, are used to develop likelihood functions that better reflect the combined nature of all error sources. To this end, by far the most straightforward and common approach is to assume that the model errors are Gaussian distributed and thus characterized by some mean vector and covariance matrix, both of which are estimated from the realizations (e.g., [14, 21, 22, 23, 24]). Alternatively, customized parametric likelihood functions have been developed, most notably in the fields of catchment and urban hydrology, to reflect the non-Gaussian, strongly correlated, and often heteroscedastic nature of residuals in some problems (e.g., [25, 26, 27]). In all of these studies, it has been shown that inclusion of model-error statistical characteristics into the Bayesian likelihood function results in a broadening of posterior distributions along with, in many cases, a reduction in posterior bias. A key concern, however, is the validity of the assumption that the errors can be adequately described by the specified parametric distribution. Indeed, our own experience with high-dimensional spatially distributed inverse problems in geophysics and hydrology suggests that it is more often the case that model errors exhibit highly complex statistics and correlations that change significantly not only over the data space, but also as a function of the input model parameters. Note that this in part has led to greatly increased interest in alternative likelihood methods such as generalized likelihood uncertainty estimation (GLUE) (e.g., [28]) and approximate Bayesian computation (ABC) (e.g., [29]).

Another avenue of research to account for the discrepancy between approximate and detailed forward solvers in Bayesian stochastic inversions, which addresses the latter point above, focuses on the development of "local" error models that describe, either statistically or deterministically, the discrepancy between the approximate and detailed forward solutions over the model parameter space. O'Sullivan and Christie [30], for example, use a small number of coarse-grid versus fine-grid model-error realizations, computed over a low-dimensional model-parameter space, to characterize through interpolation how the model-error mean and covariance matrix change as a function of the input parameters. Kennedy and O'Hagan [31] present a comprehensive theoretical framework for dealing with model errors where the error statistics are described by a Gaussian process conditioned to the points in the parameter space where the model error is known. Xu and Valocchi [32] also represent the model error as a Gaussian process that is trained during the Bayesian inversion with spatially and temporally distributed observations. Doherty and Christensen [33] and Josset et al. [12] propose the use of regression models to predict the results of the detailed solver from the approximate solution, with the latter study making use of functional principal components analysis and dimension reduction to facilitate the analysis. Finally, Cui et al. [34] and Laloy et al. [35] assume that the model error obtained from the last detailed forward simulation during two-stage MCMC (discussed below) is a valid approximation of the model error for the current set of input parameters, and use it to correct the approximate solution before computing the likelihood. In all of this work, local error models are effectively constructed by interpolating between a limited number of model-error realizations, under the implicit assumptions that the model response surface is smooth enough to do so and that the parameter space has been adequately sampled. While this may be perfectly valid for low-dimensional inverse problems, it becomes extremely difficult in high dimensions.

Yet another means of addressing the issue of model error when using approximate forward solvers in Bayesian stochastic inversions is the two-stage MCMC approach. With this method, model errors are not explicitly accounted for, but instead are avoided altogether because the approximate solver is used only in a first accept/reject stage to prevent unpromising sets of model parameters from being tested with the computationally expensive detailed solution (e.g., [36, 37, 38]). In order to realize computational gains with this technique, the approximate solver needs to be a “good” approximation in the sense that it provides results that are relatively close to the detailed one [36]. For this reason, a number of researchers have paired the approximate solver with a local error model to improve its accuracy [34, 39, 35]. The advantage of two-stage MCMC is that the effects of model errors in the Bayesian posterior distribution can be avoided. The significant disadvantage, however, is that the computational gains of the approach may still not be enough to render the inverse problem computationally tractable since each posterior realization must still pass through the detailed forward solver, in addition to other parameter sets that have passed the first stage but are later rejected.

In this paper, we attempt to address the above-mentioned challenges and present a new methodology for dealing with model errors that is geared towards inverse problems where these errors cannot be effectively characterized globally through some parametric statistical distribution or locally based on interpolation between a small number of computed realizations. Rather than focusing on the construction of a global or local error model, we instead work towards identification of the model-error component of the residual through a projection-based approach. In this regard, pairs of approximate and detailed model runs are stored in a dictionary that grows at a specified rate during the MCMC inversion procedure. At each iteration, a local model-error basis is constructed for the current test set of model parameters using the K-nearest neighbor (KNN) entries in the dictionary, which is then used to separate the model error from the other error sources. We begin in section 2 with a brief review of Bayesian-MCMC methods followed by development of our modified approach to account for model error. We then show in section 3 the application of our methodology to three example inversions involving crosshole ground-penetrating radar (GPR) travel-time tomography, where in each case the different subsurface model parameterizations. In each example, posterior parameter distributions are compared for the cases where: (i) there is no model error present; (ii) model error is present but not accounted for; and (iii) model error is accounted for using our developed approach.

2. Methodology

2.1. Bayesian inversion using MCMC

Consider the general forward problem linking a set of observed geophysical or hydrological data $\mathbf{d}_{obs}$ to a set of subsurface model parameters of interest $\mathbf{m}_{true}$:

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

where forward operator $F(\cdot)$ contains the physics and geometry of the measurements and $\mathbf{e}_d$ is a vector of data measurement errors. The corresponding inverse problem involves estimating $\mathbf{m}_{true}$ given $\mathbf{d}_{obs}$, which requires knowledge of $F(\cdot)$ along with prior information about the model parameters. Within a probabilistic framework, this can be formulated using Bayes' theorem, whereby an initial prior model parameter distribution $p(\mathbf{m})$ is updated into a more refined posterior parameter distribution $p(\mathbf{m}|\mathbf{d}_{obs})$ taking into account the observed data (e.g., [40]). That is,

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})\, p(\mathbf{m})}{p(\mathbf{d}_{obs})}, \tag{2}$$

where $p(\mathbf{d}_{obs}|\mathbf{m})$ is the likelihood function and $p(\mathbf{d}_{obs})$, which does not depend on the model parameters, acts as a normalization constant. Assuming that the data measurement errors are independent and identically normally distributed with mean zero and standard deviation σ_d, the likelihood is multi-Gaussian and can be expressed as

$$p(\mathbf{d}_{obs}|\mathbf{m}) = \frac{1}{(2\pi\sigma_d^2)^{N/2}} \exp\left[-\frac{||\mathbf{r}(\mathbf{m})||^2}{2\sigma_d^2}\right], \tag{3}$$

where $||\cdot||$ denotes the ℓ^2-norm, N is the number of data, and

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

is the residual vector, which describes the misfit between the observed data and those predicted by applying the forward operator to parameter set $\mathbf{m}$. We see that the likelihood will be maximized for a particular set of model parameters when the ℓ^2-norm of the residual is minimized, which correponds to the case where $\mathbf{m} = \mathbf{m}_{true}$ and the parameter error defined in equation (4) is equal to zero.

Equations (2) through (4) together provide a means of calculating the posterior probability of a particular set of model parameters $\mathbf{m}$. This is commonly used within MCMC sampling procedures to quantify posterior uncertainty and thus solve the inverse problem, since performing the multi-dimensional integrations necessary to obtain the statistical moments of $p(\mathbf{m}|\mathbf{d}_{obs})$ is generally not possible. In this regard, Algorithm 1 describes a basic Metropolis-Hastings MCMC code [41, 42] that is guaranteed, after burn-in, to generate a Markov chain of samples $\{\mathbf{m}_1, ..., \mathbf{m}_k\}$ from the Bayesian posterior distribution. Starting from an initial parameter set $\mathbf{m}_1$ drawn from the prior distribution, in each iteration a new parameter set $\mathbf{m}'$ is drawn from the proposal distribution $Q(\mathbf{m}'|\mathbf{m}_i)$. The likelihood of the proposed parameter set $p(\mathbf{d}_{obs}|\mathbf{m}')$ is computed using equation (3) and the probability of accepting it is evaluated using

$$p_{acc} = \min\left\{1, \frac{p(\mathbf{m}'|\mathbf{d}_{obs})Q(\mathbf{m}_i|\mathbf{m}')}{p(\mathbf{m}_i|\mathbf{d}_{obs})Q(\mathbf{m}'|\mathbf{m}_i)}\right\}. \tag{5}$$

If the new parameter set is probabilistically accepted, it becomes the new state of the chain. Otherwise, if it is rejected, the chain remains at the last accepted parameter set.

Algorithm 1: Metropolis-Hastings MCMC

$i = 1$
draw initial model parameter set $\mathbf{m}_1$ from prior distribution $p(\mathbf{m})$
compute likelihood $p(\mathbf{d}_{obs}|\mathbf{m}_1)$ using equation (3)
while $i < i_{max}$ **do**
    draw new parameter set $\mathbf{m}'$ from proposal distribution $Q(\mathbf{m}'|\mathbf{m}_i)$
    compute likelihood $p(\mathbf{d}_{obs}|\mathbf{m}')$ using equation (3)
    compute acceptance probability p_{acc} using equation (5)
    generate random number $u \sim \mathcal{U}(0, 1)$
    $i = i + 1$
    if $u \leq p_{acc}$ **then**
        $\mathbf{m}_i = \mathbf{m}'$
    else
        $\mathbf{m}_i = \mathbf{m}_{i-1}$
    end
end

2.2. Accounting for model error

Employing approximate forward solvers $\hat{F}(\cdot)$ in Bayesian-MCMC inversions in place of the true or detailed forward operator $F(\cdot)$ introduces model error which, as mentioned earlier, has the potential to strongly bias posterior statistics if not accounted for. In this case, the residual is given by the following equation:

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

where we see that the additional model-error component means that $||\mathbf{r}(\mathbf{m})||$ will not necessarily be minimized when $\mathbf{m} = \mathbf{m}_{true}$, and that feasible sets of model parameters may be mapped to extremely low likelihoods if equation (3) is directly employed. To address this issue, researchers have typically used small numbers of detailed and approximate model pairs to develop global or local error models, as described previously. However, for many inverse problems in geophysics and hydrology involving spatially distributed model parameters, non-linear forward solvers, and/or large numbers of data: (i) the model-error distribution will be too complex to characterize globally in a meaningful way using parametric statistical distributions; and (ii) the size of the model-parameter space combined with the variability of the response surface will not be conducive to effective error model development based on regression/interpolation techniques.

To overcome these challenges, we seek in this work to develop a strategy for dealing with model errors that does not depend on their accurate statistical characterization or the construction of an error model, but rather focuses on identification of the model-error component of the residual during MCMC such that it can be subtracted prior to calculation of the likelihood using equation (3). To this end, in each MCMC iteration, we use a small number of model-error realizations, all corresponding to points in the model-parameter space that are close to the parameter set being tested $\mathbf{m}'$, to build an orthogonal basis for the model error. The model-error realizations come from a dictionary that is constructed during the inversion procedure and grows over time at a specified rate as the iterations proceed. We assume that this basis, which is local as it represents the span of the KNN points to $\mathbf{m}'$, can be used to approximate the model error at $\mathbf{m}'$. At the same time, we assume that the other components of the residual at $\mathbf{m}'$, namely the parameter and data-measurement errors, cannot be well represented by the model-error basis and lie largely orthogonal to it. As a result, under these assumptions, projection of the residual onto the basis yields an estimate of the model error.

Algorithm 2 shows the steps involved in our modified MCMC procedure to generate samples from the Bayesian posterior distribution in the presence of model error. The algorithm is the same as the standard Metropolis-Hastings approach presented in Algorithm 1 with the exception of two important additions: (i) a new function `likelihood` on lines 25-33 to compute the likelihood of the proposed set of model parameters $\mathbf{m}'$ with a correction for model error, which replaces its direct computation on line 6 using equation (3); and (ii) code on lines 15–23 to build and grow the model-parameter and corresponding model-error dictionaries $\mathbf{M}_\delta$ and $\mathbf{E}_\delta$, respectively, which are used by function `likelihood` to construct the local model-error basis. To reflect these additions, new inputs required by the code are K, the number of nearest-neighbor points to consider when creating the basis, and p_{dict}, the probability during each MCMC iteration of running the detailed forward solver and adding the model parameter set and corresponding model-error realization to $\mathbf{M}_\delta$ and $\mathbf{E}_\delta$.

Algorithm 2: Modified Metropolis-Hastings MCMC to account for model error

$i = 1$, $\delta = K$
draw initial model parameter set $\mathbf{m}_1$ from prior distribution $p(\mathbf{m})$
compute $p(\mathbf{d}_{obs}|\mathbf{m}_1) =$ `likelihood(`$\mathbf{m}_1$, $\mathbf{d}_{obs}$, $\mathbf{M}_\delta$, $\mathbf{E}_\delta$`)`
while $i < i_{max}$ **do**
  draw new parameter set $\mathbf{m}'$ from proposal distribution $Q(\mathbf{m}'|\mathbf{m}_i)$
  compute $p(\mathbf{d}_{obs}|\mathbf{m}') =$ `likelihood(`$\mathbf{m}'$, $\mathbf{d}_{obs}$, $\mathbf{M}_\delta$, $\mathbf{E}_\delta$`)`
  compute acceptance probability p_{acc} using equation (5)
  generate random number $u \sim \mathcal{U}(0, 1)$
  $i = i + 1$
  if $u \leq p_{acc}$ **then**
    $\mathbf{m}_i = \mathbf{m}'$
  else
    $\mathbf{m}_i = \mathbf{m}_{i-1}$
  end
  generate random number $v \sim \mathcal{U}(0, 1)$
  if $v \leq p_{dict}$ **then**
    $\delta = \delta + 1$
    set $\mathbf{m}^*_\delta = \mathbf{m}'$
    compute model error $\mathbf{e}(\mathbf{m}^*_\delta) = \hat{F}(\mathbf{m}^*_\delta) - F(\mathbf{m}^*_\delta)$
    add $\mathbf{m}^*_\delta$ to model parameter dictionary $\mathbf{M}_\delta = \{\mathbf{m}^*_1, ..., \mathbf{m}^*_\delta\}$
    add $\mathbf{e}(\mathbf{m}^*_\delta)$ to model error dictionary $\mathbf{E}_\delta = \{\mathbf{e}(\mathbf{m}^*_1), ..., \mathbf{e}(\mathbf{m}^*_\delta)\}$
    recompute $p(\mathbf{d}_{obs}|\mathbf{m}_i) =$ `likelihood(`$\mathbf{m}_i$, $\mathbf{d}_{obs}$, $\mathbf{M}_\delta$, $\mathbf{E}_\delta$`)`
  end
end

function `likelihood(`$\mathbf{m}$, $\mathbf{d}_{obs}$, $\mathbf{M}_\delta$, $\mathbf{E}_\delta$`)`
  search dictionary $\mathbf{M}_\delta$ for K-nearest neighbors to $\mathbf{m}$
  take K corresponding model error realizations from $\mathbf{E}_\delta$ and place in set $\mathbf{E}_K(\mathbf{m})$
  build orthonormal basis $\mathbf{B}$ having span$\{\mathbf{E}_K(\mathbf{m})\}$
  compute residual $\mathbf{r}(\mathbf{m}) = \hat{F}(\mathbf{m}) - \mathbf{d}_{obs}$
  project $\mathbf{r}(\mathbf{m})$ onto $\mathbf{B}$ to estimate model error $\tilde{\mathbf{e}}(\mathbf{m}) = \mathbf{B} \cdot \mathbf{B}^T \cdot \mathbf{r}(\mathbf{m})$
  subtract estimated model error $\tilde{\mathbf{r}}(\mathbf{m}) = \mathbf{r}(\mathbf{m}) - \tilde{\mathbf{e}}(\mathbf{m})$
  compute likelihood $p(\mathbf{d}_{obs}|\mathbf{m})$ using equation (3) and replacing $\mathbf{r}(\mathbf{m})$ with $\tilde{\mathbf{r}}(\mathbf{m})$
return

With respect to addition (i) above, the modified likelihood computation for some generic model parameter set $\mathbf{m}$ proceeds as follows. First, the current model-parameter dictionary $\mathbf{M}_\delta$ is searched for the KNN parameter sets to $\mathbf{m}$, which are determined using a standard Euclidean distance measure (e.g., [43]).

Next, the K corresponding entries from the model-error-realization dictionary $\mathbf{E}_\delta$ are placed into the set $\mathbf{E}_K(\mathbf{m})$ and used to build an orthonormal basis $\mathbf{B}$ for the model error at $\mathbf{m}$ such that $\mathrm{span}\{\mathbf{B}\} = \mathrm{span}\{\mathbf{E}_K(\mathbf{m})\}$. We accomplish this using the Gram-Schmidt procedure. Assuming that the data-measurement and parameter-error components of the residual at $\mathbf{m}$ cannot be represented by, and indeed lie orthogonal to, this basis, the model error $\tilde{\mathbf{e}}(\mathbf{m})$ can then be estimated by projecting $\mathbf{r}(\mathbf{m})$ from equation (6) onto $\mathbf{B}$. That is,

$$\tilde{\mathbf{e}}(\mathbf{m}) = \mathbf{B} \cdot \mathbf{B}^T \cdot \mathbf{r}(\mathbf{m}). \tag{7}$$

Finally, the estimated model error is subtracted from the residual to yield remainder

$$\tilde{\mathbf{r}}(\mathbf{m}) = \mathbf{r}(\mathbf{m}) - \tilde{\mathbf{e}}(\mathbf{m}), \tag{8}$$

which is now largely suitable for calculation of the likelihood using equation (3) assuming independent and identically normally distributed data-measurement errors.

With respect to addition (ii) on lines 15–23 of Algorithm 2, parameter p_{dict} controls how often the detailed forward solver is run during MCMC in order to grow the model-parameter and model-error dictionaries $\mathbf{M}_\delta$ and $\mathbf{E}_\delta$, where δ denotes the current number of entries. Before starting the inversion procedure, these dictionaries are set to contain K entries consisting of unrealistically large values for the model parameters and zero values for the model-error realizations. This ensures that the KNN search in function `likelihood` can be performed; however it means that the estimated model error in the first few iterations of our procedure will be zero and thus that the returned likelihood is given by equation (3). As the MCMC iterations continue, the option to perform a dictionary update will be periodically accepted, whereby the detailed forward solver will be run alongside the approximate solver and $\mathbf{M}_\delta$ and $\mathbf{E}_\delta$ will be augmented with entries around the current state of the Markov chain. As a result, these dictionaries will become increasingly representative of the local model error, and the capacity of the computed orthogonal basis to identify the model-error component of the residual will improve over time. It is important to note that a critical step in the dictionary enrichment part of Algorithm 2 is line 22 where, after a dictionary update is performed, the likelihood is recomputed for the current state of the Markov chain. This step is necessary to maintain a consistent use of the same dictionary while estimating the acceptance probability at the subsequent steps.

Following the procedure described above, we are able to effectively reduce posterior bias due to model error using a limited number of detailed forward solver runs. In the initial stages of our algorithm when the model error is difficult to characterize due to a small number of dictionary entries, a relatively large portion of the data mismatch tends to be removed from the residual by the projection procedure, which leads to greater exploration of the model parameter space and avoids early convergence to a biased posterior distribution. As the MCMC iterations proceed and the dictionaries grow, the model error becomes more effectively identified and the algorithm begins to sample from the bias-free posterior distribution. The success of our procedure does, however, hinge on the validity of the assumptions that (i) the model error can eventually be effectively represented by the KNN-derived basis; and (ii) parameter and data-measurement errors lie orthogonal to this basis. With regard to (i), it is reasonable to think that a basis derived from nearest-neighbor model-error realizations should include in its span the model error at the current point. With regard to (ii), it is highly unlikely that the model-error basis functions, which tend to possess a high degree of spatial correlation, are capable of representing random data-measurement errors. Thus these errors tend to be largely attenuated through projection of the residual onto $\mathbf{B}$. In the case of parameter errors, our experience with the algorithm suggests that, although it cannot be proven that these errors should lie orthogonal to the model error, they usually possess vastly different spatial characteristics and are not well captured by the basis. Nevertheless, there may exist situations where some or all of the parameter error strongly resembles the model error and thus may be identified as such in projecting onto $\mathbf{B}$. In these cases, the parameter error will be subtracted along with the model error from the residual, resulting in the corresponding parameter set being more likely to be accepted in MCMC. The algorithm will therefore deliver broadened posterior distributions to reflect the fact that the model error cannot be distinguished from parameter error.

3. Application to crosshole GPR tomography

3.1. Experimental setup and forward solvers

To demonstrate the above presented model-error approach, we now apply it to several crosshole GPR tomographic examples. Crosshole GPR travel-time tomography is a popular technique in near-surface geophysical and hydrological studies whereby the travel times of radar energy between a transmitter and receiver antenna, located at various depths in two adjacent boreholes, are used to estimate the spatial distribution of radar wave velocity between the holes. The latter quantity is strongly related to soil water content, meaning that the method provides estimates of porosity below the water table and information on soil texture and water retention characteristics in the unsaturated zone.

Because the crosshole GPR travel-time inverse problem is relatively straightforward but at the same time represents a challenging test case involving spatial distributions of subsurface model parameters, is has been popular in previous stochastic inverse studies (e.g., [44, 13, 10, 21]). Here, it is of particular interest because of the variety of methods with which the forward problem can be solved, each representing a different degree of accuracy and computational speed. The most precise and computationally expensive method of determining the travel time of radar energy between the transmitter and receiver antennas, for example, involves wave propagation modeling based on Maxwell's equations, where the first-arrival times are picked from the output waveforms. Assuming that wave propagation can be adequately described by ray theory, the eikonal equation (e.g., [45]) delivers a less accurate but orders-of-magnitude cheaper solution to the travel-time computation problem, whereby the path of the first-arriving energy depends on the subsurface GPR velocity distribution but the effects of frequency are ignored. Going even further, we can also assume that the ray paths are straight lines connecting the transmitter and receiver antennas (e.g., [46]). The latter straight-ray approximation is strictly valid only in the case of a homogeneous subsurface; however it is commonly employed when velocity contrasts are less than 10%.

For all of the inversions considered in this paper, we consider an experimental configuration involving two boreholes 4 m apart and 8 m deep with transmitter and receiver positions distributed equally every 0.2 m in the left and right boreholes, respectively. Consideration of energy traveling between every combination of transmitter and receiver location leads to 1600 travel-time data. To keep the inverse problem as straightforward as possible, we focus on the estimation of subsurface slowness (inverse of velocity) rather than velocity itself, meaning that the straight-ray problem is linear. The eikonal equation serves as our detailed forward solver $F(\cdot)$ and is used to generate the "true" travel-time data for each considered example. Gaussian random noise with a standard deviation equal to $\sigma_d = 0.2$ ns is added to these data to simulate the effects of measurement errors. The straight-ray solution serves as our approximate forward model $\hat{F}(\cdot)$, which is utilized as a "cheap" alternative to the eikonal equation in the MCMC inversion procedure. Note that our choice of detailed and approximate solvers in this paper was largely made to keep computational costs reasonable for testing purposes, and importantly to allow results to be obtained for the case where there is no model error. That is, had we chosen full-waveform simulation as the detailed forward model in our examples, it would not have been possible to compare the results of our algorithm with those for the case where this model is used within standard Metropolis-Hastings MCMC.

Figure 1a shows an example subsurface slowness field for which the corresponding first-arrival GPR travel-time data, calculated using the approximate straight-ray solution and detailed eikonal equation solution, are shown in Figure 1b and 1c, respectively. The latter are visualized as a function of the trans-

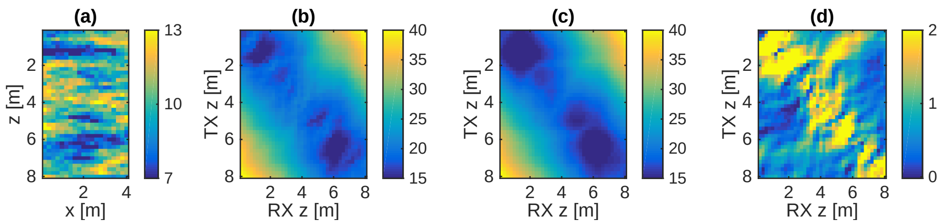

Figure 1: (a) Example GPR subsurface slowness field [ns/m]. (b) Corresponding straight-ray (approximate model) travel times [ns] plotted as a function of the transmitter (TX) and receiver (RX) antenna depths. (c) Corresponding curved-ray (detailed model) travel times [ns]; (d) Model error [ns] obtained by subtracting (c) from those in (b).

mitter and receiver antenna depths. The model error, being defined as the discrepancy between the approximate and detailed solutions $\hat{F}(\mathbf{m}) - F(\mathbf{m})$, is shown in Figure 1d. Note that, although the simulated data corresponding to each solver are visually similar, the differences between them, which in this case are on the order of 5% of the magnitude of the GPR travel times, can lead to significant posterior parameter bias in a Bayesian stochastic inversion. As the slowness field in Figure 1a has a greater correlation length in the horizontal than in the vertical direction, the largest errors are seen to occur for horizontal raypaths corresponding to the main diagonal in Figure 1d.

3.2. Model parameterization and priors

For our example inversions, we consider three different means of parameterizing the GPR slowness field between the boreholes, which leads to inverse problems of varying degrees of field complexity with different numbers of model parameters to be estimated. In the first example, we consider a simple subsurface environment consisting of 5 homogeneous horizontal layers with layer-interface positions fixed at 1, 4, 5, and 7 m. The inverse problem consists of estimating the 5 layer slowness values with the interface positions assumed known. Flat priors between 5 ns/m and 15 ns/m are prescribed for each slowness value.

Figure 2a shows three random slowness realizations that were generated from the prior for this example. In Figure 3a, the corresponding model-error realizations are shown. We see that, overall, the model error is close to zero with the exception of a few large errors located near the main diagonal of each

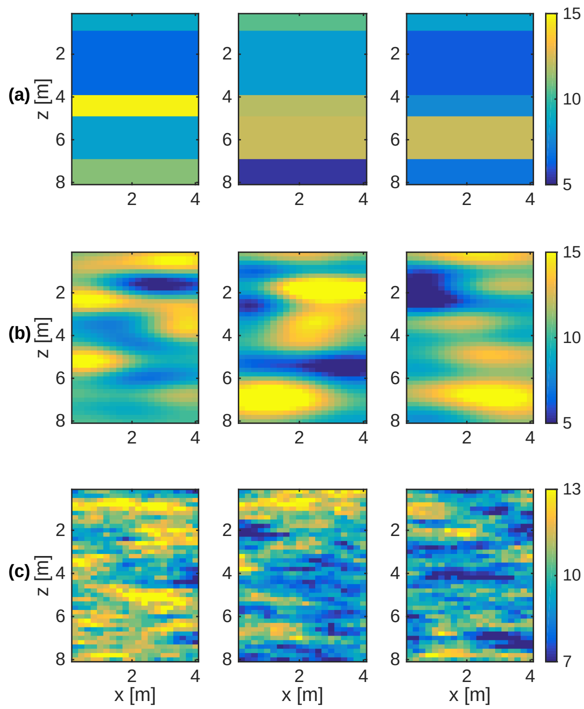

Figure 2: Example GPR slowness fields [ns/m] generated from the Bayesian prior distribution for (a) 5-layer, (b) 20-KLE-weight, and (c) 20×40 pixel-based parameterizations.

image, the latter of which correspond to transmitter and receiver positions at approximately the same depth and located close to layer interfaces across which there is a large change in slowness. This is to be expected because, at these locations, the eikonal equation will allow first-arriving energy to do most of its travel through low-slowness (high-velocity) layers, whereas the straight-ray solution forces this energy to pass through high-slowness (low-velocity) layers.

In the second example inversion, we allow for variability in both the horizontal and vertical directions by considering that the GPR slowness field is parameterized using a truncated Karhunen-Loève expansion (KLE). The truncated KLE has been utilized extensively in stochastic inverse studies to efficiently represent Gaussian random fields using a small number of parameters

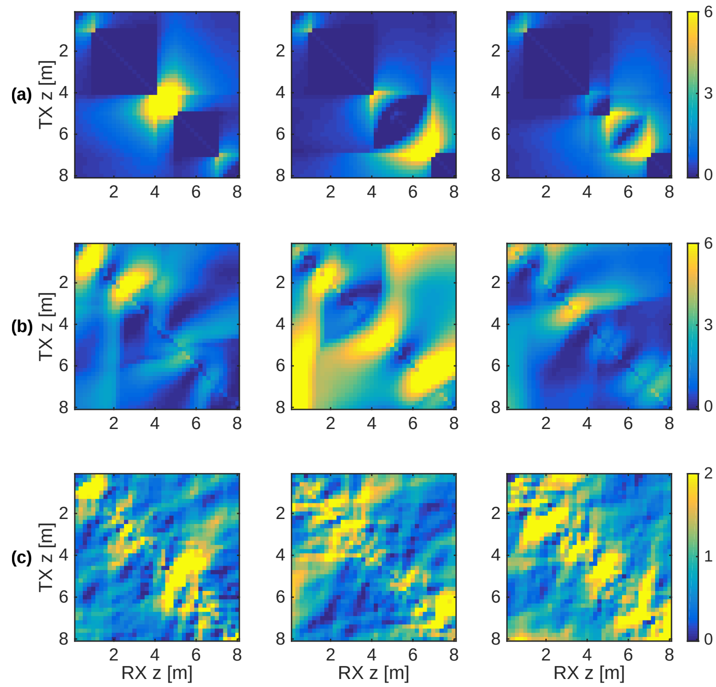

Figure 3: Travel-time model-error [ns] corresponding to the GPR slowness fields in Figure 2.

(e.g., [47, 48, 37, 49, 35]). In two spatial dimensions, it can be expressed as

$$S(x,z) = \mu(x,z) + \sum_{i=1}^{M} m_i \sqrt{\lambda_i}\, \varphi_i(x,z), \tag{9}$$

where $S(x,z)$ is the random field, $\mu(x,z)$ is its mean function, m_i are a series of independent standard normal variables, and λ_i and $\varphi_i(x,z)$ are the eigenvalues and eigenfunctions of the field's autocovariance kernel, respectively, which have been sorted in decreasing order according to the eigenvalues. Only the first M terms of the infinite KLE sum are retained in equation (9), meaning that $S(x,z)$ provides a smooth approximation to the underlying Gaussian random field that improves as the dimension M increases. In our case, the truncation limit is set to $M = 20$, meaning that 20 coefficients $\{m_1, ..., m_{20}\}$ parameterize the slowness distribution and represent the target of the inversion procedure. The prior distribution for these coefficients is Gaussian with mean zero and covariance equal to the identity matrix. For the autocovariance kernel, a squared exponential model is assumed having standard deviation equal to 4 ns/m, and horizontal and vertical correlation lengths equal to 0.8 m and 0.3 m, respectively. The mean slowness value is set equal to 10 ns/m. The domain between the boreholes is discretized using $\Delta x = \Delta z = 0.2$ m.

Figure 2b shows three random subsurface slowness fields that were generated from the prior for this example, whereas Figure 3b shows the corresponding model-error realizations. Again, we see that large model errors predominantly occur close to the main diagonal in each image, where the transmitter and receiver are located at the same depth and close to regions having a strong slowness contrast. In comparison with Figure 3a, however, note that the lack of interface constraints in this case means that the errors can occur anywhere along this diagonal. The 2-D nature of the heterogeneity also means that model errors are possible in other parts of the image space as well.

Although the truncated KLE allows for efficient parameterization of Gaussian random fields, it leads to overly smooth representations that are still far from reality. To incorporate more realism into our final inversion example, we consider a pixel-based parameterization of the subsurface whereby the domain between the boreholes is discretized into 20×40 constant-slowness square cells having side length 0.2 m, yielding 800 model parameters to be estimated. For this example, an exponential autocovariance kernel is assumed having standard deviation equal to 1.7 ns/m, and horizontal and vertical correlation lengths equal to 6 m and 1.5 m, respectively. The mean slowness is again set to 10 ns/m.

Figure 2c shows three random slowness fields generated from the Gaussian prior for the pixel-based parameterization case using the sequential Gaussian simulation code from the GSLIB software package [50]. The fields show many small-scale heterogeneities compared with those generated using the truncated KLE in Figure 2b, and are clearly more geologically plausible subsurface representations. In the corresponding model-error realizations in Figure 3c, we observe a correspondingly greater amount of small-scale variation compared to Figures 3a and 3b. Again, however, the model errors tend to be concentrated near the diagonal of these images.

All of the model-error realizations presented in Figure 3 exhibit structures that are highly correlated in the data space. Quite importantly, the error realizations are also non-Gaussian-distributed, meaning that attempts to deal with these errors as Gaussian in the inversion procedure will lead to an incorrect quantification of posterior uncertainty. To see this latter point, we generated 10,000 model error realizations for each parameterization example. For each combination of transmitter and receiver position, a quantile-quantile (Q-Q) plot was created, comparing the model-error distribution at that location with a standard normal distribution. Figure 4 shows the Q-Q plots for five data locations chosen completely at random. We observe that, for each example parameterization, the model error is strongly non-Gaussian and cannot even be roughly approximated using simple Gaussian statistics.

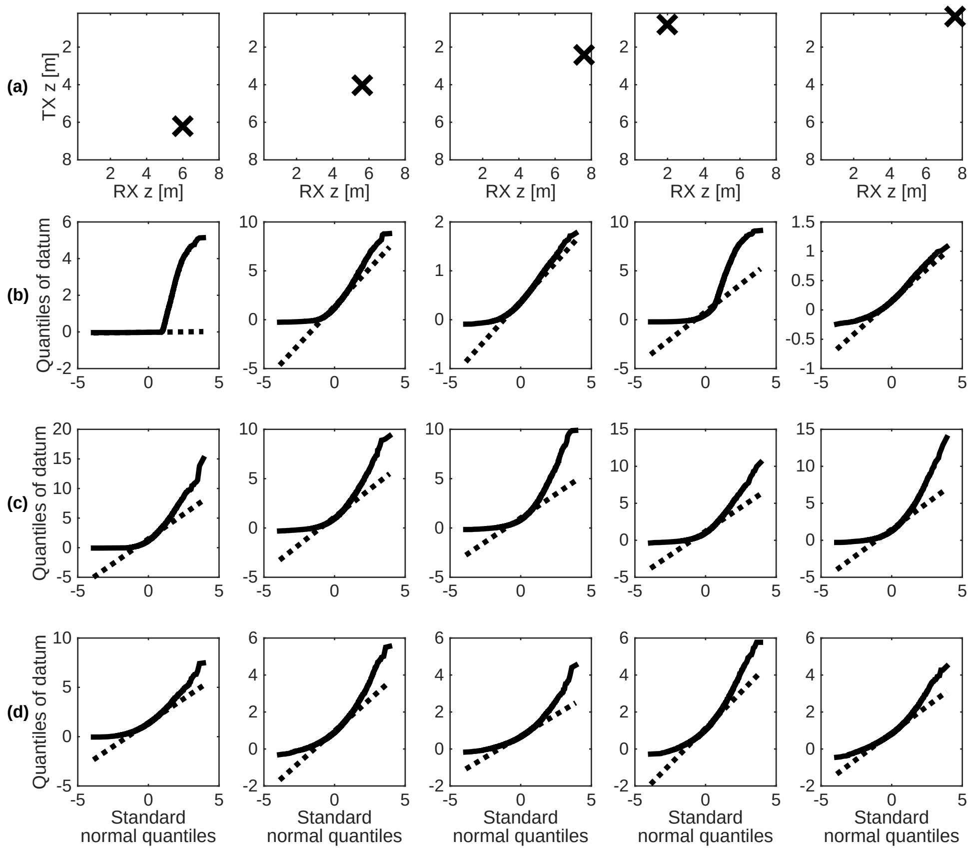

Figure 4: (a) Transmitter (TX) and receiver (RX) antennas positions of five randomly selected travel-time data. (b-d) Quantile-quantile plots (solid lines) of the model-error distribution at these locations compared with a standard normal distribution for the (b) 5-layer, (c) 20-KLE-weight, and (d) 20×40 pixel-based parameterizations. The dotted lines show the relationship to be expected if the model errors were Gaussian distributed.

3.3. Inversion settings and results

We present below the results of MCMC inversions for the three previously described slowness model parameterizations. For each parameterization, inversions were performed for: (i) the case of no model error, where the synthetic data were generated and inverted using the same detailed eikonal equation solver and standard Metropolis-Hastings was employed; (ii) the case where model error is present but not accounted for through the use of the standard Metropolis-Hastings approach; and (iii) the case where model error is present and accounted for using our proposed methodology. With regard to our method, 20 KNN were considered in every inversion to generate the model-error basis. This number was found to offer a good balance between having enough KNN to allow for flexibility in the basis to accurately represent the model error for the proposed set of parameters in MCMC, and not having too many KNN such that the basis was capable of representing other error sources in the residual. Parameter p_{dict}, which again controls the frequency with which the detailed forward solver is run to augment the model-error dictionary, was set in each inversion to 0.1% for the first 40,000 iterations, after which it was gradually reduced to a value of 0.005% after 100,000 iterations. This ensured that, at the beginning of the algorithm, focus was placed on building the model error dictionary, whereas in later iterations the detailed forward model was run less frequently to minimize computational costs. For an inversion involving 600,000 iterations, this meant that only approximately 100 complex model runs were required. Each example parameterization outlined in Section 3.2 requires a specific proposal mechanism in MCMC which is presented in the following subsections along with the inversion results.

3.3.1. Layered parameterization

For the layered subsurface example, a simple uniform proposal mechanism was used to generate new models to be tested in each MCMC iteration. This is

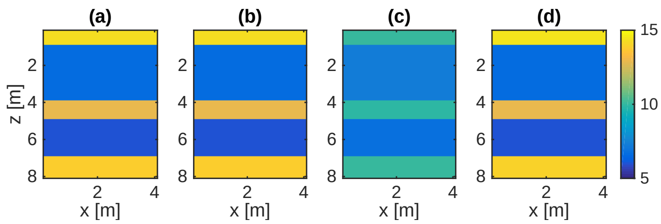

Figure 5: (a) "True" GPR slowness field [ns/m] for the 5-layer parameterization test case. (b-d) Most probable slowness fields obtained from the suite of posterior MCMC realizations when (b) there is no model error; (c) model error is present but not accounted for; and (d) model error is present and accounted for using our proposed methodology.

given by

$$\mathbf{m}' = \mathbf{m}_i + \beta\xi, \tag{10}$$

where $\mathbf{m}'$ is the proposed set of model parameters, $\mathbf{m}_i$ is the current state of the Markov chain, β is a scaling coefficient that determines the proposal width, and ξ is a vector of independent uniform random numbers drawn from $\mathcal{U}(-0.5, 0.5)$. We chose $\beta = 0.05$ for each inversion, which provided a model acceptance rate of approximately 30%. A total of 600,000 iterations were run in each case, from which the first 50,000 iterations were discarded as burn-in and the remaining samples were used to generate the posterior results.

Figure 5a shows the "true" subsurface slowness field that was used to generate the synthetic travel-time data for the 5-layer parameterization case. In Figure 5b-d, the most probable slowness fields obtained from the suite of posterior MCMC realizations are shown for the cases where (i) there is no model error, (ii) model error is present but not accounted for, and (iii) model error is present and accounted for, respectively. Figure 6, on the other hand, shows the marginal posterior parameter distributions obtained from the MCMC results for these three cases, along with the flat prior distributions for reference.

We observe in Figures 5 and 6 that, in the case where model error is not present and the only contribution to the residual is therefore data measurement error, the most probable slowness field resembles the truth and the posterior distributions are focused on the true parameter set, as could be expected. Con-

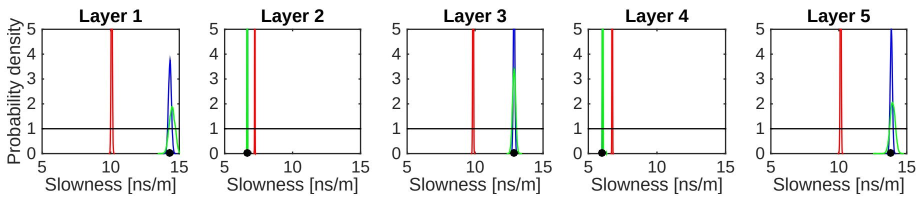

Figure 6: Prior (black) and posterior densities for the 5-layer parameterization test case when there is no model error (blue); when model error is present and not accounted for (red); and when model error is present and accounted for using our proposed methodology (green). The black dots indicate the true parameter values.

versely, when model error is present but disregarded, the posterior distributions are biased and overconfident, and the most probable slowness field deviates significantly from the truth. In this latter case, the inverted model parameters are compensating for the model error and conclusions based on the results will be misleading. Employing the model-error approach presented in Section 2.2, we see that the bias is reduced significantly and the most probable slowness field is again close to the true configuration. Note, however, that the posterior distributions are slightly broader than in the case when there is no model error, which is not surprising as some amount of parameter error may be captured by the model-error basis during the inversion procedure.

3.3.2. KLE parameterization

The increased dimensionality in representing the subsurface with a series of truncated KLE coefficients instead of layer slowness values requires, in general, more iterations in order to obtain independent samples in MCMC. The preconditioned Crank-Nicolson (pCN) technique [51, 52] allows for sampling that is robust with respect to dimension and can make MCMC considerably more efficient. Another approach for increasing efficiency is the adaptive MCMC technique [4], whereby posterior information gained from previous MCMC iterations is gradually introduced into the proposal mechanism. For the KLE parameterization example, we implemented the dimension-independent adaptive Metropolis (DIAM) MCMC algorithm proposed by Chen et al. [53], where

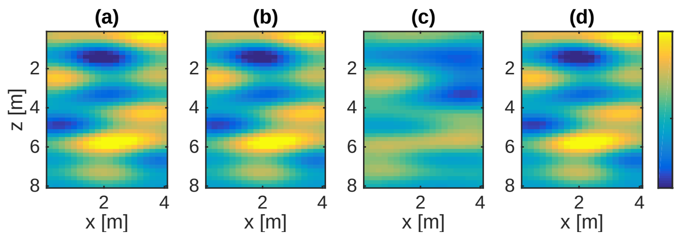

Figure 7: (a) "True" GPR slowness field [ns/m] for the 20-KLE-weight parameterization test case. (b-d) Most probable slowness fields obtained from the suite of posterior MCMC realizations when (b) there is no model error; (c) model error is present but not accounted for; and (d) model error is present and accounted for using our proposed methodology.

the proposal mechanism is described by

$$\mathbf{m}' = \bar{\mathbf{m}} + \sqrt{(1-\beta^2)}(\mathbf{m}_i - \bar{\mathbf{m}}) + \beta\xi, \tag{11}$$

where β is again a scaling coefficient that determines the proposal width, ξ is a vector of normally distributed random numbers drawn from $\mathcal{N}(0, \mathbf{C})$, and $\bar{\mathbf{m}}$ and $\mathbf{C}$ are the proposal mean and covariance matrix, respectively, defined as

$$\bar{\mathbf{m}} = (1-\epsilon)\bar{\mathbf{m}}_{post} + \epsilon\bar{\mathbf{m}}_{prior} \tag{12}$$

$$\mathbf{C} = (1-\epsilon)\mathbf{C}_{post} + \epsilon\mathbf{C}_{prior}. \tag{13}$$

Here, $\bar{\mathbf{m}}_{prior}$ and $\mathbf{C}_{prior}$ represent the prior mean and covariance, and $\bar{\mathbf{m}}_{post}$ and $\mathbf{C}_{post}$ are the corresponding posterior quantities that are estimated from the sample history. The latter were updated in our inversions every 1000 iterations, as suggested by Haario et al. [4]. We set factor ϵ to gradually decrease after 10,000 iterations from 1 to 0.5 in order to lead the proposal distribution from the prior towards the posterior. The proposal width was chosen to be $\beta = 0.01$, which yielded a model acceptance rate of around 30%. Employing the DIAM approach resulted in an order-of-magnitude decrease in the autocorrelation of the parameter history compared to standard Metropolis-Hastings. A total of 700,000 iterations were carried out for each inversion, with the first 100,000 iterations discarded as burn-in.

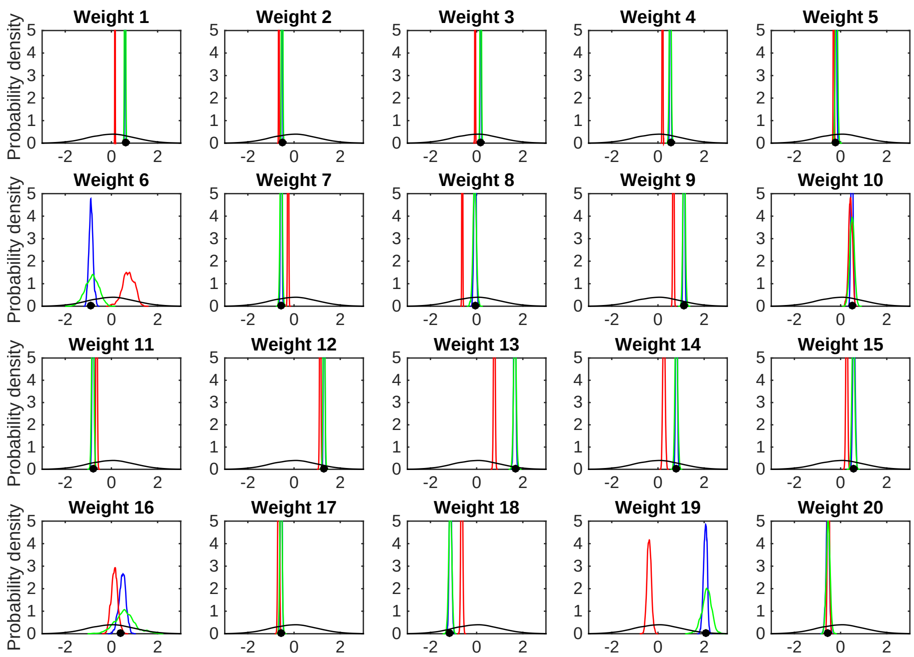

Figure 8: Prior (black) and posterior densities for the 20-KLE-weight parameterization test case when there is no model error (blue); when model error is present and not accounted for (red); and when model error is present and accounted for using our proposed methodology (green). The black dots indicate the true parameter values.

Figure 7a shows the subsurface slowness field that was used to generate the synthetic travel-time data for the 20-KLE-coefficient parameterization example, whereas Figure 7b-d present the most probable slowness fields for the three different inversion cases. In Figure 8 we show the corresponding marginal posterior parameter distributions. In accordance with what was observed previously we see that, for the case of no model error, the most probable slowness field and posterior distributions reflect very well the truth. When model error is present but disregarded, however, the posterior distributions become strongly biased and the most probable slowness field deviates significantly from the true configuration. Applying the model-error approach developed in this paper, we are able to remove this bias and better identify the true slowness configuration, again at the expense of slightly broadened distributions.

3.3.3. Pixel-based parameterization

Pixel-based parameterizations introduce additional complications into the inversion process as the dimension of the problem can be extremely large depending on the chosen discretization. One means of alleviating this issue involves introducing geostatistical prior information into the MCMC proposal mechanism, thereby reducing the number of potential model configurations to be tested. In this regard, sequential geostatistical resampling (SGR) operates by perturbing a small number of randomly chosen pixels at each MCMC iteration, where the pixel values are simulated conditional to the values at the surrounding (fixed) points assuming a prior geostatistical model. SGR has been successfully employed in a variety of spatially distributed geophysical and hydrological inverse problems to date (e.g., [54, 55, 56, 57, 1]). For further theoretical details and practical information on its implementation, please refer to these references. Here, we chose to resample a randomly selected block of 2×2 pixels in each MCMC iteration, which again yielded a model acceptance rate of approximately 30%. A total of 100,000 iterations were run in each inversion for this example. Note that, although this number is certainly not enough to provide a sufficient number of independent samples for accurate posterior inference (e.g., [1]), it

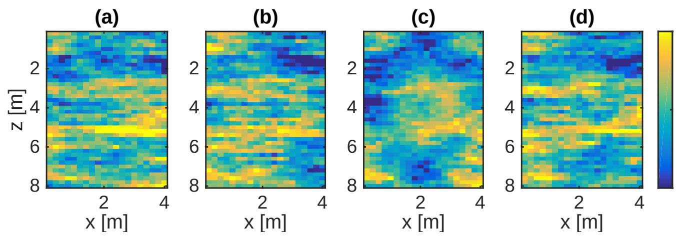

Figure 9: (a) "True" GPR slowness field [ns/m] for the 20×40 pixel-based parameterization test case. (b-d) Most probable slowness fields obtained from the suite of posterior MCMC realizations when (b) there is no model error; (c) model error is present but not accounted for; and (d) model error is present and accounted for using our proposed methodology.

importantly allows us to evaluate whether our model-error approach can be effectively employed in such a high-dimensional inverse problem.

Because of the high-dimension of the model parameter space, it is not practical to present posterior distributions for this example. As a result, in Figure 9 we show only the true subsurface slowness field along with the three best-fitting slowness fields obtained from the posterior MCMC realizations for the cases of (i) no model error; (ii) model error present but disregarded; and (ii) model error present and accounted for using our approach. Again, we see that the presence of model error leads to significant errors in the identified subsurface structures, as the model parameters are attempting to account for the model error through their spatial distribution. Applying the developed model-error approach reduces the posterior bias and the subsurface slowness field is again seen to resemble the true configuration.

4. Conclusions

We have presented in this paper a new methodology for addressing the issue of model error in Bayesian stochastic inversions that allows for a significant reduction in posterior parameter bias when using approximate forward solvers. Quite importantly, our approach is based on the identification of model-error component of the residual during MCMC, rather than on the construction of a global or local error model, the latter of which can be tremendously difficult if not impossible when dealing with high-dimensional parameter spaces and nonlinear problems. With our method, the discrepancy between the approximate and detailed forward solvers is periodically computed during the inversion procedure and the results stored in a dictionary. A local orthonormal basis is then generated in each MCMC iteration using a specified number of KNN dictionary entries, which allows us to identify and subtract the model error from the residual before computing the likelihood. The proposed methodology is highly flexible and does not depend on the model error having well defined statistical characteristics or smooth variation as a function of the input model parameters. Further, no prior information about the model error is needed before running the algorithm.

As an example, we applied our approach to the crosshole GPR travel-time tomography problem, where synthetic data were computed using the eikonal equation (detailed model) and a straight-ray assumption was made in the inversion procedure (approximate model). Using only roughly 100 detailed model calculations, the method allowed for a considerable reduction in posterior parameter bias for three different parameterizations of the subsurface slowness field: (i) 5 homogeneous horizontal layers; (ii) 20 KLE coefficients; and (iii) a grid of 20×40 pixels. For low dimensional problems it may be possible to even further reduce the computational cost by reducing the probability of enriching the model error dictionary. The choice of KNN could also be optimized in a detailed analysis that would depend on the forward solvers and parameterizations considered.

Note that, in order to identify the model-error component in the residual with our method, we make the important assumption that it lies largely orthogonal to both data measurement noise and errors related to the wrong choice of model parameters. Although the latter condition is likely to be not fully satisfied in every iteration, experience suggests that the model- and parameter-error structures are typically distinct enough such that the model error can be adequately identified. In the worst case where this is not possible, the consequence is broadened posterior distributions that include sets of model parameters whose discrepancies cannot be distinguished from model error. Future work will include the application and testing of this methodology on other inverse problems, as well as in the context of other iterative inversion techniques such as ensemble Kalman smoothing.

Acknowledgements

This work was supported by a grant to J. Irving from the Swiss National Science Foundation.

References

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

[2] M. Sambridge, A parallel tempering algorithm for probabilistic sampling and multimodal optimization, Geophysical Journal International (2013) ggt342.

[3] J. A. Vrugt, Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation, Environmental Modelling & Software 75 (2016) 273–316.

[4] H. Haario, E. Saksman, J. Tamminen, An adaptive Metropolis algorithm, Bernoulli (2001) 223–242.

[5] A. M. Stuart, J. Voss, P. Wilberg, et al., Conditional path sampling of SDEs and the Langevin MCMC method, Communications in Mathematical Sciences 2 (2004) 685–697.

[6] T. Marshall, G. Roberts, An adaptive approach to Langevin MCMC, Statistics and Computing 22 (2012) 1041–1057.

[7] R. M. Neal, MCMC using Hamiltonian dynamics, Chapter 5 in Handbook of Markov Chain Monte Carlo, Chapman & Hall / CRC Press, 2011.

[8] K. Davis, Y. Li, Fast solution of geophysical inversion using adaptive mesh, space-filling curves and wavelet compression, Geophysical Journal International 185 (2011) 157–166.

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

[10] N. Linde, J. A. Vrugt, Distributed soil moisture from crosshole ground-penetrating radar travel times using stochastic inversion, Vadose Zone Journal 12 (2013).

[11] E. Oware, S. Moysey, T. Khan, Physically based regularization of hydrogeophysical inverse problems for improved imaging of process-driven systems, Water Resources Research 49 (2013) 6238–6247.

[12] L. Josset, D. Ginsbourger, I. Lunati, Functional error modeling for uncertainty quantification in hydrogeology, Water Resources Research 51 (2015) 1050–1068.

[13] M. Scholer, J. Irving, M. C. Looms, L. Nielsen, K. Holliger, 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 (2012).

[14] S. Arridge, J. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, M. Vauhkonen, Approximation errors and model reduction with an application in optical diffusion tomography, Inverse Problems 22 (2006) 175.

[15] D. Calvetti, O. Ernst, E. Somersalo, Dynamic updating of numerical model discrepancy using sequential sampling, Inverse Problems 30 (2014) 114019.

[16] J. Goh, D. Bingham, J. P. Holloway, M. J. Grosskopf, C. C. Kuranz, E. Rutter, Prediction and computer model calibration using outputs from multi-fidelity simulators, Technometrics 55 (2013) 501–512.

[17] S.-T. Khu, M. G. Werner, Reduction of Monte-Carlo simulation runs for uncertainty estimation in hydrological modelling, Hydrology and Earth System Sciences Discussions 7 (2003) 680–692.

[18] Y. Marzouk, D. Xiu, A stochastic collocation approach to Bayesian inference in inverse problems, Communications in Computational Physics 6 (2009) 826–847.

[19] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes in Machine Learning, the MIT Press, 2006.

[20] J. Brynjarsdóttir, A. O'Hagan, Learning about physical parameters: The importance of model discrepancy, Inverse Problems 30 (2014) 114007.

[21] T. M. Hansen, K. S. Cordua, B. H. Jacobsen, K. Mosegaard, Accounting for imperfect forward modeling in geophysical inverse problems - Exemplified for crosshole tomography, Geophysics 79 (2014) H1–H21.

[22] J. Kaipio, E. Somersalo, Statistical inverse problems: Discretization, model reduction and inverse crimes, Journal of Computational and Applied Mathematics 198 (2007) 493–504.

[23] A. Lehikoinen, J. Huttunen, S. Finsterle, M. Kowalsky, J. Kaipio, Dynamic inversion for hydrological process monitoring with electrical resistance tomography under model uncertainties, Water Resources Research 46 (2010).

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

[25] D. Del Giudice, M. Honti, A. Scheidegger, C. Albert, P. Reichert, J. Rieckermann, Improving uncertainty estimation in urban hydrological modeling by statistically describing bias, Hydrology and Earth System Sciences 17 (2013) 4209–4225.

[26] G. Schoups, J. A. Vrugt, A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors, Water Resources Research 46 (2010).

[27] T. Smith, A. Sharma, L. Marshall, R. Mehrotra, S. Sisson, Development of a formal likelihood function for improved Bayesian inference of ephemeral catchments, Water Resources Research 46 (2010).

[28] K. Beven, A. Binley, The future of distributed models: model calibration and uncertainty prediction, Hydrological processes 6 (1992) 279–298.

[29] J. A. Vrugt, M. Sadegh, Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, Water Resources Research 49 (2013) 4335–4345.

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

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

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

[33] J. Doherty, S. Christensen, Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, Water Resources Research 47 (2011).

[34] T. Cui, C. Fox, M. O'Sullivan, Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance Metropolis Hastings algorithm, Water Resources Research 47 (2011).

[35] E. Laloy, B. Rogiers, J. A. Vrugt, D. Mallants, D. Jacques, Efficient posterior exploration of a high-dimensional groundwater model from two-stage

Markov chain Monte Carlo simulation and polynomial chaos expansion, Water Resources Research 49 (2013) 2664–2682.

[36] J. A. Christen, C. Fox, Markov chain Monte Carlo using an approximation, Journal of Computational and Graphical statistics 14 (2005) 795–810.

[37] Y. Efendiev, A. Datta-Gupta, X. Ma, B. Mallick, Efficient sampling techniques for uncertainty quantification in history matching using nonlinear error models and ensemble level upscaling techniques, Water Resources Research 45 (2009).

[38] X. Ma, M. Al-Harbi, A. Datta-Gupta, Y. Efendiev, et al., An efficient two-stage sampling method for uncertainty quantification in history matching geological models, SPE Journal 13 (2008) 77–87.

[39] L. Josset, V. Demyanov, A. H. Elsheikh, I. Lunati, Accelerating Monte Carlo Markov chains with proxy and error models, Computers & Geosciences 85 (2015) 38–48.

[40] A. Tarantola, Inverse problem theory and methods for model parameter estimation, SIAM, 2005.

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

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

[43] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning, Data Mining, Inference and Prediction, Second Edition, Springer series in statistics New York, 2009.

[44] M. C. Looms, K. H. Jensen, A. Binley, L. Nielsen, Monitoring unsaturated flow and transport using cross-borehole geophysical methods, Vadose Zone Journal 7 (2008) 227–237.

[45] R. L. Nowack, Wavefronts and solutions of the eikonal equation, Geophysical Journal International 110 (1992) 55–62.

[46] K. S. Cordua, M. C. Looms, L. Nielsen, Accounting for correlated data errors during inversion of cross-borehole ground penetrating radar data, Vadose Zone Journal 7 (2008) 263–271.

[47] D. Zhang, Z. Lu, An efficient, high-order perturbation approach for flow in random porous media via Karhunen-Loève and polynomial expansions, Journal of Computational Physics 194 (2004) 773–794.

[48] P. Dostert, Y. Efendiev, T. Y. Hou, W. Luo, Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification, Journal of Computational Physics 217 (2006) 123–142.

[49] A. H. Elsheikh, M. D. Jackson, T. C. Laforce, Bayesian reservoir history matching considering model and parameter uncertainties, Mathematical Geosciences 44 (2012) 515–543.

[50] C. V. Deutsch, A. G. Journel, GSLIB: Geostatistical Software Library and User's Guide, Oxford University Press, 1992.

[51] S. L. Cotter, G. O. Roberts, A. M. Stuart, D. White, et al., MCMC methods for functions: modifying old algorithms to make them faster, Statistical Science 28 (2013) 424–446.

[52] A. Beskos, M. Girolami, S. Lan, P. E. Farrell, A. M. Stuart, Geometric MCMC for infinite-dimensional inverse problems, Journal of Computational Physics 335 (2017) 327–351.

[53] Y. Chen, D. Keyes, K. J. Law, H. Ltaief, Accelerated dimension-independent adaptive Metropolis, SIAM Journal on Scientific Computing 38 (2016) S539–S565.

[54] J. Fu, J. J. Gómez-Hernández, A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling, Mathematical Geosciences 41 (2009) 105–128.

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

[56] T. M. Hansen, K. S. Cordua, K. Mosegaard, Inverse problems with non-trivial priors: Efficient solution through sequential Gibbs sampling, Computational Geosciences 16 (2012) 593–611.

[57] K. S. Cordua, T. M. Hansen, K. Mosegaard, Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information, Geophysics 77 (2012) H19–H31.