A Bayesian approach to multiscale inverse problems using the sequential Monte Carlo method

2011 Inverse Problems 27 105004

(http://iopscience.iop.org/0266-5611/27/10/105004)

Inverse Problems **27** (2011) 105004 (25pp) doi:10.1088/0266-5611/27/10/105004

A Bayesian approach to multiscale inverse problems using the sequential Monte Carlo method

Jiang Wan and Nicholas Zabaras[1]

Materials Process Design and Control Laboratory, Sibley School of Mechanical and Aerospace Engineering, 101 Frank H T Rhodes Hall, Cornell University, Ithaca, NY 14853-3801, USA

E-mail: zabaras@cornell.edu

Received 20 September 2010, in final form 26 May 2011
Published 16 September 2011
Online at stacks.iop.org/IP/27/105004

Abstract

A new Bayesian computational approach is developed to estimate spatially varying parameters. The sparse grid collocation method is adopted to parameterize the spatial field. Based on a hierarchically structured sparse grid, a multiscale representation of the spatial field is constructed. An adaptive refinement strategy is then used for computing the spatially varying parameter. A sequential Monte Carlo (SMC) sampler is used to explore the posterior distributions defined on multiple scales. The SMC sampling is directly parallelizable and is superior to conventional Markov chain Monte Carlo methods for multi-modal target distributions. The samples obtained at coarser levels of resolution are used to provide prior information for the estimation at finer levels. This Bayesian computational approach is rather general and applicable to various spatially varying parameter estimation problems. The method is demonstrated with the estimation of permeability in flows through porous media.

(Some figures in this article are in colour only in the electronic version)

1. Introduction

An important category of inverse problems is the identification of spatially varying parameters using indirect data. A typical example is the permeability estimation of the aquifer from flow data. The measurement error and inadequacy of models for complicated physical phenomena can reduce the accuracy of the estimation [1, 2]. The deterministic approaches address these problems based on exact matching or least-squares optimization without quantifying the uncertainty of the solution [3]. Other alternative approaches based on the spectral stochastic method or Bayesian inference [4–8] take into account the statistical nature of inverse problems and provide a full probabilistic description of the computed fields.

[1] Author to whom any correspondence should be addressed.

In Bayesian inference of spatially varying parameters, finite element methods are often used to discretize the unknown field [9, 10]. Standard models for spatial data, such as the Markov random field or Gaussian process (GP), are then used to model the spatially varying parameters [2, 9–12]. To increase the flexibility of the model, the process convolution approach is used as an alternative [2, 10]. By convolving white noise with a smoothing kernel, the unknown parameter field is approximated as a superposition of kernel-type functions centered at various locations. The inverse problem is then transformed to one that infers the coefficients of the expansion. In this scheme, the choice of kernels is important [11]. However, these methods are based on assumptions about the spatial correlation and are not suitable for complicated spatial fields. This challenge motivated us to develop a new way of modeling the unknown field of spatially varying parameters. To this end, we introduce a hierarchical representation of the parameter field based on sparse grid interpolation. The sparse grid collocation method uses the Smolyak algorithm to construct an interpolation of the target function with hierarchical grids and basis functions [13, 14]. The collocation points (nodes) are selected in a nested fashion to obtain many recurring points and basis functions with increasing sparse grid levels [14–16]. In other words, the basis functions are constructed on multiple scales. An adaptive refinement strategy is proposed for optimal choice of collocation points using hierarchical surplus as an error indicator to place more points around non-smooth regions [16]. In this way, rapid changes in the spatial field can be effectively captured and an optimal representation of the spatially varying parameter with minimum requirement of collocation points is achieved. In this paper, we start from a coarse level of sparse grid representation of the unknown parameter field and make inference about the surpluses. The magnitude of the posterior mean of the surpluses controls any further refinement.

The Bayesian inference is performed using Monte Carlo (MC) methods. Standard MC methods such as Markov chain Monte Carlo (MCMC) generally perform poorly for multi-modal target distributions. In this paper, we propose a sequential Monte Carlo (SMC) algorithm for sampling from the posterior distributions on various levels of the sparse grid. For each target distribution defined on a given sparse grid, a sequence of auxiliary distributions is constructed. A set of random samples is propagated according to these distributions using important sampling and resampling techniques [2, 17–20]. Great robustness to multimodality is then achieved. Moreover, the SMC method naturally fits the hierarchical representation of the spatially varying parameters. The SMC method enables a hierarchical estimation based on the nested property of the sparse grid. The estimates at coarse levels provide plenty of prior information for the estimation on finer levels, which results in more efficient exploration of the state space. In addition, this algorithm is directly parallelizable.

The rest of the paper is structured as follows. In section 2, the parameterization of the unknown parameter field using sparse grids is introduced. In section 3, a one-scale Bayesian model for the spatially varying parameter is built. The SMC method is utilized in the exploration of the posterior state space. In section 4, a strategy that bridges the estimation of surpluses on different scales is proposed. Section 5 presents a case study on permeability estimation. Finally, conclusions are given in section 6.

2. Parameterization of the unknown parameter field

The spatially varying parameter of a physical system belongs to an infinite dimensional space. In a Bayesian framework, it is usually reduced to a finite space and the inference is performed on a finite set of random variables. A simple way of implementing this task is to discretize the spatial domain into finite elements. The value of the spatially varying parameter is assumed constant within each element [9]. However, if improper resolution of discretization

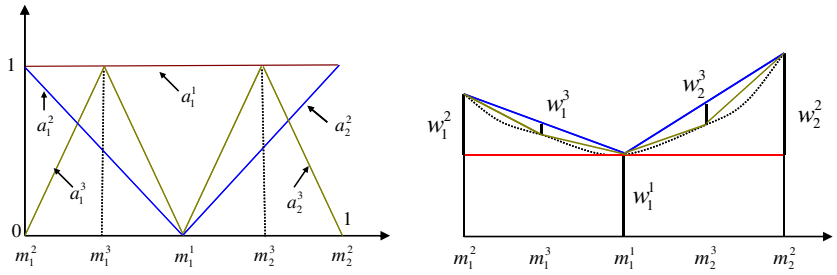

Figure 1. Hierarchical basis functions a_j^i with the support nodes (left) and the hierarchical surpluses (right). The surplus w_j^i is defined as the difference between the function value computed at a newly added point on the current sparse grid and the interpolated value at this point from the previous interpolation level.

is selected, either overfitting or a waste of computational resources takes place [2]. To increase the flexibility of the model, several researchers consider a basis function approach, such as in the truncated Karhumen–Loève expansion or using Gaussian kernels, to represent the unknown parameter field [2, 10, 21, 22]. However, these methods might require prior knowledge of the correlation or covariance functions of the stochastic process. In addition, an optimal choice of the basis could be a difficult task. Although the number of basis functions is not fixed and treated as a random variable in the trans-dimensional MCMC method, it is difficult to alter the dimension significantly while ensuring a reasonable acceptance ratio [23, 24].

In this work, we propose a hierarchical basis representation of the spatially varying parameter based on the sparse grid interpolation method. The basic idea is to have a hierarchical structure of representation that ranges from coarse to fine scales of discretization. Thus, we can perform sequential Bayesian estimation from a coarse scale with a few collocation points until an optimal choice of collocation points and basis functions is achieved. The method is still based on a finite element discretization of the spatial domain for the purpose of solving the underlying differential equations, but the unknown parameter field is approximated using interpolating functions on a set of collocation points.

The sparse grid method is a special discretization technique which constructs hierarchical interpolation based on the Smolyak algorithm [14–16]. The standard sparse grid is defined with collocation points and basis functions completely fixed at different levels of resolution. Given a sparse grid, the unknown function for the spatially varying parameter can be approximated using basis functions associated with the grid. For example, consider a parameter θ in 1D; the function $\theta = f(x)$ is approximated by the nodal basis of interpolation level 2 as (figure 1)

$$\theta = f(x) \approx w_1^1 a_1^1 + w_1^2 a_1^2 + w_2^2 a_2^2. \tag{1}$$

On level 3, it is approximated as

$$\theta = f(x) \approx w_1^1 a_1^1 + w_1^2 a_1^2 + w_2^2 a_2^2 + w_1^3 a_1^3 + w_2^3 a_2^3, \tag{2}$$

by hierarchically adding two basis functions.

The sparse grid provides several advantages in Bayesian inference of spatially varying parameters. First, the accuracy of the interpolation is increased without discarding previous results. The reusability of the collocation points and basis functions enables a sequential estimation of the surpluses from coarse scales. One does not have to set up a fine grid and estimate a large number of surpluses simultaneously. Furthermore, the definition of the surplus and the nested fashion of the collocation points potentially enable an adaptive refinement of

the grid. The collocation points of the adaptive sparse grid are case determined and are only a subset of the nodes of the standard sparse grid at the same level, which further reduces the dimensionality of the inverse problem. Detailed discussion about these topics will be given in later sections as part of our discussion of the hierarchical Bayesian model.

3. Bayesian inference

In the last section, the spatially varying parameter is parameterized by the sparse grid method. In the Bayesian framework, the unknown surpluses are treated as random variables and inferred from the observation data. In this section, a complete one-scale Bayesian model is developed on a pre-determined sparse grid. The SMC sampler is utilized for an efficient exploration of the posterior state space.

3.1. Bayesian formulation

Consider a generalized forward problem defined as

$$\boldsymbol{d} \approx \boldsymbol{F}(\theta), \tag{3}$$

where $\boldsymbol{d}$ denotes the observation data and θ denotes the model parameter which is considered as a random variable or a random vector. Based on Bayes' theorem, the posterior probability density for θ is

$$p(\theta|\boldsymbol{d}) \propto p(\boldsymbol{d}|\theta)p(\theta). \tag{4}$$

Here, $p(\boldsymbol{d}|\theta)$ is the likelihood distribution and $p(\theta)$ is the prior probability density. The forward solver $\boldsymbol{F}$ gives predictions by solving parameterized partial differential equations with numerical methods such as the finite element method.

3.1.1. Prior specification. When the model parameter θ is a spatially varying parameter, the forward problem can be reformulated as

$$\boldsymbol{d} \approx \boldsymbol{F}(f(\boldsymbol{x})), \tag{5}$$

where $\boldsymbol{x}$ refers to the location and $f(\boldsymbol{x})$ is the specific field of the spatially varying parameter, i.e. $\theta = f(\boldsymbol{x})$. On a pre-determined sparse grid at level q, the parameter field can be represented as a weighted sum of the basis functions for all collocation points in the sparse grid from level 0 to q due to the hierarchical structure of the grid, that is

$$\theta \approx \theta_q = f_q(\boldsymbol{x}) = \sum_{i=0}^{q} \sum_{j=1}^{k_i} w_j^i \cdot a_j^i(\boldsymbol{x}). \tag{6}$$

Hence, our problem is reduced to the inference of the surpluses, w_j^i. A simple non-informative prior for the surpluses is the multivariate Gaussian distribution which assumes that the surpluses are identically independently distributed as

$$\boldsymbol{w}_q|\sigma_{w,q} \sim \mathcal{N}\big(0, \sigma_{w,q}^2 \mathbf{I}_k\big), \tag{7}$$

where $\boldsymbol{w}_q$ is the vector of all surpluses up to level q for θ_q and $\mathbf{I}_k$ is an identity matrix. Suppose the hyperparameter $\sigma_{w,q}^{-2} \sim \Gamma(\alpha_0, \beta_0)$; we can premarginalize the variance from equation (7) and obtain the prior

$$p(\boldsymbol{w}_q) \propto \frac{\Gamma\big(\alpha_0 + \frac{k}{2}\big)}{\big(\beta_0 + \frac{1}{2}\|\boldsymbol{w}_q\|_2^2\big)^{\alpha_0 + \frac{k}{2}}}, \tag{8}$$

where k is the length of the vector $\boldsymbol{w}_q$.

3.1.2. Likelihood. To evaluate the likelihood, the discrepancy between the forward solver prediction and observation data should be estimated and modeled. Two primary sources of error are considered in the modeling. One is the measurement noise ζ, which is generally assumed to be an independent additive Gaussian random error with mean zero, i.e.

$$\zeta_i \sim \mathcal{N}\big(0, \sigma_\zeta^2\big), \tag{9}$$

where ζ_i is an element of the vector ζ. We postulate the following relationship:

$$\boldsymbol{d} = \boldsymbol{F}(\theta) + \zeta, \tag{10}$$

i.e. the observation is obtained from accurate forward solver predictions plus measurement noise.

The other source is the model error δ, or referred to as the approximation error, which results from the inadequacy of the forward model to represent the real physical process [25, 26]. To minimize this effect, the forward solver operates at a fixed fine scale in this work. We restrict our focus on the part of the model error that results from the parameterization of the spatially varying parameters.

With a sparse grid representation of the spatially varying parameter θ, the accurate parameter field $f(\boldsymbol{x})$ is transformed into a reduced approximative model described by equation (6). This approximation error will propagate in the forward solver and result in inaccurate predictions. Thus, the observation is formulated as

$$\begin{aligned} \boldsymbol{d} &= \boldsymbol{F}(\theta) + \zeta \\ &= \boldsymbol{F}_q(\theta_q) + (\boldsymbol{F}(\theta) - \boldsymbol{F}_q(\theta_q)) + \zeta \\ &= \boldsymbol{F}_q(\theta_q) + \delta + \zeta, \end{aligned} \tag{11}$$

where q is the level of the sparse grid, and $\boldsymbol{F}_q(\theta_q)$ refers to the forward solver using the approximate model θ_q for the spatially varying parameter θ. The model error is defined as $\delta = \boldsymbol{F}(\theta) - \boldsymbol{F}_q(\theta_q)$, the discrepancy of forward solver predictions using precise spatially varying parameters and using an approximate model as in equation (6).

In Bayesian inference, the model error δ can be treated as a random field. A Markov random field is employed as in [11]. The basic idea is that the model error at a particular location is correlated with the model errors at neighboring locations. Suppose a random vector $e = \delta + \zeta$; we assume

$$e \sim \mathcal{N}(\delta_0, \Sigma_e), \tag{12}$$

where $\delta_0 = \delta_q \mathbf{1}$ and the covariance matrix Σ_e is exponential formulated as

$$\Sigma_e = \sigma_\zeta^2 \mathbf{I} + \kappa H(\phi). \tag{13}$$

The parameter σ_ζ^2 represents the variance of the measurement error. $H(\phi) = \{H_{ij}(\phi)\}$ where $H_{ij}(\phi) = \exp\big(-\frac{||s_i - s_j||}{\phi}\big)$ and $||s_i - s_j||$ is the Euclidean distance between locations s_i and s_j. This is a measure of the magnitude of spatial dependence. κ and ϕ indicate the scale and the range of the spatial dependence, respectively [27]. Thus, the likelihood for general data modeling is

$$\boldsymbol{d}|\boldsymbol{w}_q, \delta_q, \Sigma_e \sim \mathcal{N}(\boldsymbol{F}_q(\boldsymbol{w}_q) + \delta_q \mathbf{1}, \Sigma_e). \tag{14}$$

To further reduce the dimensionality of the problem as well as the complexity of computation, we also adopt a much simplified model. We simply assume that δ is a random field (vector) with independent elements subject to a multivariate Gaussian distribution:

$$\delta \sim \mathcal{N}\big(\delta_0, \sigma_\delta^2 \mathbf{I}\big). \tag{15}$$

Since $\zeta \sim \mathcal{N}(0, \sigma_\zeta^2 \mathbf{I})$ according to equation (9), e is also subject to a Gaussian distribution $e \sim \mathcal{N}(\delta_0, \Sigma_e)$ where $\Sigma_e = \left(\sigma_\delta^2 + \sigma_\zeta^2\right)\mathbf{I}$. Thus, the likelihood is simply

$$\boldsymbol{d}|\boldsymbol{w}_q, \delta_q, \sigma_e \sim \mathcal{N}\left(\boldsymbol{F}_q(\boldsymbol{w}_q) + \delta_q \mathbf{1}, \sigma_e^2 \mathbf{I}\right), \tag{16}$$

where $\sigma_e^2 = \sigma_\delta^2 + \sigma_\zeta^2$. In this model, a gamma distribution $\Gamma(\alpha_1, \beta_1)$ is taken for σ_e^{-2}. By premarginalization, the unknown variance can be integrated out. The likelihood function is simply

$$p(\boldsymbol{d}|\boldsymbol{w}_q, \delta_q) \propto \frac{\Gamma\left(\alpha_1 + \frac{m}{2}\right)}{\left(\beta_1 + \frac{1}{2}\|\boldsymbol{d} - \boldsymbol{F}_q(\boldsymbol{w}_q) - \delta_q \mathbf{1}\|_2^2\right)^{\alpha_1 + \frac{m}{2}}}, \tag{17}$$

where m is the number of observation data. In the following sections, we refer to the former model as model *I* and to the latter one as model *II*.

3.1.3. Complete Bayesian model. Consider a pre-determined sparse grid at level q and let $\psi_q = \{\boldsymbol{w}_q, \delta_q, \sigma_\zeta, \kappa, \phi\}$ denote the vector containing all unknown parameters of the Bayesian model on this grid. The prior for the hyperparameter δ_q is taken to be $\delta_q \sim \mathcal{N}(0, \sigma_{\delta_q})$. A gamma distribution $\Gamma(\alpha_{\sigma_\zeta}, \beta_{\sigma_\zeta})$ is chosen as the prior for σ_ζ^{-2}. Since $\kappa > 0, \phi > 0$, two gamma distributions $\Gamma(\alpha_\kappa, \beta_\kappa)$ and $\Gamma(\alpha_\phi, \beta_\phi)$ are adopted as priors of κ^{-1} and ϕ^{-1}, respectively. Combining the priors for surpluses (equation (8)) and hyperparameters and the likelihood from equation (14), the posterior distribution for model *I* is

$$\pi_q(\psi_q) = p(\psi_q|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{w}_q, \delta_q) p(\boldsymbol{w}_q) p(\delta_q) p(\sigma_\zeta) p(\kappa) p(\phi). \tag{18}$$

The posterior distribution for model *II* with the likelihood from equation (17) is similar except that the hyperparameters $\sigma_\zeta, \kappa, \phi$ are removed.

3.2. Exploring the posterior state space

The Bayesian posterior distribution derived in the previous section is in general analytically intractable. Standard MCMC methods, e.g. Metropolis–Hastings (MH) sampler and Gibbs sampler, have been extensively used for such problems and their versatility and power have been proved in practical applications. However, the Markov chains might be easily trapped by local modes and long mixing times may be required. Moreover, MCMC methods estimate all unknown parameters in the Bayesian model simultaneously. This is not suitable for the hierarchical multiscale Bayesian model.

In order to bridge the gap between scales and explore multi-modal posteriors efficiently, a SMC method is employed [17, 19, 20, 28]. First, the idea of annealing/tempering is introduced. Given the target posterior distribution in equation (18), a sequence of auxiliary distributions $\{\pi_0, \ldots, \pi_n\}$ is proposed to move smoothly from a tractable distribution π_0 to the target distribution $\pi_n \equiv \pi_q(\psi_q)$. We adopt the following auxiliary distributions:

$$\pi_t(\psi_q) \propto \mathcal{L}^{\gamma_t}(\psi_q|\boldsymbol{d}) p(\psi_q), \tag{19}$$

where $t = 0, 1, \ldots, n$ and $0 = \gamma_0 < \gamma_1 < \cdots < \gamma_n = 1$ are tempering parameters. Here, $\mathcal{L}(\psi_q|\boldsymbol{d})$ is the likelihood function, $p(\psi_q)$ is the prior distribution and γ_t serves as the power exponent of the likelihood function.

The SMC method takes samples from such a sequence of probability distributions based on importance sampling and resampling techniques and constructs a sequential Bayesian inference. At step t, the basic idea is to obtain a large collection of N

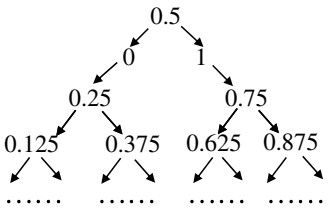

Figure 2. One-dimensional tree-like structure of the sparse grid.

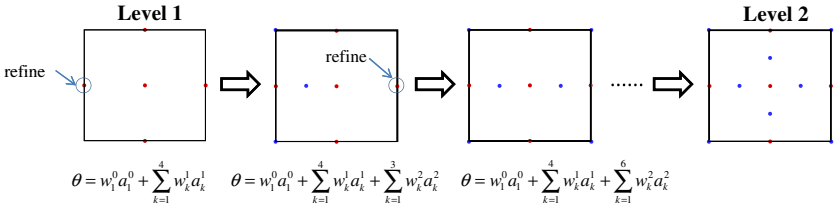

Figure 3. An example of refining the sparse grid in a two-dimensional domain.

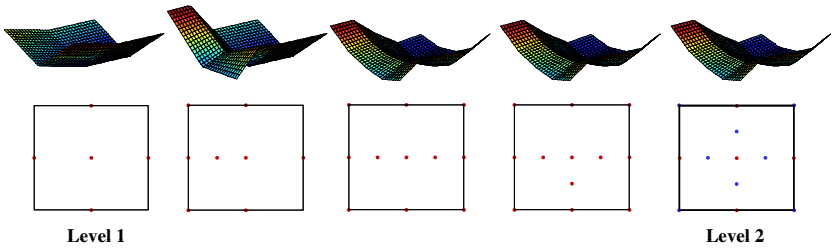

Figure 4. A hierarchical representation of the spatially varying parameter on different scales.

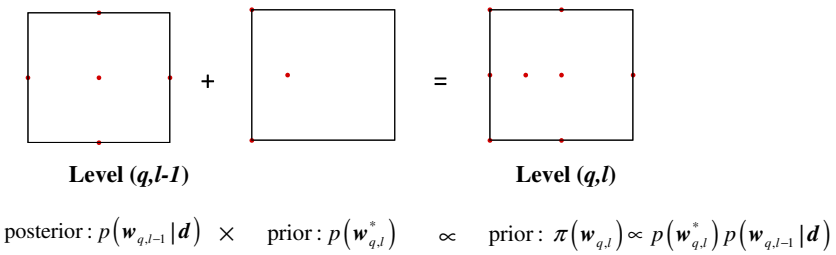

Figure 5. A hierarchical inference by taking the posterior $p(\boldsymbol{w}_{q,l-1}|\boldsymbol{d})$ as the prior on a refined grid.

weighted random samples $\{\psi_{q,t}^{(i)}, \Omega_t^{(i)}\}$ $(i = 1, \ldots, N)$ (also referred to as particles) whose empirical distribution converges asymptotically to the current target distribution π_t. Each particle can be considered as a possible configuration of the system's state [2] with an importance weight.

According to equation (19), it is easy to sample directly from π_0, the prior distribution, at the initial step. The importance sampling technique is performed sequentially to the auxiliary

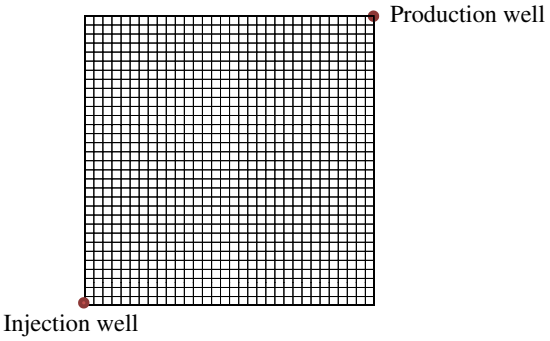

Figure 6. Schematic of the quarter five-spot problem.

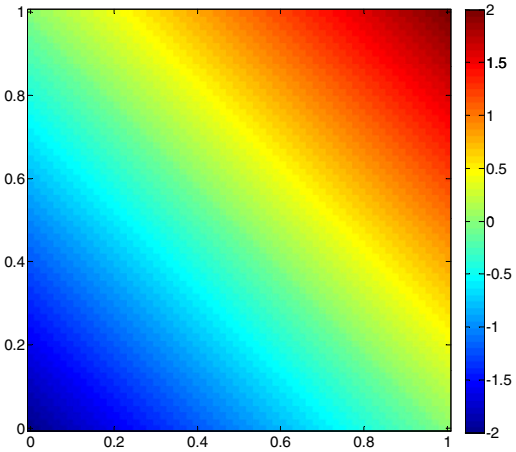

Figure 7. True permeability (logarithm) in example 1.

distributions, which is called the sequential importance sampling (SIS) [17]. In this work, we move these particles using a predetermined Markov transition kernel [19]. Suppose that at step $t-1$ we have N samples $\{\psi_{q,t-1}^{(i)}\}$ distributed in η_{t-1}. A Markov transition kernel K_t with invariant distribution π_t is proposed and new samples are marginally distributed as

$$\eta_t(\psi_q') = \int \eta_{t-1}(\psi_q) K_t(\psi_q, \psi_q') \mathrm{d}\psi_q. \tag{20}$$

We use the MH kernel with invariant distribution π_t based on a random walker sampler to move the particles $\psi_{q,t-1}^{(i)}$. The $\boldsymbol{w}_q, \delta_q, \sigma_\zeta, \kappa, \phi$ in the random vector ψ_q are updated individually via a MH kernel with a normal random walk proposal.

The importance weight $\omega^{(i)}$ estimates the discrepancy between the proposal distribution $\eta_t(\psi_q)$ and the current auxiliary distribution $\pi_t(\psi_q)$. Using the MCMC transition kernel, a recursive form for the calculation of the importance weight is [19]

$$\omega_t^{(i)} = \omega_{t-1}^{(i)} \frac{\pi_t\left(\psi_{q,t-1}^{(i)}\right)}{\pi_{t-1}\left(\psi_{q,t-1}^{(i)}\right)}. \tag{21}$$

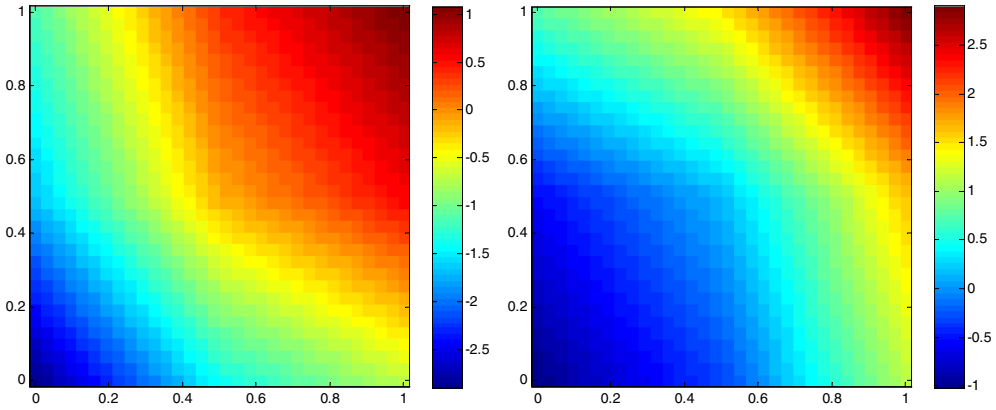

Figure 8. Posterior quantiles of the log-permeability (2% noise): 5% quantile (left) and 95% quantile (right).

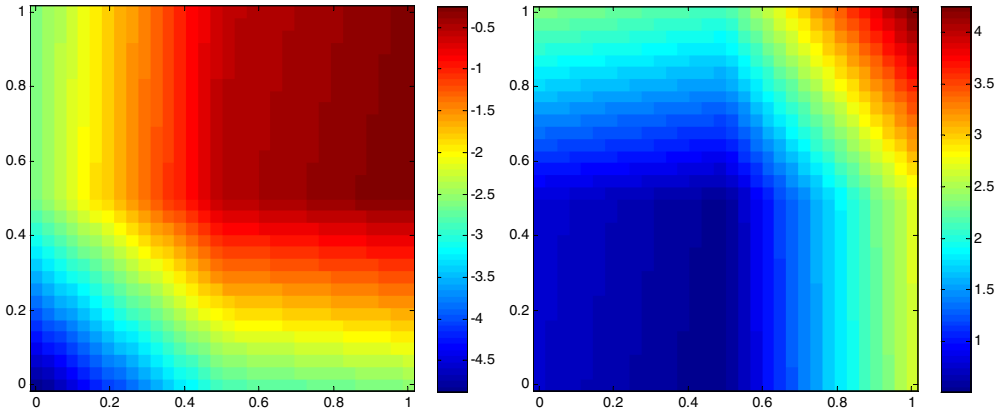

Figure 9. Posterior quantiles of the log-permeability (5% noise): 5% quantile (left) and 95% quantile (right).

It is inevitable that the SIS algorithm will degenerate and the variance of the importance weights stochastically will increase with t [29, 30]. We measure the degeneracy using the effective sample size (ESS) calculated from the normalized importance weights [30] as

$$\mathrm{ESS} = \left(\sum_{i=1}^{N} (\Omega^{(i)})^2 \right)^{-1}. \tag{22}$$

We define a threshold $\mathrm{ESS}_{\min} = \xi N \quad (\xi < 1)$. If $\mathrm{ESS} < \mathrm{ESS}_{\min}$, we carry out resampling to relieve the degeneracy of the algorithm. The simplest approach is the multinomial resampling which draws N new samples from $\{\psi_{q,t}^{(i)}\}_{i=1:N}$ according to the corresponding normalized weights $\{\Omega^{(i)}\}_{i=1:N}$ [19, 31]. After resampling, we obtain N weighted particles for the target posterior distribution $\pi_n(\psi_q)$. Obviously, the SMC method is directly parallelizable.

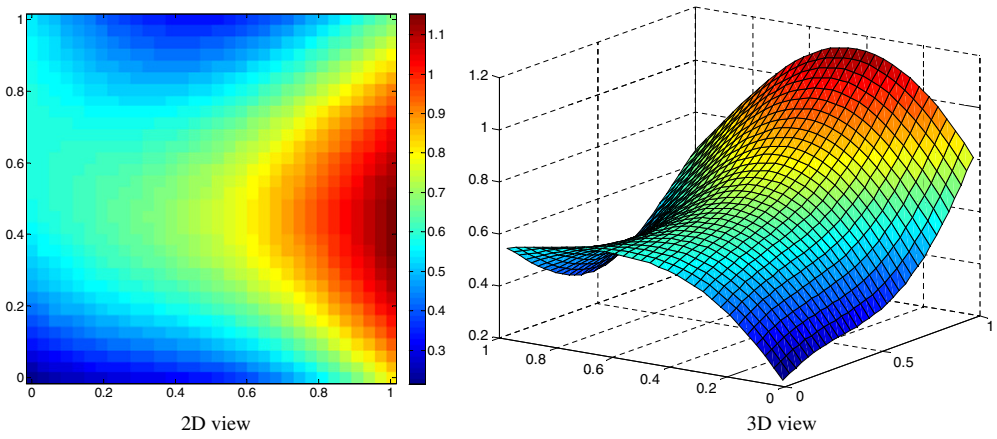

Figure 10. True permeability (logarithm) generated from a GP.

A summary of the MH kernel with a random walker sampler and of the SMC algorithm is given below in algorithms I and II, respectively.

Algorithm I: Update of w_q: MH kernel with a random walker proposal.

(i) Sample $u \sim \mathcal{U}(0, 1)$.
(ii) Sample $\widetilde{\boldsymbol{w}}_q \sim \mathcal{N}(\boldsymbol{w}_{q,t-1}^{(i)}, \sigma_m^2)$.
(iii) If $u < \min\{1, \frac{\pi_t(\widetilde{\boldsymbol{w}}_q)}{\pi_t(\boldsymbol{w}_{q,t-1}^{(i)})}\}$, set $\boldsymbol{w}_{q,t}^{(i)} = \widetilde{\boldsymbol{w}}_q$.
(iv) Else set $\boldsymbol{w}_{q,t}^{(i)} = \boldsymbol{w}_{q,t-1}^{(i)}$.

Algorithm II: SMC algorithm.

(i) Initialization: for $i = 1, \ldots, N$, sample $X_0^{(i)} \sim \pi_0(x)$ and set the importance weight $\omega_0(\psi_{q,0}^{(i)}) = \frac{1}{N}$.
(ii) Updating: at time t, for $i = 1, \ldots, N$, sample $\psi_{q,t}^{(i)} \sim K_t(\psi_{q,t-1}^{(i)}, \cdot)$ and set the importance weight according to equation (21). Then, normalize the importance weight by $\Omega_t^{(i)} = \frac{\omega_t^{(i)}}{\sum_{k=1}^N \omega_k^{(i)}}$.
(iii) Resampling: Calculate the ESS by equation (22). If $\mathrm{ESS} < \mathrm{ESS}_{\min}$, resample the particles $\{\psi_{q,1:n}^{(i)}, \Omega_n^{(i)}\}$ according to $\{\Omega_n^{(i)}\}$ to obtain a new population $\{\psi_{q,1:n}^{(i)}, 1/N\}$.
(iv) Repeat the above steps until $t = n$, i.e. the particles are distributed in the last distribution in the sequence.

4. Hierarchical Bayesian model

In the earlier sections, we defined the standard sparse grid on various levels of resolution. Given a sparse grid, the spatially varying parameter can be represented by the basis functions associated with the collocation points. Then the unknown surpluses are treated as random variables and a Bayesian model can be constructed to make inference from the observation data. Based on the one-scale Bayesian model on a single grid, we propose a hierarchical, multiscale Bayesian model in this section. A set of sparse grids from coarse to fine scales is constructed through adaptive refinement. On each grid, a one-scale Bayesian model is

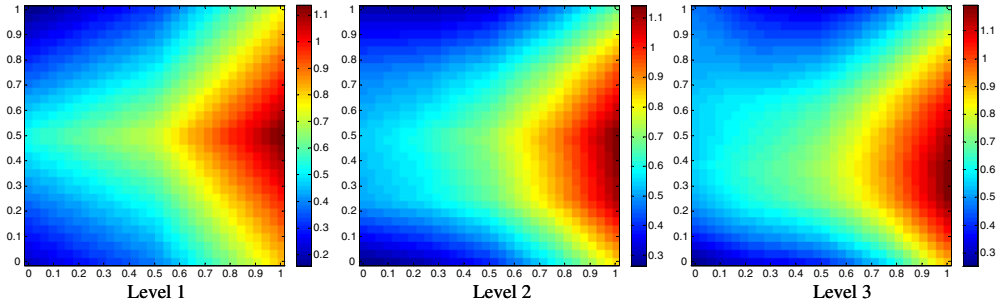

Figure 11. Posterior means estimated on three levels of the sparse grid (model *I*).

defined. Due to the hierarchical structure of the collocation points, a sequential estimation of the surpluses can be performed from coarse to fine scales. Only part of the surpluses need to be estimated on a certain grid, which significantly reduces the dimensionality of the inverse problem. Also, the surpluses can serve as an indicator of the smoothness of the parameter field. More support nodes will be added on non-smooth regions and rapid changes in the parameter field can be effectively captured. In this way, an optimal choice of the basis functions can be achieved.

4.1. Adaptive sparse grid

For the Newton–Cotes sparse grid at level q, the set of points can be obtained by refining the grid of level $q-1$ in a principled way. In fact, the 1D equidistant points of the sparse grid can be considered as a tree-like data structure as shown in figure 2. We can consider the interpolation level of a grid point m as the depth of the tree. We denote the father of a grid point as $F(m)$, where the father of the root 0.5 is itself, i.e. $F(0.5) = 0.5$. There are two sons for each grid point in each dimension. For a grid point in an N-dimensional space (here $N = 1, 2, 3$ for spatial fields), there are $2N$ sons. The sons are also the *neighbor points* of the father. The neighbor points are just the support nodes of the hierarchical basis functions in the next interpolation level. By adding the neighbor points, we actually add the support nodes from the next interpolation level. Therefore, we refine the grid locally while not violating the developments of the Smolyak algorithm [10].

In this way, a set of intermediate grids can be obtained between two standard levels of sparse grid. Figure 3 presents a 2D example in which the standard sparse grid at level 2 is constructed from the grid at level 1 by pointwise refinement. We arbitrarily pick up a point at level 1 and add its $2N$ neighbor points to the sparse grid to obtain a finer, intermediate grid. This procedure continues until we go through all collocation points that belong to the level 1 sparse grid. Since it is possible that the neighbors of one point have already been generated by other points, the difference in the number of collocation points of two successive intermediate grids is no more than $2N$, e.g. four in a 2D case. Each intermediate grid provides a set of basis functions for parameterizing the spatially varying parameter. Rather than working on the standard sparse grid, we define Bayesian models on intermediate grids and make sequential inference about the surpluses. More details will be given in the next section.

Furthermore, an adaptive refinement strategy can be used to reduce the collocation points on the sparse grid and thus the dimensionality of the inverse problem. In fact, the hierarchical surplus is a natural candidate for detecting non-smooth regions [16]. Here, the basic idea

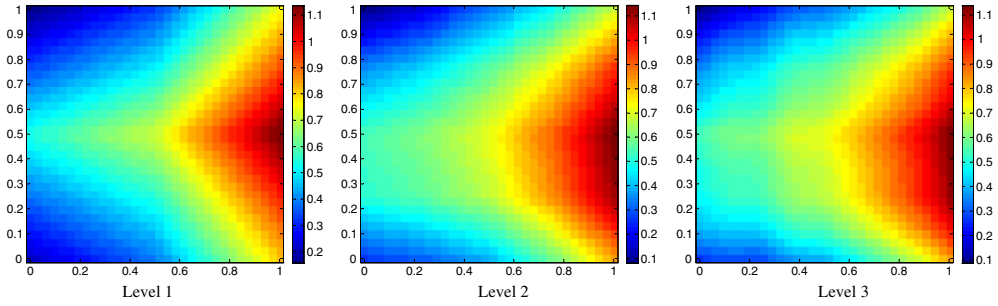

Figure 12. Posterior means estimated on three levels of the sparse grid (model II).

for adaptivity is to use the posterior mean of hierarchical surpluses as an error indicator to detect the smoothness of the spatially varying parameter field estimated on the current grid. We only refine the hierarchical basis functions a_j^i whose magnitude of the surplus satisfies $|w_j^i| \geqslant \varepsilon$. If the criterion is satisfied, the $2N$ neighbor points are added into the current sparse grid. Otherwise, we assume that the local region is smooth enough and further refinement on this basis is not required. With $\varepsilon > 0$ the parameter that controls the adaptive refinement, we introduce algorithm III as follows:

Algorithm III : Adaptive sparse grid representation of the unknown parameter.

repeat

1. Set the initial level of sparse grid as q and construct a standard sparse grid.

Construct a full Bayesian model on this level and infer all the surpluses from observation data (section 3).

2. Calculate the posterior mean of the surpluses.

3. Put the collocation points with surpluses $\left|w_j^i\right| > \varepsilon$ in the active node set which denotes the set of points whose 'sons' would be added to refine the original grid.

while the active node set is not empty **do**

(a) Pick up a point and add its $2N$ neighbor points to the sparse grid. An intermediate grid is obtained.

(b) Construct a full Bayesian model and infer all the surpluses at the current grid.

(c) Remove the point from the active node set.

end while

4. Set $q = q + 1$.

5. Place all newly added collocation points into the empty active node set

until $q = q_{\max}$ or all the points in the active node set have surpluses $\left|w_j^i\right| < \varepsilon$, i.e. the refinement of the grid terminates.

4.2. Hierarchical Bayesian inference

The hierarchically structured sparse grid provides a multiscale representation of the spatially varying parameter (figure 4). Once a grid is defined, we can construct a Bayesian model for the full surpluses and estimate them from observation data. In section 3.2, the SMC method was

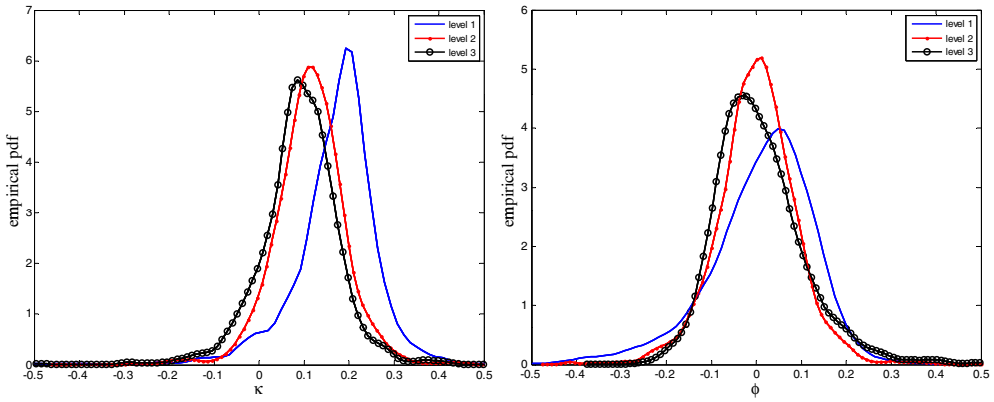

Figure 13. Empirical pdfs of κ (left) and ϕ (right) at different levels of the sparse grid.

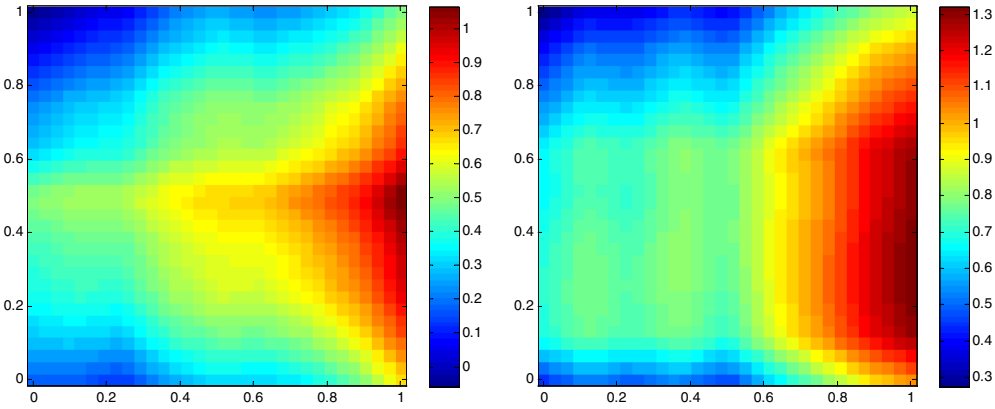

Figure 14. Posterior quantiles of the log-permeability (model *II*): 5% quantile (left) and 95% quantile (right).

Table 1. Posterior mean of the model errors $\delta_q^{(1)}$, $\delta_q^{(2)}$ and true values $\delta_q^\star$.

	Level 1	Level 2	Level 3
$\delta_q^{(1)}$	0.187	0.0281	0.0099
$\delta_q^{(2)}$	0.113	0.0205	0.0116
$\delta_q^\star$	0.0821	0.0184	0.00846

used to explore the posterior distribution on a single grid. In this section, we propose a strategy that bridges the estimation of surpluses on different scales. As a result, the dimensionality of the inverse problem is further reduced and informative priors are developed on a fine grid by incorporating the information from the posterior at the coarser grid.

Consider a sparse grid indexed by (q, l) where l denotes the lth intermediate grid between levels q and $q + 1$. Due to the nested fashion of the grid, the collocation points are composed of two parts: the inherited points from the immediate coarser grid and the new points obtained by refinement. Thus, the full surpluses can be written as

$$\boldsymbol{w}_{q,l} = \{\boldsymbol{w}_{q,l-1}, \boldsymbol{w}_{q,l}^\star\}, \tag{23}$$

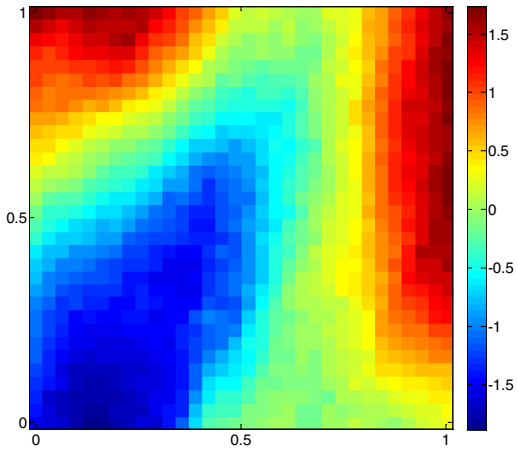

Figure 15. True permeability (logarithm) generated using the software *snesim*.

Table 2. Posterior mean of the model errors $\delta_q^{(1)}$, $\delta_q^{(2)}$ and true values $\delta_q^\star$.

	Level 1	Level 2	Level 3	Level 4
$\delta_q^{(1)}$	−0.870	−0.0931	−0.0154	−0.0073
$\delta_q^{(2)}$	−0.621	−0.0710	−0.0215	−0.0103
$\delta_q^\star$	−0.734	−0.0807	−0.0174	−0.00862

where $\boldsymbol{w}_{q,l-1}$ is a vector of surpluses which have been estimated on the coarse grid and $\boldsymbol{w}_{q,l}^\star$ is a vector of new surpluses in the current grid. In the one-scale Bayesian model in section 3, a non-informative prior is assumed for the surpluses. However, the posterior estimation of $\boldsymbol{w}_{q,l-1}$ on the coarse grid $(q, l-1)$ provides plenty of information for the re-estimation of the surpluses on grid (q, l). In fact, in a direct sparse grid interpolation at level q, we keep the surpluses estimated from levels 0 to $q-1$ and only estimate the surpluses of new collocation points at level q [15, 16]. Now the prior for $\boldsymbol{w}_{q,l}$ is defined as

$$p(\boldsymbol{w}_{q,l}) = p(\boldsymbol{w}_{q,l-1})p(\boldsymbol{w}_{q,l}^\star) \propto p(\boldsymbol{w}_{q,l-1}|\boldsymbol{d})p(\boldsymbol{w}_{q,l}^\star), \tag{24}$$

where a multivariate Gaussian distribution is assumed for the prior $p(\boldsymbol{w}_{q,l}^\star)$, i.e. $p(\boldsymbol{w}_{q,l}^\star) \sim \mathcal{N}(0, \sigma_{w,ql}^2\mathbf{I})$. The posterior distribution $p(\boldsymbol{w}_{q,l-1}|\boldsymbol{d})$ at the coarse grid $(q, l-1)$ is taken as the prior. The samples from the posterior are directly used as the samples from the prior $p(\boldsymbol{w}_{q,l-1})$ at the current grid. This hierarchical inference strategy is depicted in figure 5.

For the surpluses from points of the previous levels of interpolation, $\boldsymbol{w}_{q,l-1}$, not all elements need to be re-estimated on the grid (q, l). When a surplus satisfies $|w_j^i| < \varepsilon$, no re-estimation is required since the corresponding basis function makes negligible contribution to the interpolation. For other surpluses, if the posterior mean after re-estimation is close to that before re-estimation, no further re-estimation is required. In this study, we use the same ε as a criterion to determine convergence. After we finish estimating the surpluses on the current grid, the adaptive refinement strategy introduced in algorithm III is carried out to refine the grid. Then, the hierarchical Bayesian inference is performed on the finer grid.

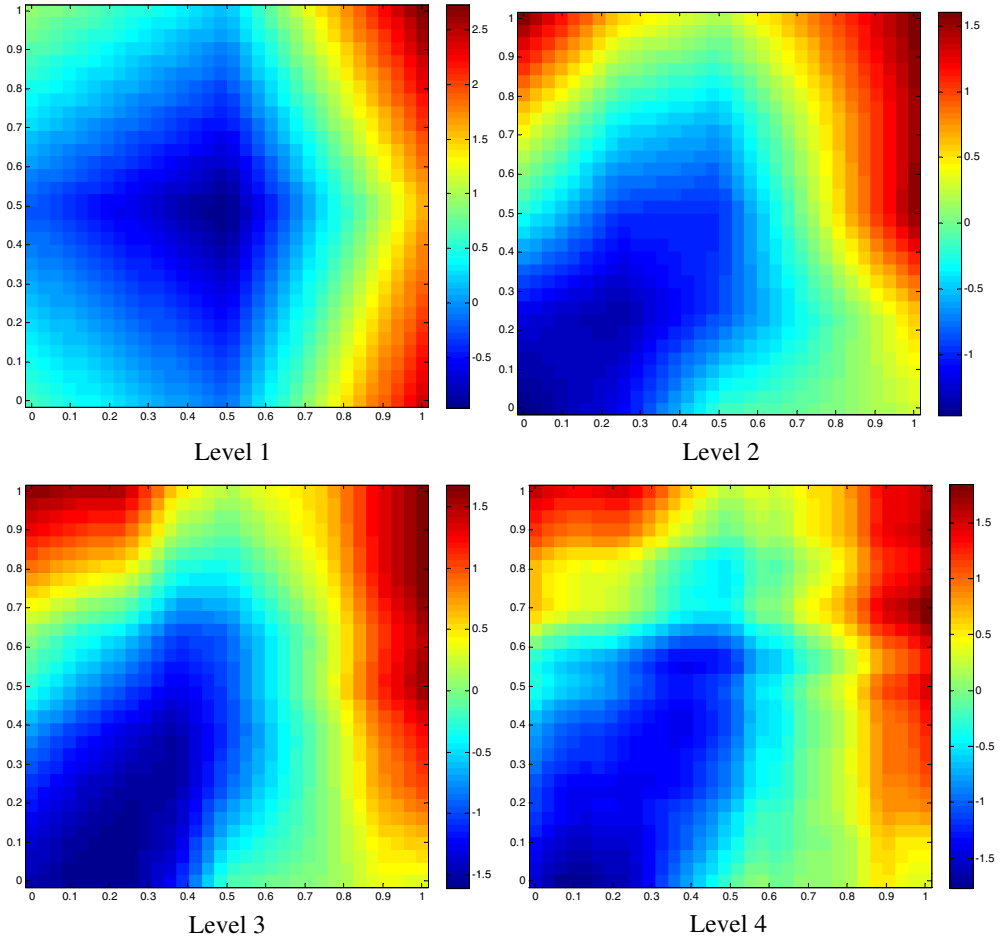

Figure 16. Posterior means estimated on four levels of the sparse grid (model *I*).

5. Numerical examples

5.1. Problem definition

We consider the nonlinear inverse problem of estimating the permeability field in flow in porous media. First, a physical model is built for corner-to-corner flow in a 2D unit square domain $D = [0, 1]^2$ [10, 32]. The injection and production wells are located in diagonally opposite vertices of the grid (figure 6).

The governing equations for the flow velocity and pressure are

$$\nabla \cdot \boldsymbol{u}(\boldsymbol{x}) = f(\boldsymbol{x}), \tag{25}$$

$$\boldsymbol{u}(\boldsymbol{x}) = -k(\boldsymbol{x})\nabla p(\boldsymbol{x}), \tag{26}$$

where $f(\boldsymbol{x})$ is the source/sink term, $\boldsymbol{u}$ is the velocity given by Darcy's law and p is the pressure. An isotropic permeability field is assumed and denoted by $k(\boldsymbol{x})$. All boundaries are no-flow boundaries.

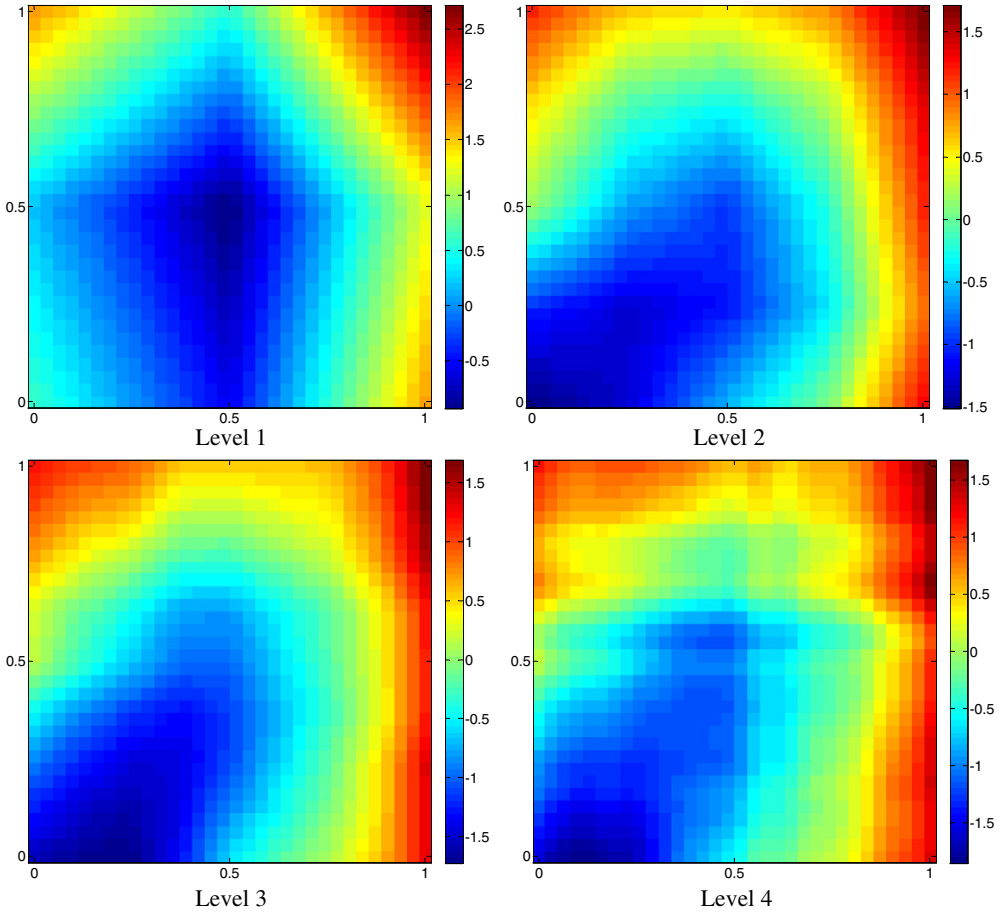

Figure 17. Posterior means estimated on four levels of the sparse grid (model II).

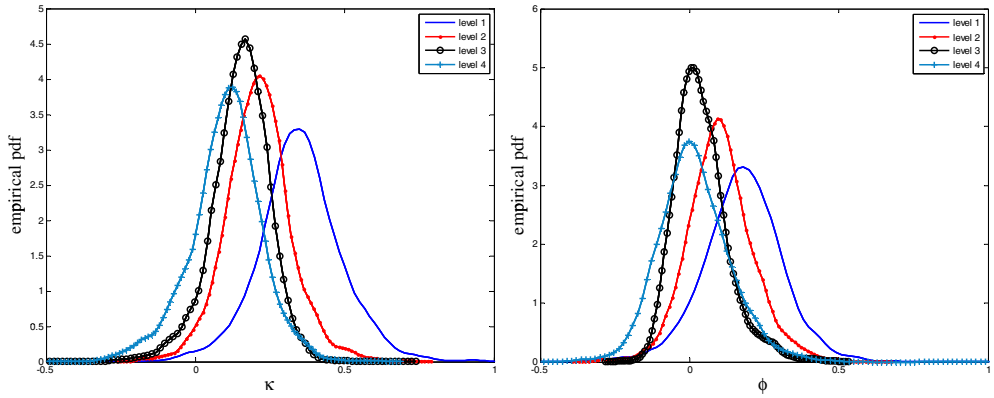

Figure 18. Empirical pdfs of κ (left) and ϕ (right) at different levels of the sparse grid.

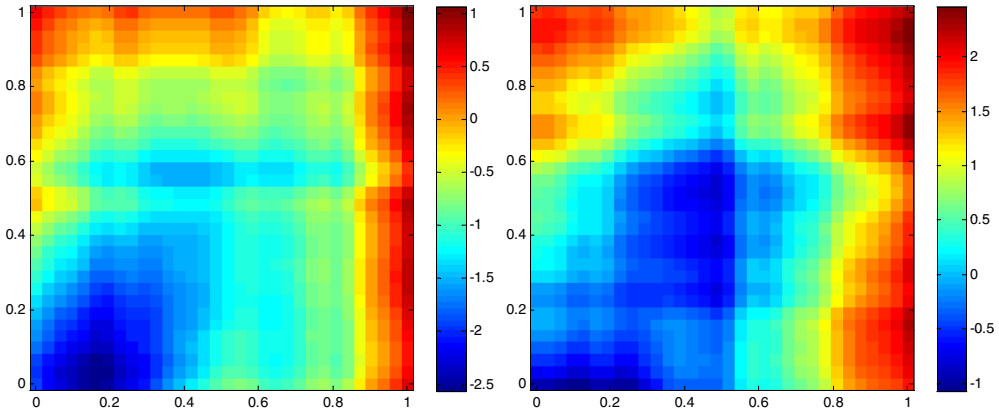

Figure 19. Posterior quantiles of the log-permeability: 5% quantile (left) and 95% quantile (right).

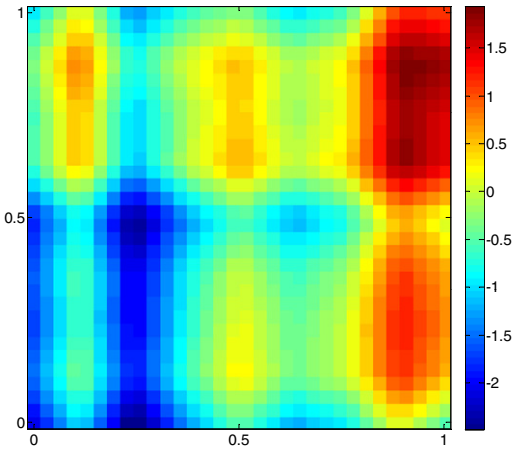

Figure 20. Posterior mean of log-permeability estimation by MCMC.

In the inverse problem of interest, the permeability field is the unknown parameter to be inferred from flow or pressure data at finite number of sensor locations. A mixed finite element method is used to obtain the numerical solution on a 32×32 grid. The observation data are generated from the numerical solutions by adding simulated noises.

To keep the permeability non-negative, we will treat the logarithm of the permeability, $\log(k)$, as the main unknown of the inverse problem. $N = 1200$ particles are employed in the implementation of the SMC algorithm. The threshold of ESS is set to be $\text{ESS}_{\min} = 0.85N$. A linear cooling schedule is selected for γ_t in equation (19). For 1500 time steps, the sequence $\{\gamma_0, \ldots, \gamma_{1500}\}$ increases uniformly from 0 to 1. We take the maximum level of sparse grid $q_{\max} = 5$ and the parameter for the adaptive refinement $\varepsilon = 0.05$ (in section 4.1). The hyperparameters used in the prior and likelihood are $\alpha_0 = 0.1$, $\beta_0 = 10$, $\alpha_1 = 0.01$, $\beta_1 = 100$, $\alpha_{\sigma_\zeta} = 0.1$, $\beta_{\sigma_\zeta} = 10$, $\alpha_\kappa = 0.01$, $\beta_\kappa = 100$, $\alpha_\phi = 0.01$, $\beta_\phi = 100$. The initial values for the unknown surpluses are generated from the priors, while the initial values for the mean of the model error vary with the level of the sparse grid.

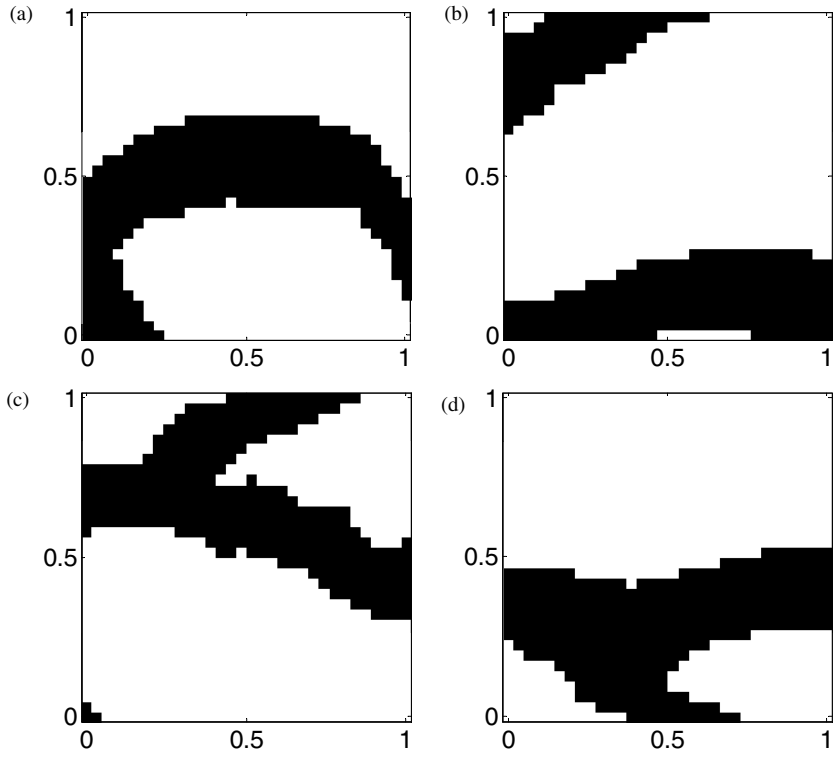

Figure 21. Examples of channelized permeabilities. The log-permeability values are 1 in black regions and 0 in the white regions.

5.1.1. Example 1. In the first example, we consider a simple permeability field of the following form [9, 10]:

$$\log k(x, y) = 2(x - 0.5) + 2(y - 0.5). \tag{27}$$

We apply a one-scale Bayesian inference to estimate the permeability. The objective of this test example is to examine the efficiency of sparse grid representation as well as the SMC algorithm proposed. For the representation of smooth functions, the sparse grid method is superior since it requires a small number of basis functions. In this example, the sparse grid of level 1 with only five collocation points is capable of representing the log-permeability field exactly. Thus, we make the inference on this sparse grid and the model error is not taken into account. The pressure is measured at a 5×5 evenly distributed sensor network and examine two cases with 2% and 5% relative noise level.

Posterior quantiles of the log-permeability inferred from the two datasets are plotted in figures 8 and 9. It is seen that the basic Bayesian framework based on the sparse grid and SMC provides rather good estimates of such a smooth permeability field. When the level of measurement noise is reduced, the inferred estimates are improved. The same problem was studied using MCMC in [10] where the permeability field was approximated by Gaussian kernels. Twenty five kernels were used to provide reasonable estimates. To make sure the Markov chain converges, 50 000 iterations were carried out. In this work, much fewer basis functions are required for a good representation of the permeability field. Besides, the particles can run in parallel, each with 1500 MCMC updates. Clearly, this approach largely reduces the computation cost.

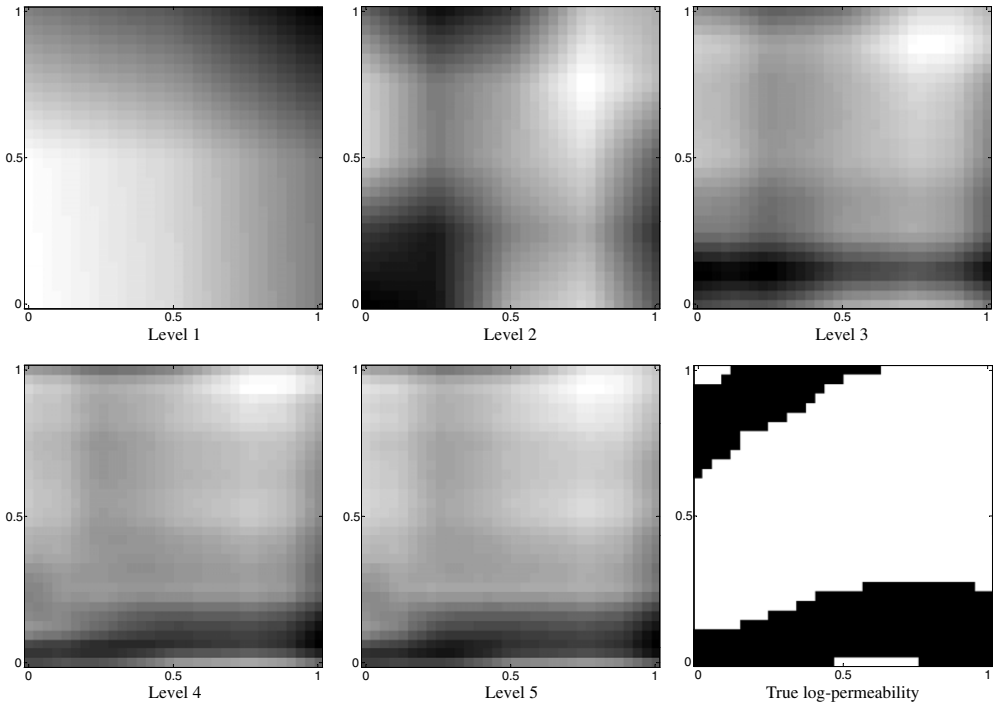

Figure 22. Posterior means estimated on the standard sparse grid from level 1 to level 5 (2% noise in data).

Table 3. Posterior mean of the model error δ_q and true values $\delta_q^\star$ (2% noise).

	Level 2	Level 3	Level 4	Level 5
δ_q	−0.189	−0.155	−0.0632	−0.0118
$\delta_q^\star$	−0.139	−0.132	−0.0514	−0.0133

Table 4. Posterior mean of the model error δ_q and true values $\delta_q^\star$ (5% noise).

	Level 2	Level 3	Level 4	Level 5
δ_q	−0.292	−0.279	−0.156	−0.0407
$\delta_q^\star$	−0.312	−0.253	−0.131	−0.0350

5.1.2. Example 2. In this example, the logarithm of the true permeability (figure 10) is generated from a GP with a correlation function $\rho(r) = \exp(-r^2)$ where r is the distance between two locations. The pressure is measured at a 5×5 sensor network with 2% noise. The initial values for the mean of the model error, δ_q, are set to be $\frac{2}{2^l}\%$ of the mean of observation data, where l is the level of sparse grid interpolation.

Two models for the model error δ as discussed in section 3.1.2 are applied here. The posterior means computed for different levels of the sparse grid are depicted in figures 11 and 12. For such a smooth permeability field, a sparse grid of level 2 is enough to provide a good approximation and the Bayesian inference is close to the true permeability field. In

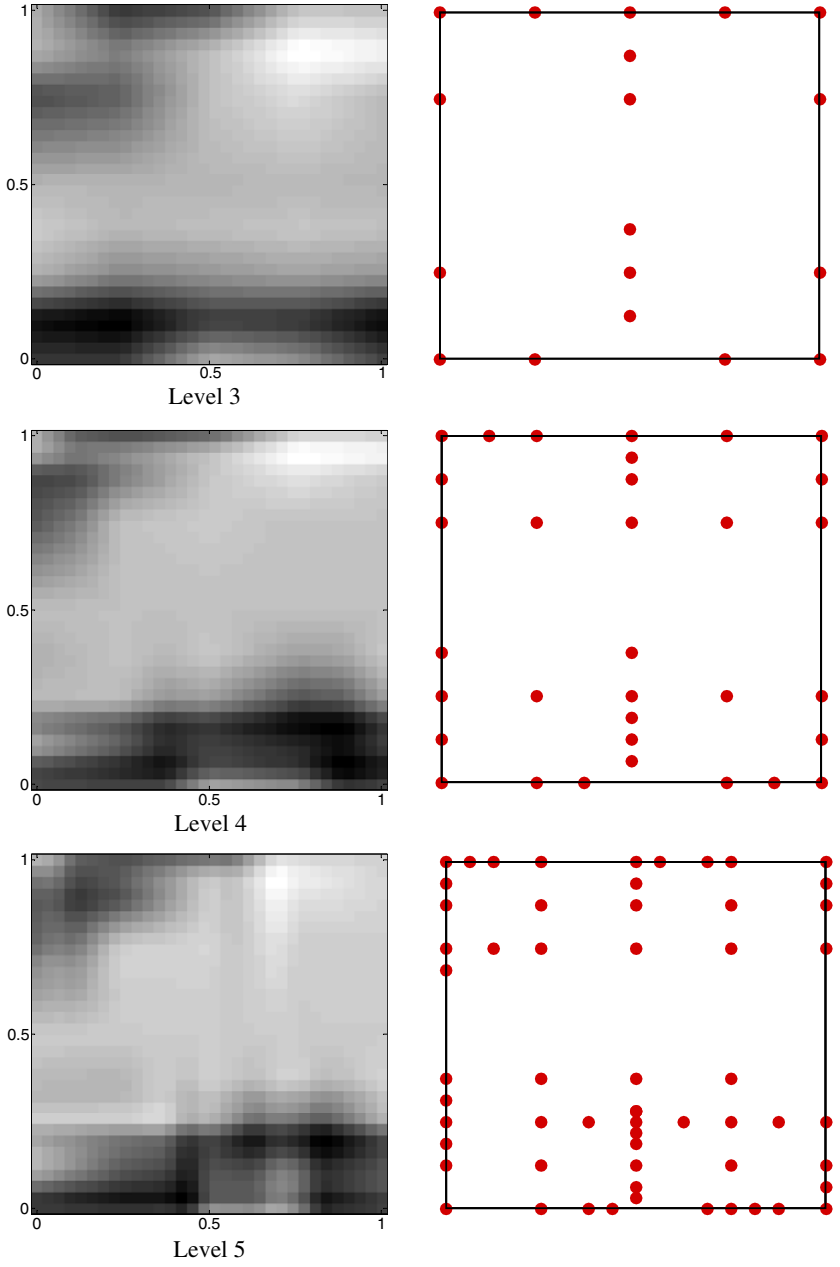

Figure 23. Posterior means estimated on sparse grid with adaptive refinement (2% noise in data).

table 1, the posterior means of the estimated model errors $\delta_q^{(1)}$ and $\delta_q^{(2)}$ corresponding to the two models are compared with the true value $\delta_q^\star$ which is taken as the mean of the difference between the observation data and those predicted by the approximate model $\boldsymbol{F}_q$. We can see that both models give similar results except that model I gives better inference at the upper-left corner of the permeability field. This can easily be understood because the dependence among model errors at different locations is considered in model I. However, the evaluation of

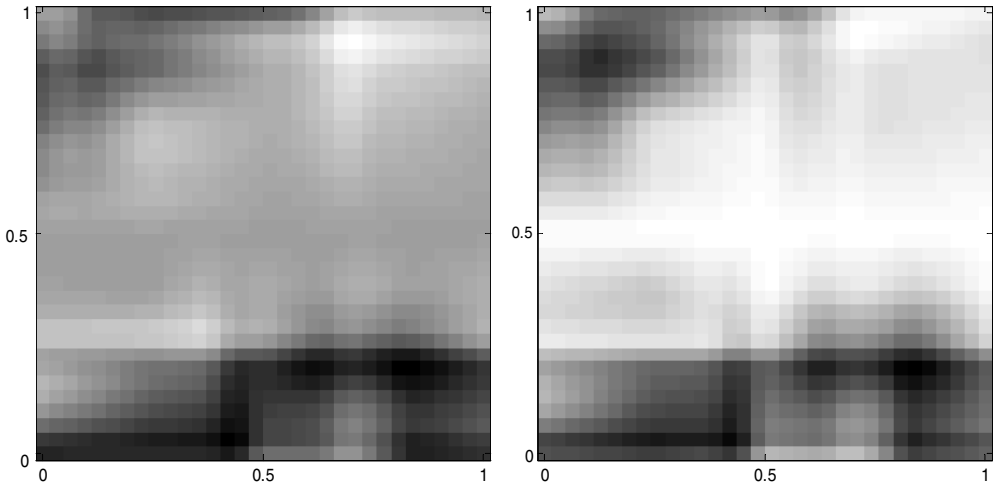

Figure 24. Posterior quantiles of the log-permeability (2% noise): 5% quantile (left) and 95% quantile (right).

hyperparameters κ and ϕ in figure 13 shows that the spatial variance is restricted in a relatively small range and scale, which implies small $\kappa H(\phi)$ in equation (13). While intuitively the model errors are correlated, the adaptivity of the sparse grid in the hierarchical Bayesian model weakens the correlation. Suppose a location where the model error is large, more collocation points would be added around it. As a result, the model error is somehow localized and there is only weak correlation between model errors at different locations. In figure 14, the posterior quantiles on level 3 obtained from the simplified model II are presented. This shows that even when model errors are uncorrelated, we can still obtain reasonable inference from observation data.

5.1.3. Example 3. In this example, the log-permeability field is generated based on a variogram model using the software *snesim* [33]. The field is defined on a 32×32 grid with a constant value of permeability in each element. The true log-permeability is plotted in figure 15. The pressure is measured on a 5×5 sensor network with 2% noise. The initial values for the mean of the model error, δ_q, are set as in example 2.

The multiscale Bayesian inference is performed on four levels of the sparse grid. The posterior means with respect to the two models (model error) are depicted in figures 16 and 17. The estimated model errors are listed in table 2. It is seen that even on a coarse scale (level 2), the main features of the true permeability are efficiently captured by the sparse grid and correctly inferred from the limited observation data. The empirical pdfs of hyperparameters κ and ϕ in model I are given in figure 18. As in example 2, the two parameters indicate a weak correlation among model errors at different locations. The posterior quantiles on level 4 obtained from model II are shown in figure 19.

In examples 2 and 3, good estimates can be obtained at a relatively coarse level of resolution. However, due to the high nonlinearity of the forward model, the forward problem is multi-modal and the standard MCMC performs poorly in these cases. For demonstration, the MH algorithm is directly applied for inference from accurate data (no noise included) on the sparse grid at level 3. The initial values of surpluses are set to be zero. In figure 20, the posterior mean after 50 000 iterations is plotted. We can see that the Markov chain is trapped by a local mode.

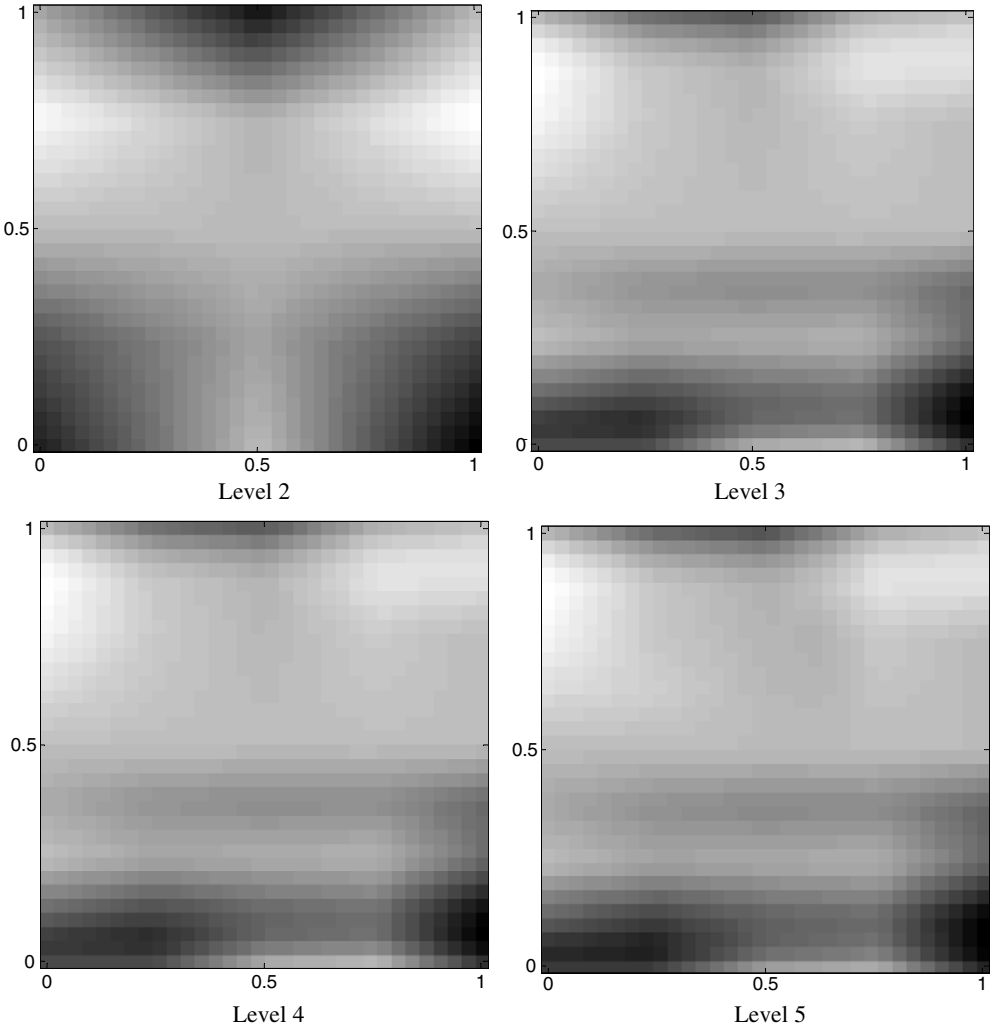

Figure 25. Posterior means estimated on the sparse grid with adaptive refinement (5% noise in data).

5.1.4. Example 4. In this example, we apply the multiscale Bayesian inference to a channelized permeability field. Channelized permeability is generally difficult to resolve by conventional models, such as the GP model and the K–L (Karhunen–Loève) expansion. Figure 21 gives four realizations of the channelized permeability created by the software *snesim*. The log-permeability values are at 1 in black regions and 0 in white regions. In this example, the permeability plotted in figure 21(b) is set to be the true permeability. The pressure is measured on a 20×20 evenly distributed sensor network with 2% and 5% noise. The initial values for the mean of the model error, δ_q, are set to be $\frac{5}{2^l}\%$ of the mean of observation data. All of the permeability fields in this example are defined on 32×32 grid blocks. According to the weak correlation among model errors shown in examples 2 and 3, we only consider the simplified model II in this example.

At first, we make Bayesian inference on the standard sparse grid, i.e. no adaptive strategies are used to refine the grid. The posterior means estimated from coarse to fine scales are

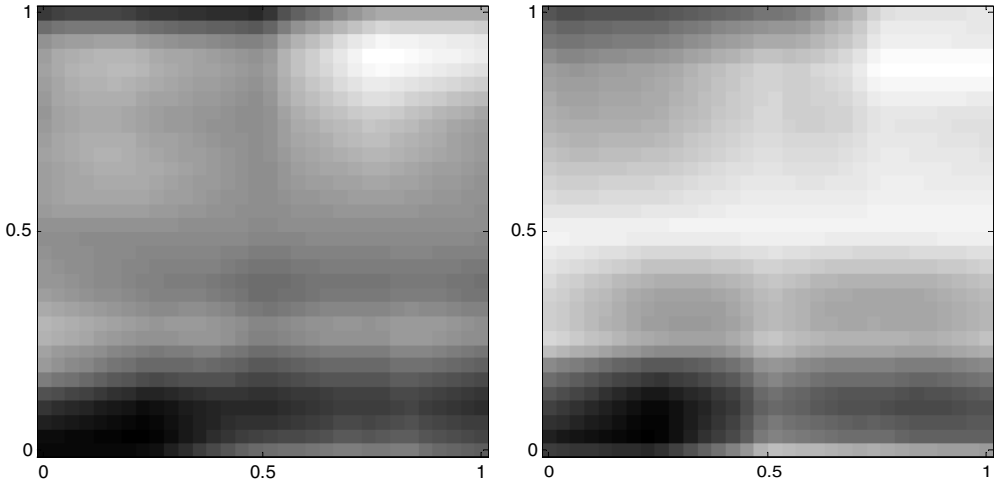

Figure 26. Posterior quantiles of the log-permeability (5% noise): 5% quantile (left) and 95% quantile (right).

presented in figure 22. The number of collocation points of the five levels are 5, 13, 29, 65 and 145. With the increase of collocation points, more local features of the true permeability field are captured in the estimation. The black strip on the bottom is well captured after level 3. However, although there is some trend for the black strip on the top-left corner on level 2, it soon disappears on subsequent finer grids.

We next apply the adaptive strategy in the refinement of the sparse grid. In this case, we start the inference on a standard sparse grid of level 2. The collocation points with (posterior mean) surpluses whose magnitude is less than 0.05 are removed from the grid and thus rejected from further refinement. As a result, the dimensionality of the inverse problem is largely reduced. The estimated posterior means and corresponding adaptive sparse grids are depicted in figure 23. The posterior quantiles on level 5 are shown in figure 24. The estimated model errors are listed in table 3. We can see that the results are improved after reducing the number of collocation points. The two black strips in the true permeability field are captured. The unsatisfactory results from the standard sparse grid representation may be due to overfitting. In this example, the true permeability field is composed of several smooth segmentations. The discontinuity only occurs on the sharp edges. Although the Smolyak algorithm has made general optimization in the choice of collocations points, there are still redundant points for this specific problem. This may produce erroneous values on the unnecessary points and thus affect the overall configuration of the estimated parameter field. The same strategy is also applied to make inference from data with 5% noise (figures 25 and 26). It should come as no surprise that the estimates are not as good as those inferred from data with a lower level of noise.

6. Conclusions

A multiscale Bayesian framework for the identification of spatially varying parameters was introduced. The parameter field was discretized using a sparse grid and represented by local basis functions associated with the collocation points. Based on the hierarchical property

of the sparse grid, a multiscale representation of the parameter field was introduced and a sequence of hierarchical Bayesian models from coarse to fine scales.

The sparse grid provides an effective way of finding an optimal choice of basis functions to approximate the spatially varying parameter. The estimated coefficients (surpluses) of the basis functions serve as indicators of the interpolation error. Accordingly, the refinement of the grid can be performed in an adaptive way. This reduces the dimensionality of the inverse problem and the computational cost of Bayesian inference. The adaptive refinement strategy removes unimportant points from estimation, and thus avoids possible overfitting and leads to improved results. The multiscale Bayesian framework proposed is a generalized way of estimating unknown spatially varying parameters from observation data. Without any prior information, the sparse grid method provides an effective strategy to construct an optimal Bayesian model.

The SMC algorithm used here is directly parallelizable and well suited for multi-modal problems. Using standard MCMC, the estimation of the surpluses can be trapped by many local modes at coarse levels of resolution, which could result in failure of further inference on fine scales. The samples from the posterior distributions at a coarse grid can provide prior information for inference of the surpluses of these collocation points on a finer grid, which helps to improve the quality of the calculations and to speed up the convergence rate.

Acknowledgments

This research was supported by the US Department of Energy, Office of Science, Advanced Scientific Computing Research, the Computational Mathematics program of the National Science Foundation (NSF) (award DMS- 0809062) and an OSD/AFOSR MURI09 award on uncertainty quantification. The computing for this research was supported by the NSF through TeraGrid resources provided by NCSA under grant number TG-DMS090007.

References

[1] Marzouk Y M, Najm H N and Rahn L A 2007 Stochastic spectral methods for efficient Bayesian solution of inverse problems *J. Comput. Phys.* **224** 560–86
[2] Koutsourelakis P S 2009 A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters *J. Comput. Phys.* **228** 6184–211
[3] Tikhonov A N 1985 *Solution of Ill-Posed Problems* (Washington: Halster Press)
[4] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (New York: Springer)
[5] Velamur Asokan B and Zabaras N 2004 Stochastic inverse heat conduction using a spectral approach *Int. J. Numer. Methods Eng.* **60** 1569–93
[6] Calvetti D and Somersalo E 2007 *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing* (Berlin: Springer)
[7] Wang J and Zabaras N 2004 A Bayesian inference approach to the inverse heat conduction problem *Int. J. Heat Mass Transfer* **47** 3927–41
[8] Wang J and Zabaras N 2005 Hierarchical Bayesian models for inverse problems in heat conduction *Inverse Problems* **21** 183–206
[9] Lee H K H, Higdon D, Bi Z, Ferreira M and West M 2002 Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media *Technical Report* Technometrics
[10] Ma X and Zabaras N 2009 An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method *Inverse Problems* **25** 035013
[11] Ferreira M A R and Lee H K H 2007 *Multiscale Modeling: A Bayesian Perspective* (Berlin: Springer)
[12] Besag J, Green P, Higdon D and Mengersen K 1995 Bayesian computation and stochastic systems *Stat. Sci.* **10** 3–41
[13] Smolyak S 1963 Quadrature and interpolation formulas for tensor products of certain classes of functions *Dokl. Akad. Nauk. SSSR* **4** 240–3

[14] Bungartz H-J and Griebel M 2004 Sparse grids *Acta Numer.* **13** 147–269
[15] Klimke A and Wohlmuth B 2005 Algorithm 847: Spinterp: piecewise multilinear hierarchical sparse grid interpolation in MATLAB *ACM Trans. Math. Softw.* **31** 561–79
[16] Ma X and Zabaras N 2009 An adaptive hierarchical sparse grid collocation algorithm for the solution of stochastic differential equations *J. Comput. Phys.* **228** 3084–113
[17] Liu J S 2008 *Monte Carlo Strategies in Scientific Computing* (New York: Springer)
[18] Rudoy D and Wolfe P J 2006 Monte Carlo methods for multi-modal distributions *Proc. 40th Asilomar Conf. Signals, Systems and Computers (Monterey, CA, 29 Oct–1 Nov 2006)* pp 2019–23
[19] Del Moral P, Doucet A and Jasra A 2006 Sequential Monte Carlo samplers *J. R. Stat. Soc.* B **68** 411–36
[20] Jasra A, Doucet A, Stephens D A and Holmes C C 2008 Interacting sequential Monte Carlo samplers for trans-dimensional simulation *Comput. Stat. Data Anal.* **52** 1765–91
[21] Dostert P, Efendiev Y and Hou T Y 2008 Multiscale finite element methods for stochastic porous media flow equations and application to uncertainty quantification *Comput. Methods Appl. Mech. Eng.* **197** 3445–55
[22] Dostert P, Efendiev Y and Mohanty B 2009 Efficient uncertainty quantification techniques in inverse problems for Richards' equation using coarse-scale simulation models *Adv. Water Resour.* **32** 329–39
[23] Green P J 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination *Biometrika* **82** 711–32
[24] Richardson S and Green P J 1997 On Bayesian analysis of mixtures with an unknown number of components *J. R. Stat. Soc.* B **59** 731–92
[25] Kennedy M C and O'Hagan A 2001 Bayesian calibration of computer models *J. R. Stat. Soc.* B **63** 425–64
[26] Higdon D, Nakhleh C, Gattiker J and Williams B 2008 A Bayesian calibration approach to the thermal problem *Comput. Methods Appl. Mech. Eng.* **197** 2431–41
[27] Haran M 2010 Gaussian random field models for spatial data *Handbook of Markov Chain Monte Carlo* ed S R Brooks, A Gelman, G L Jones and X L Meng (to appear)
[28] Billy A, Bois F Y, Parent E and Robert C P 2006 Bayesian-optimal design via interacting particle systems *J. Am. Stat. Assoc.* **101** 773–85
[29] Doucet A, Godsill S and Andrieu C 2000 On sequential Monte Carlo sampling methods for Bayesian filtering *Stat. Comput.* **10** 197–208
[30] Liu J S and Chen R 1995 Blind deconvolution via sequential imputations *J. Am. Stat. Assoc.* **90** 567–76
[31] Liu J S and Chen R 1998 Sequential Monte Carlo methods for dynamic systems *J. Am. Stat. Assoc.* **93** 1032–44
[32] Juanes R and Dub F-X 2008 A locally conservative variational multiscale method for the simulation of porous media flow with multiscale source terms *Comput. Geosci.* **12** 273–95
[33] Remy N 2005 S-GeMS: the Stanford geostatistical modeling software: a tool for new algorithms development *Quant. Geol. Geostat.* **14** 865–71