

The use of coarse-scale models in uncertainty quantification for subsurface flows

Yalchin Efendiev

Department of Mathematics & ISC

Texas A&M University

Collaborators: P. Dostert (Arizona), T. Hou (Caltech), B. Mallick (TAMU), A. Datta-Gupta
(TAMU)

Introduction

- We consider a problem of characterizing subsurface properties given coarse-scale measurements.
- The particular application is two-phase flow and transport in heterogeneous porous formations and the measured data is the volume of produced fluid (called production data or fractional flow) at the outer boundaries.
- Production data (usually measured with some precision) describes an integrated response. Trying to obtain the hydraulic conductivity (permeability) samples based on this integrated response is an ill-posed problem. The problem can be regularized using priors.
- The problem reduces to sampling from a complicated distribution involving the solutions of coupled nonlinear partial differential equations.
- Metropolis-Hastings Markov chain Monte Carlo (MCMC) methods can be used as an umbrella sampling method. MCMC used in a straightforward way is very CPU demanding.
- We propose and analyze approaches for efficient sampling which employ multi-scale models.

Prototypical model

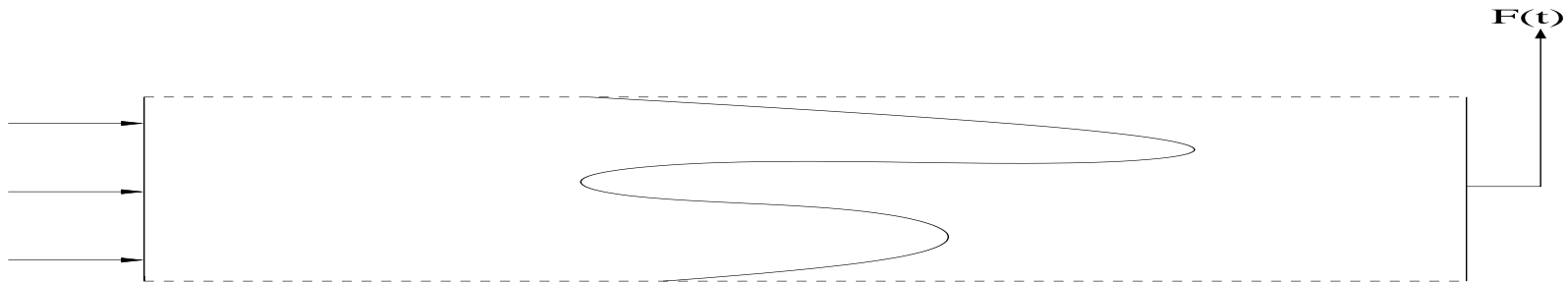
We consider two-phase flow in a heterogeneous porous formation under the assumption that the displacement is dominated by viscous effects.

$$\nabla \cdot (\lambda(S)k\nabla p) = h,$$

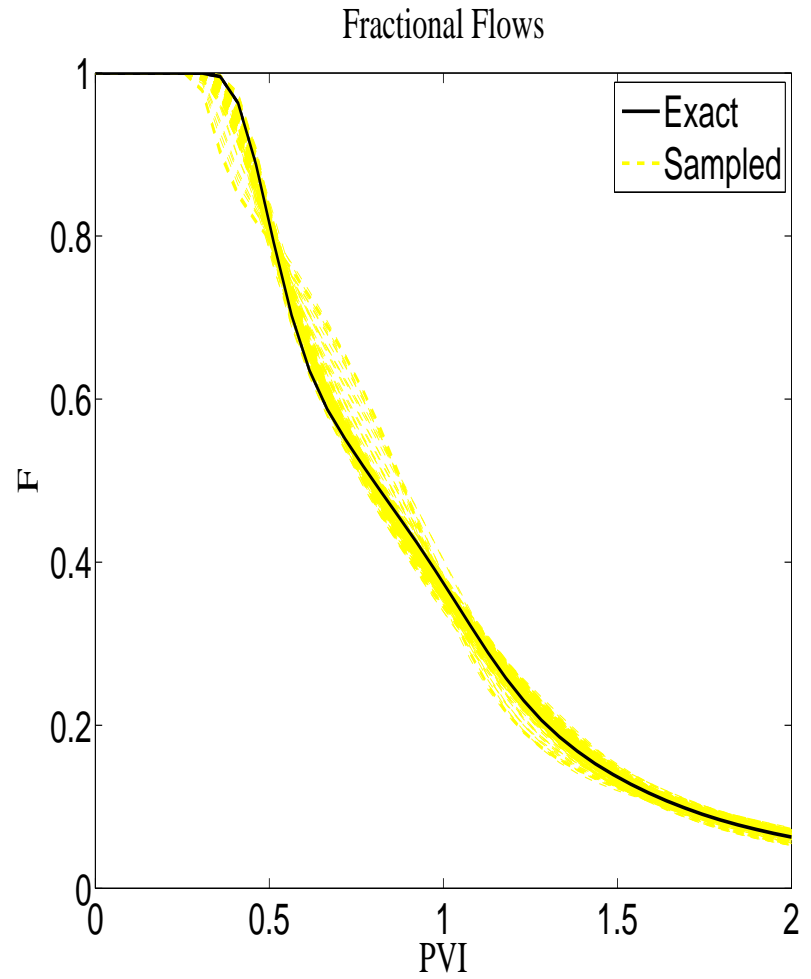
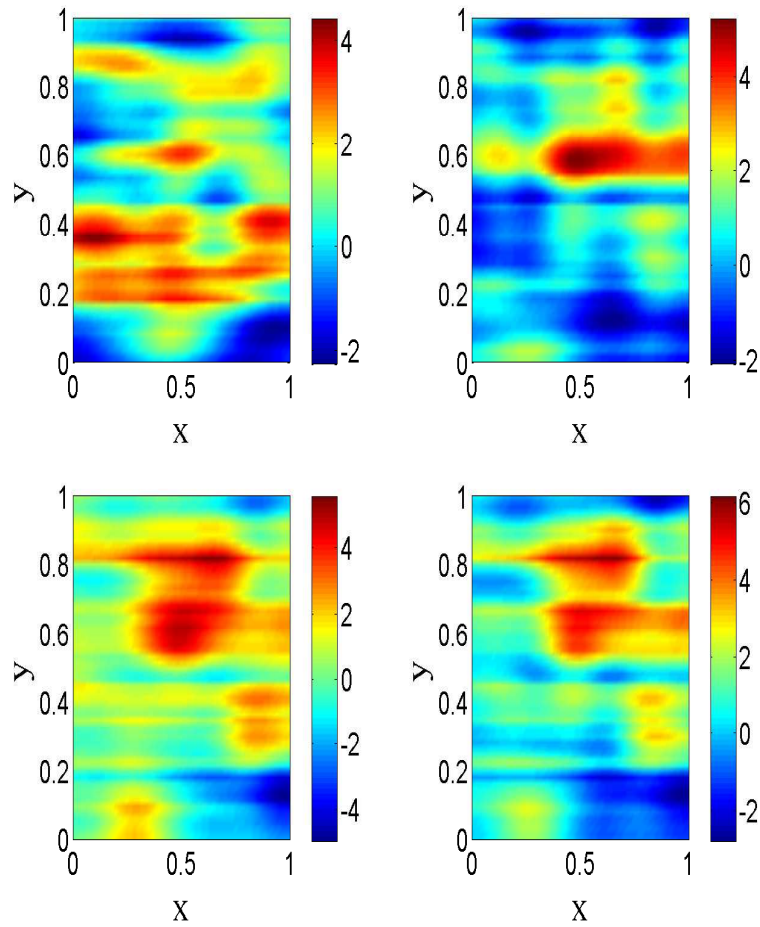
$$\frac{\partial S}{\partial t} + v \cdot \nabla f(S) = 0, \quad v = -\lambda(S)k\nabla p.$$

Measured coarse-scale data:

$$F(t) = \frac{\int_{out} v f(S) dl}{\int_{out} v dl}$$



Illustration



Problem setting

- Given the fractional flow information (coarse-scale data) $F(t)$ and some precision, we would like to sample k from $P(k|F)$.
- From Bayes theorem

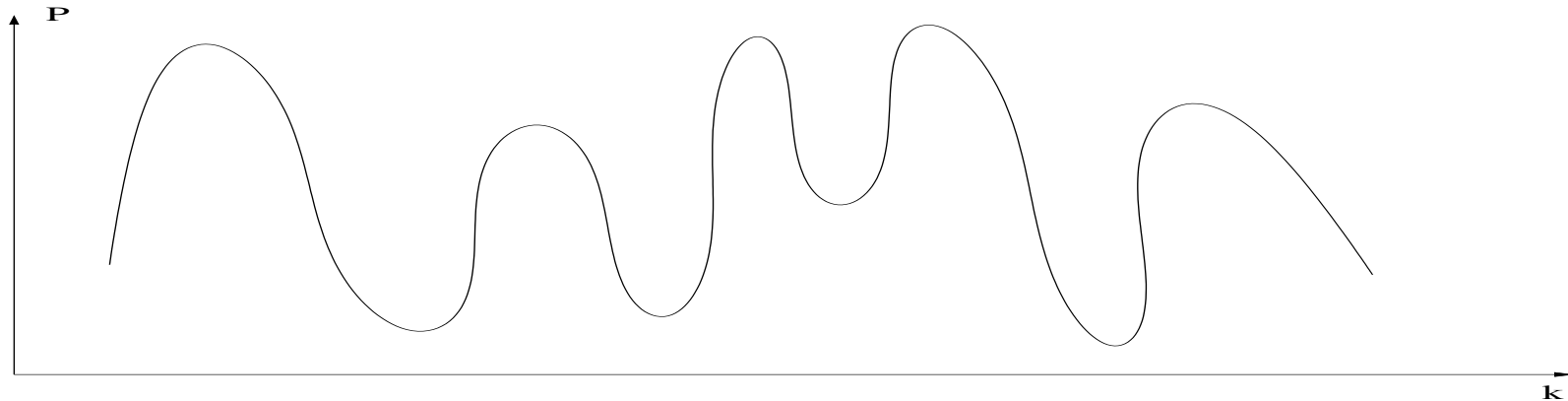
$$P(k|F) \propto P(F|k)P(k).$$

- Here $P(k)$ is the prior information, $P(F|k)$ is the likelihood and assumed given by $P(F|k) = \exp(-\frac{\|F_k(t) - F^{obs}(t)\|^2}{\sigma_f^2})$.
- A typical prior can be $P(k) = \exp(-\frac{\|\log(k) - \log(k_{obs})\|^2}{\sigma_k^2})$. Thus, the posterior distribution is

$$P(k|F) \propto \exp(-\frac{\|F_k(t) - F^{obs}(t)\|^2}{\sigma_f^2}) \exp(-\frac{\|\log(k) - \log(k_{obs})\|^2}{\sigma_k^2}).$$

Difficulties

- $\pi(k) = P(k|F)$ can be multi-modal and high dimensional.
- $\pi(k) = P(k|F)$ is not given analytically and involves the solution of nonlinear pde system.



Metropolis-Hastings MCMC

Algorithm (Metropolis-Hastings MCMC)

- Step 1. At k_n generate k from $q(k|k_n)$.
- Step 2. Accept k as a sample with probability

$$p(k_n, k) = \min \left(1, \frac{q(k_n|k)\pi(k)}{q(k|k_n)\pi(k_n)} \right),$$

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

Here $\pi(k)$ is the distribution we would like to sample.

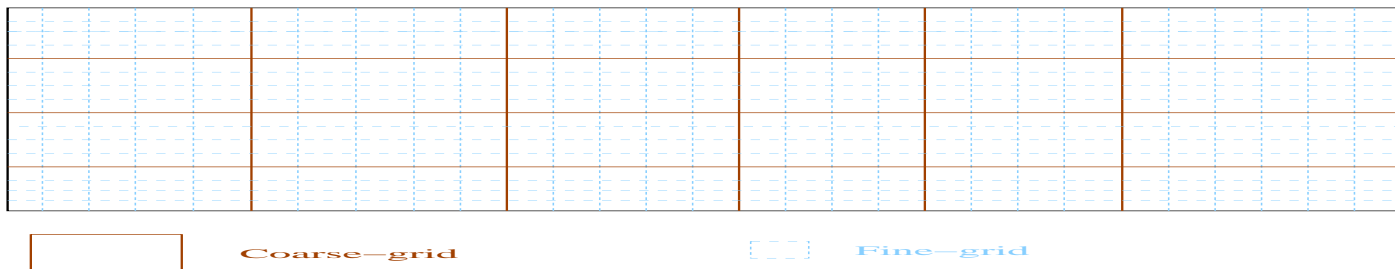
- Direct (full) MCMC simulations are usually prohibitively expensive, because each proposal requires a fine-scale computation.
- We propose an algorithm, where the proposal distribution is modified using coarse-scale spatial models.

Coarse-scale spatial models

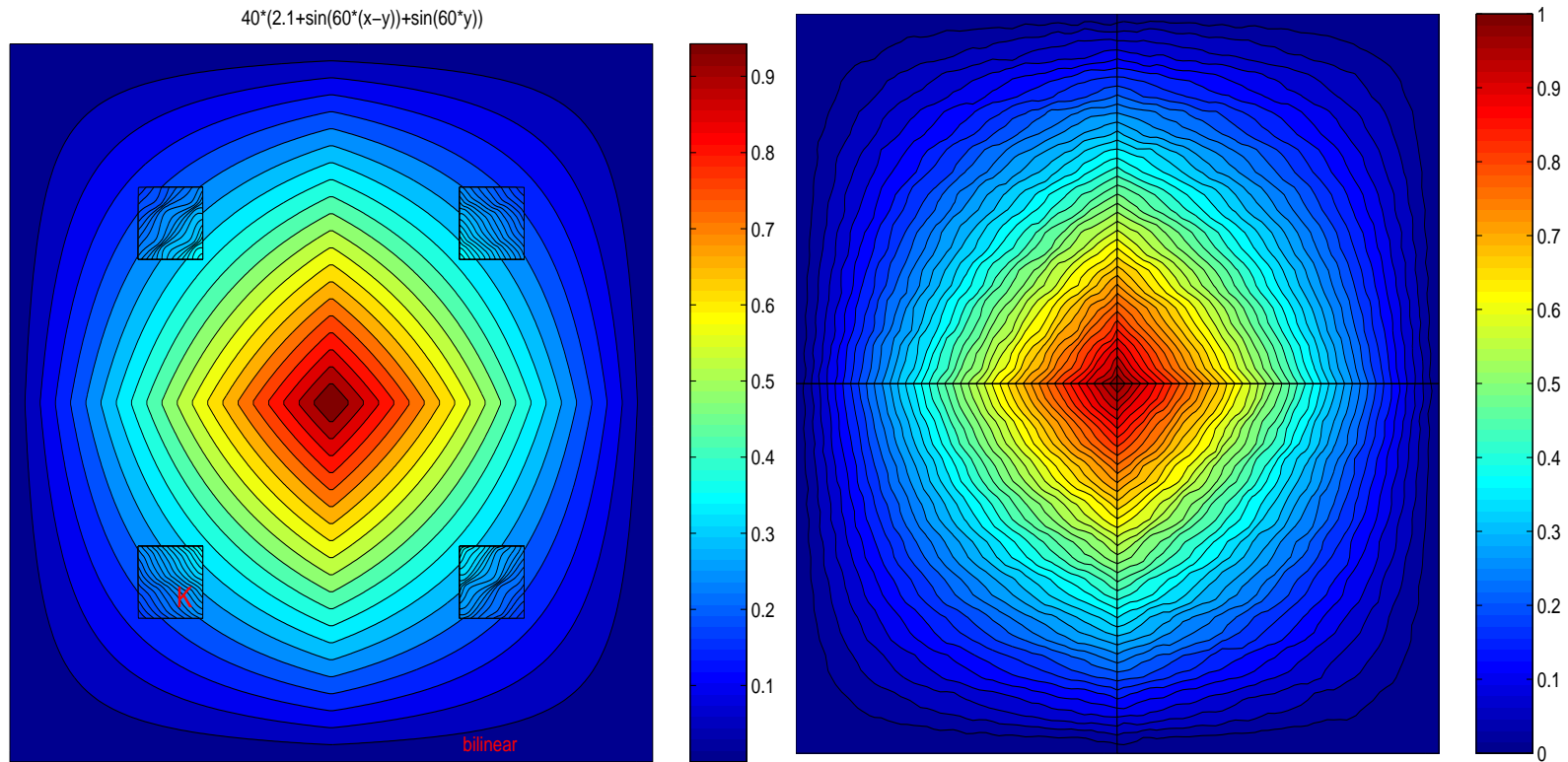
- We employ mixed multiscale finite element methods (MsFEM) for solving flow (elliptic) equations.
- The solution is represented as a linear combination of multiscale basis functions constructed *a priori* in a coarse block (or RVE). $p = \sum p_i \phi_i$, where p_i are coarse values of the solution and ϕ_i are basis functions defined by

$$\operatorname{div}(k(x)\nabla\phi^i) = 0 \quad \text{in } K$$

- Boundary conditions are very important for accuracy of subgrid capturing error. Choices: (1) local boundary conditions (the information only within the target coarse block is taken into account); (2) oversampling (the information in slightly larger than the target coarse block domain is taken into account).



Basis functions



$\phi^i = \phi_0^i$ on ∂K , where ϕ_0^i are standard bilinear basis functions.

Stochastic multiscale finite element methods

- It is known that the local approaches suffer from the resonance errors expressed as the ratio *characteristic length scale/coarse mesh size*. Limited global information (important global information) can be used to remove the resonance errors.
- Multiscale methods using limited global information rely on the assumption that the solution smoothly depends on some global fields in two-phase flow simulations

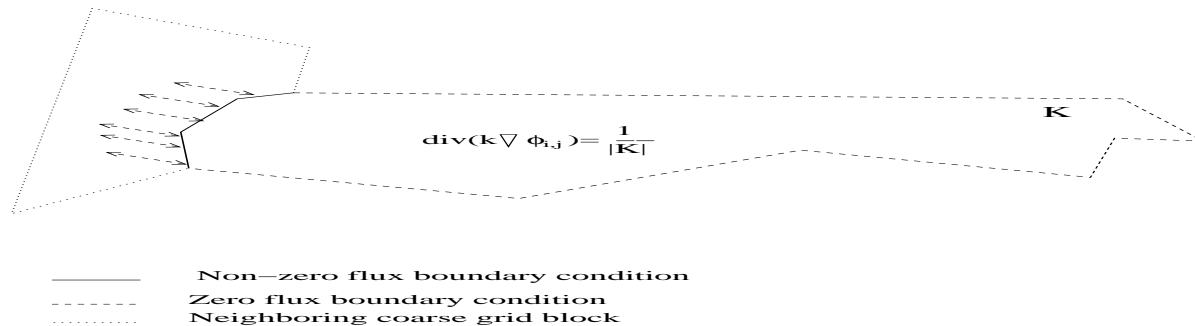
$$\|p - G(p_1, \dots, p_N)\|_{H^1} \leq \delta,$$

where δ is small. These global fields can be (1) solution of single-phase flow (Efendiev et al., 2006), (2) directional solutions of linear equations (Owhadi and Zhang, 2006) (3) local solutions.

- Basis functions need to span p_1, \dots, p_N .
- One can use some selected realizations, $k(x, \omega_1), \dots, k(x, \omega_k)$, to construct basis functions for the ensemble.
- The pre-computed basis functions are constructed based on selected realizations of the stochastic permeability field, and thus span both spatial scales and uncertainties (Aarnes and Efendiev, SIAM SC 2007)

Multiscale basis functions

- The velocity basis functions are given by $\Psi_{e_i} = k(x)\nabla\phi$, where $\text{div}(k(x)\nabla\phi) = \frac{1}{|K|}$ in K and $k(x)\nabla\phi \cdot n = \frac{v_i \cdot n}{\int_{e_l} v_i \cdot n ds}$ on e_i and 0 otherwise, where $v_i = k\nabla p_i$.
- These approaches do not require interpolation in “uncertainty space”.
- The overall approach solves the flow equation on the coarse-grid with fixed set of basis functions and solves the transport equation on the coarse-grid.



Langevin Algorithms

An important type of proposal distribution can be derived from the Langevin diffusion. The Langevin diffusion is defined by the stochastic differential equation

$$dk(\tau) = \frac{1}{2} \nabla \log \pi(k(\tau)) d\tau + dW_\tau,$$

where W_τ is the standard Brownian motion vector with independent components. The solutions at large times are samples from $\pi(k)$.

A discretization of the equation,

$$k_{n+1} = k_n + \frac{\Delta\tau}{2} \nabla \log \pi(k_n) + \sqrt{\Delta\tau} \epsilon_n,$$

where ϵ_n are independent standard normal distributions.

The proposal is chosen to be

$$Y = k_n + \frac{\Delta\tau}{2} \nabla \log \pi(k_n) + \sqrt{\Delta\tau} \epsilon_n,$$

Langevin Algorithms

The transition distribution of the proposal is

$$q(Y|k_n) \propto \exp\left(-\frac{\|Y - k_n - \frac{\Delta\tau}{2} \nabla \log \pi(k_n)\|^2}{2\Delta\tau}\right),$$
$$q(k_n|Y) \propto \exp\left(-\frac{\|k_n - Y - \frac{\Delta\tau}{2} \nabla \log \pi(Y)\|^2}{2\Delta\tau}\right).$$

The reasons for using Langevin:

- Pde's describing the physical model allow us to compute the gradients.
- The use of gradients is common in “stochastic” subsurface applications, e.g., Randomized Maximum Likelihood (RML). This approach samples the measurement data and the prior information independently and then minimize the posterior functional with these samples.
- The use of Langevin proposals usually yields higher mixing rates compared to e.g., random walk sampler.

Preconditioned coarse-gradient Langevin algorithm

The main idea: (1) use coarse-scale simulations to compute the gradient and make a proposal; (2) run the coarse-scale simulation code and check the “appropriateness” of the sample; (3) run the “fine-scale” simulation.

- Step 1. At k_n , generate a trial proposal Y from the coarse Langevin distribution $q^*(Y|k_n)$.
- Step 2. Take the proposal k as

$$k = \begin{cases} Y & \text{with probability } g(k_n, Y), \\ k_n & \text{with probability } 1 - g(k_n, Y), \end{cases}$$

where

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

- Step 3. Accept k as a sample with probability

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

where Q is the effective proposal distribution.

Preconditioned coarse-gradient Langevin algorithm

- The transition distribution of the coarse-grid proposal is

$$q^*(Y|k_n) \propto \exp\left(-\frac{\|Y - k_n - \frac{\Delta\tau}{2} \nabla \log \pi^*(k_n)\|^2}{2\Delta\tau}\right),$$
$$q^*(k_n|Y) \propto \exp\left(-\frac{\|k_n - Y - \frac{\Delta\tau}{2} \nabla \log \pi^*(Y)\|^2}{2\Delta\tau}\right).$$

- One can also construct a map from coarse-scale data to fine-scale data based on prior simulations and use it in π^* . We have used this in conjunction with more approximate upscaling methods (ensemble level upscaling) in black oil simulations.

Convergence of modified Markov Chain

Denote

$$\mathcal{E} = \{k; \pi(k) > [0]\},$$

$$\mathcal{E}^* = \{k; \pi^*(k) > [0]\},$$

$$\mathcal{D} = \{k; q(k|k_n) > [0] \text{ for some } k_n \in \mathcal{E}\},$$

To sample from $\pi(k)$ correctly, it is necessary that $\mathcal{E} \subseteq \mathcal{E}^*$. Otherwise, there will exist a subset $A \subset (\mathcal{E} \setminus \mathcal{E}^*)$ such that

$$\pi(A) = \int_A \pi(x) dx > 0 \quad \text{and} \quad \pi^*(A) = \int_A \pi^*(x) dx = 0.$$

As a result, the chain $\{k_n\}$ will never visit (sample from) A since the element of A will never be accepted for fine-scale run in Step 2. For the same reason, we should require that $\mathcal{E} \subseteq \Omega$.

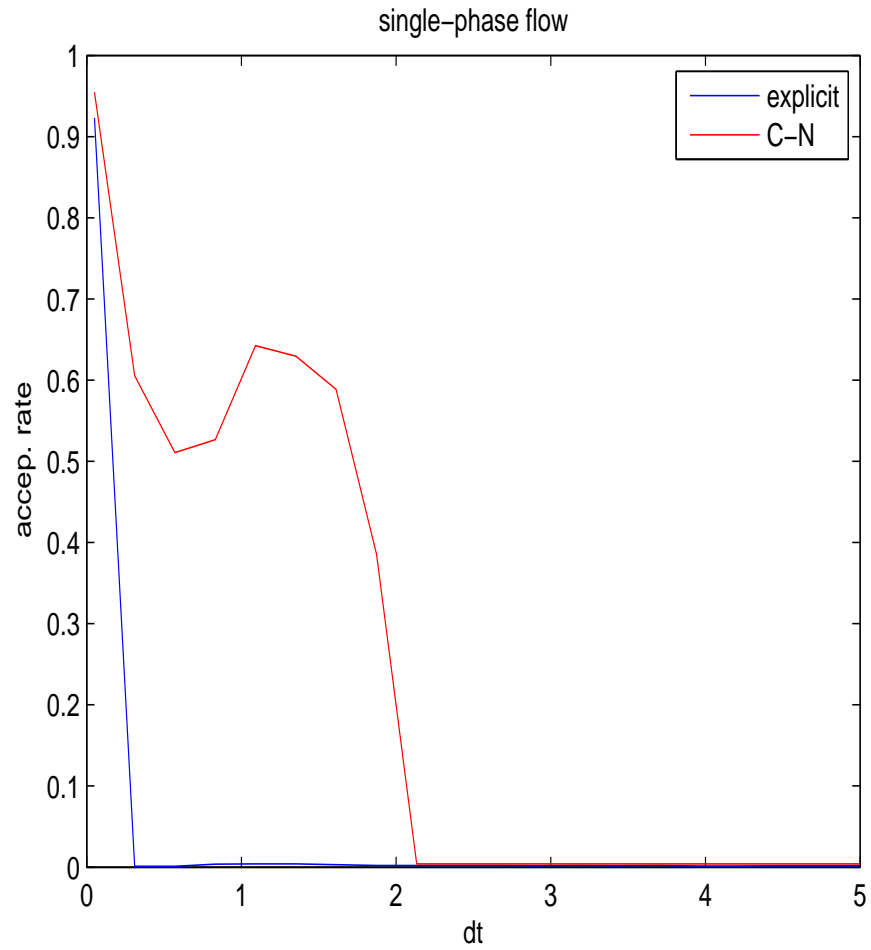
Langevin algorithms

- Langevin proposal $x \rightarrow y$ is a discretization of Langevin SDE.
- Choose largest scaling of $\Delta t(n)$ (size of move) such that $\liminf_{n \rightarrow \infty} E(\text{accep.pr}) > 0$.
- Denote the number of iterations for mixing by $M(n) = O(1/\Delta t(n))$.
- Computational cost is $C(n) = M(n) \times (\text{cost of each step})$.
- (Roberts et al 97, 98) (product of 1-d distributions): random walk sample $M(n) = O(n)$, Langevin $M(n) = O(n^{1/3})$.
- Let $\pi(k) = \exp(-F(k) - \langle k, Lk \rangle)$. We can make the proposal (A. Stuart et al. 2007, A. Datta-Gupta et al. 2007)

$$\frac{Y - k}{\Delta t} = -F(k) - \frac{1}{2}(L(k) + L(Y)) + \sqrt{1/\Delta t}N(0, 1).$$

- Under some conditions on $F(k)$ (A. Stuart et al. 2007), $M(n) = O(1)$.

Numerical results



Some remarks

- If $\pi^*(k)$ is a smooth surface, then instead of using $\pi^*(k)$, it can be interpolated employing sparse collocation techniques

$$\tilde{\pi}^* = \sum_i \pi^*(k_i) L_i(k),$$

where k_i are sparse collocation points and $L_i(k)$ are polynomials (similar ideas have been used in literature, e.g., B. Ganapathysubramanian and N. Zabaras, JCP, 2007).

- For more complex models and simple upscaling methods, we have used nonparameteric methods to construct a nonlinear relation between coarse and fine responses - $R^c \approx G(R^f)$ based on off-line simulations. G is then used in speeding-up the sampling (Efendiev, A. Datta-Gupta, X. Ma, WRR, 2008).
- The analysis of MH in the homogenization setting can be carried out. Assuming that the media properties are strictly stationary and have scale separation, then the media properties can be homogenized.
- Homogenized media properties are independent of space and random variable which is invariant w.r.t. to the dynamical systems representing the spatial media.

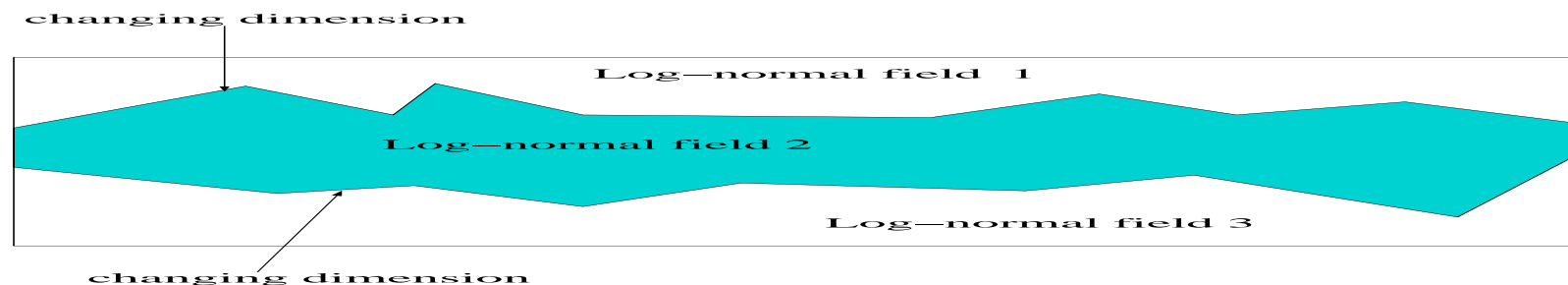
Numerical setting

- We have mostly considered log-normal conductivity fields $k(x) = \exp(Y(x))$, where $Y(x)$ is prescribed with a covariance matrix (e.g., normal or exponential). The conductivity field is parameterized via Karhunen-Loève Expansion

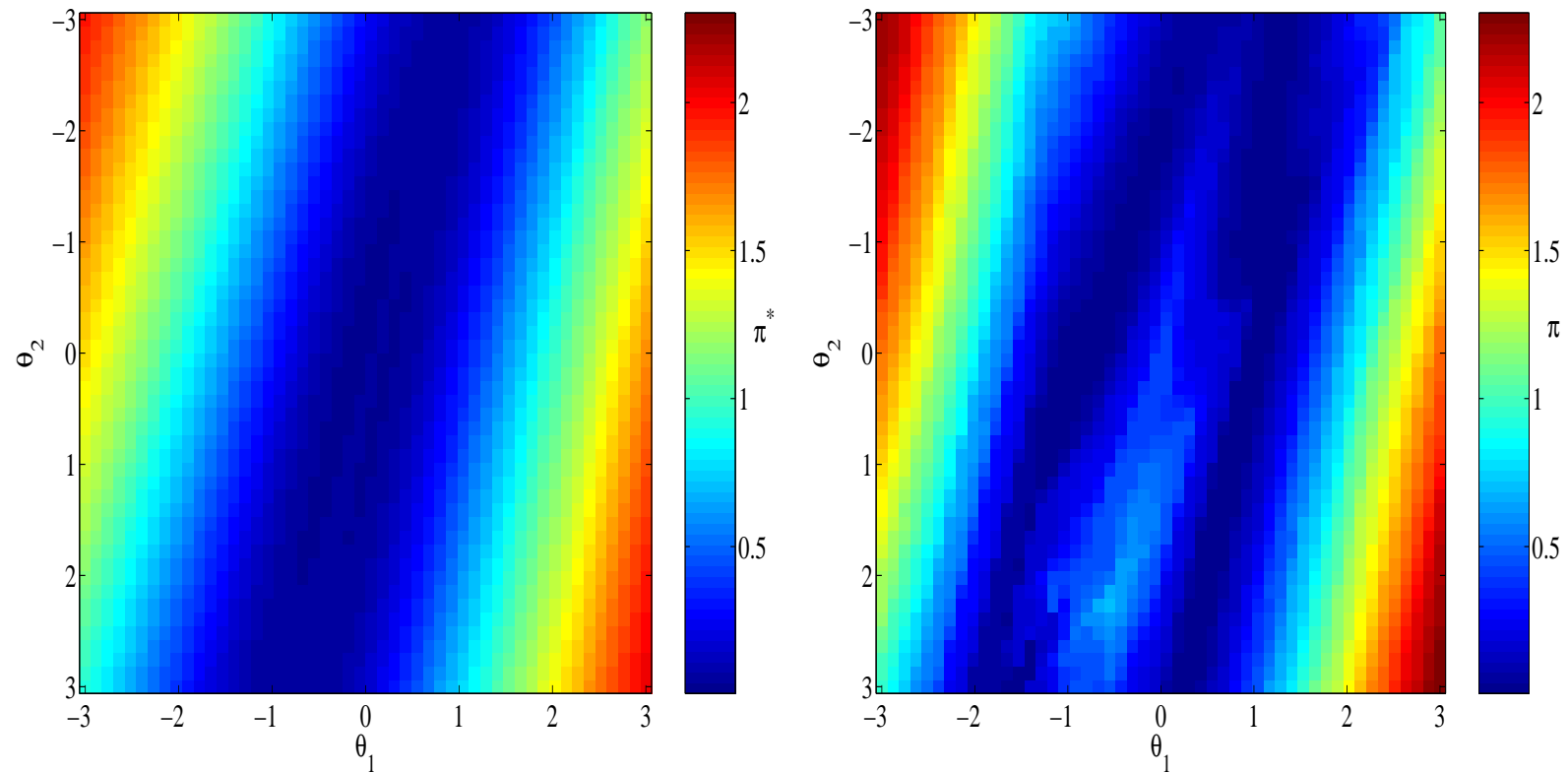
$$Y(x, \omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \theta_k(\omega) \phi_k(x),$$

where $E(\theta_k) = 0$, $E(\theta_i \theta_j) = \delta_{ij}$, λ_k and $\phi_k(x)$ are eigenvalues and eigenvectors of covariance matrix. The conductivity field can be conditioned at some locations.

- We have also considered hierarchical models representing facies. Stochastic level set equations are used to characterize the “movement” of facies’ boundaries. Reversible jump MCMC is used.

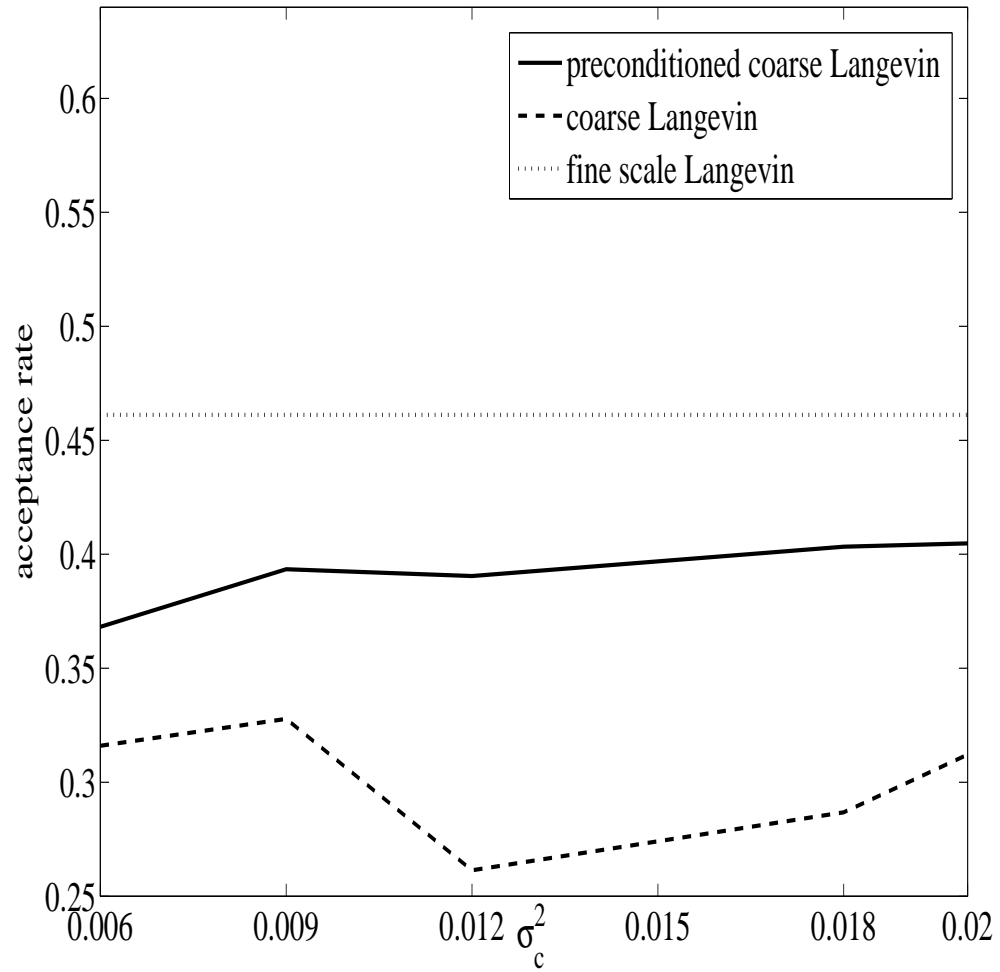


Numerical Results



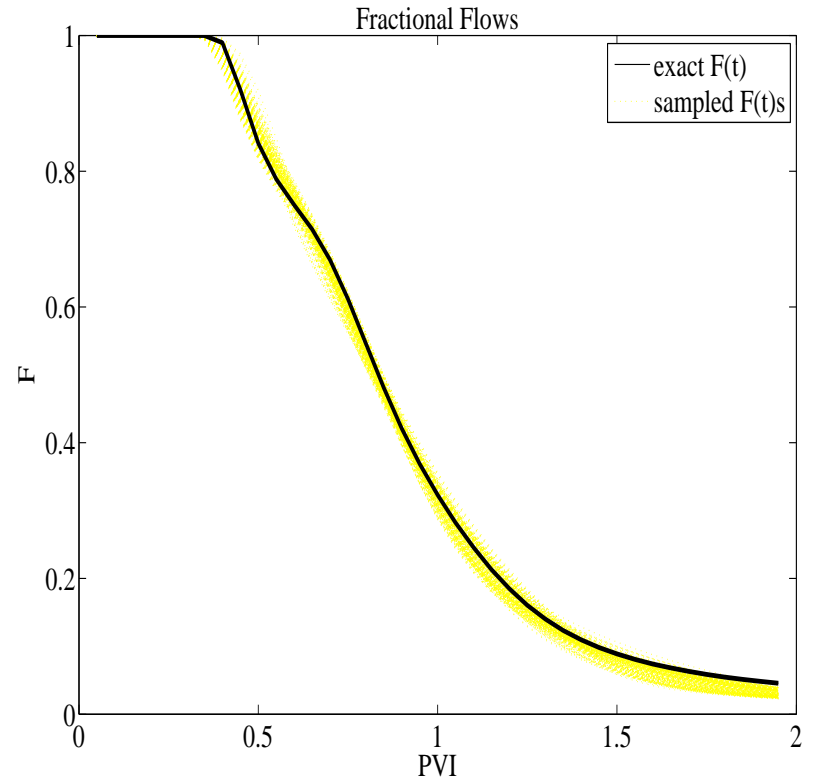
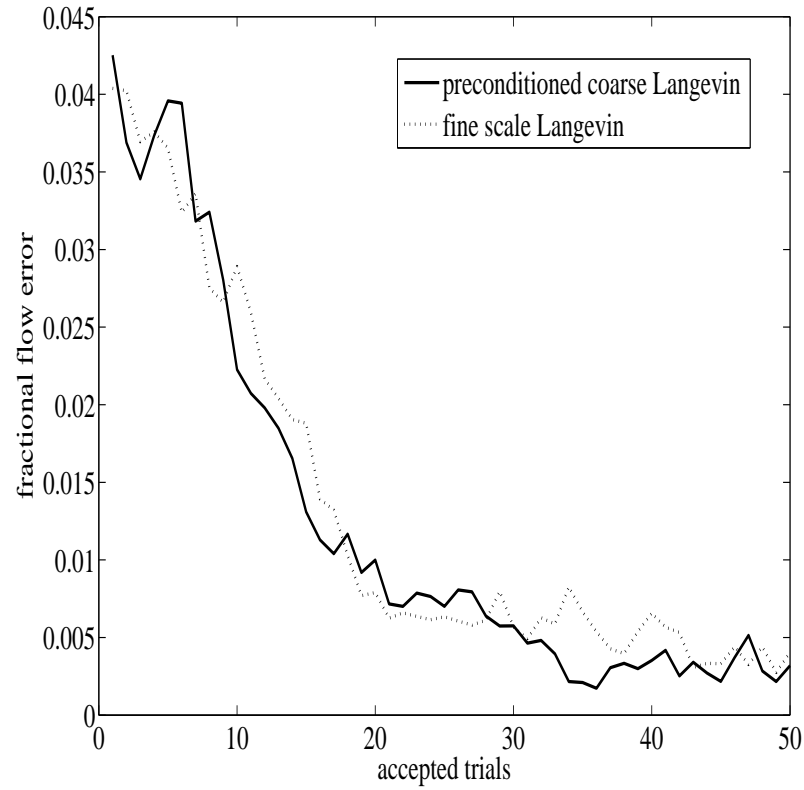
Left: Coarse-scale response surface π^* restricted to a 2-D hyperplane. Right: Fine-scale response surface π restricted to the same 2-D hyperplane.

Numerical results

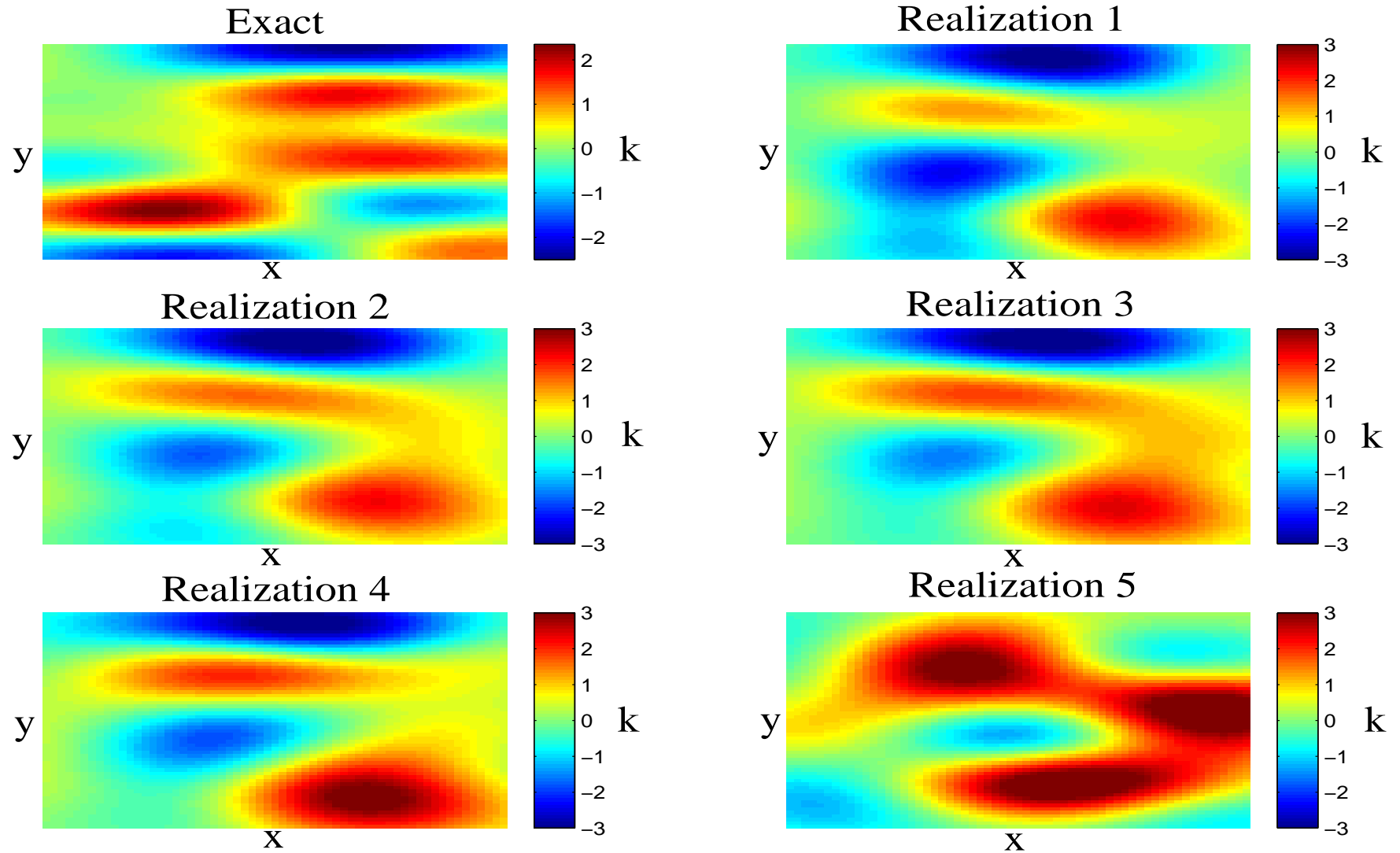


Acceptance rate comparison, $\delta = 0.1$.

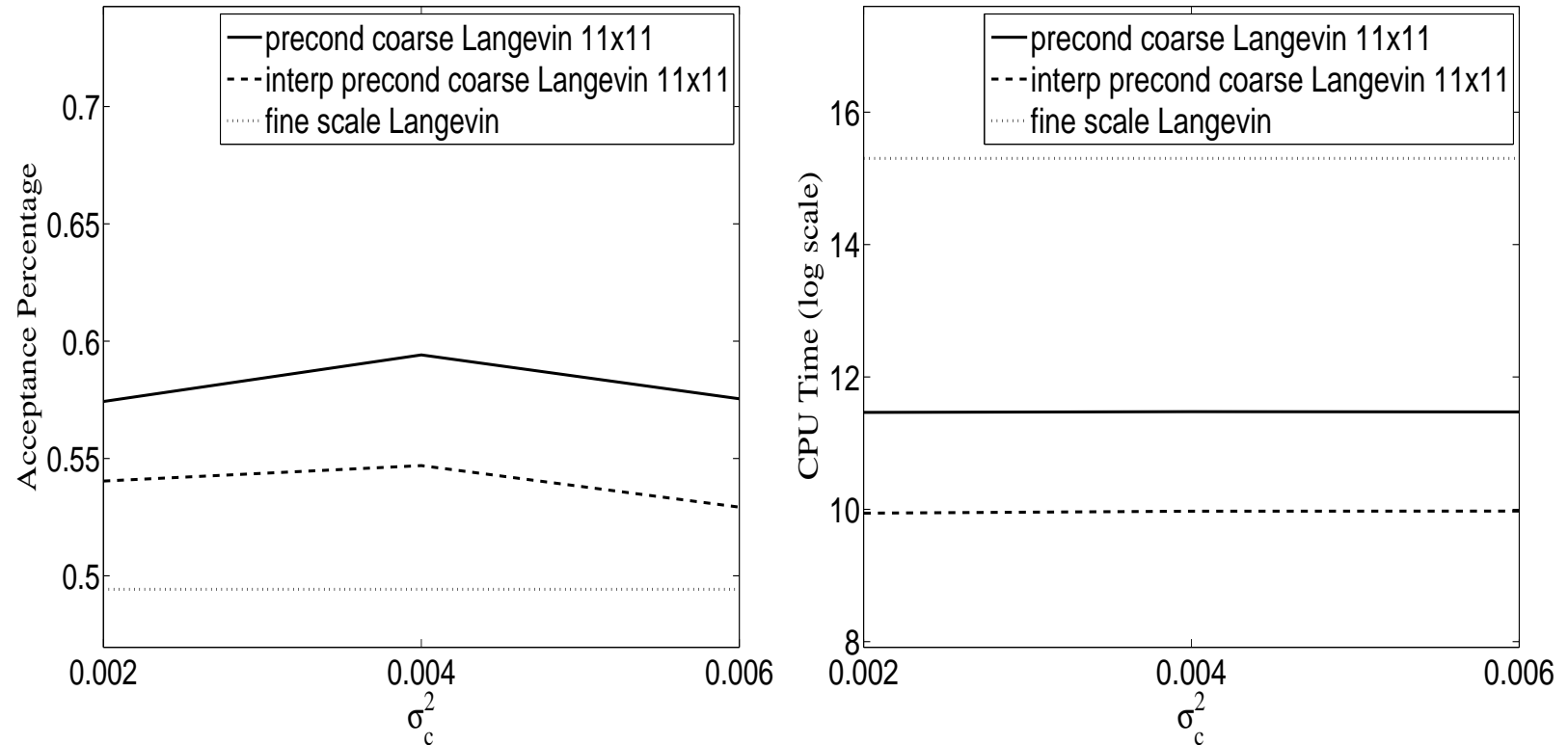
Numerical results



Numerical results

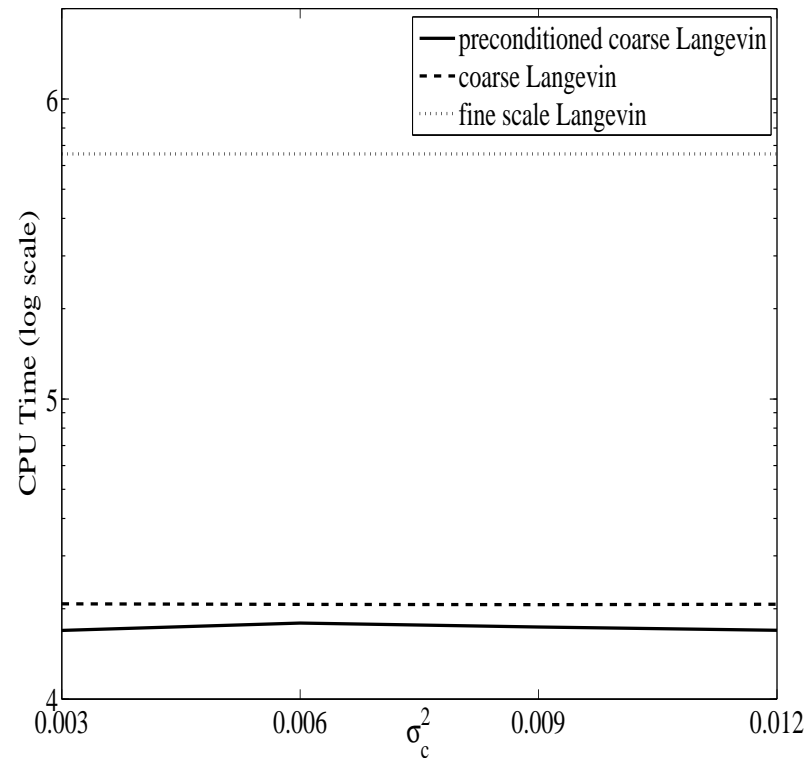


Numerical Results



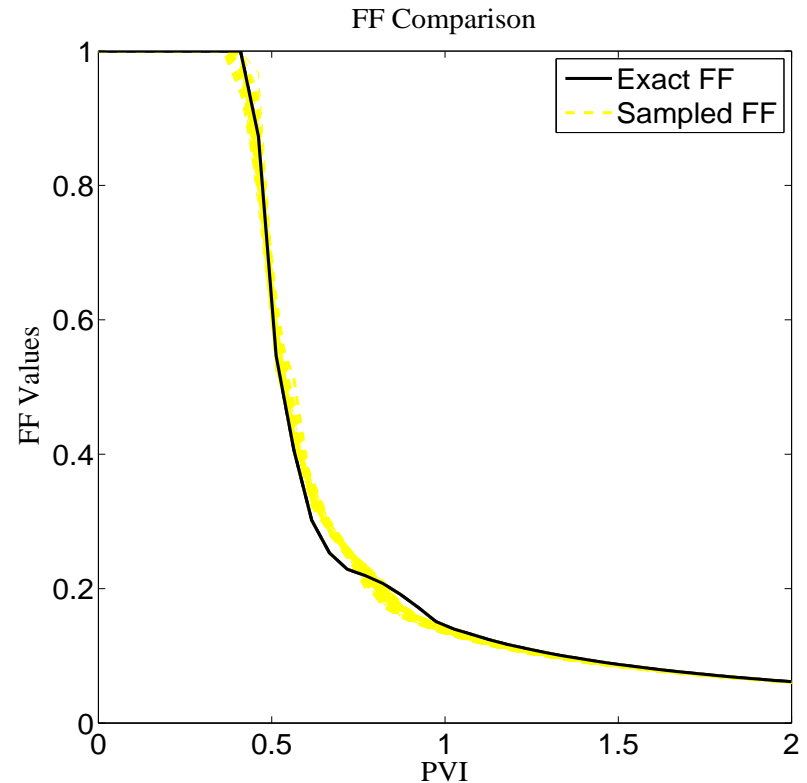
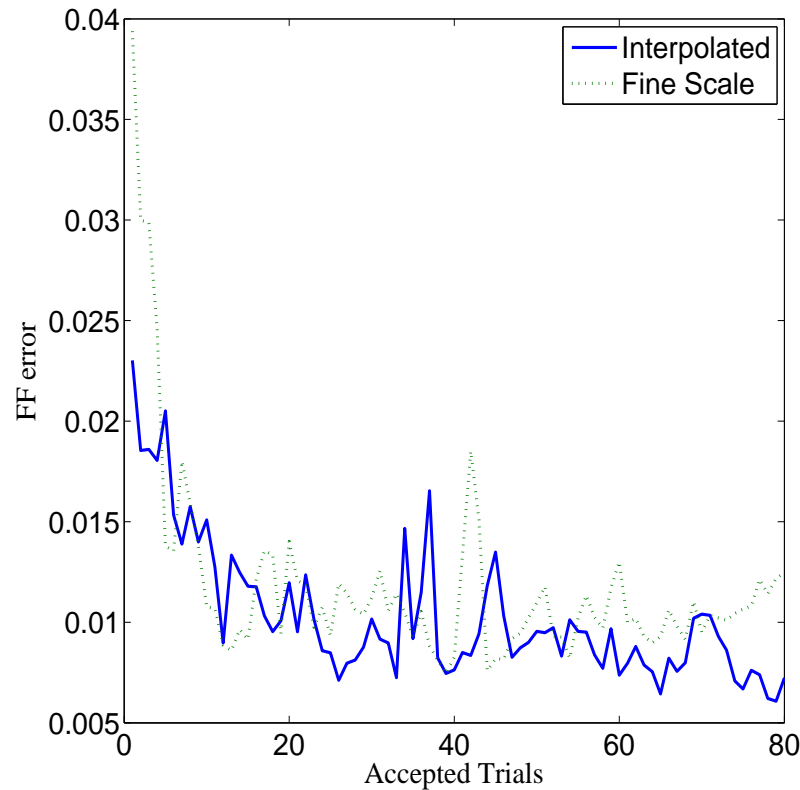
Left: Acceptance rate comparison. Right: Natural log of CPU time (seconds) comparison. In each plot $\delta = 0.05$ and $\sigma_f^2 = 0.002$.

Numerical Results



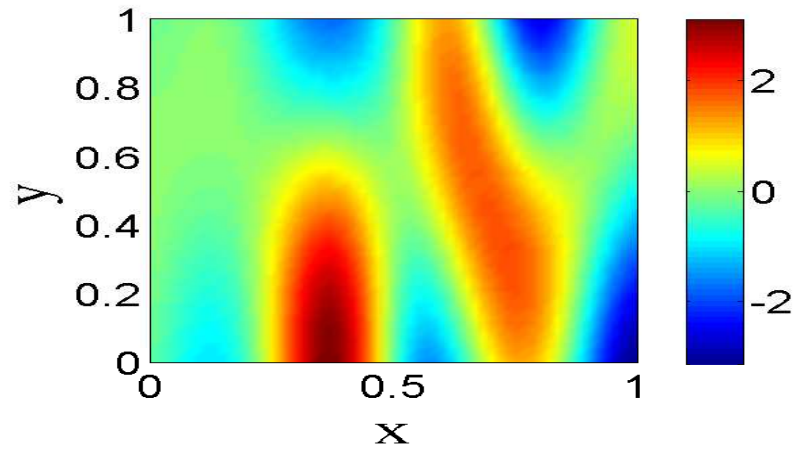
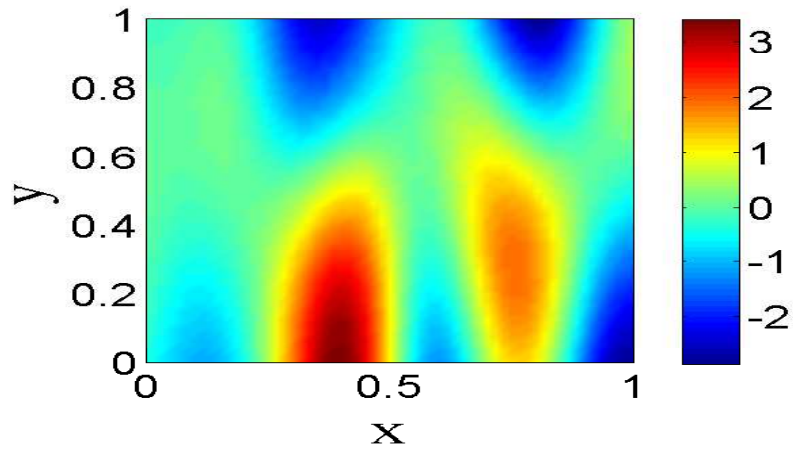
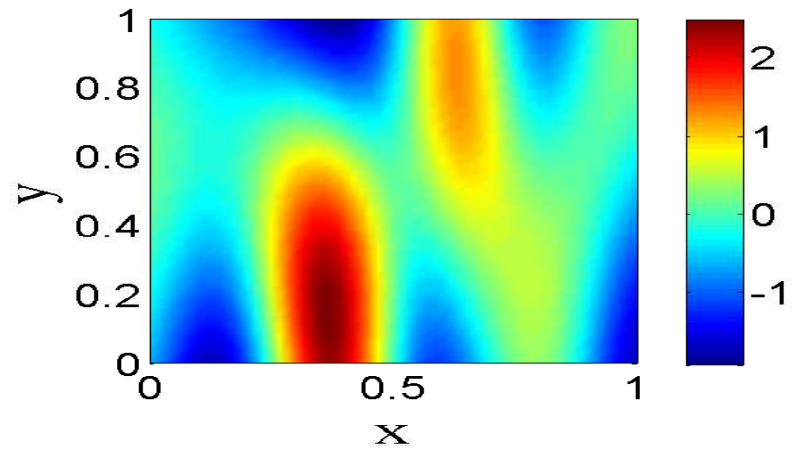
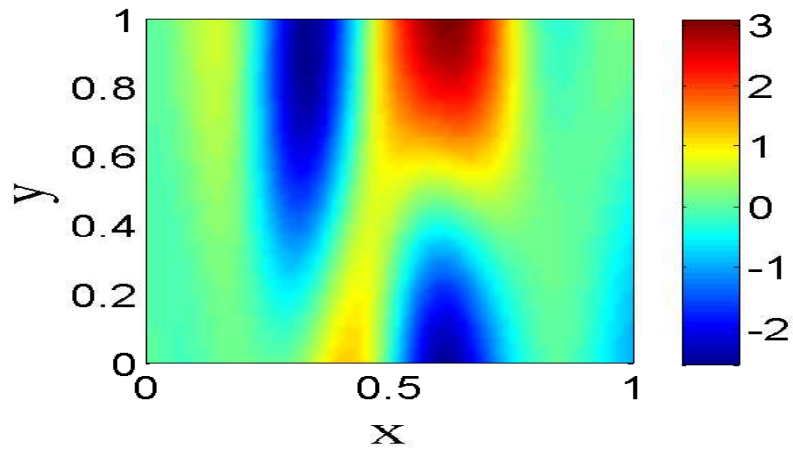
CPU times (seconds) for Langevin algorithms. $\sigma_f^2 = 0.003$, $\delta = 0.05$, 7×7 coarse-grid.

Numerical Results



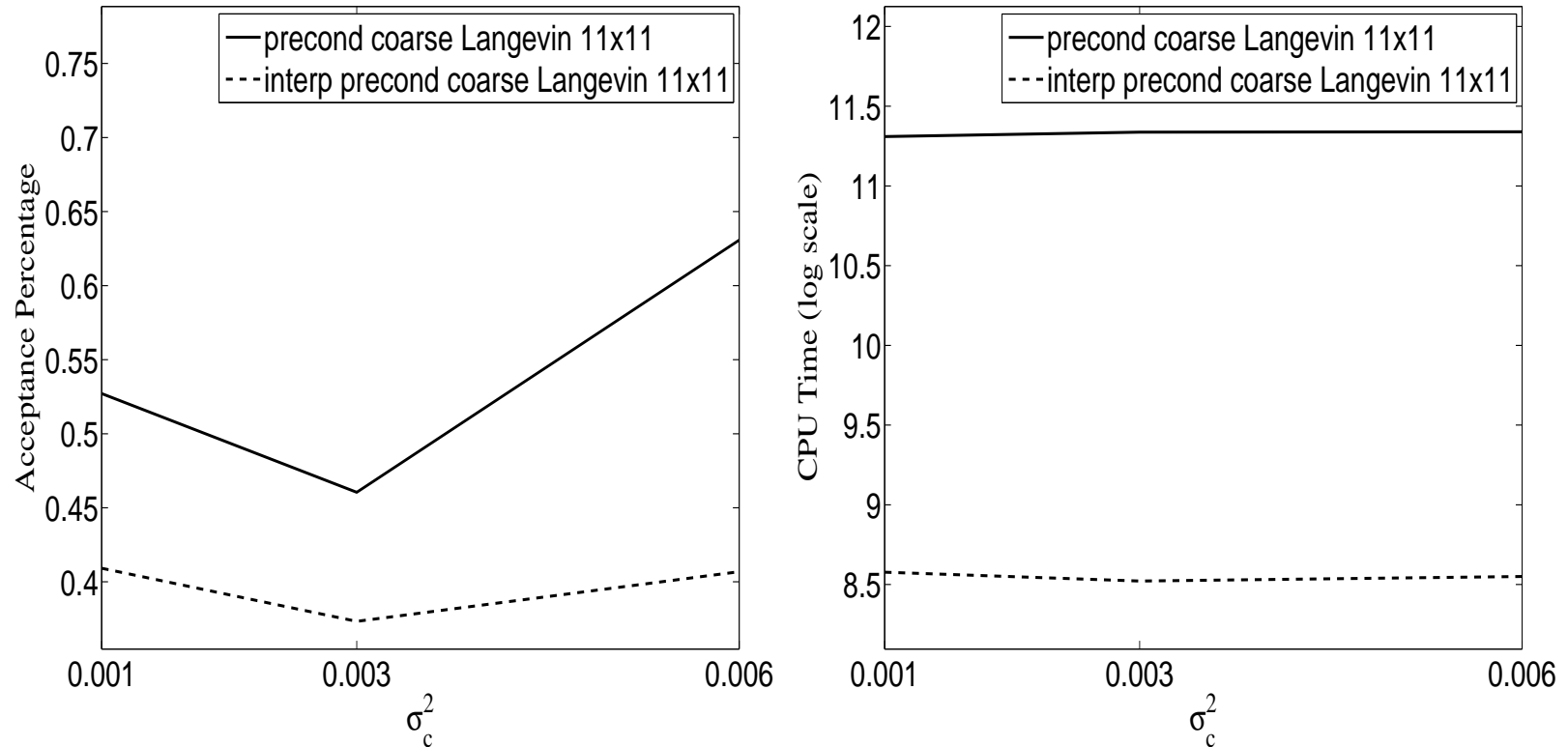
Left: The fractional flow errors for the fine Langevin algorithm compared with interpolated Langevin algorithm. Right: The fractional flows of sampled realizations and the reference fractional flow. In these numerical tests, $\delta = 0.05$ and $\sigma_f^2 = 0.002$.

Numerical Results



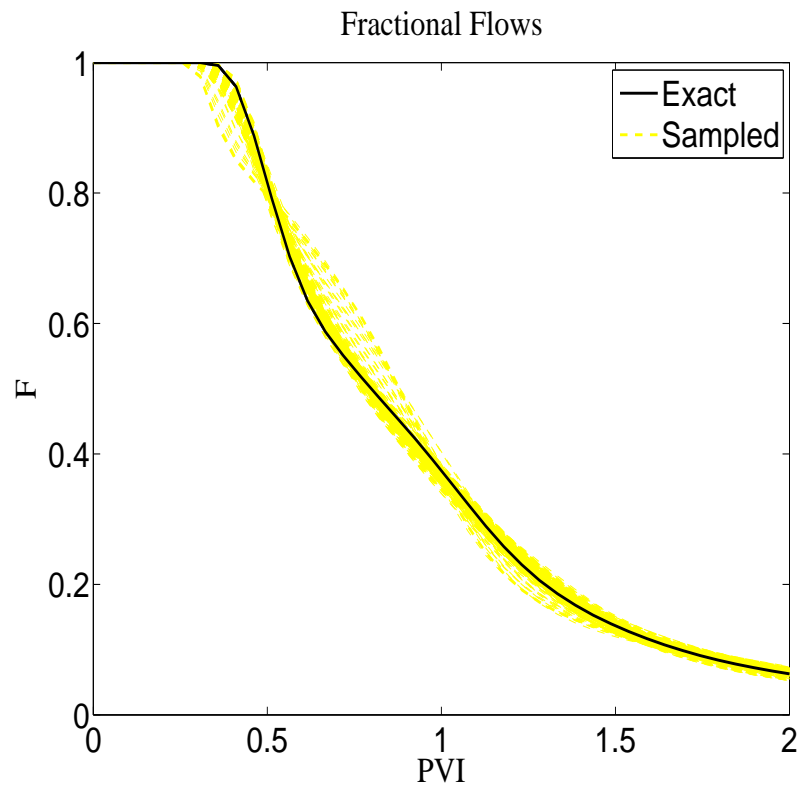
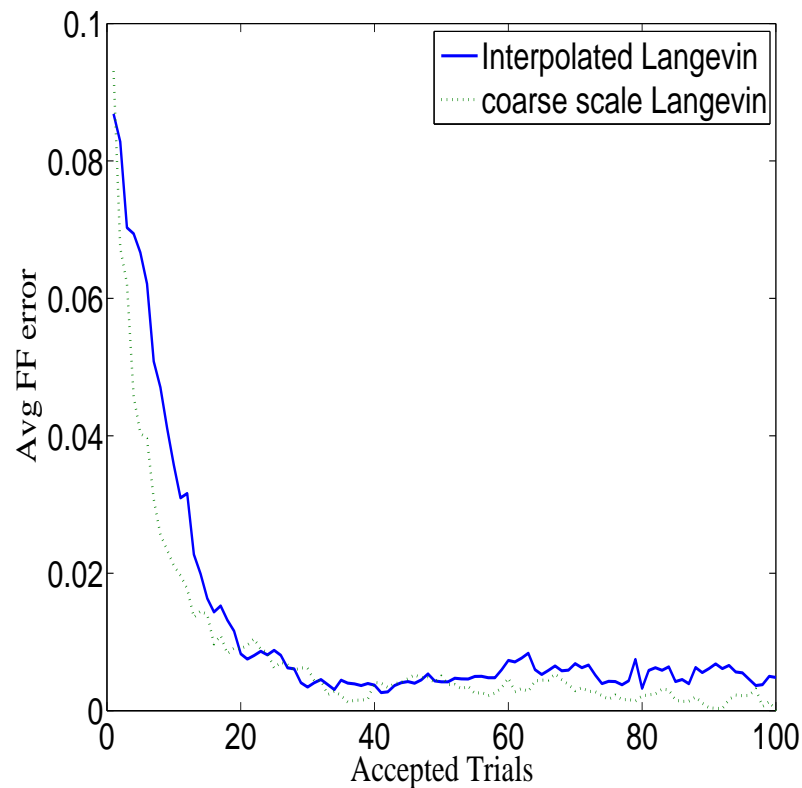
Upper left plot is the reference conductivity. The other three plots are examples of accepted conductivity realizations.

Numerical Results



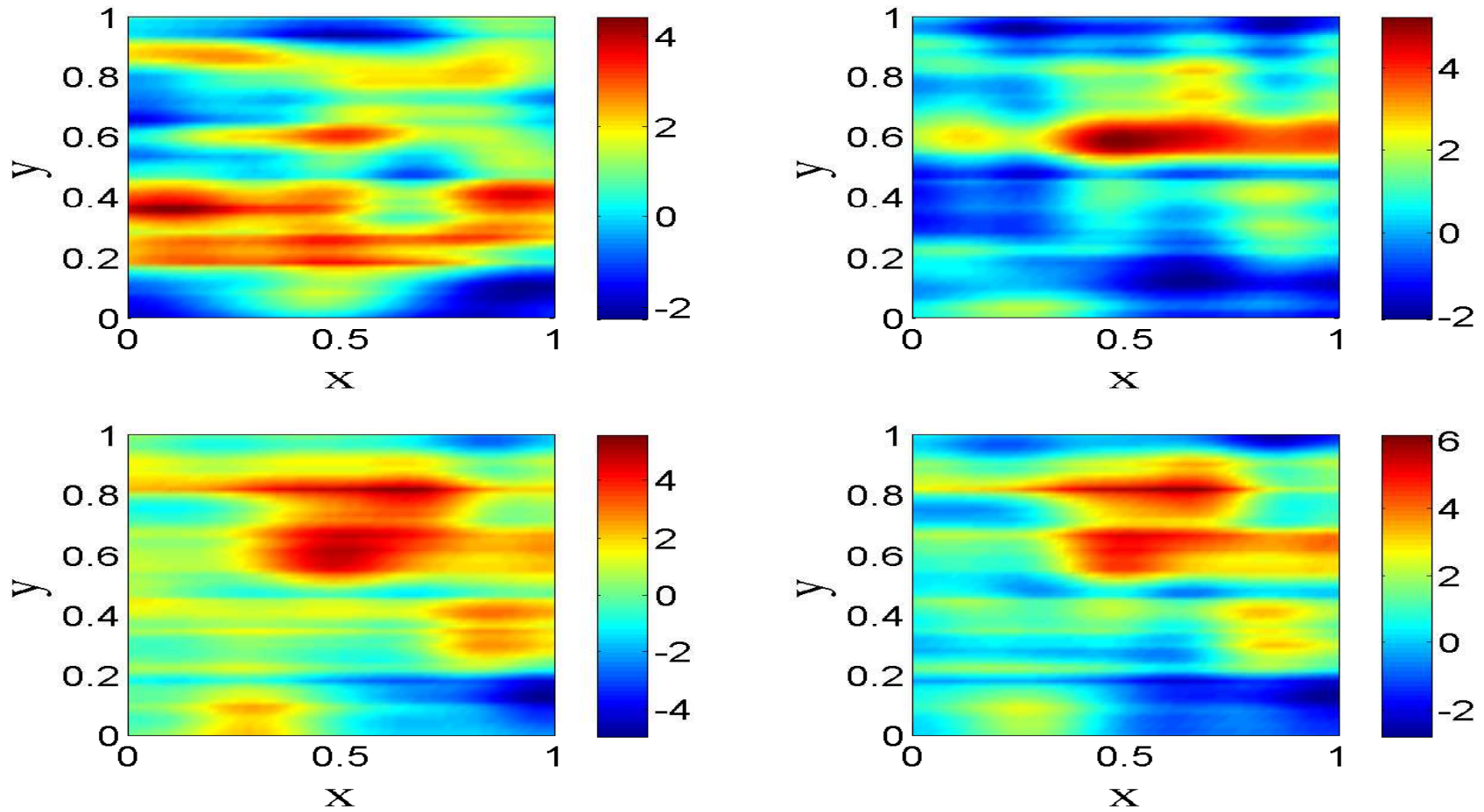
Left: Acceptance rate comparison. Right: Natural log of CPU time (seconds) comparison. In each plot we use $\delta = 0.05$ and $\sigma_f^2 = 0.001$.

Numerical Results



Left: The fractional flow errors for coarse Langevin compared with interpolated Langevin. Right: The fractional flows of sampled realizations and the reference fractional flow. In these numerical tests, $\delta = 0.05$, $\sigma_f^2 = 0.001$.

Numerical results



Upper left plot is the reference conductivity. The other three plots are examples of accepted conductivity realizations.

Conclusions

- Direct sampling using MH MCMC approaches is expensive
- Inexpensive coarse-scale models can be used to precondition Langevin MH simulations.
- Coarse-scale simulations are based on multiscale finite element methods.
- Multiscale basis functions can be constructed to represent an ensemble of conductivity fields.
- Numerical discretization of Langevin equation can help to improve the mixing of MC.
- Numerical results demonstrate CPU time can be reduced by two orders of magnitude.
- We have extended it to take into account the fields which are locally stationary.
- We have applied these techniques to soil moisture predictions where the data at different scales (from satellite measurements) are obtained.