Multiple Markov Chains Monte Carlo Approach for Flow Forecasting in Porous Media

V. Ginting, F. Pereira, A. Rahunanthan

Department of Mathematics and Center for Fundamentals of Subsurface Flow, School of Energy Resources, University of Wyoming, Laramie, WY 82071

Abstract

Predictions in subsurface formations consists of two steps: characterization and prediction using the characterization. In the characterization, we reconstruct the subsurface properties, such as distributions of permeability and porosity, with a set of limited data. A Bayesian approach using Markov Chain Monte Carlo (MCMC) methods is well suited for reconstructing permeability and porosity fields. This statistical approach aims at generating a Markov chain from which a stationary, posterior distribution of the characteristics of the subsurface may be constructed. A crucial step in this framework is the calculation of the likelihood information which can be computationally very demanding. This limitation hinders the application of the Bayesian framework for the flow predictions in porous media in a practical period of time. The parallel computation of multiple MCMCs can substantially reduce computation time and can make the framework more suitable to subsurface flows. In this paper, we consider multi–MCMC and compare the multi–MCMC with the MCMCs for the predictions of subsurface flows.

Keywords:
Bayesian Inference, Forecasting in Porous Media, Multiple MCMCs

1. Introduction

Oil recovery and energy security, CO₂ sequestration, and monitoring and remediation of subsurface aquifer contamination each may be viewed as forecasting problems wherein we want to predict quantities—fractions of oil and water, CO₂ concentration, or concentration of radioactive particles—using subsurface fluid flow models with expertise and limited data.

Our strategy consists of establishing a complete statistical description of subsurface properties, such as permeability and porosity that are conditioned to existing measurement data. A Bayesian approach using Markov Chain Monte Carlo (MCMC) methods is well suited for reconstructing permeability and porosity fields. This statistical approach aims at generating a Markov chain from which a stationary, posterior distribution of the characteristics of
the subsurface may be constructed. A crucial step in this framework is the calculation of the likelihood information which involves solving coupled partial differential equations with permeability and porosity as input parameters. The flow simulator computation time and the sequential nature of MCMC simulation limit the posterior exploration in a practical period of time. To make things worse, we often explore a high-dimensional posterior distribution in these applications.

It is imperative to address this limitation appropriately using parallel MCMCs that make the posterior exploration tractable. An MCMC algorithm can be parallelized in two ways: (1) running multiple MCMCs in parallel, or (2) parallelization of a single MCMC by pre-fetching algorithm [6]. The debate on one long chain or several shorter chains is still very loud among statisticians [1, 2]. In a single long MCMC, the chain will be “closer” to the target distribution at the end of the run compared to the many shorter MCMC runs. High-dimensional problems often have longer burn-in periods and that several shorter runs may be a waste of resources since the burn-in period must be discarded from each shorter chain. Geyer [1] argues that if you cannot get a good answer with one long run, then you cannot get a good answer with many shorter runs either. In contrast, the camp which supports several multiple chains argues that, although a single run will eventually cover the entire sample space, by taking a number of parallel MCMCs, we can make sure that no portion of the sample space is unexplored. They also argue that several shorter runs can better diagnose non-convergence.

Since the forecasting production in porous media requires long simulation times for likelihood computations and the convergence rate is very slow, we want to speed-up generation of a single MCMC. The regeneration technique [3, 4], which is suitable typically for problems with low-dimensional state-space, is not applicable in this case, because our application entails exploration of very high-dimensional posterior distribution. When the state-space is high-dimensional, we can divide the state-space into blocks, and then for each iteration of the MCMC we can send each block to a separate GPU [5]. This approach helps to speed-up the generation of single MCMC. However, it requires additional effort to carry out analysis of the limiting distribution in order to come up with appropriate blocks. When the conditional dependence structure in a limiting distribution is complicated, it can be very challenging. Brockwell in [6] presented a case study in which this approach is not applicable. As an alternative, he presented a pre-fetching algorithm which does not require any particular analysis of the limiting distribution of the MCMC to speed-up the generation of a single MCMC by parallelization. In this algorithm, the multiple likelihoods are computed ahead of time and in turn only the ones needed are used to proceed. This algorithm obviously has limitations in terms of parallel efficiency. Strid showed how to use available information to make better predictions of the future states of the MCMC chain and substantially increase the efficiency of the pre-fetching algorithm [7]. As indicated in a recent study by the authors [8], it helps to speed-up the generation of a single MCMC. However, for more challenging real problems this speed-up may not be adequate to offer a reliable prediction in a practical period of time.

One would like to have a single MCMC for the exploration. However, considering the simulation time needed for the applications in porous media, there is no hope of exploring all the possibilities in a practical period of time in a single MCMC. For the prediction, we want to explore all the possibilities (if we can) of the productions that may occur given the partial production curves. Based on the available production data, we collect all the possible profiles of the parameters that give almost similar production curves with the specified precision associated with the measurement and the simulated numerical solution. For this purpose, multiple MCMCs can be very attractive. In the approach of multiple MCMCs, if we start with a different seed for each MCMC, we can increase the chance of exploring more possibilities. This is more on the desirable side of the prediction. Also, the multi-MCMC uses the hardware more efficiently and consequently, is very applicable to more computationally demanding problems. In this paper, we consider multi-MCMC for the prediction of production for a two-phase flow problem.

This paper is organized as follows. We discuss the physical and mathematical modeling of the problem at hand in section 2. The parametrization of uncertainty using the Karhunen-Loève expansion for unknown permeability and porosity fields is discussed in section 3. In section 4, we discuss a Bayesian approach for quantifying uncertainty in both permeability and porosity fields. In the numerical simulations in section 5, we implement multi-MCMC to predict the partial production curves in an oil reservoir and compare the multi-MCMC to the MCMCs. Section 6 contains our conclusions.
2. Physical and Mathematical Modeling

The dynamics of the motion of fluids in a heterogeneous reservoir $\Omega$, which is considered here, is categorized as an immiscible two-phase system with water and oil (denoted by $w$ and $o$, respectively) that is incompressible. To simplify the model, capillary pressure and gravity are not included, there is no mass transfer between phases, and the two fluids fill the pore space. The reservoir is equipped with an injection well from which water is discharged to displace the trapped oil towards the production wells. The injection well is located in one of the corners. We have two production wells: one well is situated along the diagonal, opposite to the injection well, and the other is situated at the center of a side which is one of the two sides that enclose the production well at the corner (see Figure 2). The wells are modeled through appropriate boundary conditions. The governing equations of flow and transport are

$$\nabla \cdot \mathbf{v} = 0, \quad \text{where} \quad \mathbf{v} = -\lambda(s)k(x)\nabla p, \quad x \in \Omega,$$

and

$$\phi(x) \frac{\partial s}{\partial t} + \nabla \cdot (f(s)\mathbf{v}) = 0,$$

where $\mathbf{v}$ is the Darcy velocity, $s$ is the water saturation, $k$ is the absolute permeability and $\phi$ is the porosity. The total mobility $\lambda(s)$ and the flux function $f(s)$ are respectively given by:

$$\lambda(s) = \frac{k_{rw}}{\mu_w} + \frac{k_{ro}}{\mu_o}, \quad f(s) = \frac{k_{rw}(s)/\mu_w}{\lambda(s)},$$

where $k_{rj}, j = w, o$, is the relative permeability of the phase $j$ [9].

We consider sampling of two important physical quantities—permeability and porosity—of the oil reservoir using partial fractional curves as governing measurements. For each production boundary the fractional flow $F(t)$ is defined as the fraction of oil in the produced fluid, i.e.,

$$F(t) = 1 - \frac{\int_{\partial \Omega_{out}} v_n f(s) \, dl}{\int_{\partial \Omega_{out}} v_n \, dl},$$

where $\partial \Omega_{out}$ denotes outflow boundary, $v_n$ is normal velocity field and $t$ is the dimensionless time measured in Pore Volume Injected (PVI), which is computed as

$$\text{PVI} = \int_0^T V_p^{-1} \int_{\partial \Omega_{out}} v_n \, dl \, d\tau,$$

where $V_p$ is the total pore-volume of the reservoir and $T$ denotes the time taken for injection of water. The system (2.1) is simulated using an efficient and reliable physics-based operator splitting technique (see for example [10, 11] for further discussion), which is implemented on GPU machines.

3. Parametrization of Uncertainty

In the current work, we employ the Karhunen-Loève expansion (KLE) [12, 13] to reduce the potentially large dimension of the uncertainty space describing the permeability and porosity which is accomplished through appropriate parametrization inherent in the expansion (see for example [14, 15, 16, 17, 18] for similar applications). A standard assumption in geostatistics is to model the permeability to follow a log-normal distribution [19], i.e.,

$$\log[k(x, \omega)] = Y^k(x, \omega),$$

where $x \in \Omega \subset \mathbb{R}^2$, and $\omega$ is a random element in a probability space, and $Y^k(x, \omega)$ is a field possessing a Gaussian distribution and a covariance function

$$R(x_1, x_2) = \sigma_y^2 \exp\left(-\frac{|x_1 - x_2|^2}{2L_x^2} - \frac{|y_1 - y_2|^2}{2L_y^2}\right),$$

$$= \sigma_y^2 \exp\left(-\frac{1}{2}|L^{-1}(x_1 - x_2)|^2\right),$$

(3.1)
0.0 0.3 0.6 0.9

value

0 10 20 30 40 50
eigenvalue

Figure 1: Eigenvalues of the KLE for the Gaussian covariance with \( L_x = L_y = 0.2 \) and \( \sigma_Y^2 = 4.0 \).

Here, the correlation length is the distance from a point in \( \Omega \) beyond which there is no further correlation of a physical property (permeability or porosity) associated with that point. The series representation of \( Y_k(x, \omega) \) is

\[
Y_k(x, \omega) = \sum_{i=1}^{\infty} \theta_k^i(\omega) \varphi_i(x), \quad \text{with} \quad \theta_k^i(\omega) = \int_{\Omega} Y_k(x, \omega) \varphi_i(x) dx
\]

(3.2)

being functions of a random variable, and \( \varphi_i \) a set of basis functions satisfying

\[
\int_{\Omega} R(x_1; x_2) \varphi_i(x_2) dx_2 = \lambda_i \varphi_i(x_1), \quad i = 1, 2, ..., \tag{3.3}
\]

that makes \( Y_k^i \) uncorrelated, and \( \lambda_i = E[(Y_k^i)^2] > 0 \). Denoting \( \theta_k^j = Y_k^j / \sqrt{\lambda_j} \), then \( \theta_k^j \) satisfies \( E(\theta_k^j) = 0 \) and \( E(\theta_k^j \theta_k^j) = \delta_{ij} \), and thus

\[
Y_k(x, \omega) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \theta_k^i(\omega) \varphi_i(x) \approx \sum_{i=1}^{N_k} \sqrt{\lambda_i} \theta_k^i(\omega) \varphi_i(x). \tag{3.4}
\]

We assume that eigenvalues \( \{\lambda_i\}_{i=1}^{\infty} \) are ordered so that \( \lambda_1 \geq \lambda_2 \geq \cdots \). The basis functions \( \varphi_i(x) \) in (3.3) are deterministic and resolve the spatial dependence of the permeability field and in particular its correlation structure. The uncertainty is represented by the scalar random variables \( \theta_i^k \). In general, we only need to keep the leading order terms (quantified by the magnitude of \( \lambda_i \)) and still capture most of the energy of the stochastic process \( Y_k(x, \omega) \).

With respect to the porosity field, we make use of the standard assumption that the porosity exhibits a similar spatial correlation structure to the permeability. In turn, this allows us to employ (3.4). The dependence of porosity to the expansion is expressed as

\[
\phi(x) = \frac{\phi_{\min} + \phi_{\max} e^{Y_\phi}}{1 + e^{Y_\phi}}, \quad \phi_{\min} \text{ and } \phi_{\max} \in (0, 1),
\]

(3.5)

where \( Y_\phi \) is KLE for porosity as in (3.4), and \( \phi_{\min} \) and \( \phi_{\max} \) are the lower and upper limits of the porosity of the reservoir.

4. Bayesian Inference

As alluded to earlier, we want to sample the permeability and porosity fields conditioned on the available fractional flow data \( F_m \). This is translated into sampling from the conditional distribution \( P(\psi|F_m) \), where \( \psi = [\theta^k \theta^\phi] \) with \( \theta^k \)
Figure 2: Top: Left to right the permeability (in log) and porosity distributions of the underlying field, respectively. Bottom: Left to right water saturation plots at $t = 0.4$ PVI and $t = 1.4$ PVI, respectively.
and $\theta^k$ vectors containing the random coefficients in the KL expansions. According to Bayes’ theorem this distribution satisfies the proportionality relation

$$P(\psi|F_m) \propto P(F_m|\psi)P(\psi),$$

(4.1)

where $P(F_m|\psi)$ represents the likelihood function (that requires the forward solution of the two-phase flow) and $P(\psi)$ is the prior distribution of $\psi$. The normalizing constant in this expression is not important, because we use an iterative updating procedure. We assume that the likelihood function follows a Gaussian distribution, i.e.,

$$P(F_m|\psi) \propto \exp\left(-\frac{(F_m - F_\psi)^T \Sigma (F_m - F_\psi)}{2\sigma^2_F}\right),$$

(4.2)

where $F_\psi$ is the simulated fractional flow curve that is obtained by solving the forward problem with known permeability $k$ and porosity $\phi$, in other words with known $\psi$, and $\Sigma$ is the covariance matrix representing the measurement errors. We take $\Sigma = I/2\sigma^2_F$, where $I$ is the identity matrix and $\sigma^2_F$ is the precision associated with the measurement $F_m$ and numerical solution $F_\psi$.

In practice, both porosity and permeability might be dependent on each other, and that is the reason we take the same KL expansion structure for both fields. However, when we make a proposal, we assume that $\theta^k$ and $\theta^\phi$ are independent of each other and thus avoiding ad-hoc use of a correlation between porosity and permeability.

We use the Metropolis-Hasting MCMC to sample from the posterior distribution. At each iteration, $\psi_p = [\theta^k_p \theta^\phi_p]$ is proposed using an instrumental distribution $q(\psi_p|\psi)$, where $\psi$ represents the previously accepted state/parameters in the chain, and then the forward problem is solved to determine the acceptance probability,

$$\alpha(\psi, \psi_p) = \min\left(1, \frac{q(\psi_p|\psi)P(\psi_p|F_m)}{q(\psi|\psi_p)P(\psi|F_m)}\right),$$

(4.3)

i.e., $\psi_p$ is accepted with probability $\alpha(\psi, \psi_p)$.

5. Numerical Simulations

We now discuss the simulations of the two-phase flow problem in an oil reservoir as illustrated and present the associated numerical results. We have two production wells: one well is situated along the diagonal, opposite to the
Figure 4: Prediction curves: MCMC 1, MCMC 2 and MCMC 3 used 4500 proposals for the prediction. The multi–MCMC used 1500 proposals from each chain for the prediction. The vertical line marks $t = 0.4$ PVI beyond which no data were used.

Figure 5: Prediction curves: MCMC 1, MCMC 2 and MCMC 3 used 9000 proposals for the prediction. The multi–MCMC used 3000 proposals from each chain for the prediction. The vertical line marks $t = 0.4$ PVI beyond which no data were used.
Figure 6: Prediction curves: MCMC 1, MCMC 2 and MCMC 3 used 18000 proposals for the prediction. The multi-MCMC used 6000 proposals from each chain for the prediction. The vertical line marks \( t = 0.4 \) PVI beyond which no data were used.

Figure 7: Prediction curves: MCMC 1, MCMC 2 and MCMC 3 used 36000 proposals for the prediction. The multi-MCMC used 12000 proposals from each chain for the prediction. The vertical line marks \( t = 0.4 \) PVI beyond which no data were used.
injection well, and the other is situated at the center of a side which is one of the two sides that enclose the production well at the corner. We refer to them as corner and center wells, respectively (see Figure 2). The relative permeability functions of water and oil take the form of \( s^2 \) and \((1 - s)^2\), respectively, and the viscosity ratio between water and oil is 1:20. We assume that at \( t = 0 \), the reservoir is saturated by oil without any water, i.e., \( s = 0 \). The water is then injected at the injection well at the rate of one pore-volume every five years.

The \( \sigma_F^2 \) in the likelihood function (4.2) must be fixed a priori [20], because treating \( \sigma_F^2 \) as an unknown parameter results in an unacceptably large estimate. In our application, we specify \( \sigma_F^2 = 10^{-4} \).

For KLE in (3.4), we select the correlation length \( L_x = L_y = 0.2 \) and variance \( \sigma_Y^2 = 4 \). Figure 1 shows that the eigenvalues decay very fast for these values, and it is enough to consider the first twenty eigenvalues in the KLE. Since we assume that the permeability and porosity share the same spatial structure, we share the same KLE structure for the permeability and porosity fields, with \( N_k = N_\phi = 20 \). Twenty eigenvalues lead to the exploration of a high-dimensional posterior in the MCMC framework. Therefore, we employ a component-wise tuning for the posterior exploration.

We use synthetic profiles for permeability and porosity distributions. These profiles enable us to create realistic behavior of subsurfaces. Then, using these profiles we run the flow simulator and produce the production curves. We use these production curves as the reference production curves in our simulation study.

Figure 3 shows a comparison of computing times for MCMC and multi–MCMC. The average computing time is computed using all three MCMCs. In this case the speed–up is about 3 when we use the multi–MCMC.

The forecasting of production curves in the oil reservoir consists of two steps: characterization and prediction. In the first step, we characterize the underlying field using the available data (in our case, the production curves until 0.4 PVI). Because of the nature of the inverse problem at hand, it is expected that we cannot recover a unique profile for permeability and porosity nor is it necessary. In practice, it is more relevant to gather a set of probable profiles that has been rigorously selected by the Bayesian MCMC, run the forward problem (in our case, the production curves until 1.4 PVI) using those accepted profiles, and then, aggregate the results of the forward problem to predict the rest of the production curve. We refer this curve as the prediction of production curve.

Obviously, all the accepted profiles give a very similar set of partial production curves. However, there is no way we can select the exact profile for the underlying field. In this case, we consider all the profiles and predict the production using those profiles. This way we make the predictions more reliable. The more the samples, the more reliable the predictions.

Next we look at the prediction of production curves. We consider three MCMCs generated from different seeds. For the multi–MCMCs, we consider an equal number of proposals from each chain. Figure 4 shows the predictions using 4500 proposals in each chain and the combination of three chains considering 1500 proposals from each chain. In each chain, 10% of total proposals are discarded for the prediction. The vertical line separates the measured production curves and the predicted production curves. Figure 5, Figure 6, and Figure 7 show the predictions using 9000, 18000 and 36000 proposals, respectively. As illustrated by the numerical results reported in the above mentioned figures, we have an indication that the reliability of the predictions based on a single chain is comparable to the reliability of predictions based on multiple chains. Our results suggest that we can consider multi–MCMC approach for prediction in porous media flows in a practical period of time.

6. Conclusion

We considered a Bayesian statistical approach for the prediction of flows in an oil reservoir. In this approach we need to compute the likelihood function for each proposal in an MCMC method. The computation of the likelihood function is computationally very demanding for the applications in porous media. It often limits the applicability of the Bayesian framework to these applications. In this paper, we showed that the multi–MCMC approach can alleviate this limitation and this approach makes the Bayesian framework more attractive to produce flow predictions in porous media.

References


