Efficient sampling techniques for uncertainty quantification in history matching using nonlinear error models and ensemble level upscaling techniques

Y. Efendiev* A. Datta-Gupta† X. Ma‡ B. Mallick§

Abstract

The Markov Chain Monte Carlo (MCMC) is a rigorous sampling method to quantify uncertainty in subsurface characterization. However, the MCMC usually requires many flow and transport simulations in evaluating the posterior distribution and can be computationally expensive for fine-scale geological models. We propose a methodology that combines coarse- and fine-scale information to improve the efficiency of MCMC methods. The proposed method employs offline computations for modeling the relation between coarse- and fine-scale error responses. This relation is modeled using nonlinear statistical maps which are used in efficient sampling within MCMC framework. We propose two-stage MCMC where inexpensive coarse-scale simulations are performed to determine whether or not to run the fine-scale (resolved) simulations. The latter is determined based on a statistical model developed offline. The proposed method is an extension of the approaches considered earlier (e.g., [10]) where linear relations are used for modeling the response between coarse-scale and fine-scale models. The

*Department of Mathematics, Texas A&M University, College Station, TX 77843
†Department of Petroleum Engineering, Texas A&M University, College Station, TX 77843
‡Department of Petroleum Engineering, Texas A&M University, College Station, TX 77843
§Department of Statistics, Texas A&M University, College Station, TX 77843

approach considered here does not rely on the proximity of approximate and resolved models and can employ much coarser and inexpensive models to guide the fine-scale simulations. Numerical results for three-phase flow and transport demonstrate the advantages, efficiency and utility of the method for uncertainty assessment in the history matching.

1 Introduction

Uncertainties on the detailed description of reservoir lithofacies, porosity, and permeability are major contributors to uncertainty in reservoir performance forecasting. Reducing this uncertainty can be achieved by integrating additional data in subsurface modeling. With the increasing interest in accurate prediction of subsurface properties, subsurface characterization based on dynamic data, such as production data including gas/oil ratio becomes more important.

To predict future reservoir performance, the reservoir properties, such as porosity and permeability, need to be conditioned to production data. It is essential that the permeability (and porosity) realizations adequately reflect the uncertainty in the reservoir properties, i.e., the probability distribution is sampled correctly. The uncertainty quantification is typically carried out in a Bayesian framework where multiple realizations are sampled from a posterior distribution that incorporates the prior information (e.g., [18]). This problem is challenging because the permeability field is a function defined on a large number of grid blocks and the production data nonlinearly depends on permeability. The Markov Chain Monte Carlo (MCMC) method and its modifications have been used in previous findings to sample the posterior distribution (e.g., [24]).

The direct MCMC simulations are generally very CPU demanding because each proposal requires solving a forward coupled nonlinear partial differential equations over a large time interval. The forward fine-scale problem is usually formulated on a large number of grid

blocks, which makes it prohibitively expensive to perform sufficient number of MCMC simulations. There have been a few attempts to propose MCMC methods with high acceptance rate (e.g., [18, 24, 25]). For example, the randomized maximum likelihood method uses unconditional realizations of the production and permeability 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. Though efficient in many cases, this method may not properly sample the posterior distribution [22]. Thus, developing efficient rigorous MCMC calculations with high acceptance rate remains a challenging problem.

In this paper, we extend two-stage MCMC methods considered before (e.g., [10, 11]). The two-stage MCMC involves a pre-screening stage where the proposals are screened using approximate models. If the proposal is accepted in the first stage (screening stage), then the resolved computations are performed to compute the acceptance probability. The novelty of the proposed approach is two-fold. First, we employ error modeling ([16, 28, 15, 4, 26, 20]) which allows mapping the coarse-scale data to the fine-scale data via a nonlinear map. The main goal of the error modeling is to construct a map from the coarse-scale errors between computed and observed data to the fine-scale errors based on some prior (offline) computations. The mapping between these low dimensional quantities is often easy to construct based on fewer samples. Secondly, we consider inexpensive ensemble level upscaling type methods for coarse-scale modeling (cf. [2]). To our best knowledge the error models have not yet been used in rigorous sampling methods. Previous approaches within two-stage MCMC methods [11, 10] have used only linear relation between coarse- and fine-scale responses. Though it is found to be effective in many cases, linear relations may not be very reliable when coarse-scale models both across spatial scales and uncertainties (in highly nonlinear equations) are considered. In particular, for the examples presented here we have found poor performance when linear models are used.

Ensemble level upscaling methods compute upscaled quantities which represent not only subgrid variations, but also variations across the ensemble. These approaches compute upscaled coefficients based on some sampled realizations. In this paper, we use some selected realizations based on the idea of sparse interpolation techniques to compute reference points for upscaled permeabilities. Furthermore, for an arbitrary realization, the upscaled permeability is interpolated using the pre-computed values of upscaled permeabilities at reference points. This procedure saves the computational time by avoiding the computation of upscaled quantities. Though for upscaled permeabilities such a saving may not be very important, for the computations of pseudo relative permeabilities or when using global upscaling methods, these computational saving is important. We note that one can consider more general ensemble level upscaling methods such as those presented in [2].

Two-stage MCMC methods considered here modifies the instrumental probability distribution by filtering the proposals via simplified models. Our numerical results show that using inexpensive coarse-scale computations one can increase the acceptance rate of MCMC calculations. Here the acceptance rate refers to the ratio between the number of accepted permeability samples and the number of times of solving the fine-scale nonlinear PDE system. In offline computational stage, we use several hundreds realizations of the permeability field to construct an error model and develop ensemble level upscaling. For the error modeling, we use piece-wise linear functions to fit the scattered data representing coarse-scale errors vs. fine-scale errors. At the first stage, using coarse-scale runs we determine whether or not to run the fine-scale simulations. To compute the approximation of the fine-scale error, offline statistical models are used. If the proposal is accepted at the first-stage, then a fine-scale simulation is performed at the second stage to determine the acceptance probability of the proposal. The first stage of the MCMC method modifies the proposal distribution. It is easy to show that the modified Markov chain satisfies the detailed balance condition for the correct distribution. In the paper, we also discuss the efficiency of two-stage MCMC methods.

We would like to note that two-stage MCMC algorithms have been used previously (e.g., [3, 19, 26, 17]) in different situations.

Numerical results for permeability fields generated using two-point geostatistics are presented in the paper. Using the Karhunen-Loève expansion, we can represent the high dimensional permeability field by a number of parameters. Furthermore, static data (the values of permeability field at some sparse locations) can be easily incorporated into the Karhunen-Loève expansion to further reduce the dimension of the parameter space. Numerical results are presented for black oil model (three phase flow and transport) with 8 production wells and 1 injection well. In all the simulations, we observe nearly two times increase in the acceptance rate. In other words, the preconditioned MCMC method can accept the same number of samples with much less fine-scale runs.

The paper is organized in the following way. In the next section, we briefly describe the model equations and their upscaling. Section 3 is devoted to the description of MCMC methods. Numerical results are presented in Section 4.

2 Fine and coarse models

In this section we briefly introduce the fine- and coarse-scale models used in the simulations. We consider black oil model in a subsurface formation (denoted by Ω) under the assumption that the displacement is dominated by viscous effects. We neglect the effects of gravity and capillary pressure, although our proposed approach is independent of the choice of physical mechanisms. Porosity will be considered to be constant. The phases will be referred to as water, oil and gas, designated by subscripts w, o, and g, respectively. Simultaneous flow of three phases is governed by the following three equations (e.g., [5])

$$\frac{\partial}{\partial t}\left(\frac{S_j}{B_j}\right) + \nabla \cdot \left(\frac{k(x)k_{rj}}{\mu_j B_j}\nabla p_j\right) = q_j^s, \quad j = w, o, \tag{1}$$

$$\frac{\partial}{\partial t}\left(\frac{S_g}{B_g}+\frac{S_o R_{so}}{B_o}\right)+\nabla\cdot\left(\frac{k(x)k_{ro}R_{so}}{\mu_o B_o}\nabla p_o+\frac{k(x)k_{rg}}{\mu_g B_g}\nabla p_g\right)=q_g^s, \tag{2}$$

where B_j $(j=w,o)$ is the formation volume factor of phase j, $k(x)$ is heterogeneous absolute permeability field, q^s is the source terms, p_j is the pressure of the phase j, k_{rj} is the relative permeabilities, S_j is the saturation of the phase j and R_{so} is the solubility of gas in oil.

Next, we will briefly describe single-phase flow upscaling procedure. These types of approaches for upscaling are discussed by many authors; see e.g., [9]. The main idea of this approach is to upscale the absolute permeability field $k(x)$ on the coarse-grid (see Figure 1), then solve the original system on the coarse-grid with upscaled permeability field. Below, we discuss briefly the upscaling of absolute permeability and ensemble level upscaling methods used in our simulations.

Consider the fine-scale permeability that is defined in the domain with underlying fine grid as shown in Figure 1. On the same graph we illustrate a coarse-scale partition of the domain. To calculate the upscaled permeability field at the coarse-level, we use the solutions of local pressure equations. The main idea of the calculation of a coarse-scale permeability is that it delivers the same average response as that of the underlying fine-scale problem locally. For each coarse domain K, we solve the local problems

$$div(k(x)\nabla\phi_j)=0, \tag{3}$$

with some coarse-scale boundary conditions. Here $k(x)$ denotes the fine-scale permeability field. We will use the boundary conditions which are given by $\phi_j=1$ and $\phi_j=0$ on the opposite sides along the direction e_j and no flow boundary conditions on all other sides. For these boundary conditions, the coarse-scale permeability tensor is given by

$$(k^* e_j, e_l)=\frac{1}{|K|}\int_K (k(x)\nabla\phi_j(x), e_l)dx, \tag{4}$$

where ϕ_j is the solution of (3) with prescribed boundary conditions. Various boundary condition can have some influence on the accuracy of the calculations, including periodic, Dirichlet and etc. These issues have been discussed in e.g., [30]. In particular, for determining the coarse-scale permeability field one can choose the local domains that are larger than the target coarse block, K, for (3). Once the upscaled absolute permeability is computed, the original equation is solved on the coarse-grid, without changing the form of relative permeability curves.

Ensemble level upscaling methods compute the upscaled permeabilities based on values of upscaled permeability fields for some realizations which are computed offline. To demonstrate the concept, we assume that the permeability field is computed for realizations ω_1, ..., ω_N and the values are $k^*(x,\omega_1)$, ..., $k^*(x,\omega_N)$. In general, ω is infinite dimensional, though in applications considered in this paper, the permeability fields are characterized by two-point correlation functions on a fine grid and ω will be taken to be finite dimensional. Ensemble level upscaling attempts to approximate $k(x,\omega)$ for any value of ω using $k^*(x,\omega_1)$, ..., $k^*(x,\omega_N)$. First work in this direction ([2]) uses statistical approach for this approximation. In this paper, we employ deterministic interpolation theory to approximate $k^*(x,\omega)$ given $k^*(x,\omega_1)$, ..., $k^*(x,\omega_N)$. In our simulations, we will be using linear relations for log of permeabilities and sparse interpolation techniques in high dimensional space (e.g., [31]) to approximate $k^*(x,\omega)$. In particular,

$$\log(k^*(x,\omega)) = \sum_i L_i(\omega)\log(k^*(x,\omega_i)),$$

where $L_i(\omega)$ are interpolation weights which are readily available for interpolations considered here.

As for the quantities which will be used to condition the permeability field, we take gas/oil ratio (commonly abbreviated GOR). When oil is brought to surface conditions it is

usual for some gas to come out of solution. GOR is the ratio of the gas that comes out of the solution, to the volume of oil. Our goal in this paper, is to sample the fine-scale permeability field based on GOR which is a function of time in each producing well. The relation between GOR and the permeability field is highly nonlinear and can not be accurately described via linear relations.

3 Methodology

To find the permeability field given GOR information, we assume that an observed GOR, $F^{ref}(t)$, is given. Consequently, one can consider this problem as a sampling from the conditional distribution $P(k|F^{ref})$. Using Bayes theorem we can write

$$P(k|F^{ref}) \propto P(F^{ref}|k)P(k). \tag{5}$$

The normalizing constant in this expression is not important, because we use iterative updating procedure. In (5), $P(F^{ref}|k)$ represents the likelihood function and requires the forward solution of black oil model. We will be using Metropolis-Hasting MCMC (see [27]) to sample from the posterior distribution $P(k|F)$. The main idea of MCMC is to generate a Markov chain whose stationary distribution is given by $P(k|F^{ref})$. At each iteration, a permeability field, k, is proposed using instrumental distribution $q(k|k_n)$ (where k_n is previously accepted permeability field), and then forward problem is solved to determine the acceptance probability,

$$Pr(k_n, k) = \min\left(1, \frac{q(k_n|k)P(k|F^{ref})}{q(k|k_n)P(k_n|F^{ref})}\right), \tag{6}$$

i.e. $k_{n+1} = k$ with probability $Pr(k_n, k)$, and $k_{n+1} = k_n$ with probability $1 - Pr(k_n, k)$.

Since each proposal requires the fine-scale computation, direct (full) MCMC is expensive. Typically, direct MCMC requires many iterations for the convergence to a steady state, where

each iteration involves the computation of the fine-scale solution over a large time interval. One way to achieve efficiency is to propose an algorithm that increases the acceptance rate of MCMC. This minimizes rejection of proposals after detailed flow and transport calculations. In this paper, we use coarse-scale solutions based on single-phase upscaling to increase the acceptance rate. The main idea of this algorithm is to compare GOR that correspond to the coarse-scale models to determine whether or not to run fine-scale simulations.

To formulate the algorithm, we introduce several notations. For our numerical results, we will sample the likelihood

$$\pi(k) = P(F^{ref}|k) \propto \exp(-\frac{\|F^{ref} - F_k\|^2}{\sigma_f}), \tag{7}$$

where F^{ref} is GOR, F_k is GOR that is obtained from the simulations with permeability k, and $\Sigma = \sigma_f I$ is covariance matrix representing the measurement errors. Here $\|F^{ref} - F_k\| = \left(\int_0^T |F^{ref} - F_k|^2 dt\right)^{1/2}$ is the fine-scale error between the simulated and the observed data. We note that GOR is a function of time at producing wells. In our simulations, we start with a number of realizations (usually 100-200) of the permeability field and construct the error model between $\|F_k - F^{ref}\|$ and $\|F_{k^*} - F^{ref}\|$ (see Figure 7). Using known statistical methods we derive nonlinear relation between these quantities

$$\|F_k - F^{ref}\| \approx G(\|F_{k^*} - F^{ref}\|),$$

where G is a nonlinear function which is estimated based on a limited number of realizations of the permeability field. G can be assumed to be random as it is done in our simulations. In our simulations, we use piece-wise Gaussian processes to fit the relation $\|F_k - F^{ref}\|$ vs. $\|F_{k^*} - F^{ref}\|$. In this case, the surrogate probability distribution used in the simulations is

$$\pi^*(k) = P(F^{ref}|k^*) \propto \exp(-\frac{1}{\sigma_f}\frac{G_0(\|F_{k^*} - F^{ref}\|)}{\sigma_{k^*}}),$$

where G_0 and σ_{k^*} are the mean and the variance of the piece-wise Gaussian for a given k^* (see Figure 7). The subscript k^* is used in σ to indicate that the variance of piece-wise Gaussian model is a function of k^*. We note that one can use more general probabilistic models to model error relations and also use the full covariance matrix corresponding to the measurement errors.

Algorithm (two-stage MCMC with nonlinear error model)

- **Offline:** Start with offline computations of GOR cross-plot by computing $\|F_k - F^{ref}\|$ and $\|F_{k^*} - F^{ref}\|$. Estimate the operator G, such that $\|F_k - F^{ref}\| \approx G(\|F_{k^*} - F^{ref}\|)$.

- **Offline:** Compute $k^*(x, \omega_i)$ for some realizations ω_i.

- Step 1. At k_n generate k from $q(k|k_n)$.

- Step 2. Accept k for *the fine-scale run* with probability

$$g(k_n, k) = \min\left(1, \frac{q(k_n|k)\pi^*(k)}{q(k|k_n)\pi^*(k_n)}\right), \tag{8}$$

  i.e. $k_{n+1} = k$ (conditionally) with probability $h(k_n, k)$, and $k_{n+1} = k_n$ (conditionally) with probability $1 - g(k_n, k)$. If rejected go to step 1.

- Step 3. Accept k with probability

$$Pr(k_n, k) = \min\left(1, \frac{Q(k_n|k)\pi(k)}{Q(k|k_n)\pi(k_n)}\right), \tag{9}$$

  i.e. $k_{n+1} = k$ with probability $Pr(k_n, k)$, and $k_{n+1} = k_n$ with probability $1 - Pr(k_n, k)$. The proposal function $Q(k|k_n)$ satisfies

$$Q(k|k_n) = g(k_n, k)q(k|k_n) + \left(1 - \int g(k_n, k)q(k|k_n)dk\right)\delta_{k_n}(k). \tag{10}$$

The expression for Q can be simplified (e.g., [11])

$$Pr(k_n, k) = \min\left(1, \frac{\pi(k)\pi^*(k_n)}{\pi(k_n)\pi^*(k)}\right). \tag{11}$$

One can also add new data to improve the estimates; however, this can introduce bias and will be avoided in our simulations.

In [11], it was shown that the detailed balance condition holds and MCMC converges to the correct distribution. This proof applies here. We would like to note that from (11) one obtains (see [10, 11]) that

$$Pr(k_n, k) > \exp\left(-\frac{\|F_k - F^{ref}\| - \frac{G_0(\|F_{k^*} - F^{ref}\|)}{\sigma_{k^*}}}{\sigma_f} + \frac{\|F_{k_n} - F^{ref}\| - \frac{G_0(\|F_{k_n^*} - F^{ref}\|)}{\sigma_{k^*}}}{\sigma_f}\right).$$

It is clear from this expression that if the error in $\|F_k - F^{ref}\| - G_0(\|F_{k^*} - F^{ref}\|)/\sigma_{k^*}$ is small for a generic k, the acceptance probability is close to 1. Thus, if the approximation with G is accurate and the spread of this approximation is low, then one can achieve high acceptance rate. Clearly, if the coarse-scale model approximates the results of the fine-scale simulations, then one can achieve high acceptance rate. However, in general, single-phase upscaling techniques do not provide accurate approximations of the fine-scale results for three-phase systems and thus modeling the relation between coarse- and fine-scale models are needed. The errors associated with this modeling will affect the efficiency of two-stage MCMC methods. In this paper, we explore this via numerical simulations.

We again note this approach extends previous approaches within two-stage MCMC methods [11, 10] where linear relations between coarse- and fine-scale responses are used. In the numerical examples considered in this paper, we have found poor performance when linear models are used.

4 Numerical results

For our numerical tests, we use the Karhunen-Loève expansion (KLE) [21, 29] to obtain the permeability field in terms of an optimal L^2 basis. By truncating KLE, we can represent the permeability matrix by a small number of random parameters. To impose the hard constraints (the values of the permeability at prescribed locations), one can find a linear subspace of our parameter space (a hyperplane) which yields the corresponding values of the permeability field. First, we briefly recall the facts of the KLE. Denote $Y(x,\omega) = \log[k(x,\omega)]$, where the random element ω is included to remind us that k is a random field. We assume that $E[Y(x,\omega)] = 0$. Suppose $Y(x,\omega)$ is a second order stochastic process with $E\int_\Omega Y^2(x,\omega)dx < \infty$, where E is the expectation operator. Given an orthonormal basis $\{\phi_k\}$ in $L^2(\Omega)$, we can expand $Y(x,\omega)$ as

$$Y(x,\omega) = \sum_{k=1}^{\infty} Y_k(\omega)\phi_k(x), \qquad Y_k(\omega) = \int_\Omega Y(x,\omega)\phi_k(x)dx.$$

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

$$E[Y_iY_j] = \int_\Omega \phi_i(x)dx \int_\Omega R(x,y)\phi_j(y)dy = 0, \qquad i \neq j.$$

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

$$\int_\Omega R(x,y)\phi_k(y)dy = \lambda_k\phi_k(x), \qquad k = 1,2,\ldots, \tag{12}$$

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

$$R(x, y) = \sum_{k=1}^{\infty} \lambda_k \phi_k(x) \phi_k(y). \tag{13}$$

Denote $\eta_k = Y_k/\sqrt{\lambda_k}$, then η_k satisfy $E(\eta_k) = 0$ and $E(\eta_i \eta_j) = \delta_{ij}$. It follows that

$$Y(x, \omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \eta_k(\omega) \phi_k(x), \tag{14}$$

where ϕ_k and λ_k satisfy (12). We assume that the eigenvalues λ_k are ordered as $\lambda_1 \geq \lambda_2 \geq \ldots$. The expansion (14) is called the Karhunen-Loève expansion. In the KLE (14), the L^2 basis functions $\phi_k(x)$ are deterministic and resolve the spatial dependence of the permeability field. The randomness is represented by the scalar random variables η_k.

After we discretize the domain Ω by a rectangular mesh, the continuous KLE (14) is reduced to a finite number of terms. In our paper, we work with finite dimensional covariance matrices defined over the square domain with 50×50 resolution. As a consequence, the covariance matrix is 2500×2500. Note that 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 $Y_N = \sum_{k=1}^{N} \sqrt{\lambda_k} \eta_k \phi_k$, define the energy ratio of the approximation as

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

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.

It is common to use variogram instead of covariance functions for stochastic permeability fields. The relation between them can be easily written as $\gamma(x, y) = C - R(x, y)$, where $C = E(Y(x, \omega)^2)$ is constant for stationary processes and $\gamma(x, y)$ denotes the variogram.

Typical variograms used in the modeling of subsurface processes are exponential, normal, and spherical [6]. In this paper, we will use spherical variogram and denote correlation lengths by l_1, l_2 and the variance of $\log(k)$ by $\sigma_{\log(k)}$. We first solve the eigenvalue problem (12) numerically on the rectangular mesh and obtain the eigenpairs $\{\lambda_k, \phi_k\}$. Further, we can sample $Y(x, \omega)$ from the truncated KLE (14) by generating Gaussian random variables η_k.

In all numerical simulations, we will assume that the reservoir is filled with oil and water is injected to displace the oil. We consider 9 spot pattern of water flooding. There is one water injector, and 8 producers (see Figure 2). The domain is taken to be square with 50×50 fine grid resolution with dimensions 50×50. The coarse grid is taken to be uniform 5×5 in all cases. As for ensemble level upscaling, we use lowest order Smolyak interpolation which provides $2d + 1$ nodes in d-dimensional space. We refer to [31] for the description of Smolyak interpolation and to [7] the results on upscaling using Smolyak interpolation. As we mentioned earlier that ensemble level upscaling techniques used here simply interpolate the log of upscaled permeability based on the values of log of upscaled permeability at some regularly spaced points in high dimensions. Because the points are regularly spaced the interpolation formula and interpolation weights can be easily derived (see [31]).

Solution gas/oil ratio and gas formation volume factor are shown in Figure 3 (left) and Figure 3 (right). From right top figure, the bubble point pressure of the reservoir is 3000 psia. Relative permeability of water, oil, and gas are shown in Figure 4 (left) and in Figure 4 (right). Modified Stones II second three-phase relative permeability model was used to compute oil relative permeability [1]. We note that if reservoir pressure is above the bubble point pressure, the flow is two-phase (water and oil); if the pressure drops below the bubble point pressure, then the gas evolves into a liquid phase and a gas phase. The flow is three-phases: water, oil and gas. There are maximum of three components, water, oil and gas. In the black oil model, it is assumed that no mass transfer occurs between the water phase

and other two phases. Moreover, the mass fractions of the oil and gas components in the oil phase can be determined by gas solubility.

In the first example, we use a reference permeability field with the correlation lengths $l_x = 30$, $l_y = 2$, and the variance of $\log(k)$ is 2. A realization of this permeability field is used to generate a reference permeability field. In the sampling procedure, we choose the permeability fields with different correlation lengths (still on 50×50 fine grid resolution). In particular, we choose $l_x = 18$, $l_y = 3$ and keep only 50 eigenvalues/eigenvectors in KLE. In Figure 5, GOR misfit vs. iterations are plotted. We compare the use of two coarse scale models. In the first case, we use the coarse-scale model where no approximation is made in computing upscaled permeabilities. In the second example, we use ensemble level upscaled permeability which is computed based on offline computations. We note that the second case is less accurate though less expensive to compute. We observe from this figure that all three curves show that the two-stage MCMC methods have similar convergence as the full MCMC. These results are based on 1000 proposals. The acceptance rates are given as following, full MCMC - 8.8 %, two-stage MCMC with upscaling is 22 %, and two-stage MCMC with ensemble level upscaling is 17 %. We observe that two-stage MCMC methods provide nearly two times higher acceptance rates. Since the computational cost associated with coarse-scale simulations is negligble, two-stage MCMC methods improve CPU required for sampling the posterior distribution by nearly two-fold. In all simulations, random walk instrumental distribution for $q(k|k_n)$ is used, where $k = k_n + \epsilon_n$ with ϵ_n being a random perturbation with prescribed distribution. The formal convergence diagnosis can performed using multiple chains method based convergence diagnosis ([14]). In this paper, our goal is to compare modified chain with the chain obtained via direct MCMC, and thus we restrict ourselves to only showing RMS vs. 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. Figure 6 shows the GOR

matches of the producers. As we see that the sampled realizations match GOR very well. In these figures, reference GOR is designated by green color, and initial GOR is designated by blue, and the sampled GORs are designated by red color. The data which is used to model the error is presented in Figure 7. We plot both mean as well as mean plus/minus standard deviation. In our simulations, we have used piecewise linear relation to model the mean behavior and constant variation within each bin. Furthermore, Gaussian distribution is used for each bin. In Figure 8, the samples of the permeability field are depicted. One can observe that the initial model does not have the high flow channel located at upper side of the domain, while the corrected permeability models have this high conductivity region.

In the second example, the reference permeability field is chosen with the correlation lengths $l_x = 25$ and $l_y = 6$ and 100 eigenvalues/eigenvectors are kept in KLE. For sampling purposes, the permeability fields are generated with 50×50 fine resolution and with $l_x = 20$ and $l_z = 4$ and 40 eigenvalues/eigenvectors are kept in KLE. In Figure 9, GOR misfit vs. iterations are plotted. As before, we compare the use of two coarse scale models. In the first case, we use coarse-scale model where no approximation is made in computing upscaled permeabilities. In the second example, we use ensemble level upscaled permeability which is computed based on offline computations. We observe from this figure that all three curves show that the two-stage MCMC has similar convergence as the full MCMC. These results are based on 1000 proposals. The acceptance rates are given as following, full MCMC - 12 %, two-stage MCMC with upscaling is 24 %, and two-stage MCMC with ensemble level upscaling is 20 %. Figure 10 shows the GOR matches of the producers. As we see that the sampled realizations match GOR very well. As before, reference GOR is designated by green color, and initial GOR is designated by blue, and the sampled GORs are designated by red color. The data which is used to model the error is presented in Figure 11. In our simulations, we have used piecewise linear relation to model the mean behavior and constant variation within each segment.

5 Conclusions

In this paper, we study two-stage MCMC methods which use simplified and inexpensive upscaled models to speed-up the sampling of the posterior distribution. Our underlying equations describe the flow and transport of three-phase flow system in heterogeneous porous media. We employ single-phase upscaling methods for coarsening the flow and transport equations. Because single-phase upscaling techniques are not very accurate for complex black oil models (see e.g., Figure 11), the proposed algorithm requires offline computations where a nonlinear relation between the coarse- and fine-scale models are constructed. This relation is used within the context of two-stage MCMC to perform rigorous sampling. The proposed method generalizes the existing multi-stage MCMC methods where a linear relation between coarse- and fine-scale models are used. In our coarse-scale models, we also use ensemble level upscaling techniques for inexpensive coarse-scale computations. The efficiency of two-stage MCMC methods depends on the accuracy of the nonlinear approximation and the spread in this approximation. The modeling errors in this approximation affect the efficiency of two-stage MCMC methods. In this paper, we explore this via numerical simulations. Numerical results study the advantages, efficiency and utility of the method for uncertainty assessment in the history matching. We show that two-stage MCMC methods provide two times higher acceptance rates, and thus improve CPU required for sampling the posterior distribution by nearly two-fold.

References

[1] K. Aziz, *Petroleum reservoir simulation*, Chapman & Hall, 1979.

[2] Y. Chen and L. Durlofsky, *An ensemble level upscaling approach for efficient estimation of fine-scale production statistics using coarse-scale simulations*, SPE paper

106086, presented at the SPE Reservoir Simulation Symposium, Houston, Feb. 26-28 (2007).

[3] A. Christen and C. Fox, *MCMC using an approximation*, Technical report, Department of Mathematics, The University of Auckland, New Zealand.

[4] M. Christie, J. Glimm, J. Grive, D. Highon, D. Sharp, and M. Wood-Schultz, *Error analysis ans simulations of complex phenomena*, Los Alamos Sci. 29 (2005) 625.

[5] A. Datta-Gupta and M.J. King, *Streamline Simulation: Theory and Practice*, Publisher: Society of Petroleum Engineers, 2007

[6] C. V. Deutsch and A. G. Journel, *GSLIB: Geostatistical software library and user's guide, 2nd edition*, Oxford University Press, New York, 1998.

[7] P. Dostert, *Uncertainty Quantification Using Multiscale Methods for Porous Media Flows*, Ph.D. thesis, Texas A&M University, 2007

[8] P. Dostert, Y. Efendiev, T. Hou, and W. Luo, *Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification*, 217 (1), pp.123-142, 2006

[9] L. J. Durlofsky, *Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media*, Water Resour. Res., 27 (1991), pp. 699–708.

[10] Y. Efendiev, A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick, *An efficient two-stage Markov chain Monte Carlo method for dynamic data integration*, 41, W12423, doi:=10.1029/2004WR003764

[11] Y. Efendiev, T. Hou and W. Luo, *Preconditioning Markov chain Monte Carlo simulations using coarse-scale models*, SIAM. Sci. Comp. 28(2), pp. 776-803.

[12] V. Ginting, *Analysis of two-scale finite volume element method for elliptic problem*, Journal of Numerical Mathematics, 12(2) (2004), pp. 119–142.

[13] P. Frauenfelder, C. Schwab and R. A. Todor, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228.

[14] Gelman, A., and Rubin, B., 1992, *Inference from iterative simulation using multiple sequences* , Statistical Science, 7, pp. 457-511.

[15] J. Glimm and D. H. Sharp, *Prediction and the quantification of uncertainty*, Phys. D, 133 (1999), pp. 152–170. Predictability: quantifying uncertainty in models of complex phenomena (Los Alamos, NM, 1998).

[16] J. Glimm, S. Hou, Y.H. Yee, D.H. Sharp and K. Ye, *Solution error models for uncertainty quantification*, AMS Advances in Differential Equations and Mathematical Physics (2002).

[17] D. Higdon, H. Lee and Z. Bi, *A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information*, IEEE Transactions on Signal Processing, 50(2) (2002), pp. 388-399.

[18] P. Kitanidis, *Quasi-linear geostatistical theory for inversing*, Water Resour. Res., 31 (1995), pp. 2411–2419.

[19] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*, Springer, New-York, 2001.

[20] O. Lodoen, H. Omre, L. Durlofsky, and Y. Chen, *Assessment of uncertainty in reservoir production forecasts using upscaled models*, in proceedings of the 7th International Geostatistics Congress, Banff, Canada, September 26 - October 1, 2004.

[21] M. Loeve, *Probability Theory*, 4th ed., Springer, Berlin, 1977.

[22] X. Ma, M. Al-Harbi, A. Datta-Gupta, and Y. Efendiev *An Efficient Two-Stage Sampling Method for Uncertainty Quantification in History Matching Geological Models*, to appear in SPEJ.

[23] S. P. Meyn, R. L. Tweedie, *Markov Chains and Stochastic Stability*, Springer-Verlag, London, 1996.

[24] D. Oliver, L. Cunha, and A. Reynolds, *Markov chain Monte Carlo methods for conditioning a permeability field to pressure data*, Mathematical Geology, 29 (1997).

[25] D. Oliver, N. He, and A. Reynolds, *Conditioning permeability fields to pressure data*, 5th European conference on the mathematics of oil recovery, Leoben, Austria, 3-6 September, 1996.

[26] H. Omre and O. P. Lodoen, *Improved production forecasts and history matching using approximate fluid flow simulators*, SPE Journal, September 2004, pp. 339-351.

[27] C. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, New-York, 1999.

[28] A. O'Sullivan, *Modelling Simulation Error For Improved Reservoir Prediction*, PhD Thesis, (Heriot Watt University, 2004).

[29] E. Wong, *Stochastic Processes in Information and Dynamical Systems*, MCGraw-Hill, 1971.

[30] X. H. Wu, Y. Efendiev, and T. Y. Hou, *Analysis of upscaling absolute permeability*, Discrete and Continuous Dynamical Systems, Series B, 2 (2002), pp. 185–204.

[31] D. Xiu and J. Hesthaven, *High-Order Collocation Methods for Differential Equations with Random Inputs*, SIAM J. Sci. Comput. Wol. 27, No. 3, 2007, pp. 1118-1139.

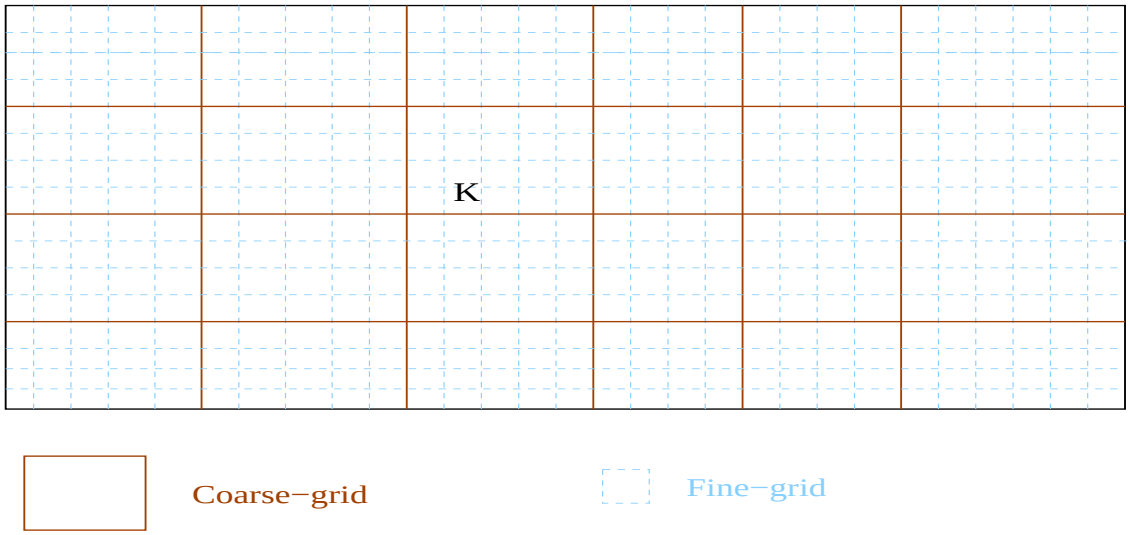

Figure 1: Schematic description of fine- and coarse-grids. Bold lines illustrate a coarse-scale partitioning, while thin lines show a fine-scale partitioning within coarse-grid cells.

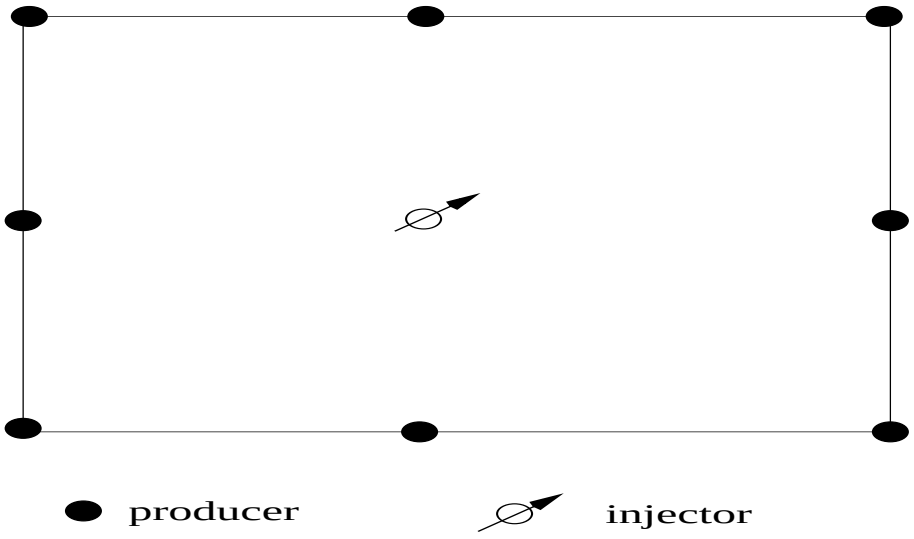

Figure 2: Well configuration used in the simulations.

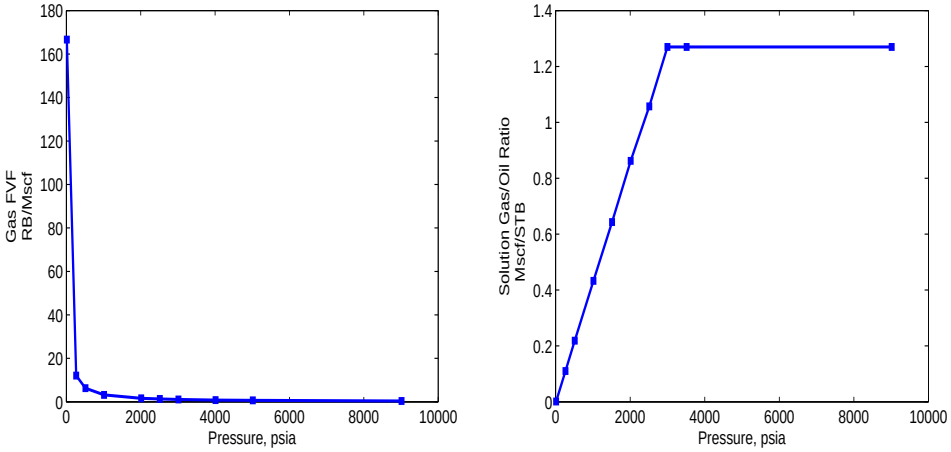

Figure 3: Solution gas/oil ratio and gas formation volume factor.

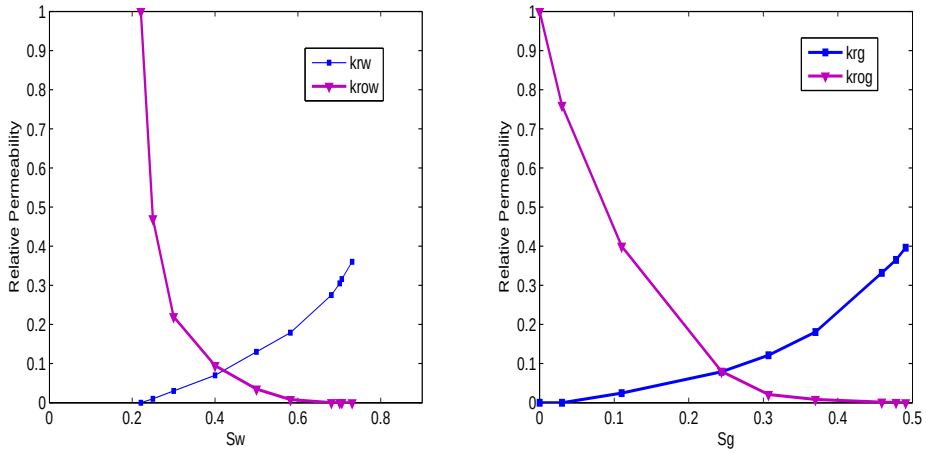

Figure 4: Relative permeabilities used in the simulations.

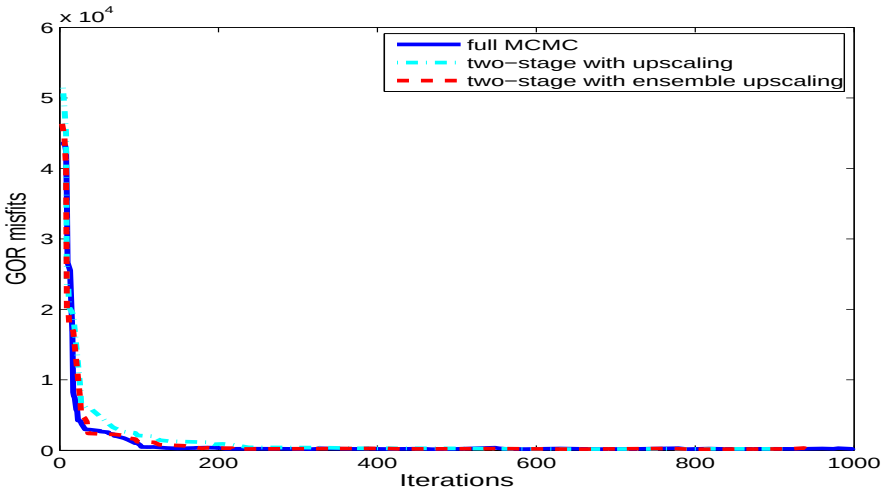

Figure 5: Error vs. proposal iterations.

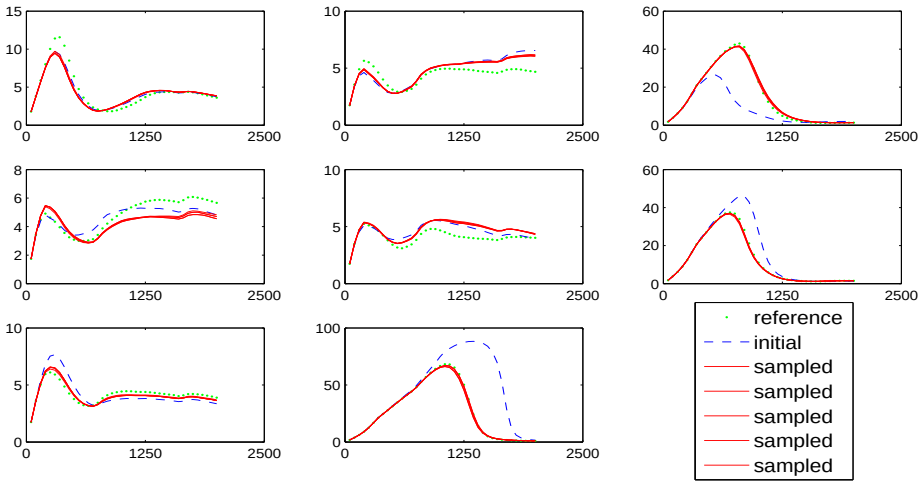

Figure 6: GOR data at producers.

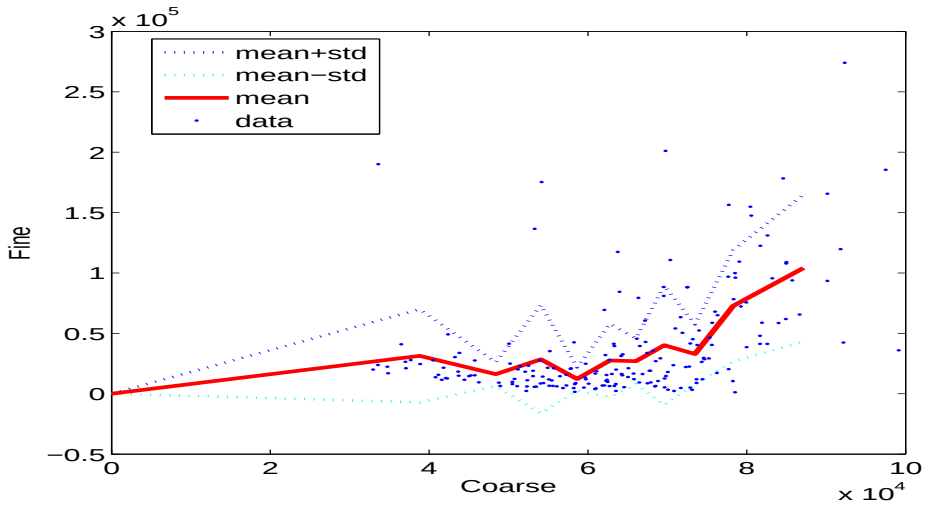

Figure 7: Error model.

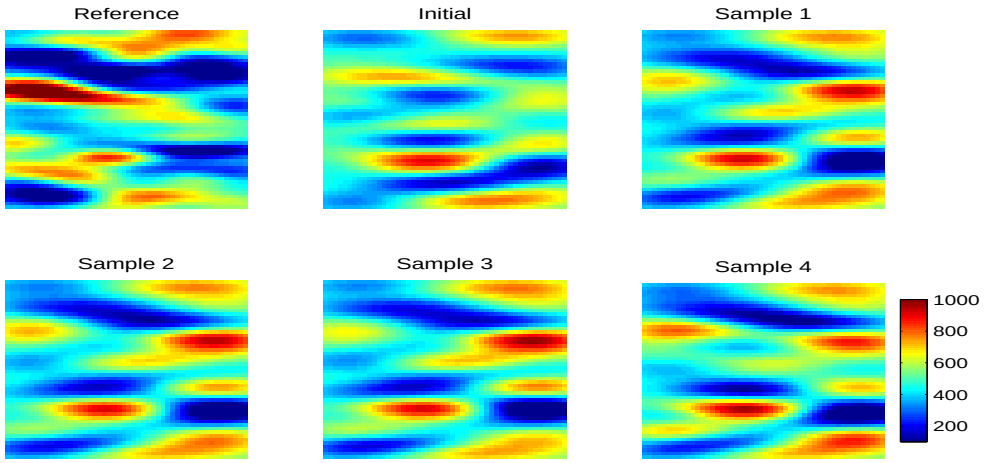

Figure 8: Permeability realizations.

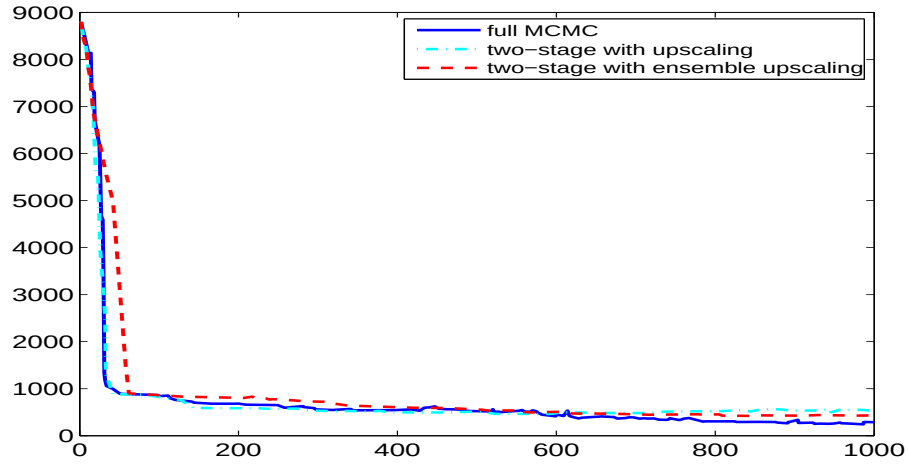

Figure 9: Error vs. proposal iterations.

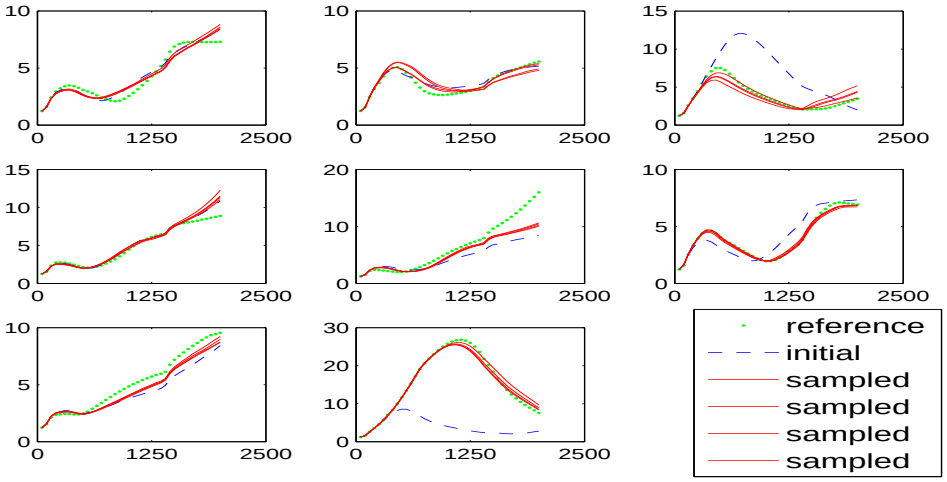

Figure 10: GOR data at producers.

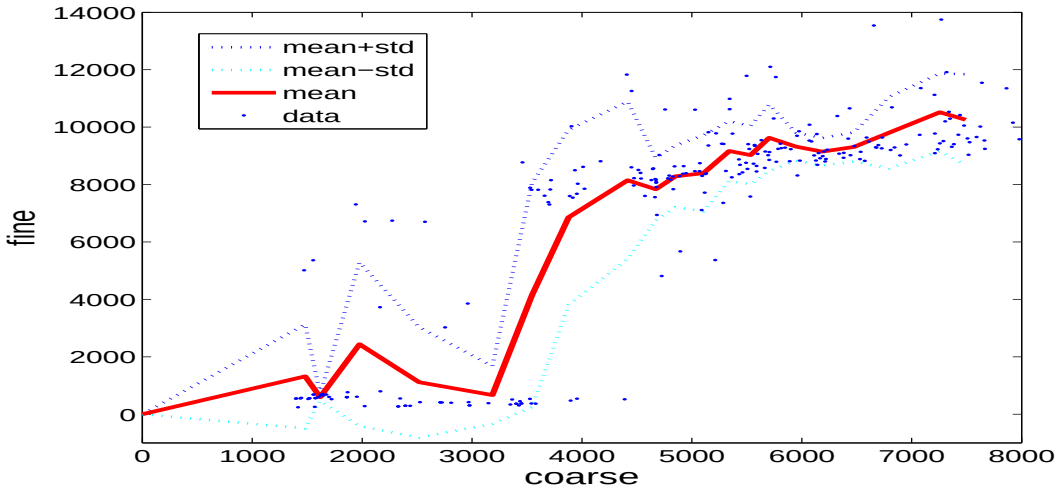

Figure 11: Error model.