Advances in Water Resources 32 (2009) 329–339

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Efficient uncertainty quantification techniques in inverse problems for Richards' equation using coarse-scale simulation models

P. Dostert [a,*], Y. Efendiev [b], B. Mohanty [c]

[a] Department of Mathematics, Program in Applied Mathematics, The University of Arizona, 617 N. Santa Rita Ave., Tucson, AZ 85721-0089, United States
[b] Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, United States
[c] Biological and Agricultural Engineering, Texas A&M University, College Station, TX 77843, United States

ARTICLE INFO

Article history:
Received 27 January 2008
Received in revised form 13 November 2008
Accepted 19 November 2008
Available online 3 December 2008

Keywords:
Uncertainty Quantification
MCMC
Langevin
Multiscale
Richards' equation
Sparse grid collocation
KLE
Hydraulic conductivity

ABSTRACT

This paper concerns efficient uncertainty quantification techniques in inverse problems for Richards' equation which use coarse-scale simulation models. We consider the problem of determining saturated hydraulic conductivity fields conditioned to some integrated response. We use a stochastic parameterization of the saturated hydraulic conductivity and sample using Markov chain Monte Carlo methods (MCMC). The main advantage of the method presented in this paper is the use of multiscale methods within an MCMC method based on Langevin diffusion. Additionally, we discuss techniques to combine multiscale methods with stochastic solution techniques, specifically sparse grid collocation methods. We show that the proposed algorithms dramatically reduce the computational cost associated with traditional Langevin MCMC methods while providing similar sampling performance.

1. Introduction

We consider efficient uncertainty quantification techniques in inverse problems for Richards' equation, which describes the infiltration of water into a porous media whose pore space is filled with air and water. These problems are motivated by applications to soil moisture predictions. Soil moisture conditions are important in determining the amount of infiltration and ground water recharge. Soil moisture is controlled by factors such as soil type, topography, vegetation, and climate. Soil moisture is typically measured at different scales varying from point scale (*in-situ*) to remote sensing scale (of order several kilometers). The objective is to predict the soil moisture at different resolutions via prediction of the saturated hydraulic conductivity field.

The uncertainty quantification problem considered in this paper consists of the prediction of the saturated hydraulic conductivity fields conditioned to some average dynamic data (e.g., average flux). In a probabilistic context, this problem can be regarded as conditioning the saturated conductivity fields, K_s, to the dynamic data, F, with measurement errors. One would like to sample from the conditional distribution $P(K_s|F)$. Using the Bayes formula we can write $P(K_s|F) \propto P(F|K_s)P(K_s)$, where $P(K_s)$ is the unconditioned (prior) distribution of the conductivity field. We denote the sampling target distribution as $\pi(K_s) = P(K_s|F)$.

Sampling from this distribution is difficult since the process of predicting soil moisture is nonlinear. Instead, it is possible to estimate the probability distribution from the outcomes of predictions for numerous realizations of the field. To compute a realization of the conductivity field, one needs a computation on a scale fine enough to determine the dynamic data fields accurately. Markov chain Monte Carlo (MCMC) methods have been used previously in this context (e.g., [24,3]), but the need for fine-scale computations at each step of the MCMC algorithm makes these methods very expensive. Some attempts to propose MCMC methods with high acceptance rates have been made. One such example is the randomized maximum likelihood method [27,28]. This approach uses unconditional realizations of the dynamic and conductivity data and solves a deterministic gradient-based inverse problem. The solution of this minimization problem is taken as a proposal, and is accepted with probability one, because the rigorous acceptance probability is very difficult to estimate. In addition to solving a gradient-based inverse problem, this method may not properly sample the posterior distribution. Developing efficient rigorous MCMC calculations with high acceptance rate remains a challenging problem. We note that the approaches considered in this paper are not sequential. Various sequential approaches for soil moisture assimilation are studied in the literature (e.g., [25,20]). In general,

* Corresponding author.
E-mail address: dostert@math.arizona.edu (P. Dostert).

0309-1708/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2008.11.009

sequential approaches based Ensemble Kalman Filter techniques [10,16,22] can be used for hydraulic conductivity predictions; however, these approaches may not sample multi-modal probability distributions correctly and various treatments are needed.

In this paper, sampling is performed using MCMC methods with Langevin instrumental probability distribution. Langevin algorithms generally provide high acceptance rates for MCMC problems, but they require one to compute the gradient of the target distribution. This cannot be calculated analytically; thus one must perform a realization in each stochastic dimension to determine this gradient. For large stochastic dimensions, this can be prohibitively expensive. We employ inexpensive coarse-scale simulations for the computations of the gradients of the posterior distribution. We note that Glimm and Sharp employed error models between coarse- and fine-scale simulations to quantify the uncertainty in a pioneering work [14]. Using coarse-scale simulations, the approximation for the gradient of the target distribution is computed. These coarse-scale gradients may not be very accurate, and for this reason, the computed results are first tested with coarse-scale distributions within a two-stage MCMC [5]. If the result is accepted at the first stage, then a fine-scale simulation is performed at the second stage to determine the acceptance probability.

To speed up the computations further, the approximation of the target distribution is computed based on few realizations of the permeability field. Using these realizations and coarse-scale simulations, we compute a few observation points on the posterior response surface. Furthermore, using sparse interpolation techniques, we approximate this response surface. In this approach, the gradient computations are analytical, and the entire sampling procedure can be implemented with little computational effort. We note that approaches which combine sparse collocation techniques and multiscale methods are not new. In a recent paper [11], the authors propose an approach which uses variational multiscale method as well as multiscale finite element methods to solve a stochastic parabolic equation.

Numerical results for sampling saturated conductivity fields are presented. Using the Karhunen–Loève expansion (KLE) [35,23], the high-dimensional conductivity field can be represented by a small number of parameters. The number of parameters are related to the decay in the spectrum of the covariance matrix. Samples of log-permeability can be written as a linear combination of eigenvectors of the covariance matrix times the square root of the eigenvalues. We refer to [38] (see also [36]) where more detailed discussions on KLE including some analytical results are discussed. Static data (the values of the conductivity fields at some sparse locations) are incorporated into the KLE to further reduce the dimension of the parameter space. The proposed Langevin algorithms are tested for both isotropic and anisotropic hydraulic conductivity fields. The isotropic field results in a very small dimensional parameter space, while the anisotropic field results in a significantly higher dimensional parameter space. Anisotropic fields are often more applicable to real world data and simulation, while the isotropic field allows us to better analyze the numerical methods. Numerical results are presented for Richards' equation for a single typical isotropic field and another typical anisotropic field. We show that the preconditioned coarse-scale Langevin algorithm and its interpolated variant provide similar performance to the fine-scale Langevin algorithm while using significantly less CPU time for these typical cases.

The paper is organized as follows. In the next section, we present fine-and coarse-scale equations. Section 3 is devoted to the problem setting and methodology. In Section 4, we present the numerical implementation and results.

2. Fine- and coarse-scale simulation models

In this section the fine- and coarse-scale models for Richards' equation are introduced. The formula describing Richards' equation under some assumptions is given by

$$\frac{\mathrm{d}\theta(\psi)}{\mathrm{d}t} - \mathrm{div}(K(x,\psi)\nabla(\psi + x_3)) = 0 \quad \text{in } \Omega, \tag{2.1}$$

where $\theta(\psi)$ is volumetric water content and ψ is the pressure head.

The following are assumed ([29]) for Eq. (2.1): (1) the porous medium and water are incompressible; (2) the temporal variation of the water saturation is significantly larger than the temporal variation of the water pressure; (3) air phase is infinitely mobile so that the air pressure remains constant; in this case it is atmospheric pressure which equals zero and (4) the source/sink terms are neglected.

Constitutive relations between θ and ψ and between K and ψ are developed appropriately, which consequently gives nonlinearity behavior in Eq. (2.1). The relation between the water content and pressure is referred to as the moisture retention function. The equation written in Eq. (2.1) is called the *coupled-form* of Richards' equation. In other literature this equation is also called the mixed form of Richards' equation, due to the fact that there are two variables involved in it, namely, the water content θ and the pressure head ψ. Taking advantage of the differentiability of the soil retention function, one may rewrite Eq. (2.1) as follows:

$$C(u)\frac{\mathrm{d}\psi}{\mathrm{d}t} - \mathrm{div}(K(x,\psi)\nabla(\psi + x_3)) = 0 \quad \text{in } \Omega, \tag{2.2}$$

where $C(u) = d\theta/d\psi$ is the specific moisture capacity. This version is referred to as the *head-form* (*h-form*) of Richards' equation. Another formulation of the Richards' equation is based on the water content θ

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} - \mathrm{div}D(x,\theta)\nabla\theta - \frac{\partial K}{\partial x_3} = 0 \quad \text{in } \Omega, \tag{2.3}$$

where $D(\theta) = K(\theta)/(\mathrm{d}\theta/\mathrm{d}\psi)$ defines the diffusivity. This form is called the θ-*form* of Richards' equation. The sources of nonlinearity of Richards' equation come from the moisture retention and relative hydraulic conductivity functions, $\theta(\psi)$ and $K(x,\psi)$, respectively.

Perhaps the most widely used empirical constitutive relations for the moisture content and hydraulic conductivity are due to the work of van Genuchten [32]. He proposed a method of determining the functional relation of relative hydraulic conductivity to the pressure head by using the field observation knowledge of the moisture retention. In turn, the procedure would require curve-fitting to the proposed moisture retention function with the experimental/observational data to establish certain parameters inherent to the resulting hydraulic conductivity model. There are several widely known constitutive relations such as the Haverkamp model [17] and the Exponential model [33]. In this work, we focus on the Exponential model, given by

$$\theta(\psi) = \theta_s\, \mathrm{e}^{\beta\psi}, \quad K(x,\psi) = K_s(x)\, \mathrm{e}^{\alpha\psi}. \tag{2.4}$$

The variable K_s in the above model is known as the saturated hydraulic conductivity. It has been observed that the hydraulic conductivity has a broad range of values which, together with the functional forms presented above, confirm the nonlinear behavior of the process. Furthermore, the water content and hydraulic conductivity approach zero as the pressure head goes to very large negative values. Because we are interested in mass conservative schemes, a finite volume formulation of the global problem will be used. We ignore gravity terms and impose it via pressure drop boundary conditions. The implementation of gravity effects within multiscale methods is currently under investigation.

The fine-scale model is simply defined as the finite volume formulation of the Richards' equation on a fine-scale mesh. For the coarse-scale model a multiscale finite volume element method (MsFVEM) introduced in [8], is used. MsFVEM can be applied to Richards' equation in the general form as it was shown in [8]; however, MsFVEM offers a great advantage when the nonlinearity and heterogeneity of $K(x,\psi)$ are separable, i.e.

$$K(x,\psi) = K_s(x)k_r(\psi). \tag{2.5}$$

In this case, the local basis functions become linear, and the corresponding space V_ϵ^h is a linear space. Next, we briefly describe the method. Let $\mathscr{K}^h$ denote the collection of coarse elements/rectangles K. Consider a coarse element K, and let ξ_K be its center. The element K is divided into four rectangles of equal area by connecting ξ_K to the midpoints of the element's edges. We denote these quadrilaterals by K_ξ, where $\xi \in Z_h(K)$ are the vertices of K. Also, we denote $Z_h = \bigcup_K Z_h(K)$ and $Z_h^0 \subset Z_h$, the vertices which do not lie on the Dirichlet boundary of Ω. The control volume V_ξ is defined as the union of the quadrilaterals K_ξ sharing the vertex ξ.

The key idea of the method is the construction of basis functions on the coarse grids such that these basis functions capture the small-scale information on each of these coarse grids. The method that we use follows its finite element counterpart presented in [18]. The basis functions are constructed from the solution of the leading order homogeneous elliptic equation on each coarse element with some specified boundary conditions. Thus, if we consider a coarse element K that has d vertices, the local basis functions $\phi_i, i = 1,\ldots,d$ are set to satisfy the following elliptic problem:

$$\begin{aligned} &-\nabla\cdot(k\cdot\nabla\phi_i) = 0 \quad \text{in } K \\ &\phi_i = g_i \quad \text{on } \partial K \end{aligned} \tag{2.6}$$

for some function g_i defined on the boundary of the coarse element K. Hou et al. [18] have demonstrated that a careful choice of boundary conditions would improve the accuracy of the method. In previous findings, the function g_i for each i is chosen to vary linearly along ∂K or to be the solution of the local one-dimensional problems [19]. Similarly the solution of the problem in a slightly larger domain has also been chosen to define the boundary conditions in the past. The boundary conditions for the basis functions that are used in this paper will be discussed later. We will require $\phi_i(x_j) = \delta_{ij}$. Finally, a nodal basis function associated with the vertex x_i in the domain Ω is constructed from the combination of the local basis functions that share this x_i and are zero elsewhere. We would like to note that one can use an approximate solution of Eq. (2.6) whenever possible. For example, in the case of periodic or scale separation cases, the basis functions can be approximated using homogenization expansion (see [8]). This type of simplification is not applicable for problems considered in this paper.

Now, we may formulate the finite dimensional problem. We want to find a $\psi_h \in V^h$ with $\psi_h = \sum_{z\in Z_h^0}\psi_z\phi_i$ with z being nodal points such that

$$\int_{V_z}(\theta(\eta^{\psi_h}) - \theta^{n-1})\mathrm{d}x - \Delta t\int_{\partial V_z} K_s(x)k_r(\eta^{\psi_h})\nabla\psi_h\cdot n\ \mathrm{d}l = 0 \tag{2.7}$$

for every control volume $V_z \subset \Omega$. Here, η^{ψ_h} is the average of ψ_h over the coarse block K and θ^{n-1} is the value of $\theta(\eta^{\psi_h})$ evaluated at time step $n-1$. To this equation we can directly apply a linearization procedure, as described in [12]. The details are omitted here.

One can generalize MsFVEM to problems when there is no separation of nonlinearities and spatial scales (see Eq. (2.5)). In this case, we find $\psi_h \in S^h$, where S^h is a standard finite dimensional space (e.g., piecewise linear functions) such that

$$\int_{V_z}(\theta(\eta^{\psi_h}) - \theta^{n-1})\mathrm{d}x - \Delta t\int_{\partial V_z} K(x,\eta^{\psi_h})\nabla\psi_{\text{local}}\cdot n\ \mathrm{d}l = 0 \quad \forall z \in Z_h^0, \tag{2.8}$$

where η^{ψ_h} is the average of ψ_h over the coarse block K, θ^{n-1} is the value of $\theta(\eta^{\psi_h})$ evaluated at time step $n-1$, and ψ_{local} is the solution of the local problem in K

$$\begin{aligned} &-\mathrm{div}(K(x,\eta^{\psi_h})\nabla\psi_{\text{local}}) = 0 \quad \text{in } K \in S^h, \\ &\psi_{\text{local}} = \psi_h \quad \text{on } \partial K. \end{aligned} \tag{2.9}$$

In this paper, we will consider only separable case given by the relation Eq. (2.5). We note that one can also use upscaling or multiscale techniques (e.g., [6,7,34,3,37,39,40]) in uncertainty quantification.

3. Uncertainty quantification methods

In this section, the methods used for uncertainty quantification are presented. First, the Karhunen–Loève expansion (KLE) is introduced as a method to parameterize our uncertain saturated conductivity field K_s. Next, a traditional MCMC algorithm, along with the Langevin variant are presented. Lastly, the modifications using coarse-scale models and collocation are shown.

3.1. KLE

Since problem under consideration consists of sampling the saturated conductivity field, K_s, given some dynamic data, we need some type of parameterization. We use the KLE [23,35] to expand the saturated conductivity field in terms of an optimal L^2 basis. By truncating the expansion we can represent the conductivity matrix by a small number of random parameters. First, we briefly recall the facts of the KLE. Denote $Y(x,\omega) = \log[K_s(x,\omega)]$, where the random element ω is included to remind us that K_s is a random field. For simplicity, we assume that $E[Y(x,\omega)] = 0$. Suppose $Y(x,\omega)$ is a second order stochastic process with $E\int_\Omega Y^2(x,\omega)\mathrm{d}x < \infty$, where E is the expectation operator. Given an orthonormal basis $\{\phi_k\}$ in $L^2(\Omega)$, we can expand $Y(x,\omega)$ as a general Fourier series

$$Y(x,\omega) = \sum_{k=1}^{\infty} Y_k(\omega)\phi_k(x), \quad Y_k(\omega) = \int_\Omega Y(x,\omega)\phi_k(x)\,\mathrm{d}x. \tag{3.1}$$

We are interested in the special L^2 basis $\{\phi_k\}$ which makes the random variables Y_k uncorrelated. That is, $E(Y_iY_j) = 0$ for all $i\neq j$. Denote the covariance function of Y as $R(x,y) = E[Y(x)Y(y)]$. Then such basis functions $\{\phi_k\}$ satisfy

$$E[Y_iY_j] = \int_\Omega \phi_i(x)\,\mathrm{d}x\int_\Omega R(x,y)\phi_j(y)\mathrm{d}y = 0, \quad i\neq j. \tag{3.2}$$

Since $\{\phi_k\}$ is a complete basis in $L^2(\Omega)$, it follows that $\phi_k(x)$ are eigenfunctions of $R(x,y)$:

$$\int_\Omega R(x,y)\phi_k(y)\mathrm{d}y = \lambda_k\phi_k(x), \quad k = 1,2,\ldots, \tag{3.3}$$

where $\lambda_k = E[Y_k^2] > 0$. Furthermore, we have

$$R(x,y) = \sum_{k=1}^{\infty}\lambda_k\phi_k(x)\phi_k(y). \tag{3.4}$$

Denote $\chi_k = Y_k/\sqrt{\lambda_k}$, then χ_k satisfy $E(\chi_k) = 0$ and $E(\chi_i\chi_j) = \delta_{ij}$. It follows that

$$Y(x,\omega) = \sum_{k=1}^{\infty}\sqrt{\lambda_k}\chi_k(\omega)\phi_k(x), \tag{3.5}$$

where ϕ_k and λ_k satisfy Eq. (3.3). We assume that the eigenvalues λ_k are ordered as $\lambda_1 \geqslant \lambda_2 \geqslant \ldots$. The expansion Eq. (3.5) is called the Karhunen–Loève expansion. In the KLE Eq. (3.5), the L^2 basis func-

tions $\phi_k(x)$ are deterministic and resolve the spatial dependence of the permeability field. The randomness is represented by the scalar random variables χ_k. After we discretize the domain Ω by a rectangular mesh, the continuous KLE Eq. (3.5) is reduced to finite terms. Generally, we only need to keep the leading order terms (quantified by the magnitude of λ_k) and still capture most of the energy of the stochastic process $Y(x,\omega)$. For an N-term KLE approximation $Y_N = \sum_{k=1}^{N} \sqrt{\lambda_k}\chi_k\phi_k$, define the energy ratio of the approximation as

$$e(N) := \frac{E\|Y_N\|^2}{E\|Y\|^2} = \frac{\sum_{k=1}^{N}\lambda_k}{\sum_{k=1}^{\infty}\lambda_k}. \tag{3.6}$$

If $\lambda_k, k = 1,2,\ldots,$ decay very fast, then the truncated KLE would be a good approximation of the stochastic process in the L_2 sense. An example is the conductivity field $K_s(x,\omega)$ which is a log-normal homogeneous stochastic process. Then $Y(x,\omega)$ is a Gaussian process with the covariance kernel

$$R(x,y) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|^2}{2L_1^2} - \frac{|x_2 - y_2|^2}{2L_2^2}\right) \tag{3.7}$$

and χ_k are independent standard Gaussian random variables. In the previous formula, L_1 and L_2 are the correlation lengths in each dimension, and $\sigma^2 = E(Y^2)$ is a constant.

For implementation, we first solve the eigenvalue problem Eq. (3.3) numerically on a rectangular mesh corresponding to the fine-scale model and obtain the eigenpairs $\{\lambda_k, \phi_k\}$. The fine-scale mesh is chosen such that there are at least five points per correlation lengths. This introduces a discretization error which will be ignored in our simulations. Since the eigenvalues decay fast for our examples, the truncated KLE approximates the stochastic process $Y(x,\omega)$ fairly well in the L^2 sense. Therefore, we can sample $Y(x,\omega)$ from the truncated KLE Eq. (3.5) by generating Gaussian random variables χ_k. For all of the experiments in this paper, the KLE is truncated when approximately 95% of the energy $e(N)$ is captured. We note that the sum of eigenvalues is σ^2 for our test examples. Once we have a truncated KLE, the hard constraints (the values of the conductivity at prescribed locations) are implemented. This is done by solving a simple linear system resulting in a linear subspace of our parameter space (a hyperplane) which yields the corresponding values of the conductivity field. For example, if the KLE consists of 20 terms after truncation, the saturated conductivity is known at 11 locations, then the resulting stochastic space would be nine dimensions. In general, one needs to choose 9 of χ_k randomly and solve the system of 11 equations for 11 unknowns. There are multiple choices for these 11 variables. In our simulations, we search all 11×11 matrices in 11×20 matrix with the lowest condition number. This can be easily done using QR decomposition.

3.2. Description of dynamic data

The dynamic data under consideration is given by the average fluxes at given locations and times, which is denoted by $F(x,t)$, with x being the location where the average flux is measured and t is the time. This can be thought of as an average soil moisture measurement which is measured in practice. Note that this method is not limited to this type of measurement, and various measurements can be used in a straightforward way within the proposed techniques. For example, hydraulic head data can be used instead. In our simulations, average fluxes are used because it is more difficult for multiscale methods to accurately capture the average fluxes than the hydraulic head. The latter is due to the fact that the fluxes are less smooth compared to the hydraulic head. Typically, the prior information about the saturated conductivity field consists of its covariance matrix and the values of the saturated conductivity at some sparse locations. Since the average flux values provide an integrated response, the map from the saturated conductivity field to the fluxes is not one-to-one. Hence this problem is ill-posed in the sense that there exist many different realizations for the given data.

3.3. MCMC methods

From the probabilistic point of view, our problem can be regarded as conditioning the saturated conductivity fields to the flux data with measurement errors. Consequently, our goal is to sample from the conditional distribution $P(K_s|F)$, where K_s is the fine-scale conductivity field and F is the average flux data. Using the Bayes formula we can write

$$P(K_s|F) \propto P(F|K_s)P(K_s), \tag{3.8}$$

where $P(K_s)$ is the unconditioned (prior) distribution of the conductivity field. In practice, the measured fluxes contain measurement errors. In this paper, we assume that the measurement error satisfies a Gaussian distribution; thus, the likelihood function $P(F|K_s)$ takes the form

$$P(F|K_s) \propto \exp\left(-\frac{\|F - F_{K_s}\|^2}{\sigma_f^2}\right), \tag{3.9}$$

where F consists of the reference fluxes, F_{K_s} consists of the average fluxes for conductivity field K_s, and σ_f is the measurement precision. In practice, F_{K_s} is computed by solving Eq. (2.1) and recording the desired fluxes for the given K_s on the fine-grid. Since both F and F_{K_s} are functions of time and space (denoted by x and t), the norm $\|F - F_{K_s}\|^2$ is defined as the L_2 norm

$$\|F - F_{K_s}\|^2 = \int_0^X \int_0^T [F(x,t) - F_{K_s}(x,t)]^2 \mathrm{d}t\mathrm{d}x. \tag{3.10}$$

Denote the sampling target distribution as

$$\pi(K_s) = P(K_s|F) \propto \exp\left(-\frac{\|F - F_{K_s}\|^2}{\sigma_f^2}\right)P(K_s). \tag{3.11}$$

Since different conductivity fields may produce the same average flux response, the distribution $\pi(K_s)$ is a function of K_s with multiple local maxima. Sampling from the distribution $\pi(K_s)$ can be accomplished by the MCMC method. For a given proposal distribution $q(y|x)$, the Metropolis–Hastings MCMC algorithm (see, e.g., [30, p. 233]) consists of the following steps:

Algorithm 1. (Metropolis–Hastings MCMC, Robert and Casella [30])

- Step 1. At K_{sn} generate Y from $q(Y|K_{sn})$.
- Step 2. Accept Y as a sample with probability

$$p(K_{sn}, Y) = \min\left(1, \frac{q(K_{sn}|Y)\pi(Y)}{q(Y|K_{sn})\pi(K_{sn})}\right), \tag{3.12}$$

i.e. take $K_{sn+1} = Y$ with probability $p(K_{sn}, Y)$, and $K_{sn+1} = K_{sn}$ with probability $1 - p(K_{sn}, Y)$.

The MCMC algorithm generates a Markov chain $\{K_{sn}\}$ whose stationary distribution is $\pi(K_s)$. A remaining question is how to choose an efficient proposal distribution $q(Y|K_{sn})$.

An important type of proposal distribution can be derived from the Langevin diffusion, as proposed by Grenander and Miller [15]. Following from [5] the proposal distribution $q(Y|K_{sn})$ in Algorithm 1 is chosen as

$$Y = K_{sn} + \frac{\Delta\tau}{2}\nabla \log \pi(K_{sn}) + \sqrt{\Delta\tau}\epsilon_n. \tag{3.13}$$

Since ϵ_n are independent Gaussian vectors, the transition distribution of the proposal generator Eq. (3.13) is

$$\begin{aligned} q(Y|K_{sn}) &\propto \exp\left(-\frac{\|Y - K_{sn} - \frac{\Delta\tau}{2}\nabla \log \pi(K_{sn})\|^2}{2\Delta\tau}\right), \\ q(K_{sn}|Y) &\propto \exp\left(-\frac{\|K_{sn} - Y - \frac{\Delta\tau}{2}\nabla \log \pi(Y)\|^2}{2\Delta\tau}\right). \end{aligned} \tag{3.14}$$

The scheme Eq. (3.13) can be regarded as a problem-adapted random walk since gradient information of the target distribution is included to enforce a biased random walk.

3.3.1. Preconditioned Langevin MCMC algorithm

The major computational costs of Algorithm 1 come in the computation of the target distribution $\pi(K_s)$ and the gradient, $\nabla \log \pi(K_s)$. Since the map between the saturated conductivity and the average flux is governed by the solution to Richards' equation, there is no explicit formula for the target distribution $\pi(K_s)$. To compute the function $\pi(K_s)$, a solution to Richards' equation on the fine-scale for the given K_s is needed. For the same reason, we need to compute the gradient of $\log \pi(K_s)$ in Eq. (3.13) numerically (by finite differences), which involves solving Richards' equation multiple times. To compute the acceptance probability Eq. (3.12), Richards' equation needs to be solved one more time. While one advantage of Langevin algorithms is their higher acceptance rates, the number of simulations need for a single sample makes the direct (full) MCMC simulations with Langevin samples prohibitively expensive.

To bypass the above difficulties, we consider a coarse-grid Langevin MCMC algorithm where most of the fine-scale computations are replaced by coarse-scale computations. Based on a coarse-grid model of the stationary distribution, samples are generated from Eq. (3.13) using the coarse-scale gradient. The proposals are further filtered by an additional Metropolis acceptance–rejection test on the coarse-grid. If the sample does not pass the coarse-grid test, the sample is rejected, and no further fine-scale test is necessary. The argument for this procedure is that if a proposal is not accepted by the coarse-grid test, then it is unlikely to be accepted by the fine-scale test either. By eliminating most of the "unlikely" proposals with inexpensive coarse-scale tests, we can avoid wasting CPU time simulating the rejected samples on the fine-scale.

To model $\pi(K_s)$ on the coarse-scale, we define a coarse-grid map $F^*_{K_s}$ between the saturated conductivity field K_s and the average flux F. The map F^* is determined by coarse-scale solutions. Consequently, the target distribution $\pi(K_s)$ can be approximated by

$$\pi^*(K_s) \propto \exp\left(-\frac{\|F - F^*_{K_s}\|^2}{\sigma_c^2}\right)P(K_s), \tag{3.15}$$

where σ_c is the measurement precision for the coarse-grid distribution and usually taken to be of the same order as σ_f. Note that one can perform *apriori* simulations and determine the relation between coarse- and fine-scale flux errors. This can be done by choosing a number of independent hydraulic conductivity realizations from the prior distribution and modeling the relation between $\|F - F^*_{K_s}\|$ and $\|F - F_{K_s}\|$ via non-parametric approaches, i.e., $\|F - F^*_{K_s}\| \approx G(\|F - F_{K_s}\|)$, where the function G is estimated *apriori*. Furthermore, this relation can be used for determining π^*. This is particularly important if the coarse-scale models are not accurate. We have used this idea in petroleum applications in [9].

Then the Langevin samples are generated from Eq. (3.13) using the coarse-grid gradient of the target distribution

$$Y = K_{sn} + \frac{\Delta\tau}{2}\nabla \log \pi^*(K_{sn}) + \sqrt{\Delta\tau}\epsilon_n. \tag{3.16}$$

The transition distribution of the coarse-grid proposal Eq. (3.16) is

$$\begin{aligned} q^*(Y|K_{sn}) &\propto \exp\left(-\frac{\|Y - K_{sn} - \frac{\Delta\tau}{2}\nabla \log \pi^*(K_{sn})\|^2}{2\Delta\tau}\right), \\ q^*(K_{sn}|Y) &\propto \exp\left(-\frac{\|K_{sn} - Y - \frac{\Delta\tau}{2}\nabla \log \pi^*(Y)\|^2}{2\Delta\tau}\right). \end{aligned} \tag{3.17}$$

By replacing the fine-scale gradient with the coarse-scale gradient, we can reduce the computational cost dramatically but still direct the proposals to regions with larger probabilities.

Because of the high dimension of the problem and the discretization errors, most proposals generated by the Langevin algorithms (both Eqs. (3.13 and 3.16)) will be rejected by the Metropolis acceptance–rejection test Eq. (3.12). To avoid wasting expensive fine-scale computations on unlikely acceptable samples, we further filter the Langevin proposals by the coarse-scale acceptance criteria

$$g(K_{sn}, Y) = \min\left(1, \frac{q^*(K_{sn}|Y)\pi^*(Y)}{q^*(Y|K_{sn})\pi^*(K_{sn})}\right). \tag{3.18}$$

Combining all the discussion above, we have the following revised MCMC algorithm.

Algorithm 2. (Preconditioned coarse-gradient Langevin algorithm)

- Step 1. At K_{sn}, generate a trial proposal Y from Eq. (3.16).
- Step 2. Take the proposal K_s as

$$K_s = \begin{cases} Y \text{ with probability} g(K_{sn}, Y), \\ K_{sn} \text{ with probability} 1 - g(K_{sn}, Y), \end{cases}$$

where

$$g(K_{sn}, Y) = \min\left(1, \frac{q^*(K_{sn}|Y)\pi^*(Y)}{q^*(Y|K_{sn})\pi^*(K_{sn})}\right).$$

Therefore, the proposal K_s is generated from the effective instrumental distribution

$$\begin{aligned} Q(K_s|K_{sn}) &= g(K_{sn}, K_s)q^*(K_s|K_{sn}) \\ &\quad + \left(1 - \int g(K_{sn}, K_s)q^*(K_s|K_{sn})\mathrm{d}K_s\right)\delta_{K_{sn}}(K_s). \end{aligned} \tag{3.19}$$

- Step 3. Accept K_s as a sample with probability

$$\rho(K_{sn}, K_s) = \min\left(1, \frac{Q(K_{sn}|K_s)\pi(K_s)}{Q(K_s|K_{sn})\pi(K_{sn})}\right), \tag{3.20}$$

i.e., $K_{sn+1} = K_s$ with probability $\rho(k_n, k)$, and $K_{sn+1} = K_{sn}$ with probability $1 - \rho(K_{sn}, K_s)$.

Step 2 screens the trial proposal Y by the coarse-grid distribution before passing it to the fine-scale test. The filtering process changes the proposal distribution of the algorithm from $q^*(Y|K_{sn})$ to $Q(K_s|K_{sn})$ and serves as a preconditioner to the MCMC method. This is why it is called the preconditioned coarse-gradient Langevin algorithm. We note that testing proposals by approximate target distributions is not a very new idea. Similar strategies have been developed previously in [21,2].

Note that there is no need to compute $Q(K_s|K_{sn})$ and $Q(K_{sn}|K_s)$ in Eq. (3.20) by formula Eq. (3.19). The acceptance probability Eq. (3.20) can be simplified as

$$\rho(K_{sn}, K_s) = \min\left(1, \frac{\pi(K_s)\pi^*(K_{sn})}{\pi(K_{sn})\pi^*(K_s)}\right). \tag{3.21}$$

In fact, this is obviously true for $K_s = K_{sn}$ since $\rho(K_{sn}, K_{sn}) \equiv 1$. For $K_s \neq K_{sn}$

$$\begin{aligned} Q(K_{sn}|K_s) &= g(K_s, K_{sn}) q(K_{sn}|K_s) \\ &= \frac{1}{\pi^*(K_s)} \min\left(q(K_{sn}|K_s)\pi^*(K_s), q(K_s|K_{sn})\pi^*(K_{sn})\right) \\ &= \frac{q(K_s|K_{sn})\pi^*(K_{sn})}{\pi^*(K_s)} g(K_{sn}, K_s) = \frac{\pi^*(K_{sn})}{\pi^*(K_s)} Q(K_s|K_{sn}). \end{aligned}$$

Substituting the above formula into Eq. (3.20), we immediately get Eq. (3.21).

3.3.2. Collocation methods

Instead of using a coarse-scale model in Algorithm 2, we propose using collocation methods based on the coarse-scale models. Suppose we wish to approximate functions $f : [-1,1]^N \to \mathbb{R}$ using only the known values of f at some locations in $[-1,1]^N$. Here N refers to the stochastic dimension. One may consider two different problems in this situation: the first where the known values are given by scattered data in $[-1,1]^N$ and the other where the approximation is based on values at previously chosen points (gridded data). Only the latter problem will be considered since interpolation from scattered data in high dimensions remains a challenging problem. For simplicity, we consider approximation via Lagrange interpolation in high dimensions.

Suppose one dimensional Lagrange interpolation for each dimension is defined by $i = 1, \ldots, N$ by

$$U_j(f)(\chi) = \sum_{j=1}^{M_i} f\left(\chi_j^i\right) L_j^i(\chi), \tag{3.22}$$

where the L_j^i are traditional Lagrange basis functions and M_i is the number of nodes in the ith dimension. Then the interpolant of f in multiple dimensions can be written as

$$\begin{aligned} (U_{i_1} \otimes \cdots \otimes U_{i_N})(f) = \sum_{i_1=1}^{M_{i_1}} \cdots \sum_{i_N=1}^{M_{i_N}} f\left(\chi_{i_1}^{j_1}, \ldots, \chi_{i_N}^{j_N}\right) \\ \cdot \left(L_{i_1}^{j_1} \otimes \cdots \otimes L_{i_N}^{j_N}\right). \end{aligned} \tag{3.23}$$

If $N = 10$ and $M = 4$ then we have $4^{10} = 1{,}048{,}576$ terms in Eq. (3.23). Each of our function values is generated by a solution to a PDE system; thus full tensor product interpolation is prohibitively expensive. Sparse grid collocation methods, more specifically the Smolyak algorithm [31], can be used to alleviate this problem.

The Smolyak algorithm is a linear combination of product formulas chosen so that an interpolation property for $N = 1$ is preserved for $N > 1$. We make $|\mathbf{i}| = i_1 + \cdots + i_N$ for $\mathbf{i} \in \mathbb{N}^N$. The Smolyak algorithm is defined by

$$A(q, N) = \sum_{q-N+1 \leqslant |\mathbf{i}| \leqslant q} (-1)^{q-|\mathbf{i}|} \cdot \binom{N-1}{q-|\mathbf{i}|} \cdot (U_{i_1} \otimes \cdots \otimes U_{i_N}). \tag{3.24}$$

Note that one must evaluate f at only sparse values given by

$$H(q, N) = \bigcup_{q-N+1 \leqslant |\mathbf{i}| \leqslant q} (\Theta_{i_1} \times \ldots \times \Theta_{i_N}), \tag{3.25}$$

where $\Theta_i = \{\theta_1^i, \ldots, \theta_{M_i}^i\}$ are the set of points used by U_i. This leads us to $n(k+N, N) \approx \frac{2^k}{k!} \cdot N^k$ nodes used by $A(N+k, N)$. Here the k term determines how many nodes will be used. For a fixed N, we define $A(N+k, N)$ as kth order Smolyak interpolation.

As suggested by numerous sources [26,41,1], Smolyak formulas that are based on the extrema of Chebyshev polynomials are considered. We choose

$$\chi_j^i = -\cos\frac{\pi \cdot (j-1)}{M_i - 1}, \quad j = 1, \ldots, M_i \tag{3.26}$$

and define $\chi_1^i = 0$ for $M_i = 1$. We also choose $M_1 = 1$ and $M_i = 2^{i-1} + 1$ for $i > 1$. This has the benefit of making our nodal sets nested; thus $H(q, N) \subset H(q+1, N)$.

Using the Smolyak formulas and Lagrange interpolation, $A(N+k, N)$ is exact for all polynomials of degree k. Using techniques described in [1], the one dimensional error estimate is given by

$$\|f - U_i(f)\|_\infty \leqslant E_{M_i - 1}(f) \cdot \left(1 + \Lambda_{M_i}\right), \tag{3.27}$$

where E_M is the error of the best approximation by polynomials $p \in \mathbf{P}(M, 1)$ and Λ_M is the Lebesgue constant for the Chebyshev polynomials. We have the estimate

$$\Lambda_M \leqslant \frac{2}{\pi} \log(M-1) + 1 \tag{3.28}$$

for $M \geqslant 2$. Estimates are also obtained in multiple dimensions [1,41,26]. One can also use collocation methods to determine the weights in chaos expansion as it is often done in the literature. In our numerical simulations, we will restrict ourselves to a bounded compact set in the uncertainty space.

3.3.3. Preconditioned interpolated Langevin MCMC algorithm

We now consider the modification of the preconditioned coarse-gradient Langevin MCMC (Algorithm 2) in which the coarse-scale response is replaced by a response obtained by collocation. To model the target distribution $\pi(K_s)$ using coarse-scale collocation, we define an interpolated coarse-grid map $\tilde{\pi}^*(K_s)$ between the saturated conductivity field K_s and the target distribution π^*. The map $\tilde{\pi}^*$ is determined by coarse-scale collocation. Similarly, the Langevin samples are generated using the interpolated coarse-grid gradient of the target distribution

$$Y = K_{sn} + \frac{\Delta\tau}{2} \nabla \log \tilde{\pi}^*(K_{sn}) + \sqrt{\Delta\tau}\epsilon_n \tag{3.29}$$

and the transition distribution of the coarse-scale collocation proposal is defined by

$$\begin{aligned} \tilde{q}^*(Y|K_{sn}) &\propto \exp\left(-\frac{\|Y - K_{sn} - \frac{\Delta\tau}{2}\nabla \log \tilde{\pi}^*(K_{sn})\|^2}{2\Delta\tau}\right), \\ \tilde{q}^*(K_{sn}|Y) &\propto \exp\left(-\frac{\|K_{sn} - Y - \frac{\Delta\tau}{2}\nabla \log \tilde{\pi}^*(Y)\|^2}{2\Delta\tau}\right). \end{aligned} \tag{3.30}$$

We again further filter the Langevin proposals by the coarse-scale collocation acceptance criteria

$$g(K_{sn}, Y) = \min\left(1, \frac{\tilde{q}^*(K_{sn}|Y)\tilde{\pi}^*(Y)}{\tilde{q}^*(Y|K_{sn})\tilde{\pi}^*(K_{sn})}\right). \tag{3.31}$$

Combining all the discussion above, we now have the revised Langevin MCMC algorithm based on coarse-scale collocation [4].

Algorithm 3. (Preconditioned interpolated coarse-gradient Langevin algorithm)

- Step 1. At K_{sn}, generate a trial proposal Y from Eq. (3.29).
- Step 2. Take the proposal K_s as

$$K_s = \begin{cases} Y \text{ with probability} g(K_{sn}, Y), \\ K_{sn} \text{ with probability} 1 - g(K_{sn}, Y), \end{cases}.$$

Therefore, the proposal K_s is generated from the effective instrumental distribution

$$\begin{aligned} Q(K_s|K_{sn}) &= g(K_{sn}, K_s)\tilde{q}^*(K_s|K_{sn}) \\ &\quad + \left(1 - \int g(K_{sn}, K_s)\tilde{q}^*(K_s|K_{sn})\mathrm{d}K_s\right)\delta_{K_{sn}}(K_s). \end{aligned} \tag{3.32}$$

- Step 3. Accept K_s as a sample with probability

$$\rho(K_{sn}, K_s) = \min\left(1, \frac{Q(K_{sn}|K_s)\pi(K_s)}{Q(K_s|K_{sn})\pi(K_{sn})}\right), \tag{3.33}$$

4. Numerical results

In this section the numerical results for the preceding algorithms will be presented. The saturated conductivity, $K_s(x)$, is assumed to be defined on the unit square $\Omega = [0,1]^2$ and is known at some sparse locations in Ω. Additionally, we assume the covariance of $\log(K_s)$ is known. We discretize Ω with a rectangular mesh and represent K_s as a matrix. The goal in these numerical results is to show the proposed algorithms (Algorithms 2 and 3) have similar sampling performance to the traditional Langevin MCMC (Algorithm 1) with significantly less computational cost.

For the uncertainty quantification problems considered in this paper, we assume that the discrete integrated response function, F, contains the average flux on the uppermost boundary of the fine-grid domain for a given set of times. Let us suppose that for each time in a given set of k times, the flux is found across the entire upper cell boundary at n equally spaced intervals. We denote these times as $t_1, \ldots, t_k$. The response is then given by the average flux at each of the n intervals, and each of these k times, in order. In other words,

$$F = (Flux_{avg}(x_1,t_1), \ldots, Flux_{avg}(x_n,t_1), Flux_{avg}(x_1,t_2), \ldots, Flux_{avg}(x_n,t_k)) \tag{4.1}$$

As noted previously, we use the MsFVEM for our coarse-scale model. Note, when using the MsFVEM, we can reconstruct the fine-scale flux using the multiscale basis functions. Thus, the averaging used for F will always be over the fine-grid domain regardless of whether we are using a fine-scale or coarse-scale model. Additionally, the discrete average flux response surfaces are scaled down by a factor of 10. This makes the average flux values to be, on average, between 0 and 1. This does not change the overall results and the only purpose of the scaling is to keep σ_f and σ_c values at approximately the same magnitude as in our previous MCMC investigation for multiphase immiscible flow [5].

Numerical results for the exponential constitutive relation with (dimensionless) $\alpha = 0.01$, $\beta = 0.01$, and $\theta_s = 0.5$ are presented. The initial pressure head is assumed to be $\psi = -10$. Dirichlet boundary conditions are given by $\psi = 0$ on the bottom of Ω, $[0,1] \times 0$, and $\psi = -10$ on the portion of the top boundary between $x = 0.3$ and $x = 0.7$, $[0.3, 0.7] \times 1$. We assume Neumann boundary conditions otherwise. We choose these particular boundary conditions, as opposed to no flow on the sides and given pressure heads on the top and bottom, so that we can see more variations in the pressure head profiles.

4.1. KLE using normal covariance

In the first numerical results, to represent the discrete saturated conductivity field, we generate a KLE using normal covariance and correlation lengths $L_1 = 0.3$ and $L_2 = 0.3$. Only 15 terms are kept in the KLE since the eigenvalues decay very rapidly. Additionally, it is assumed that the saturated conductivity is known at eight sparse locations. This leaves us with 7 stochastic dimensions to sample from. We restrict ourselves to the hypercube $[-2.5, 2.5]^7$ in stochastic space. The reason for this restriction is due to the sparse grid collocation. For each simulation we test 5000 samples in the MCMC. A 49×49 fine-scale model and a 5×5 coarse-scale model are used. Note that these methods are vertex-based. Thus the number of cells are 48×48 and 4×4 in the fine-scale and coarse-scale model, respectively. The time step $\Delta\tau$ of the Langevin MCMC algorithm is denoted by δ.

First, comparisons between the fine-scale target distribution π, coarse-scale distribution π^*, and interpolated coarse-scale distribution $\tilde{\pi}^*$ are presented. We plot the restriction of the target distributions to a 2D hyperplane by fixing the value at 5 of the 7 stochastic dimensions. We plot the first two dimensions, θ_1 and θ_2, restricted to $[-2.5, 2.5]^2$.

In Fig. 1 it can be seen that the coarse-scale surface captures many of the large features of the fine-scale surface. This is expected, as one would not expect the small features to be captured using the coarse-scale method. The last two plots correspond to the interpolated coarse-scale surface for Smolyak order 1 and 3, respectively. In this case of $N = 7$ stochastic dimensions, 1st order has 15 nodes, 2nd order has 113 nodes, and 3rd order has 589 nodes. While the values in the interpolated surfaces do not correspond exactly, it appears the 1st order surface captures some very large scale features of the coarse-scale surface. The 3rd order method captures many more features but requires more computations

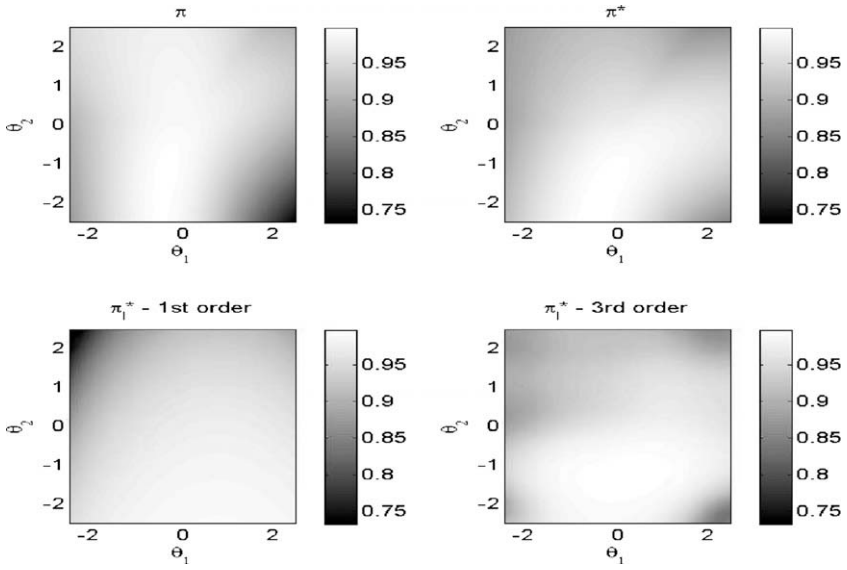

Fig. 1. Fine-scale response surface π, coarse-scale response surface π^*, and two interpolated coarse-scale response surfaces $\tilde{\pi}^*$ (of order 1 and 3) restricted to a 2D hyperplane. Here θ_1 and θ_2 correspond to the first two stochastic dimensions.

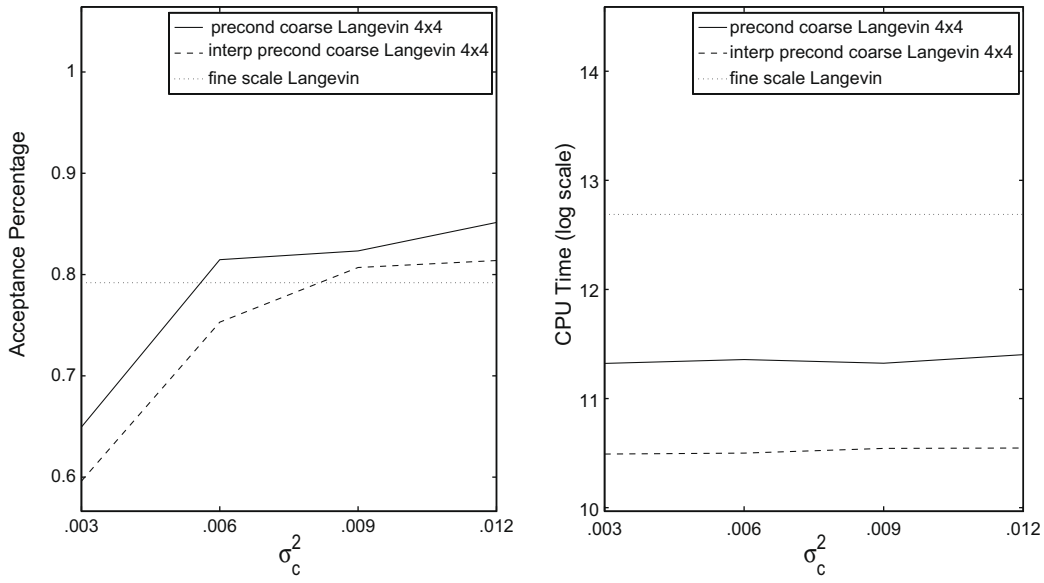

Fig. 2. Comparison between preconditioned coarse Langevin (Algorithm 2), preconditioned interpolated coarse Langevin (Algorithm 3), and fine-scale Langevin (Algorithm 1) for various σ_c values with $\sigma_f^2 = 0.003$ and $\delta = 0.05$. Left: acceptance rate. Right: natural log of CPU time (seconds).

for large stochastic dimensions. For this reason, we consider only first order interpolation for the remainder of the paper.

Next, the acceptance rates of each algorithm are compared. In the left portion of Fig. 2, the acceptance rates with different coarse-scale precision σ_c are compared. Recall the acceptance rate is the ratio between the number of accepted saturated conductivities and the number of fine-scale acceptance–rejection tests. Since Algorithm 1 does not have a coarse-scale test, it is constant for all σ_c, and its acceptance rate is simply the number of accepted saturated conductivities divided by the total number of samples. In Fig. 2, it can be seen that Algorithms 2 and 3 have a higher acceptance rate for some values σ_c than Algorithm 1. This is due to the preconditioning step, which filters proposals that are not likely to be accepted by the fine-scale test. In the right portion of Fig. 2, we compare the CPU time for each of the algorithms on a natural log scale. First, there is a significant savings in CPU time when using Algorithm 2 rather than Algorithm 1. This is primarily due to the fact that Algorithm 2 uses a coarse-scale gradient, $\nabla \log \pi^*(k)$, while Algorithm 1 uses a fine-scale gradient $\nabla \log \pi(k)$. Since $\pi(k)$ and $\pi^*(k)$ are functions determined by simulations, these gradients are computed using finite differences. Thus we must run a simulation in each of the 7 stochastic dimensions and our CPU savings is partly due to 7 coarse-scale simulations per sample as opposed to 7 fine-scale simulations. Additionally, the preconditioning step in Algorithm 2 helps filter the samples, resulting in less fine-scale acceptance–rejection tests. Looking at Algorithm 3, there is an even more remarkable savings in CPU time. This is clearly due to fact that the target distribution, $\tilde{\pi}^*(k)$, and the interpolated gradient, $\nabla \log \tilde{\pi}^*(k)$, can be computed analytically.

It is worth noting that, even for 3rd order, the computational cost in computing the Smolyak nodes is still small compared to the coarse-gradient Langevin MCMC. For example if we are to sample in 7 dimensions using coarse-gradient Langevin, then we must run 8 forward simulations for each sample. After approximately 75 samples, we have already run the coarse-scale simulations more times than if we had generated nodes for 3rd order Smolyak interpolation.

We now compare the average flux errors for each algorithms in Fig. 3. The purpose of this is to demonstrate that the mixing time (the number of samples for which it takes to converge to steady state) of each algorithm is approximately the same. At steady state we are only accepting samples that are below the measurement error in the target distribution. It is clear from this figure, that Algorithms 2 and 3 converge to steady state in slightly more iterations than Algorithm 1. Considering that Algorithm 2 is over four times faster than Algorithm 1, it seems the increased mixing time is negligible. The formal convergence diagnosis can be performed using a multiple chains method based convergence diagnosis [13]. In this paper, our goal is to compare the modified Langevin algorithm with the direct Langevin algorithm; thus we restrict ourselves to only showing errors versus the number of iterations. We note that the convergence diagnostics has nothing to do with the rate of convergence, which depends on the second largest eigenvalue of the transition matrix of the Markov chain. For the complex chains, calculation of these eigenvalues is not simple.

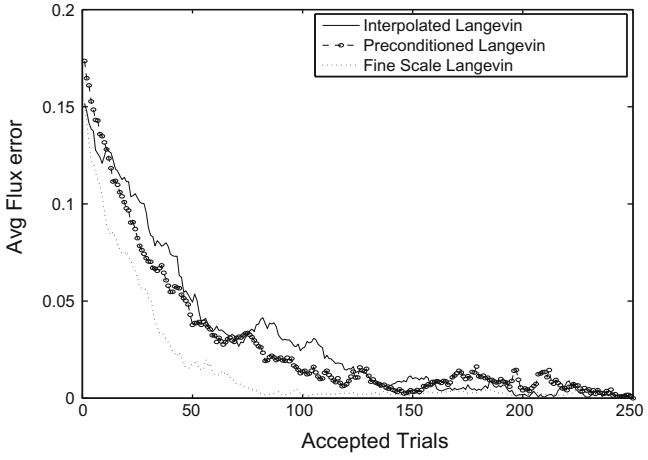

Fig. 3. Average flux errors for each algorithm.

Lastly, we present some saturated conductivity realizations sampled from the posterior distribution. In Fig. 4 the exact K_s is shown in the upper left, while three accepted conductivities from Algorithm 3 are showing in the remaining three plots. Note that due to the fact that the proposals are correlated, there is a correlation between the samples. One can use an independent sampler as an instrumental probability distribution or independent chains in

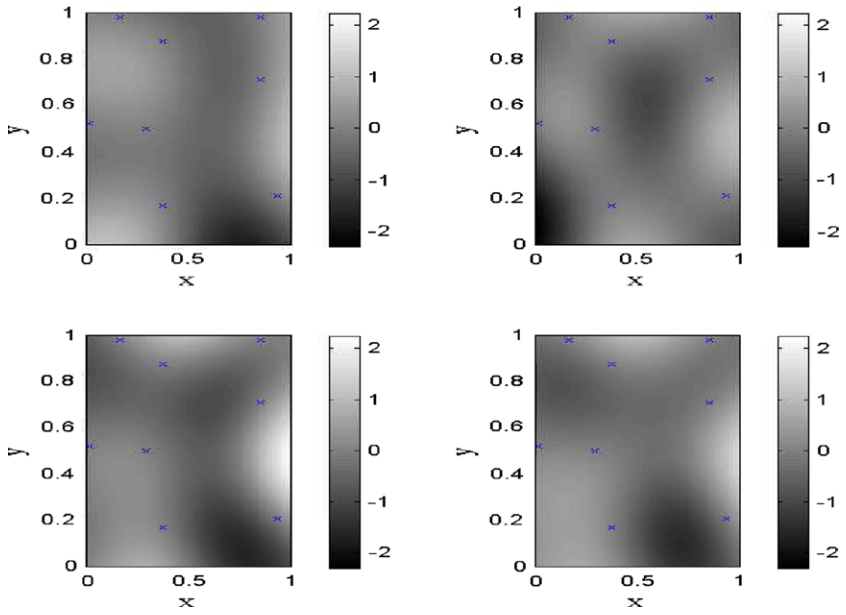

Fig. 4. Exact K_s (upper left) and three different accepted conductivities for preconditioned interpolated coarse Langevin (Algorithm 3).

order to avoid this. It can be seen that the samples capture many features of the reference saturated conductivity quite well. Note that these accepted saturated conductivity fields correspond to those with errors in Fig. 3; thus they give nearly the same average flux responses, so they are all eligible samples.

4.2. KLE using exponential covariance

We now introduce a set of numerical results which use the proposed algorithms for a conductivity generated using the KLE with exponential covariance instead of normal covariance. The exponential covariance kernel is given by

$$R(x,y) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|}{L_1} - \frac{|x_2 - y_2|}{L_2}\right). \tag{4.2}$$

The KLE with exponential covariance contains many more terms in the expansion. Moreover, the conductivity fields for exponential covariance have more fine-scale features. Our goal to show the results using the proposed algorithms in the context of exponential covariance are similar to the previous results which used normal covariance.

We generate a truncated KLE using exponential covariance with correlation lengths $L_1 = 0.5$ and $L_2 = 0.4$. 100 terms are kept, and it is assumed the conductivity field is known at 15 distinct points. Thus we have 85 stochastic dimensions. Again, we restrict ourselves to the hypercube $[-2.5, 2.5]^{85}$ in stochastic space, consider a 49×49 fine grid, and a 5×5 coarse grid. Since the gradient of the target distribution must be computed for each of the 85 dimensions, Algorithm 1 would require 85 fine-scale solutions for each proposal. This is not computationally feasible in our setting. Thus only Algorithms 2 and 3 will be compared. Additionally, only 1st order interpolation is considered, since 2nd order would require 14,621 values. Only a brief set of numerical results are presented, since many of the same ideas have been discussed previously.

In Fig. 5, π, π^* and $\tilde{\pi}^*$ are compared. As in the previous case, the coarse-scale surface matches many large features of the fine-scale

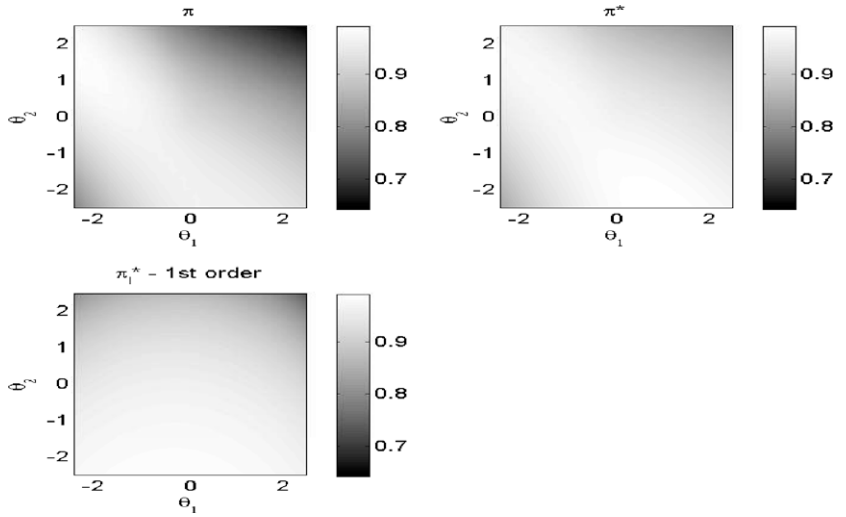

Fig. 5. Fine-scale response surface π, coarse-scale response surface π^*, and interpolated coarse-scale response surfaces $\tilde{\pi}^*$ restricted to a 2D hyperplane.

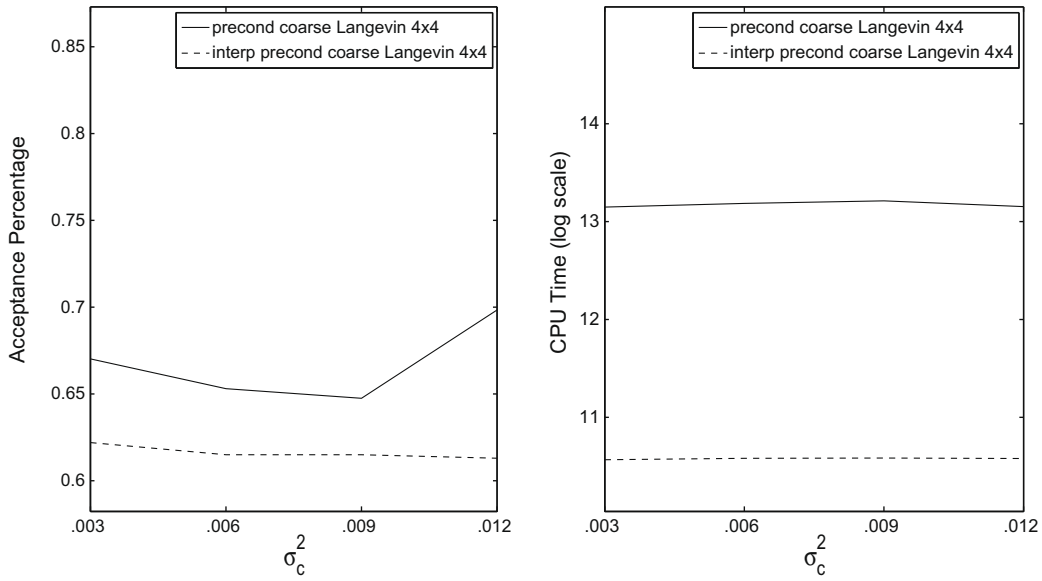

Fig. 6. Comparison between preconditioned coarse Langevin (Algorithm 2) and preconditioned interpolated coarse Langevin (Algorithm 3) for various σ_c values. We use $\sigma_f^2 = 0.003$, $\delta = 0.05$ and exponential covariance in the KLE. Left: acceptance rate. Right: natural log of CPU time (seconds).

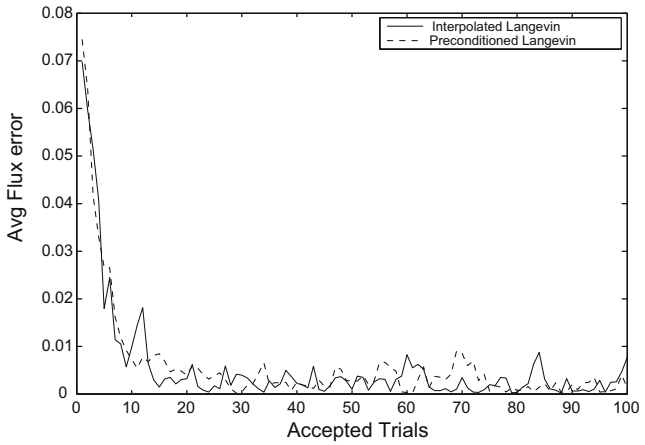

Fig. 7. Average flux errors for preconditioned interpolated coarse Langevin (Algorithm 3) compared to preconditioned coarse Langevin (Algorithm 2).

surface. Additionally, the interpolated surface matches some general features of the coarse-scale surface. In Fig. 6 the acceptance rate (left) and CPU time (right) for Algorithms 2 and 3 are shown. We find that the two algorithms have similar acceptance rates, while Algorithm 3 is over ten times faster than Algorithm 2. This vast improvement in CPU time is clearly due to the use of interpolated gradients in Algorithm 3, as opposed to coarse-scale gradients in Algorithm 2.

In Fig. 7 it is demonstrated that Algorithm 3 converges to steady state at a similar rate to Algorithm 2. In Fig. 8 we compare the reference saturated conductivity (upper left) with three different accepted conductivities found by Algorithm 3. Since the features of the conductivities in the case of the exponential covariance are less smooth, it is much harder to see the correspondence of the accepted conductivities to the reference.

As we can see from the preceding results, exponential covariance produces results very similar to normal covariance for Richards' equation. We find that Algorithm 3 provides an acceptance rate similar to Algorithm 3 while providing a dramatic savings in CPU time.

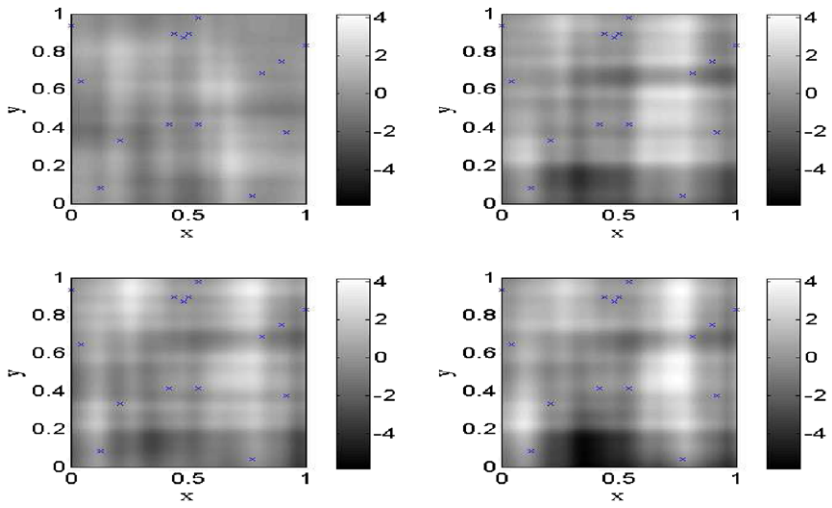

Fig. 8. Exact K_s (upper left) and three different accepted conductivities.

We note that the efficiency of the proposed approaches depends on the accuracy of the interpolation. In particular, the proposed approaches become more effective as the approximation of response surface is more accurate. Otherwise, the proposed approaches may not give any CPU saving; however, they still provide a rigorous sampling of the posterior distribution. For realistic cases considered in this paper, it is difficult to obtain analytical convergence estimates. One can perform convergence analysis for simplified cases in the limit of scale separation. This is currently under investigation. One can improve the efficiency of the proposed approaches by performing more accurate approximation of the response surfaces. For example, using adaptive interpolation techniques, one can achieve higher efficiency of the proposed MCMC. This will be explored in future.

5. Concluding remarks

In this paper, efficient uncertainty quantification techniques in inverse problems for Richards' equation using coarse-scale simulation models are proposed. The problem under consideration consists of sampling hydraulic conductivity given dynamic measure data, such as average boundary fluxes. Our goal is to use inexpensive coarse-scale simulations in the context of Langevin MCMC and compare modified Langevin MCMC algorithms to the traditional fine-scale Langevin MCMC algorithm. A preconditioned coarse-gradient Langevin MCMC algorithm is considered, where the proposals are made based on coarse-scale simulations and tested with inexpensive coarse-scale runs in order to increase the acceptance rate. Coarse-scale simulations are performed using multiscale finite volume element methods. Additionally, sparse grid collocation techniques are employed to approximate the coarse-scale simulations and an interpolated variant of the preconditioned coarse-gradient Langevin algorithm is introduced.

Each algorithm is applied to the uncertainty quantification problem for both small and large stochastic spaces. For the small-dimensional stochastic spaces, the numerical results verify that the proposed algorithms are efficient and give similar performance as the fine-scale Langevin algorithms. For large-dimensional stochastic spaces, where a fine-scale Langevin algorithm may not be computationally feasible, it is shown that both proposed algorithms can be used as an alternative to the fine-scale Langevin algorithm. In our numerical simulations, we consider hydraulic conductivity fields described using two-point correlation functions. In future work, we plan to test the proposed approaches with more complex heterogeneities.

Acknowledgement

This research is supported by NSF CMG Grant 0621113. We are grateful to reviewers for their comments which helped to improve the paper substantially.

References

[1] Barthelmann V, Novak E, Ritter K. High dimensional polynomial interpolation on sparse grids. Adv Comput Math 2000;12(4):273–88.
[2] Christen A, Fox C. MCMC using an approximation. J Comput Graph Stat 2005;14(4):795–810.
[3] Das NN, Mohanty BP, Njoku EG. A Markov chain Monte Carlo algorithm for upscaled soil-vegetation-atmosphere-transfer modeling to evaluate satellite-based soil moisture measurements. Water Resour Res 2008;44:W05416. doi:10.1029/2007WR006472.
[4] Dostert P. Uncertainty quantification using multiscale methods for porous media flows. PhD dissertation, Texas A&M University, College Station, TX; 2007.
[5] Dostert P, Efendiev Y, Hou T, Luo W. Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification. J Comput Phys 2006;217(1):123–42.
[6] Durlofsky LJ. Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media. Water Resour Res 1991;27:699–708.
[7] Efendiev Y, Ginting V, Hou T, Ewing R. Accurate multiscale finite element methods for two-phase flow simulations. J Comput Phys 2006;220(1):155–74.
[8] Efendiev Y, Hou T, Ginting V. Multiscale finite element methods for nonlinear problems and their applications. Commun Math Sci 2004;2(4):553–89.
[9] Efendiev Y, Ma X, Datta-Gupta A, Mallick B. Multi-stage MCMC using non-parameteric error estimators, submitted for publication.
[10] Evensen G. Data assimilation. The ensemble Kalman filter. Springer; 2007.
[11] Ganapathysubramanian B, Zabaras N. Modeling diffusion in random heterogeneous media: data-driven models stochastic collocation and the variational multiscale method. J Comput Phys 2007;226:326–53.
[12] Ginting V. Computational upscaled modeling of heterogeneous porous media flow utilizing finite volume method. PhD dissertation, Texas A&M University, College Station, TX; 2004.
[13] Gelman A, Rubin B. Inference from iterative simulation using multiple sequences. Stat Sci 1992;7:457–511.
[14] Glimm J, Sharp DH. Prediction the quantification of uncertainty. Physica D 1999;133:152–70. Predictability: quantifying uncertainty in models of complex phenomena (Los Alamos, NM, 1998).
[15] Grenander U, Miller MI. Representations of knowledge in complex systems with (discussion). J Roy Stat Soc Ser B (Methodological) 1994;56(4):549–603.
[16] Gu Y, Oliver DS. An iterative ensemble Kalman filter for multiphase fluid flow data assimilation. SPE J 2007;12(4):438–46.
[17] Haverkamp R, Vauclin M, Touma J, Wierenga PJ, Vachaud G. A comparison of numerical solution models for one-dimensional infiltration. Soil Sci Soc Am J 1977;41:285–94.
[18] Hou TY, Wu XH. A multiscale finite element method for elliptic problems in composite materials and porous media. J Comput Phys 1997;134(1):169–89.
[19] Jenny P, Lee SH, Tchelepi H. Multi-scale finite volume method for elliptic problems in subsurface flow simulation. J Comput Phys 2003;187(1):47–67.
[20] De Lannoy GJM, Reichle RH, Houser PR, Pauwels VRN, Verhoest NEC. Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter. Water Resour Res 2007;43:W09410. doi:10.1029/2006WR005449.
[21] Liu JS. Monte Carlo strategies in scientific computing. New York: Springer-Verlag; 2001.
[22] Liu N, Oliver DS. Critical evaluation of the ensemble Kalman filter on history matching of geologic facies. SPE Reserv Eval Eng 2005;8(6):470–7.
[23] Loève M. Probability theory. 4th ed. Berlin: Springer; 1977.
[24] Ma X, Al-Harbi M, Datta-Gupta A, Efendiev Y. A multistage sampling approach to quantifying uncertainty during history matching geological models. SPE J 2008;13(10):77–87.
[25] Margulis SA, McLaughlin D, Entekhabi D, Dunne S. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 field experiment. Water Resour Res 2002;38(12):1299. doi:10.1029/2001WR001114.
[26] Babuska I, Nobile F, Tempone R. A stochastic collocation method for elliptic partial differential equations with random input data. SIAM J Numer Anal 2007;45(3):1005–34.
[27] Oliver D, Cunha L, Reynolds A. Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. Math Geol 1997;29.
[28] Oliver D, He N, Reynolds A. Conditioning permeability fields to pressure data. In: Fifth European conference on the mathematics of oil recovery, Leoben, Austria; 3–6 September 1996.
[29] Richards LA. Capillary conduction of liquids through porous mediums. Physics 1931;1(5):318–33.
[30] Robert C, Casella G. Monte Carlo statistical methods. New York: Springer-Verlag; 1999.
[31] Smolyak SA. Quadrature and interpolation formulas for tensor products of certain classes of functions. Doklady Akademii Nauk SSSR 1963;4:240–3.
[32] van Genuchten MTh. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Sci Soc Am J 1980;44:892–8.
[33] Warrick AW. Time-dependent linearized infiltration: III. Strip and disc sources. Soil Sci Soc Am J 1976;40:639–43.
[34] Western AW, Grayson RB, Blschl G. Scaling of soil moisture: a hydrologic perspective. Annu Rev Earth Planet Sci 2002;30:149180. doi:10.1146/annurev.earth.30.091201.140434.
[35] Wong E. Stochastic processes in information and dynamical systems. New York: MCGraw-Hill; 1971.
[36] Zhang D. Stochastic methods for flow in porous media: coping with uncertainties. San Diego, CA: Academic Press; 2002. p. 350, ISBN: 012-7796215.
[37] Zhang D. Nonstationary stochastic analysis of the transient unsaturated flow in randomly heterogeneous media. Water Resour Res 1999;35:1127–41. doi:10.1029/1998WR900126.
[38] Zhang D, Lu Z. An efficient high-order perturbation approach for flow in random porous media via Karhunen–Loeve and polynomial expansions. J Comput Phys 2004;194(2):773–94.
[39] Zhu J, Mohanty BP. Spatial averaging of van Genuchten hydraulic parameters for steady-state flow in heterogeneous soils: a numerical study. Vadose Zone J 2002;1:261–72.
[40] Zhu J, Mohanty BP. Upscaling of hydraulic properties of heterogeneous soils. In: Pachepsky YA et al., editors. Scaling methods in soil physics. Boca Raton Fla.: CRC Press; 2003. p. 97–118.
[41] Xiu D, Hesthaven J. High-order collocation methods for differential equations with random inputs. SIAM J Sci Comput 2005;27(3):1118–39.