



# Markov Chain Monte Carlo Methods for Fluid Flow Forecasting in the Subsurface

Alsadig Ali<sup>1</sup>, Abdullah Al-Mamun<sup>1,2</sup>, Felipe Pereira<sup>1</sup>,  
and Arunasalam Rahunathan<sup>3</sup>✉

<sup>1</sup> Department of Mathematical Sciences, The University of Texas at Dallas,  
Richardson, TX 75080, USA

<sup>2</sup> Institute of Natural Sciences, United International University, Dhaka, Bangladesh

<sup>3</sup> Department of Mathematics and Computer Science, Central State University,  
Wilberforce, OH 45384, USA  
aRahunathan@centralstate.edu

**Abstract.** Accurate predictions in subsurface flows require the forecasting of quantities of interest by applying models of subsurface fluid flow with very little available data. In general a Bayesian statistical approach along with a Markov Chain Monte Carlo (MCMC) algorithm can be used for quantifying the uncertainties associated with subsurface parameters. However, the complex nature of flow simulators presents considerable challenges to accessing inherent uncertainty in all flow simulator parameters of interest. In this work we focus on the transport of contaminants in a heterogeneous permeability field of a aquifer. In our problem the limited data comes in the form of contaminant fractional flow curves at monitoring wells of the aquifer. We then employ a Karhunen-Loève expansion to truncate the stochastic dimension of the permeability field and thus the expansion helps reducing the computational burden. Aiming to reduce the computational burden further, we code our numerical simulator using parallel programming procedures on Graphics Processing Units (GPUs). In this paper we mainly present a comparative study of two well-known MCMC methods, namely, two-stage and Differential Evolution Adaptive Metropolis (DREAM), for the characterization of the two-dimensional aquifer. A thorough statistical analysis of ensembles of the contaminant fractional flow curves from both MCMC methods is presented. The analysis indicates that although the average fractional flow curves are quite similar, both time-dependent ensemble variances and posterior analysis are considerably distinct for both methods.

**Keywords:** Porous media · Contaminant transport · Two-stage · MCMC · DREAM · MPSRF

## 1 Introduction

In subsurface characterization and flow forecasting, one could characterize the subsurface using a Bayesian framework. The Bayesian framework consists essentially of a Markov Chain Monte Carlo (MCMC) algorithm, in which we repeatedly solve a flow numerical simulator that models the porous media problem of

interest [11]. In this paper we use a single-phase flow numerical simulator that simulates the contaminant transport in an aquifer. In the MCMC algorithm we characterize the heterogeneous permeability field of the aquifer using the simulator. Our C/C++ flow simulator takes advantage of Graphics Processing Units (GPUs), which have the computational capacity to speedup the single-phase flow simulation [12]. The GPU flow simulator runs on a computational grid of the permeability field that is divided into several thousand elements. However, populating each computational element by a random permeability value and changing those values in each MCMC iteration is impracticable. In this case the dimension of the stochastic space is that of the computational domain and the dimension reduction is achieved by a Karhunen-Loève expansion (KLE) [15]. Due to the serial nature of the MCMC algorithm, the computational burden in solving the problem recurrently using the GPU numerical simulator is still huge. This can make the Bayesian framework less attractive for our problem. Parallelizations of the MCMC algorithm to speedup the characterization were considered in [10, 12]. The simulation of several parallel MCMCs reduces the computational cost drastically. However, the convergence of such parallel MCMCs should be carefully analyzed.

In this paper we consider two-stage and Differential Evolution Adaptive Metropolis (DREAM) MCMC methods for the subsurface characterization. The two-stage MCMC was introduced in [6, 8] and it has been investigated more recently (see [7, 14] and references therein). The two-stage procedure is of particular interest to subsurface flow problems because samples can be rejected with inexpensive simulations on coarse grids. The DREAM is an extension of Differential Evolution (DE) MCMC, which integrates the essential ideas of DE genetic algorithm and MCMC algorithm [3]. Inspired by DE MCMC, the DREAM was proposed in [22] and applied to several interesting problems (see [19–21] and references therein). The DREAM is well known to converge relatively faster when compared to earlier procedures. It employs subspace sampling and outlier chain correction to accelerate the convergence towards the stationary distribution.

In the current work we run several parallel simulations for each MCMC method. In this approach, we need to determine when it is safe to stop the MCMC simulations for a reliable characterization of the permeability field. There are several convergence diagnostics available for this purpose and those diagnostics fall into two categories: the first category of diagnostics entirely depends on the output values of the MCMC simulation and those in the second category use not only the output values but also the information on the target distribution. In the first category, Brooks and Gelman [4] proposed a convergence diagnostic that uses the Multivariate Potential Scale Reduction Factor (MPSRF) to decide when to terminate MCMC simulations. Very recently in [2] we proposed a stopping criterion using a statistical analysis for single-phase flow prediction. In the analysis we considered ensembles of fractional flow curves to decide when it is safe to stop MCMC simulations for a reliable characterization and prediction. In this work we use the criterion in [2] for both MCMC methods and compare them for the characterization of the permeability field of the aquifer. Our results

show that the two-stage MCMC provides a good estimate for the average of the quantity of interest (fractional flow curves), but the DREAM MCMC method reveals that the posterior distribution is not well characterized by the two-stage method.

We organize the present study as follows. We discuss the computational modeling of the single-phase flow of the aquifer in Sect. 2. The reduction of the parameter space dimension by the KLE is discussed in Sect. 3. Section 4 contains the statistical framework using two-stage and DREAM MCMC methods for the characterization of the permeability field in the aquifer. The MPSRF, a frequently used criterion to estimate convergence of the MCMC methods, is presented in Sect. 5. Results obtained from our numerical experiments are discussed in Sect. 6. Concluding remarks appear in Sect. 7.

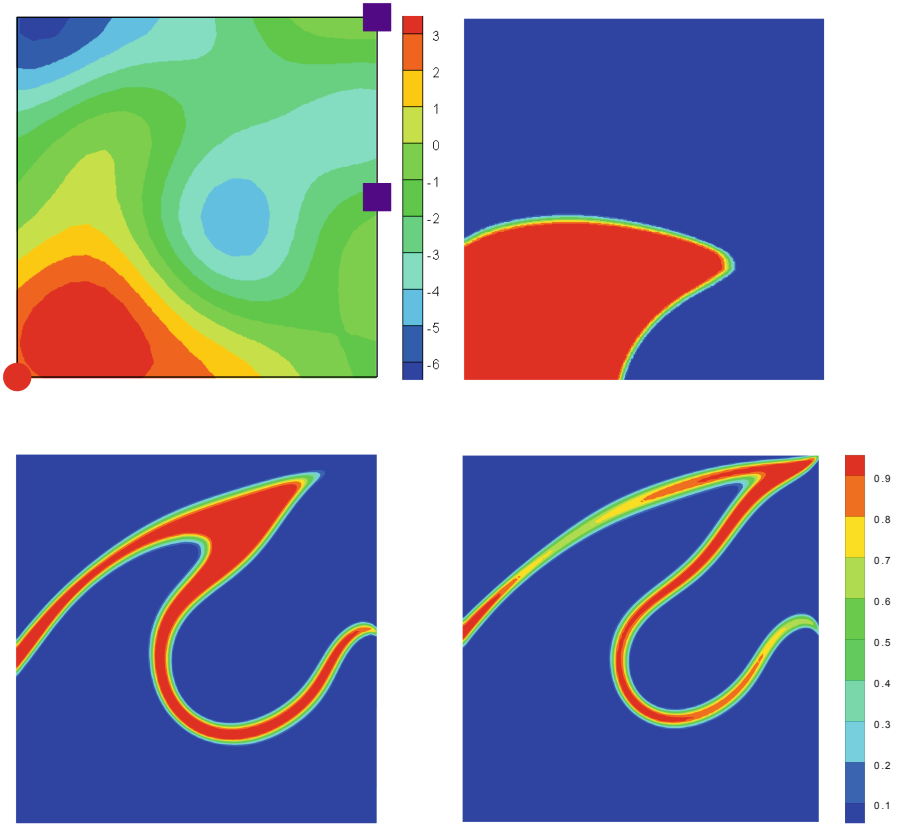
## 2 Computational Physical Model

We consider a unit square-shaped subsurface aquifer  $\Omega$  with a heterogeneous permeability field. The aquifer contains two monitoring wells: one of which is located at the top right corner of the domain (corner well) and the other well (center well) is placed at the middle of the right boundary. An accidental contamination occurs at the spill site and the contaminated water flows naturally through the aquifer. The spill well is positioned at the bottom left corner through which the contaminated (or tracer) water is discharged (Fig. 1). In this single-phase flow model we assume a relatively low concentrations of the contaminant, thus it does not affect the velocity field.

Let us denote the Darcy velocity and the pressure of the fluid by  $\mathbf{v}(\mathbf{x}, t)$ , and  $p(\mathbf{x}, t)$ , respectively, where  $\mathbf{x} \in \Omega$  is the location at time  $t$ . We also denote the absolute permeability of the rock, porosity of the rock, fluid viscosity, and contaminant concentration in the fluid by  $k(\mathbf{x})$ ,  $\phi(\mathbf{x})$ ,  $\mu$ , and  $\rho(\mathbf{x}, t)$ , respectively. We consider that the pore space of the aquifer is filled by water. Applying Darcy's law and mass conservation, the governing equations describing the single-phase flow can be written as the following [5]:

$$\begin{aligned} \nabla \cdot \mathbf{v} &= 0, \text{ where } \mathbf{v} = -\frac{k(\mathbf{x})}{\mu} \nabla p, \quad \mathbf{x} \in \Omega, \\ \phi(\mathbf{x}) \frac{\partial \rho(\mathbf{x}, t)}{\partial t} + \nabla \cdot (\rho(\mathbf{x}, t) \mathbf{v}) &= 0. \end{aligned} \tag{1}$$

In this study the porosity of the rock is considered a constant throughout the domain with  $\phi(\mathbf{x}) = 0.2$ . We assume no-flow boundary conditions and we take  $\rho(\mathbf{x}, t = 0) = 0$ . The system of partial differential equations in (1) does not contain any source or sink because all three wells are modeled through appropriate boundary conditions. The coupled system consists of an elliptic problem and a hyperbolic problem. After applying an operator splitting technique, we solve each problem separately by an appropriate numerical scheme [13, 18].



**Fig. 1.** Top: On the left, the reference permeability field for a unit square-shaped region, where circle and squares on the boundary denote the spill and monitoring wells, respectively. On the right, contaminant flow at  $t = 0.25$  PVI. Bottom: Contaminant flow at  $t = 0.75$  PVI on the left and  $t = 1.0$  PVI on the right.

The permeability field is characterized by using measured data in the form of fractional flow curves,  $F(t)$  which are defined as

$$F(t) = 1 - \frac{\int_{\partial\Omega_{\text{out}}} v_n(\mathbf{x}, t) \rho(\mathbf{x}, t) ds}{\int_{\partial\Omega_{\text{out}}} v_n(\mathbf{x}, t) ds}, \tag{2}$$

where  $\partial\Omega_{\text{out}}$  represents the well outflow boundary, and  $v_n(\mathbf{x}, t)$  is the component of the velocity field normal to the well boundary. The dimensionless time is denoted by  $t$ , which is measured in Pore Volume Injected (PVI) and is computed using the following integral:

$$\text{PVI} = \int_0^T V_p^{-1} \int_{\partial\Omega_{\text{out}}} v_n(\mathbf{x}, t) ds d\tau, \tag{3}$$

where  $V_p$  denotes the total pore volume of the reservoir and  $T$  represents the total time the contaminated water entered through the spill well.

### 3 Reduction in Parameter Space

If we consider the aquifer in a  $128 \times 128$  computational domain, we have 16,384 elements. In our Bayesian framework the numerical simulator uses the permeability value in each element. Therefore, we need to start with a random permeability value in each element and change one or more of those values in each MCMC iteration. Thus, the dimension of the parameter space is 16,384 and it presents a far-fetched framework for the characterization of the permeability field. In order to reduce the number of uncertainty parameters, we therefore use the KLE. Below we reduce the parameter space from 16,384 to  $d$ , which is 20 for our study, using the KLE. The KLE has been explained in [10–12]. A very short discussion on the KLE is being presented below for the sake of completeness.

Suppose  $\log[k(\mathbf{x})] = Y^k(\mathbf{x})$ , where  $\mathbf{x} \in \Omega \subset \mathbf{R}^2$ ,  $Y^k(\mathbf{x})$  is a Gaussian field and the covariance function  $\text{Cov}(Y^k(\mathbf{x}_1), Y^k(\mathbf{x}_2))$  is given by the following formula:

$$R(\mathbf{x}_1, \mathbf{x}_2) = \sigma_Y^2 \exp\left(-\frac{|x_1 - x_2|^2}{2L_x^2} - \frac{|y_1 - y_2|^2}{2L_y^2}\right), \quad (4)$$

where  $L_x$  and  $L_y$  are the correlation lengths in  $x$ - and  $y$ -direction, respectively, and  $\sigma_Y^2 = \text{Var}[(Y^k)^2]$ . It is assumed that  $Y^k(\mathbf{x})$  is a second-order stochastic process and  $E[(Y^k)^2] = 0$ . Thus, for a given orthonormal basis  $\{\varphi_i\}$  in  $L^2(\Omega)$ ,  $Y^k(\mathbf{x})$  can be expressed as the following:

$$Y^k(\mathbf{x}) = \sum_{i=1}^{\infty} Y_i^k \varphi_i(\mathbf{x}), \quad \text{with} \quad Y_i^k = \int_{\Omega} Y^k(\mathbf{x}) \varphi_i(\mathbf{x}) d\mathbf{x}, \quad (5)$$

where  $Y_i^k$  are random coefficients. On the other hand, since  $L^2$  is a complete space, thus  $\varphi_i(\mathbf{x})$  is an eigenfunction satisfying

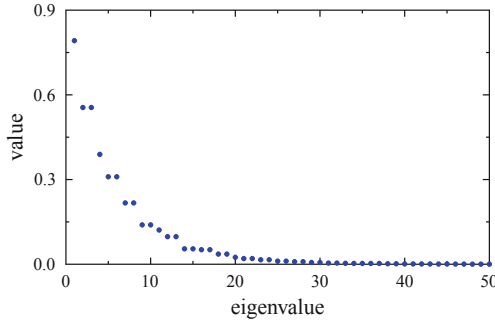
$$\int_{\Omega} R(\mathbf{x}_1, \mathbf{x}_2) \varphi_i(\mathbf{x}_2) d\mathbf{x}_2 = \lambda_i \varphi_i(\mathbf{x}_1), \quad i = 1, 2, \dots, \quad (6)$$

and the corresponding eigenvalue  $\lambda_i = E[(Y_i^k)^2] > 0$ . By using the assumption  $\theta_i = Y_i^k / \sqrt{\lambda_i}$ , the KLE in Eq. (5) can be expressed as the following:

$$Y^k(\mathbf{x}) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \theta_i \varphi_i(\mathbf{x}). \quad (7)$$

If the eigenvalues decrease, a truncated KLE can be written as

$$Y_d^k(\mathbf{x}) = \sum_{i=1}^d \sqrt{\lambda_i} \theta_i \varphi_i(\mathbf{x}). \quad (8)$$



**Fig. 2.** Decay of eigenvalues in KLE

If the eigenvalues in (8) decay quickly, the  $Y_d^k$  will be a good approximation of  $Y^k$ .

We set  $L_x = L_y = 0.2$  and  $\sigma_Y^2 = 4$  in (4). Figure 2 shows that the eigenvalues decay fast for these values. In this paper we thus consider the first twenty eigenvalues in the KLE to model the permeability field.

## 4 Bayesian Framework

### 4.1 Exploration of the Posterior

In this subsection we discuss how to characterize the permeability field using the fractional flow curves at the monitoring wells of the aquifer. Let us denote the fractional flow data by  $F_m$  and the corresponding permeability field by  $\psi$ . Using the Bayes’ theorem we can write the posterior probability

$$P(\psi|F_m) \propto P(F_m|\psi)P(\psi), \tag{9}$$

where  $P(\psi)$  denotes the prior distribution and the normalizing constant is ignored due to the iterative search in the MCMC algorithm. The  $\psi$  is generated through the KLE, for which the vector  $\theta$  is used as input in the expansion. In the remainder of the discussion we use  $\psi = \text{KLE}[\theta]$ . Moreover, we consider a Gaussian distribution as in [9] for the likelihood function, i.e.,

$$P(F_m|\psi) \propto \exp\left(- (F_m - F_\psi)^\top \Sigma (F_m - F_\psi)\right), \tag{10}$$

where the simulated fractional flow data  $F_\psi$  is obtained by the numerical solution from the GPU simulator for each permeability distribution  $\psi$  in the MCMC algorithm. We denote the covariance matrix by  $\Sigma$ , which is defined as  $\Sigma = \mathbf{I}/2\sigma_F^2$  where  $\mathbf{I}$  and  $\sigma_F^2$  are the identity matrix and the precision parameter, respectively.

We sample data from the posterior by using the Metropolis-Hasting (MH) MCMC and create a Markov chain, which has the posterior distribution as target distribution. We consider an instrumental distribution  $q(\psi_p|\psi)$ , where  $\psi$

denotes the previously accepted proposal, in order to propose  $\boldsymbol{\psi}_p = \text{KLE}[\boldsymbol{\theta}_p]$  at every iteration. We use the following acceptance probability in MH MCMC and the probability value is computed by solving the forward problem for a given permeability distribution on the numerical simulator:

$$\alpha(\boldsymbol{\psi}, \boldsymbol{\psi}_p) = \min \left( 1, \frac{q(\boldsymbol{\psi}|\boldsymbol{\psi}_p)P(\boldsymbol{\psi}_p|F_m)}{q(\boldsymbol{\psi}_p|\boldsymbol{\psi})P(\boldsymbol{\psi}|F_m)} \right). \quad (11)$$

We now describe the two-stage and DREAM MCMCs that use the MH algorithm.

## 4.2 Two-Stage MCMC

Here we present the two-stage MCMC method. The method has been widely used for porous media applications [8, 16]. The two-stage MCMC consists of a screening procedure, which relies on a coarse-scale model approximating the governing equations (1). The coarse-scale discretization is done in a similar way as in the fine-scale discretization. The main idea lies on a rigorous projection of  $k$  on the coarse-scale that is obtained from the fine-scale resolution. For this reason, an upscaling method is used so that the effective permeability values on the coarse-scale yield a similar average response as that of the underlying fine-scale problem locally [11]. We then run the numerical simulator on the coarse-scale model and get the numerical solution  $F_c$ . Now we can compute the coarse-scale and fine-scale acceptance probabilities

$$\begin{aligned} \alpha_c(\boldsymbol{\psi}, \boldsymbol{\psi}_p) &= \min \left( 1, \frac{q(\boldsymbol{\psi}|\boldsymbol{\psi}_p)P_c(\boldsymbol{\psi}_p|F_m)}{q(\boldsymbol{\psi}_p|\boldsymbol{\psi})P_c(\boldsymbol{\psi}|F_m)} \right) \text{ and} \\ \alpha_f(\boldsymbol{\psi}, \boldsymbol{\psi}_p) &= \min \left( 1, \frac{P_f(\boldsymbol{\psi}_p|F_m)P_c(\boldsymbol{\psi}|F_m)}{P_f(\boldsymbol{\psi}|F_m)P_c(\boldsymbol{\psi}_p|F_m)} \right), \text{ respectively.} \end{aligned} \quad (12)$$

In the two-stage MCMC the following Random Walk Sampler (RWS) is used:

$$\boldsymbol{\theta}_p = \beta \boldsymbol{\theta} + \sqrt{1 - \beta^2} \boldsymbol{\epsilon}, \quad (13)$$

where  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}_p$  represent the previously accepted proposal and the current proposal, respectively. The symbol  $\boldsymbol{\epsilon}$  stands for a  $\mathcal{N}(0, 1)$ -random variable and  $\beta$  ( $= 0.75$ ) is a tuning parameter [1]. The two-stage MCMC algorithm is presented in Algorithm 1. The convergence diagnostic to break the “for loop” in the algorithm is described in Sect. 5.

## 4.3 DREAM MCMC

The DREAM, which is an extension of DE MCMC [3], runs multiple MCMCs simultaneously for a thorough exploration of the posterior, and has an in-built mechanism to adapt the scale and orientation of the proposal distribution during the evolution to the posterior distribution [22]. We describe the DREAM for our

---

**Algorithm 1.** Two-stage MCMC

---

- 1: Given covariance function  $R$  generate KLE.
  - 2: **for**  $p = 1$  to  $M_{\text{mcmc}}$  **do**
  - 3:   At  $\boldsymbol{\psi} = \text{KLE}[\boldsymbol{\theta}]$  using (13) generate  $\boldsymbol{\psi}_p = \text{KLE}[\boldsymbol{\theta}_p]$ .
  - 4:   Compute the upscaled permeability on the coarse-scale using  $\boldsymbol{\psi}_p$ .
  - 5:   Solve the forward problem on the coarse-scale to get  $F_c$ .
  - 6:   Compute the coarse-scale acceptance probability  $\alpha_c(\boldsymbol{\psi}, \boldsymbol{\psi}_p)$ .
  - 7:   **if**  $\boldsymbol{\psi}_p$  is accepted **then**
  - 8:     Use  $\boldsymbol{\psi}_p$  in the fine-scale simulation to get  $F_f$ .
  - 9:     Compute the fine-scale acceptance probability  $\alpha_f(\boldsymbol{\psi}, \boldsymbol{\psi}_p)$ .
  - 10:     **if**  $\boldsymbol{\psi}_p$  is accepted **then**  $\boldsymbol{\psi} = \boldsymbol{\psi}_p$ .
  - 11:     **end if**
  - 12:   **end if**
  - 13: **end for**
- 

application below. In Algorithm 2,  $m$  denotes the number of parallel chains that we run simultaneously.

---

**Algorithm 2.** DREAM MCMC

---

- 1: Given covariance function  $R$  generate KLE.
  - 2: **for**  $c = 1$  to  $m$  **do**
  - 3:   **for**  $p = 1$  to  $M_{\text{mcmc}}$  **do**
  - 4:     At  $\boldsymbol{\psi} = \text{KLE}[\boldsymbol{\theta}]$  using equation (14) generate  $\boldsymbol{\psi}_p = \text{KLE}[\boldsymbol{\theta}_p]$ .
- 

$$\boldsymbol{\theta}_p = \boldsymbol{\theta} + (I_d + f) \gamma(\delta, d') \left( \sum_{j=1}^{\delta} \boldsymbol{\theta}^{r_1(j)} - \sum_{k=1}^{\delta} \boldsymbol{\theta}^{r_2(k)} \right) + \boldsymbol{\epsilon}, \tag{14}$$

where  $\delta$  and  $d'$  denote the number of pairs that are used to generate the proposal and the number of parameters that are updated jointly in each iteration, respectively. Two randomly chosen chains are denoted by  $r_1$  and  $r_2$ . Moreover,  $f$  and  $\boldsymbol{\epsilon}$  are drawn independently from  $\text{CU}[-b, b]$  and  $\mathcal{N}(0, b^*)$ , respectively, where  $\text{CU}[a, b]$  represents the continuous uniform distribution on the interval  $[a, b]$ .

---

**Algorithm 3.** DREAM MCMC (continued)

---

- 5:   Use  $\boldsymbol{\psi}_p$  to get  $F_f$  (on the fine-scale).
  - 6:   Compute the acceptance probability  $\alpha(\boldsymbol{\psi}, \boldsymbol{\psi}_p)$  using (11).
  - 7:   **if**  $\boldsymbol{\psi}_p$  is accepted **then**  $\boldsymbol{\psi} = \boldsymbol{\psi}_p$ .
  - 8:   **end if**
  - 9:   **end for**
  - 10: **end for**
-



We simultaneously run the DREAM MCMC with  $m = 11$  parallel chains. In (14) we set  $\delta = 5$ ,  $b = 0.1$  and  $b^* = 10^{-6}$ . The jump rate is given by  $\gamma = \frac{2.38\beta_0}{\sqrt{2\delta d'}}$ , where the constant  $\beta_0 = \frac{1}{16}$ . The user should select the value of  $\beta_0$  in such a way the MCMC method has an acceptance rate between 15–35%. In the present study we set  $d' = 5$ , i.e., we update five parameters at every iteration. In addition to setting those parameters, we also set  $\gamma = 1.0$  at every tenth iteration to encourage a jump between two disconnected posterior modes.

### 5 Convergence Diagnostics of MCMC Methods

The DREAM MCMC requires a parallel simulation of multiple chains simultaneously. However, the two-stage MCMC does not require the same. Due to the computational burden in repeatedly solving the numerical simulator, we still run the parallel simulation of several two-stage MCMCs. Now we need to investigate the convergence of multiple chains in each method for a reliable characterization of the permeability field of the aquifer. In this section we discuss convergence diagnostics for that purpose. The Potential Scale Reduction Factor (PSRF) and its multivariate extension MPSRF are used to measure the convergence of multiple MCMCs [4].

We set the number of parameters as  $d = 20$  in  $\theta$  as discussed in Sect. 3. Let us consider  $m$  chains and  $n$  posterior draws of  $\theta$  in each chain.  $\theta_i^t$  refers to the vector  $\theta$  at iteration  $t$  in the  $i$ th chain of multiple MCMCs. Then the posterior variance-covariance matrix in higher dimensions is computed by

$$\widehat{\mathbf{V}} = \frac{n-1}{n} \mathbf{W} + \left(1 + \frac{1}{m}\right) \frac{\mathbf{B}}{n}. \tag{15}$$

The within-covariance-matrix  $\mathbf{W}$  is given by

$$\mathbf{W} = \frac{\mathbf{1}}{\mathbf{m}(\mathbf{n}-1)} \sum_{i=1}^{\mathbf{m}} \sum_{t=1}^{\mathbf{n}} (\theta_i^t - \bar{\theta}_i) (\theta_i^t - \bar{\theta}_i)', \tag{16}$$

and the between-chain-covariance-matrix  $\mathbf{B}$  is computed by

$$\mathbf{B} = \frac{\mathbf{n}}{\mathbf{m}-1} \sum_{i=1}^{\mathbf{m}} (\bar{\theta}_i - \bar{\theta}_{..}) (\bar{\theta}_i - \bar{\theta}_{..})', \tag{17}$$

where  $\bar{\theta}_i$  denotes the mean of  $\theta$ s within the chain and  $\bar{\theta}_{..}$  represents the mean of  $\theta$ s in all the chains. The PSRF is defined as follows:

$$\text{PSRF}_p = \sqrt{\frac{\text{diag}(\widehat{\mathbf{V}})_{\mathbf{p}}}{\text{diag}(\mathbf{W})_{\mathbf{p}}}}, \quad \text{where } p = 1, 2, \dots, d. \tag{18}$$

The PSRF values close to one indicate that the samples in multiple chains are being generated from the same limiting distribution and thus confirm the convergence. The MPSRF is computed as follows [17]:

$$\text{MPSRF} = \sqrt{\left(\frac{n-1}{n} + \left(\frac{m+1}{m}\right) \lambda_1\right)}, \tag{19}$$

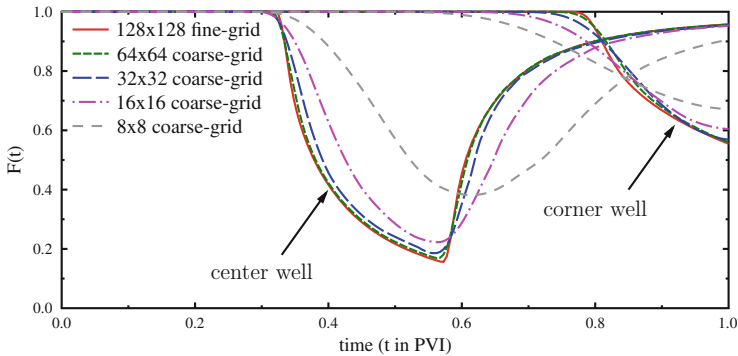
where  $\lambda_1$  is the greatest eigenvalue of  $\mathbf{W}^{-1}\mathbf{B}/\mathbf{n}$ . If the MPSRF approaches one for a reasonably large  $n$ , the convergence of the multiple chains is ensured. Moreover, a relationship between the PSRFs and MPSRF is given by [4]

$$\max \text{ of PSRFs } \leq \text{MPSRF}. \tag{20}$$

## 6 Numerical Results

### 6.1 Simulation Study

In this subsection we present a simulation study of the characterization of the heterogeneous permeability field of the aquifer. See Fig. 1. The aquifer is not contaminated initially. The contaminated water flows through a spill well into the aquifer at a rate of one pore-volume every five years. The synthetic reference permeability field is constructed on a fine-grid of size  $128 \times 128$ . The fractional flow curves for this reference field at the monitoring wells are shown in Fig. 5. These curves are generated by running the numerical simulator on the reference field until  $t = 1.0$  PVI. The time evolution of the contaminant flow on the reference permeability field is shown in Fig. 1. We use the two-stage and DREAM MCMC methods to generate samples from the posterior. A coarse-grid of size  $32 \times 32$  is chosen for the two-stage algorithm, which runs four times faster than the fine-grid simulation, but still manages to capture the general trend of the flow. See Fig. 3.



**Fig. 3.** A comparison of simulated fractional flows obtained from coarse-scale models and the fine-scale model.

### 6.2 Convergence Analysis

This subsection contains the MPSRF and PSRF analysis for both two-stage and DREAM MCMCs. We run twelve and eleven chains for two-stage and DREAM MCMCs, respectively. Using an equal number of total proposals in all chains,

we then compute the maximum of PSRFs and the MPSRF against the number of iterations. Figure 4 shows that the MPSRF for each method is the upper bound of the maximum of the PSRFs, which is consistent with the inequality shown in (20). Moreover, it is also observed that both MPSRF and the maximum of PSRFs for the DREAM MCMC have a faster downward trend than the two-stage MCMC method. Thus, we can conclude that the DREAM MCMC samples from the posterior faster than the two-stage MCMC and converges faster towards the stationary distribution. However, Fig. 4 demonstrates that both MCMC methods need a large number of iterations to achieve a complete convergence. To achieve a complete convergence, the curves should approach the numerical value of one.

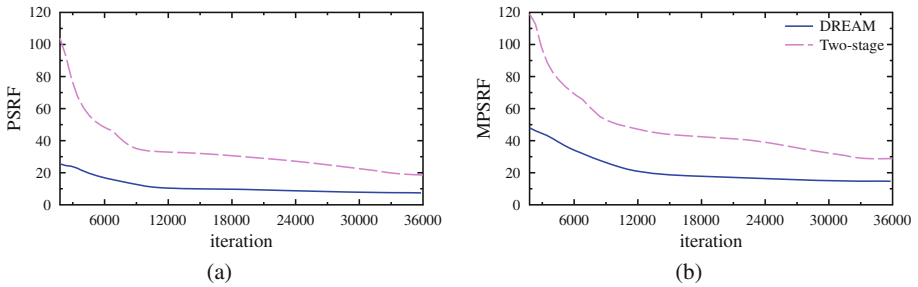


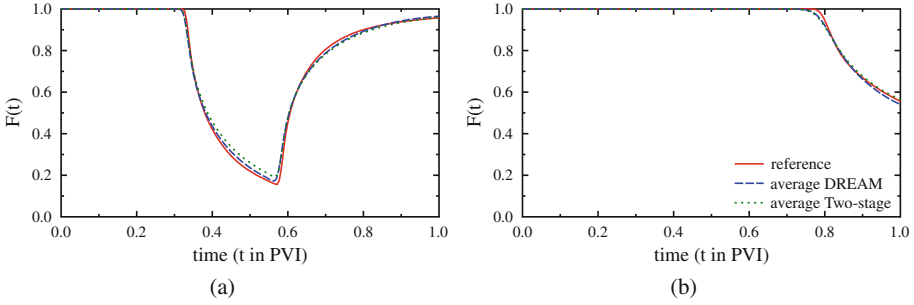
Fig. 4. The maximum of PSRFs and the MPSRF for DREAM and two-stage MCMCs

Next we focus on a statistical analysis of two ensembles of accepted fractional flow curves. In the analysis we consider the variances of the ensembles of fractional flow curves as well as the posterior distributions.

### 6.3 Variance Analysis

We compute the average production curves using 24000 accepted proposals and compare those curves with the reference fractional flow curves. See Fig. 5. From the comparison we can say that both two-stage and DREAM MCMCs produce very similar fractional flow curves. Table 1 shows the acceptance rates for both MCMCs, where  $\sigma_F^2$  and  $\sigma_C^2$  are precision parameters in (10) for coarse- and fine-scale simulations, respectively. Note that the acceptance rates for both MCMC methods are the same, however, the convergence rate in the DREAM MCMC is considerably higher than that in the two-stage MCMC (see Fig. 4).

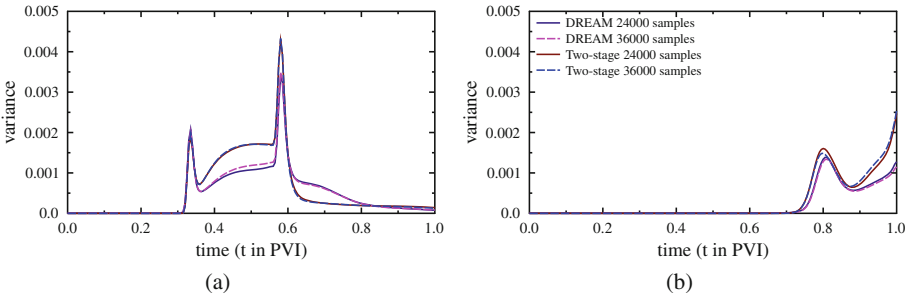
We now construct two ensembles by taking the same number of fractional flow curves in each MCMC method: The first ensemble contains 24000 samples and the second one has 36000 samples. Figure 6 displays the variances of the fractional flow curves for the center and corner wells of the aquifer. The variance curves differ between not only both ensembles but also both MCMC methods.



**Fig. 5.** Fractional flow curves of (a) center and (b) corner wells for two-stage and DREAM MCMCs.

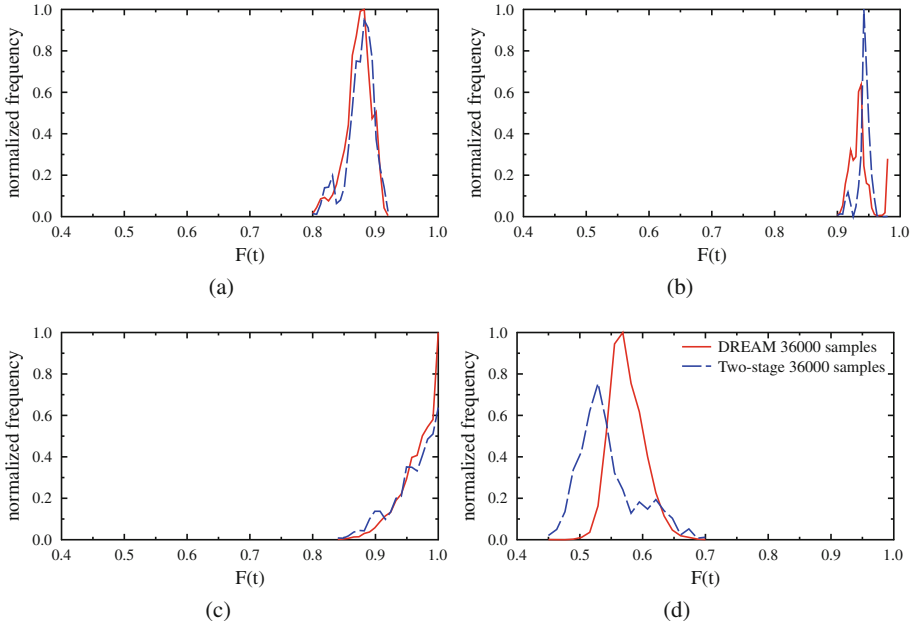
**Table 1.** A comparison of accepted proposals for two-stage and DREAM MCMCs

	Two-stage MCMC	DREAM MCMC
$\sigma_F^2$	$10^{-4}$	$10^{-4}$
$\sigma_C^2$	$2 \times 10^{-4}$	—
Acceptance rate	35%	35%



**Fig. 6.** Variance plots of fractional flow curves at (a) the center well and (b) corner well for two-stage and DREAM MCMCs.

Furthermore, we sketch the posterior densities in Fig. 7 for the same ensembles that we considered for the variance analysis. The normalized frequencies reveal that the posterior densities are also different for both MCMC methods.



**Fig. 7.** Normalized frequency plots of fractional flow curves for two-stage and DREAM MCMCs. Top: Center well. Bottom: Corner well. Left: At  $t = 0.78$  PVI. Right: At  $t = 0.98$  PVI.

## 7 Conclusions

Using a GPU-based single-phase flow simulator we have compared two frequently used MCMCs, the two-stage algorithm based on a random walk sampler and the DREAM. We have confirmed that the DREAM converges much faster than the two-stage MCMC by comparing the corresponding PRSF and MPRSF curves. Moreover, a careful statistical analysis of ensembles of accepted fractional flow curves, produced by the two MCMCs, reveals that such ensembles share essentially the same average behavior. However, significant differences have been observed in the time-dependent variance curves as well as in the posterior distributions for those ensembles. This provides an indication that, for the purpose of making Monte Carlo predictive simulations, one might observe considerable differences in the results. One could combine both MCMC methods in a two-stage version of DREAM to achieve good convergence along with reduced computational cost in line with the work of [14]. As a future work we intend to combine both methods for our problem and study the convergence.

**Acknowledgment.** A. Rahunathan's research was supported by National Science Foundation under Grant No. HRD-1600818. The computer simulations were performed on the NSF-funded GPU computing cluster housed in the Department of Mathematics and Computer Science at Central State University.

## References

1. Akbarabadi, M., Borges, M., Jan, A., Pereira, F., Piri, M.: A Bayesian framework for the validation of models for subsurface flows: synthetic experiments. *Comput. Geosci.* **19**(6), 1231–1250 (2015). <https://doi.org/10.1007/s10596-015-9538-z>
2. Al-Mamun, A., Barber, J., Ginting, V., Pereira, F., Rahunanthan, A.: Contaminant transport forecasting in the subsurface using a Bayesian framework. *Appl. Math. Comput.* (2020). <https://doi.org/10.1016/j.amc.2019.124980>
3. Braak, C.: A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. *Stat. Comput.* **16**, 239–249 (2006)
4. Brooks, S., Gelman, A.: General methods for monitoring convergence of iterative simulations. *J. Comput. Graph. Stat.* **7**, 434–455 (1998)
5. Chen, Z., Huan, G., Ma, Y.: *Computational Methods for Multiphase Flows in Porous Media*. SIAM, Philadelphia (2006)
6. Christen, J., Fox, C.: Markov chain Monte Carlo using an approximation. *J. Comput. Graph. Stat.* **14**(4), 795–810 (2005)
7. Cui, T., Fox, C., O’Sullivan, M.J.: Bayesian calibration of a large-scale geothermal reservoir model by a new adaptive delayed acceptance metropolis Hastings algorithm. *Water Resour. Res.* **47**(10) (2011)
8. Efendiev, Y., Datta-Gupta, A., Ginting, V., Ma, X., Mallick, B.: An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resour. Res.* **41**(12) (2005)
9. Efendiev, Y., Hou, T., Luo, W.: Preconditioning Markov chain Monte Carlo simulations using coarse-scale models. *SIAM J. Sci. Comput.* **28**(2), 776–803 (2006)
10. Ginting, V., Pereira, F., Rahunanthan, A.: Multiple Markov chains Monte Carlo approach for flow forecasting in porous media. *Procedia Comput. Sci.* **9**, 707–716 (2012)
11. Ginting, V., Pereira, F., Rahunanthan, A.: A multi-stage Bayesian prediction framework for subsurface flows. *Int. J. Uncertain. Quantif.* **3**(6), 499–522 (2013)
12. Ginting, V., Pereira, F., Rahunanthan, A.: A prefetching technique for prediction of porous media flows. *Comput. Geosci.* **18**(5), 661–675 (2014). <https://doi.org/10.1007/s10596-014-9413-3>
13. Ginting, V., Pereira, F., Rahunanthan, A.: Multi-physics Markov chain Monte Carlo methods for subsurface flows. *Math. Comput. Simul.* **118**, 224–238 (2015)
14. Laloy, E., Rogiers, B., Vrugt, J.A., Mallants, D., Jacques, D.: Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion. *Water Resour. Res.* **49**(5), 2664–2682 (2013)
15. Loève, M.: *Probability Theory*. Springer, Heidelberg (1997)
16. Ma, X., Al-Harbi, M., Datta-Gupta, A., Efendiev, Y.: An efficient two-stage sampling method for uncertainty quantification in history matching geological models. *SPE J.* **13**(1), 77–87 (2008)
17. Malyskhina, N.: *Markov switching models: an application to roadway safety*. Ph.D. thesis, Purdue University (2008)
18. Pereira, F., Rahunanthan, A.: A semi-discrete central scheme for the approximation of two-phase flows in three space dimensions. *Math. Comput. Simul.* **81**(10), 2296–2306 (2011)
19. Shockley, E.M., Vrugt, J.A., Lopez, C.F.: PyDREAM: high-dimensional parameter inference for biological models in python. *Bioinformatics* **34**(4), 695–697 (2017)

20. Vrugt, J.A., ter Braak, C.J.F.: Dream<sub>(D)</sub>: an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, noncontinuous, and combinatorial posterior parameter estimation problems. *Hydrol. Earth Syst. Sci.* **15**(12), 3701–3713 (2011)
21. Vrugt, J.: Markov chain Monte Carlo simulation using the DREAM software package: theory, concepts, and MATLAB implementation. *Environ. Model Softw.* **75**, 273–316 (2016)
22. Vrugt, J.A., ter Braak, C.J.F., Clark, M.P., Hyman, J.M., Robinson, B.A.: Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resour. Res.* **44**(12) (2008)