A Bayesian Approach to Characterizing Uncertainty in Inverse Problems Using Coarse and Fine Scale Information

Dave Higdon, Herbert Lee, and Zhuoxin Bi

January 30, 2001

Abstract

The Bayesian approach allows one to easily quantify uncertainty, at least in theory. In practice, however, MCMC can be computationally expensive, particularly in complicated inverse problems. Here we present methodology for improving the speed and efficiency of an MCMC analysis by combining runs on different scales. By using a coarser scale, the chain can run faster (particularly when there is an external forward simulator involved in the likelihood evaluation) and better explore the posterior, being less likely to become stuck in local maxima. We discuss methods for linking the coarse chain back to the original fine scale chain of interest. The resulting coupled chain can thus be run more efficiently without sacrificing the accuracy achieved at the finer scale.

Keywords

Stochastic simulation, Markov chain Monte Carlo, SPECT, Hydrology, Flow in porous media, Posterior simulation, Coupled MCMC

I. Introduction

ONE approach to characterizing uncertainty in inverse problems is to utilize the Bayesian paradigm so that the resulting posterior distribution describes the uncertainty in the unknown parameters. In large scale problems, such as the examples presented in this paper, the use of Markov chain Monte Carlo (MCMC) for exploring this posterior distribution has compared quite favorably to alternative approaches (c.f. [1], [2]). However, the use of posterior simulation to characterize uncertainty in such problems can be a computationally daunting task. Typically, inverse problems (e.g., tomography and image reconstruction) require numerous runs of a complicated forward simulator which takes a large vector of inputs x and then produces output which is related to the observed data y. In many applications, the computational demands of the simulator can greatly restrict the number of forward simulations that can be carried out, making an MCMC-based approach difficult or even impractical. One way around this difficulty is to speed up the forward simulation by running it on a smaller, coarsened version of the original inputs $\tilde{x}$. Of course, this improvement in speed usually results in less accurate simulations.

In this paper we propose formulating a coarsened version of the problem which yields a simpler, more tractable, posterior distribution. We then use the resulting posterior simulation output to "guide" and speed

The authors are at Duke University, Box 90251, Durham, NC, 27708, USA. E-mail: higdon@stat.duke.edu .

up the posterior simulation on the original, fine scale specification. We use an approach based on coupling the Markov chains exploring the coarse and fine scale posteriors [3]. We favor this approach because it lends itself easily to simulation in a parallel computing environment. We also briefly discuss an alternative approach which makes use of the reversible jump methodology of Green [4].

The next section lays out the details of our approach using a very simple imaging problem as a motivating example. We then go on to consider two applications, one in single photon emission tomography (SPECT) and one in hydrology. The latter application is currently being implemented in a parallel computing environment here at the Center for Multiscale Modeling and Distributed Computing at Duke University. The paper ends with a brief discussion. The notation of the paper is enumerated in Appendix A.

II. Basic formulation

At the fine scale we model the sampling density, or likelihood, for $y = (y_1, \ldots, y_n)^T$ by $L(y|x, \theta_y)$ where $x = (x_1, \ldots, x_m)^T$ denotes a vector of unknown parameter inputs for the simulator and θ_y may hold additional parameters to control the likelihood. Note that evaluation of $L(y|x, \theta_y)$ requires running the forward simulator with inputs x. The fine scale formulation is completed by specifying priors $\pi(x|\theta_x)$ and $\pi(\theta)$ where θ_x is a hyperparameter that controls the prior for x, and θ holds the parameters (θ_y, θ_x). The resulting posterior distribution $\pi(x, \theta|y)$ can then be explored via MCMC.

Under the coarse specification, we define a coarsened counterpart for the inputs x given by $\tilde{x} = (\tilde{x}_1, \ldots, \tilde{x}_{\tilde{m}})^T = Cx$ where C is the coarsening operation which maps a m-vector to a lower dimensional $\tilde{m}$-vector. Often, C is a $\tilde{m} \times m$ matrix so that Cx is a simple linear operation, such as averaging or summing over fine scale pixels within coarse pixels. However a more complicated operation is conceivable in hydrological applications where upscaling has been the focus of a large amount of research. Depending on the problem, y, θ_y, and θ_x might also have coarsened counterparts $\tilde{y}$, $\tilde{\theta}_{\tilde{y}}$, and $\tilde{\theta}_{\tilde{x}}$ which may be slightly altered from their original form. In addition, the coarse likelihood and prior models may also differ, yielding an analogous posterior density. This results in two separate posteriors, one for each scale:

$$\begin{array}{lrcl} \text{fine} & \pi(x, \theta|y) & \propto & L(y|x, \theta_y) \times \pi(x|\theta_x) \times \pi(\theta) \\ \text{coarse} & \tilde{\pi}(\tilde{x}, \tilde{\theta}|\tilde{y}) & \propto & \tilde{L}(\tilde{y}|\tilde{x}, \tilde{\theta}_{\tilde{y}}) \times \tilde{\pi}(\tilde{x}|\tilde{\theta}_{\tilde{x}}) \times \tilde{\pi}(\tilde{\theta}). \end{array}$$

For the examples used in this paper, we specify Gaussian priors for x and $\tilde{x}$ so that

$$\pi(x|\theta) \propto \theta^{\frac{m}{2}} \exp\{-\tfrac{1}{2}\theta x^T W x\} \quad (1)$$

$$\tilde{\pi}(\tilde{x}|\tilde{\theta}) \propto \tilde{\theta}^{\frac{\tilde{m}}{2}} \exp\{-\tfrac{1}{2}\tilde{\theta}\tilde{x}^T \tilde{W} \tilde{x}\} \quad (2)$$

where the sparse precision matrices W and $\tilde{W}$ determine the dependence structure for x and $\tilde{x}$ respectively.

Markov chain Monte Carlo may be used to obtain samples from the sibling posterior distributions, with the chain corresponding to the coarser formulation running four times as fast as the fine chain for a typical 2-dimensional problem. In addition to running faster, the sampling properties of the coarse chain are often superior to its fine scale counterpart. This is because local modes in the fine posterior, which can inflate

the autocorrelation in a simple single site MCMC chain, are often smoothed out or removed in the coarse posterior.

Example

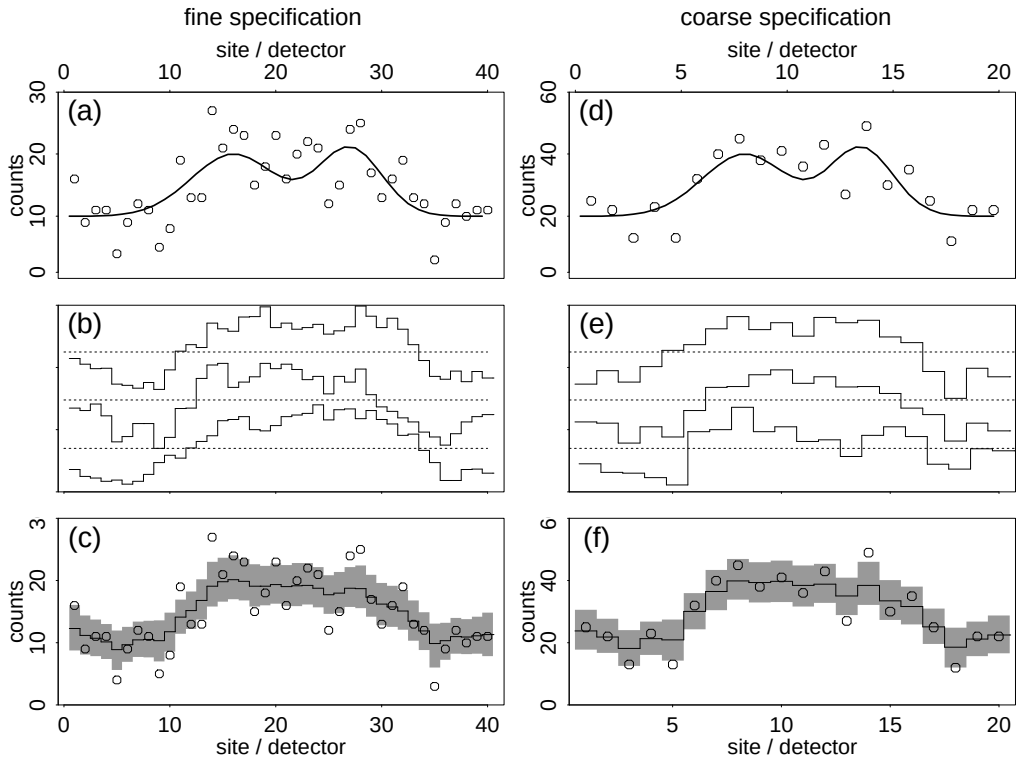

Fig. 1. Fine and coarse scale models for a simple image problem. Fine scale formulation: (a) true image intensity and observed counts under the fine scale formulation; (b) realizations of x from the fine scale posterior (dashed lines mark 15 counts for each realization); (c) posterior mean and 90% pointwise credible regions for x the fine scale image intensity. Coarse scale formulation: (d) true image intensity and observed counts under the coarse scale formulation; (e) realizations of $\tilde{x}$ from the coarse scale posterior (dashed lines mark 30 counts for each realization); (f) posterior mean and 90% pointwise credible regions for $\tilde{x}$ the coarse scale image intensity.

To demonstrate the construction of a coarse formulation, we take an artificial 1-d, blur-free, image problem. Here we have $n = 40$ detectors recording emissions from a one-dimensional source. Figure 1(a) shows the true source intensity and the recorded counts y. Under the fine scale formulation, we break x into $m = 40$ pixels which correspond to the 40 detectors. Hence we are estimating the emission intensity x_i averaged over each site i. We assume each detector counts only emissions from its corresponding image pixel and that counts follow a Poisson distribution so that the likelihood function is

$$L(y|x) \propto \prod_{i=1}^{n} x_i^{y_i} \exp\{-x_i\}. \tag{3}$$

A simple Gaussian Markov random field (MRF) prior is used for the unknown image intensities

$$\pi(x|\theta) \propto \theta^{\frac{m}{2}} \exp\{-\tfrac{1}{2}\theta \sum_{i=1}^{m-1} (x_i - x_{i+1})^2\} = \theta^{\frac{m}{2}} \exp\{-\tfrac{1}{2}\theta x^T W x\} \tag{4}$$

where $\theta = \theta_x$ controls the smoothness of x and the $m \times m$ precision matrix W is tridiagonal with diagonal elements W_{ii} equal to the number of pixels adjacent to pixel i, and off-diagonal elements $W_{i,i\pm1} = -1$. Finally a $\Gamma(a = 1, b = .005)$ prior is specified for θ so that

$$\pi(\theta) \propto \theta^{a-1} e^{-b\theta},\ \theta > 0. \tag{5}$$

This gives the resulting posterior distribution

$$\pi(x, \theta|y) \propto L(y|x) \times \pi(x|\theta) \times \pi(\theta)$$

which is explored using an MCMC scheme which updates the components of x and θ one at a time. The MCMC scheme goes as follows:

1. Initialize parameters at some value $(x, \theta)^0$.

2. Update each x_i according to Metropolis rules:

$\triangleright$ generate proposal $x_i^* \sim U[x_i - r, x_i + r]$.

$\triangleright$ set new value x_i' to x_i^* with probability $\min\{1, \frac{\pi(x^*,\theta|y)}{\pi(x,\theta|y)}\}$, else set x_i' to x_i.

3. Update θ given x.

$\triangleright$ $\theta' \sim \Gamma(a + \frac{n}{2}, b + \frac{1}{2}\sum_{i\sim j}(x_i - x_j)^2)$

4. Repeat 2. and 3. giving the partial realization $(x, \theta)^1, (x, \theta)^2, \ldots$ of a Markov chain with stationary distribution $\pi(x, \theta|y)$. Note that dropping θ from the output gives draws $x^1, x^2, \ldots$ from the marginal posterior $\pi(x|y) = \int \pi(x, \theta|y) d\theta$.

Figures 1(a–c) show the fine scale data, three realizations produced by the MCMC chain, and a posterior summary of x respectively.

Analogously, a coarse scale formulation could also constructed for this problem. One possibility is to combine the counts from adjacent detector pairs so that the coarsened data $\tilde{y}$ consist of $\tilde{n} = 20$ counts, each of which are the sum of two fine scale counts which were emitted from a coarse pixel that is now twice as wide. These coarsened counts are shown in Figure 1(d). Similarly we divide the unknown coarsened image $\tilde{x}$ into $\tilde{m} = 20$ pixels which correspond to the $\tilde{n} = 20$ "coarse detectors." In this case, $\tilde{x} = Cx$ where $C = I_{\tilde{m}} \otimes (1, 1)$ — $\tilde{x}$ is the sum over adjacent fine pixels in x. The coarse likelihood and priors are defined to be

$$\begin{aligned}
\tilde{L}(\tilde{y}|\tilde{x}) &\propto \prod_{i=1}^{\tilde{n}} \tilde{x}_i^{\tilde{y}_i} \exp\{-\tilde{x}_i\} \\
\tilde{\pi}(\tilde{x}|\tilde{\theta}) &\propto \tilde{\theta}^{\frac{\tilde{m}}{2}} \exp\{-\tfrac{1}{2}\tilde{\theta} \sum_{i=1}^{\tilde{m}-1} (\tilde{x}_i - \tilde{x}_{i+1})^2\} = \tilde{\theta}^{\frac{\tilde{m}}{2}} \exp\{-\tfrac{1}{2}\tilde{\theta}\tilde{x}^T \tilde{W} \tilde{x}\} \\
\tilde{\pi}(\tilde{\theta}) &\propto \tilde{\theta}^{\tilde{a}-1} e^{-\tilde{b}\tilde{\theta}},\ \tilde{\theta} > 0
\end{aligned}$$

where the $\tilde{m} \times \tilde{m}$ precision matrix $\tilde{W}$ is defined exactly as its fine scale counterpart W was earlier.

The resulting posterior $\tilde{\pi}(\tilde{x}, \tilde{\theta}|\tilde{y})$ can be explored with MCMC just as with the fine scale formulation (in fact, we use the same code). Figures 1(e–f) show realizations from coarse posterior and a simple summary of the posterior distribution for $\tilde{x}$. Note the coarse and fine scale posterior realizations are quite similar. □

A. Coupled Markov chain Monte Carlo

In order to link the coarse and fine scale formulations we employ the ideas from *Metropolis coupled chains* [3]. Coupling ideas have been used previously for diagnosing convergence, e.g., [5], [6], as well as for improving mixing properties of MCMC chains [7], [8]. However using coupling to link scales appears to be a novel approach which is hinted at in the rejoinder of [9]. The basic idea is this: instead of running two separate MCMC chains, one on the fine parameter space and one on the coarse parameter space, run a single chain on the product space. This *coupled* chain has stationary distribution $\pi(x,\theta|y) \times \tilde{\pi}(\tilde{x},\tilde{\theta}|\tilde{y})$ so that if we take only the fine scale realizations from this chain, they will have stationary distribution $\pi(x,\theta|y)$. The key to a useful construction of a coupled chain is to include updates which allow information to pass between the two scales.

One possible implementation of such a chain alternates standard within scale updates with "swapping" updates that allow information to move between scales as shown below.

$$\begin{array}{ccccccccccc} (x,\theta)^1 & \stackrel{\text{MCMC}}{\longrightarrow} & (x,\theta)^2 & \stackrel{\text{SWAP}}{\longrightarrow} & (x,\theta)^3 & \stackrel{\text{MCMC}}{\longrightarrow} & (x,\theta)^4 & \stackrel{\text{SWAP}}{\longrightarrow} & (x,\theta)^5 & \stackrel{\text{MCMC}}{\longrightarrow} & \cdots \\ (\tilde{x},\tilde{\theta})^1 & \stackrel{\text{MCMC}}{\longrightarrow} & (\tilde{x},\tilde{\theta})^2 & & (\tilde{x},\tilde{\theta})^3 & \stackrel{\text{MCMC}}{\longrightarrow} & (\tilde{x},\tilde{\theta})^4 & & (\tilde{x},\tilde{\theta})^5 & \stackrel{\text{MCMC}}{\longrightarrow} & \cdots \end{array}$$

Here the updates marked $\stackrel{\text{MCMC}}{\longrightarrow}$ affect the parameters within a given scale as in Section 2.1, while the updates marked $\stackrel{\text{SWAP}}{\longrightarrow}$ denote a Hastings update [10] that proposes new candidates $(x^*,\theta^*,\tilde{x}^*,\tilde{\theta}^*)$ according to the proposal kernel

$$q((x,\theta,\tilde{x},\tilde{\theta}) \to (x^*,\theta^*,\tilde{x}^*,\tilde{\theta}^*))$$

which is accepted according to the standard Hastings rule with probability

$$1 \wedge \frac{\pi(x^*,\theta^*|y)\tilde{\pi}(\tilde{x}^*,\tilde{\theta}^*|\tilde{y}) \times q((x^*,\theta^*,\tilde{x}^*,\tilde{\theta}^*) \to (x,\theta,\tilde{x},\tilde{\theta}))}{\pi(x,\theta|y)\tilde{\pi}(\tilde{x},\tilde{\theta}|\tilde{y}) \times q((x,\theta,\tilde{x},\tilde{\theta}) \to (x^*,\theta^*,\tilde{x}^*,\tilde{\theta}^*))}$$

where $a \wedge b$ is the minimum of a and b.

In this paper, we break the swapping proposal kernel into the product

$$q((x,\theta,\tilde{x},\tilde{\theta}) \to (x^*,\theta^*,\tilde{x}^*,\tilde{\theta}^*)) = q((x,\theta) \to (\tilde{x}^*,\tilde{\theta}^*)) \times q((\tilde{x},\tilde{\theta}) \to (x^*,\theta^*))$$

where $q((x,\theta) \to (\tilde{x}^*,\tilde{\theta}^*))$ generates a coarse scale proposal $(\tilde{x}^*,\tilde{\theta}^*)$ from the current fine scale state (x,θ), and the kernel $q((\tilde{x},\tilde{\theta}) \to (x^*,\theta^*))$ generates a fine scale proposal (x^*,θ^*) from the current coarse scale state $(\tilde{x},\tilde{\theta})$.

Generating the coarse scale proposal is carried out by deterministically coarsening the current fine scale state and then generating a candidate value $\tilde{\theta}^*$ by simulating from its full conditional distribution (according to the coarse scale posterior) given the new proposed value $\tilde{x}^*$. Thus

$$q((x,\theta) \to (\tilde{x}^*,\tilde{\theta}^*)) = I[\tilde{x}^* = Cx] \times \tilde{\pi}(\tilde{\theta}^*|\tilde{x}^*,\tilde{y}) \tag{6}$$

where $I[\cdot]$ is the indicator function, Cx is the coarsening operation applied to the fine scale value x, and $\tilde{\pi}(\tilde{\theta}^*|\tilde{x}^*,\tilde{y})$ is the conditional distribution of $\tilde{\theta}^*$ under the coarse posterior distribution. In the applications considered in this paper Cx simply sums or averages fine scale pixels within a coarse scale pixel.

One recipe for generating the fine scale candidate (x^*, θ^*) given the current coarse scale parameters $(\tilde{x}, \tilde{\theta})$ is to generate a draw from the prior distribution $\pi(x|\theta^\dagger)$ subject to the constraint $\tilde{x} = Cx$, where $\theta^\dagger$ is a deterministic function of $\tilde{\theta}$ chosen to be appropriate for the candidate x^* given its relationship to $\tilde{x}$. Once x^* has been generated, θ^* can then be drawn from its full conditional given x^*. Hence

$$q((\tilde{x}, \tilde{\theta}) \to (x^*, \theta^*)) = \pi(x^*|\theta^\dagger, \tilde{x} = Cx^*) \times \pi(\theta^*|x^*, y) \tag{7}$$

where $\pi(x^*|\theta^\dagger, \tilde{x} = Cx^*)$ denotes the prior density for x^* given $\theta^\dagger$ and given the constraint $\tilde{x} = Cx^*$ (i.e. the coarsened version of x^* is $\tilde{x}$).

Generation of the candidate x^* according to this above rule is only feasible in the case when $\pi(x^*|\theta)$ has a multivariate normal distribution and the constraint $\tilde{x} = Cx^*$ is linear. In this case, $\pi(x^*|\theta^\dagger, \tilde{x} = Cx^*)$ corresponds to a $m - \tilde{m}$ rank normal density whose mean and variance are derived in Appendix B. In cases where simulating directly from $\pi(x^*|\theta^\dagger, \tilde{x} = Cx^*)$ is not feasible (e.g., $\pi(x|\theta)$ is not normal, the coarsening operation is not linear, or m is large), an alternative to (7) is to generate x^* sequentially. Appendix C gives the details of a sequential approach for generating x^* when the prior for $\pi(x|\theta)$ is normal and Cx averages within coarse pixels. Similar approaches could be formulated in other situations.

We now look at an implementation of a multiscale coupled chain in our ongoing example.

Example (continued)

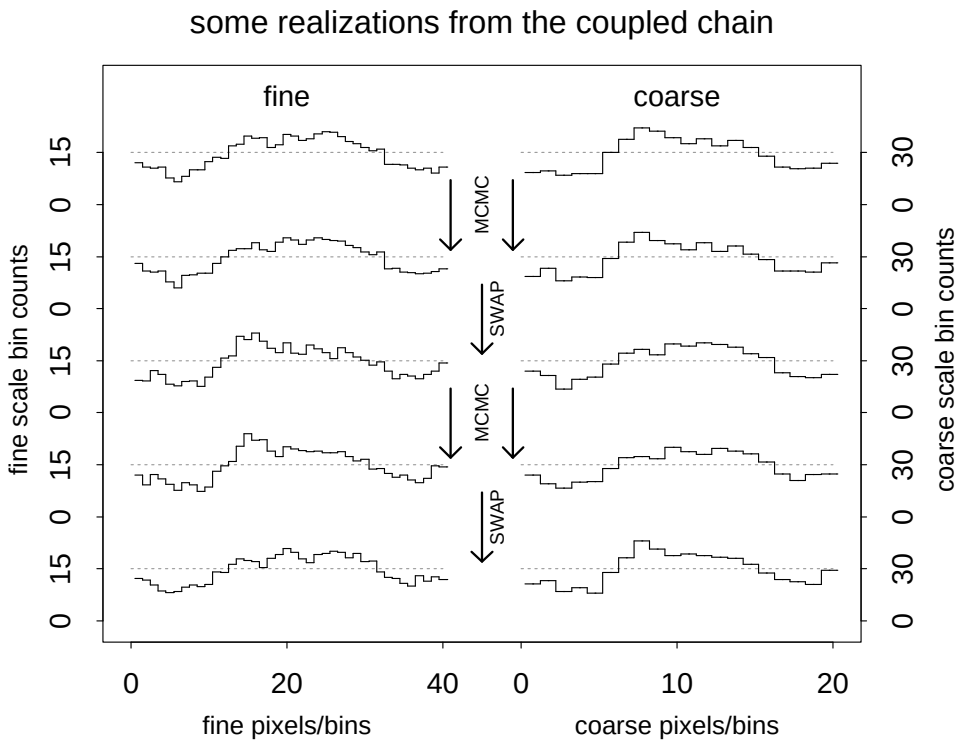

Fig. 2. Realizations from four consecutive updates from the coupled MCMC scheme. Updates alternate between standard within scale updates and Hastings swap proposals. In this sequence both proposed swaps were accepted.

Continuing with the 1-d image example, we implement this coupled chain idea by alternating within scale MCMC updates with swap proposals. The swap proposal is the product of two kernels. The first is the deterministic coarsening swap of (6) so that

$$q((x,\theta) \to (\tilde{x}^*, \tilde{\theta}^*)) = I[\tilde{x}^* = Cx] \times (\tilde{\theta}^*)^{\frac{\tilde{m}}{2}+\tilde{a}-1} \exp\{-(\tilde{b} + \tfrac{1}{2}(\tilde{x}^*)^T \tilde{W} \tilde{x}^*)\tilde{\theta}^*\} (\tilde{b} + \tfrac{1}{2}(\tilde{x}^*)^T \tilde{W} \tilde{x}^*)^{\frac{\tilde{m}}{2}+\tilde{a}-1}$$

where C is the $\tilde{m} \times m$ matrix $I_m \otimes (1,1)$. The second proposes a fine candidate given the coarse parameters according to (7):

$$\begin{aligned} q((\tilde{x},\tilde{\theta}) \to (x^*,\theta^*)) \quad &\propto \quad \exp\{-\tfrac{1}{2}(z - M\tilde{x})^T V^{-1} (z - M\tilde{x})\} (\theta^*)^{\frac{m}{2}+a-1} \\ &\quad * \exp\{-(b + \tfrac{1}{2}(x^*)^T W x^*)\theta^*\} (b + \tfrac{1}{2}(x^*)^T W x^*)^{\frac{m}{2}+a-1} \end{aligned}$$

where z is a subvector of x^* given in Appendix B, and matrices M and V are also given in Appendix B. The sequence in Figure 2 shows four successive updates – within scale updates alternated with swap proposals. This is part of a much longer MCMC run. □

B. Alternative sampling approaches

Before moving on to applications, we note that there are alternative approaches to using coarse scale samples to help explore the fine scale posterior. The most straightforward approach is importance sampling [11]. Here one could use MCMC to sample from the coarse posterior and then construct fine scale candidates according to (7) (or some other approach). These fine scale candidates could then be used as the importance sample. We tried this approach on the simple one-dimensional example and have found serious problems – the importance weights are dominated by one or two huge weights. This is because the importance sample does a poor job covering the region of highest posterior density on the fine scale. The few realizations from the importance distribution that do hit this region yield huge importance weights. Such problems can be expected since the coarse to fine proposal is not designed to use information from the likelihood.

A more appealing strategy for building an MCMC chain that shares information between scales is the reversible jump approach of Green [4]. Instead of sampling from the product distribution draws are constructed from a mixture of the fine and coarse posteriors under this approach. Note that this new target distribution is a mixture of two densities whose support is on different spaces – adding a bit of additional complexity. We refer the reader to additional references on MCMC methods that move between measures [12], [13], [14], [15].

This approach treats scale as a dynamic parameter in this simulation so that the sampling chain alternates between two types of updates. The first updates the parameters given the current scale value (fine or coarse). The second updates the scale as well as the parameter values. Updating the scale is carried out using a generalized Hastings step [4] – a proposal for the new scale is made, along with new parameter values that are consistent with this proposed scale. Preliminary experiments with the simple one-dimensional example are promising. We have elected to make use of the coupled chains in this paper for two main reasons. First, the coupled chain approach readily lends itself to parallel implementation so that the amount of processing on each scale can be easily controlled. Second, reversible jump acceptance probabilities don't simplify to the same

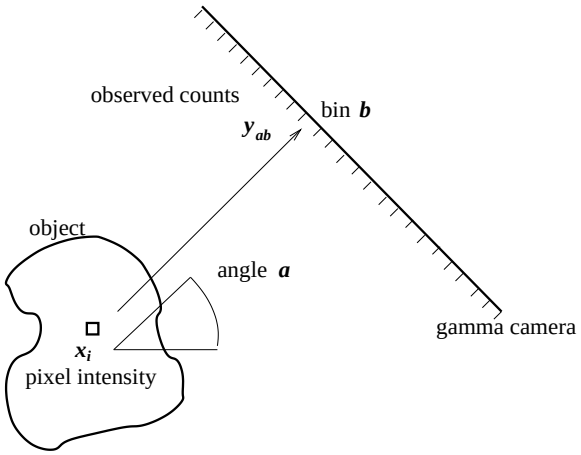

Fig. 3. SPECT reconstructions: A pixelated object emits photons with location dependent intensities x_i. The gamma camera obtains binned counts of photon emissions from various different positions controlled by the angle a. The counts from each angle a and each bin of the gamma camera b are recorded as y_{ab}. Since a photon may be scattered, absorbed, miss the gamma camera, or otherwise fail to be detected, the probability map p_{abi} gives the probability of an emission from pixel i being detected at angle a and bin b.

extent that they do under the coupled chain approach. Hence these applications require estimation of unknown normalizing constants which change with every specification, a computationally expensive proposition.

III. Examples

A. A SPECT application

Here we consider a computer generated data set from single photon emission computed tomography (SPECT). The goal is to construct a photon emission intensity map of an object using photon counts detected by a gamma camera. This application considers reconstructing a two-dimensional object. Figure 3 gives a diagram of the information obtained during a SPECT scan. As the object emits photons, the gamma camera records the locations of photon hits along the camera array. The gamma camera array is free to rotate completely around the object. At a given camera position, the counts are eventually recorded as counts over 128 bins which correspond to distance along the camera. Counts are recorded at 120 positions indexed by angle a as the camera array rotates around the object.

The data consist of counts y_{ab} obtained from bin b of the gamma camera while the it was positioned at angle a. Lead columnators on the camera absorb any photon that doesn't hit the camera at nearly a right angle. The columnators, along with the physical characteristics of the object determine the chance of a photon emitted from pixel i hits bin b while the camera is positioned at angle a. Here we treat the probabilities p_{abi} that were used to generate the data as known. The true image source intensity is shown in Fig. 4(a).

Thus the counts y_{ab} have a Poisson distribution with mean

$$\lambda_{ab} = \sum_i x_i p_{abi}$$

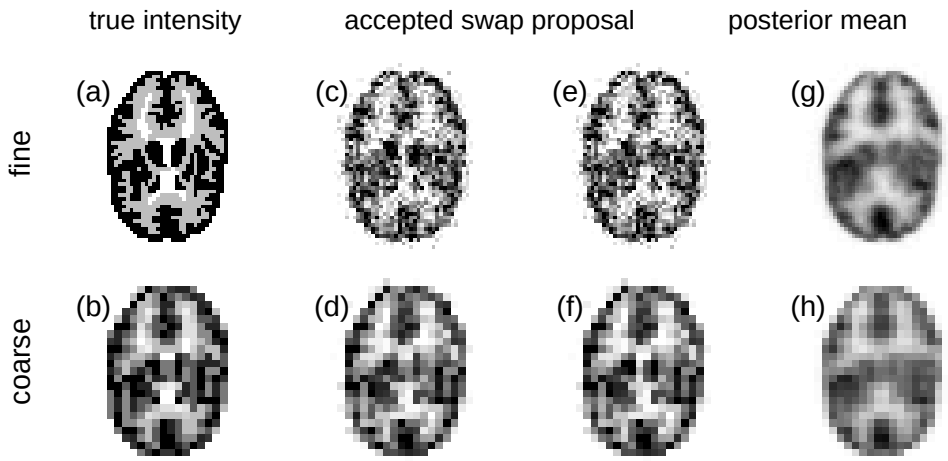

Fig. 4. Coupled fine and coarse scale MCMC for a SPECT example. (a) true emission intensities; (b) coarsened version of the true intensities; (c) & (d) current values for x and $\tilde{x}$ during the coupled MCMC run; (e) & (f) proposed fine and coarse images x^* and $\tilde{x}^*$ after swapping an interior patch of the images in (c) & (d); (g) postrior mean for x; (f) posterior mean for $\tilde{x}$.

which can be expressed in matrix form

$$\lambda_{n\times 1} = P_{n\times m} x_{m\times 1}$$

where $n = 128 \times 120$ is the number of bins times the number of angles and $m = 128 \times 128$ is the number of pixels. Given the object intensities x_i, the likelihood can be written

$$L(y|x) \propto \prod_{a,b} \lambda_{ab}^{y_{ab}} e^{-\lambda_{ab}}.$$

A far more detailed derivation of the likelihood for SPECT is given in Vardi et al. [16].

Our coarse formulation uses pixels that are four times the area of the fine specification. Hence

$$\tilde{\lambda}_{\tilde{n}\times 1} = \tilde{P}_{\tilde{n}\times\tilde{m}} \tilde{x}_{\tilde{m}\times 1} \tag{8}$$

where $\tilde{n} = n$ and $\tilde{m} = 64 \times 64$. So calculating the change in $\tilde{\lambda}$ after changing an $\tilde{x}_j$ is four times faster as compared to the fine scale calculation. We specify a simple first order Gaussian MRF prior for x and $\tilde{x}$ given in (1) and (2) where the precision matrices W and $\tilde{W}$ have diagonal elements W_{ii} equal to the number of neighboring sites, and $W_{ij} = -1$ if sites i and j are adjacent, and $W_{ij} = 0$ otherwise. Rather diffuse $\Gamma(a, b)$ priors are assigned to the parameters θ and $\tilde{\theta}$. Within a scale, standard MCMC is carried out as described in [17] or [18].

We originally used swaps as described in the 1-d imaging example of Section II. However we found that the fine scale proposals were not sufficiently accurate near the edges of the emission phantom (Fig. 4(a)). Instead we proposed to swap only interior pieces of the fine and coarse images. Fig. 5 shows how this is carried out. To construct the proposal, the same interior regions of the two images are exchanged. The region within the coarse scale exterior is then deterministically coarsened; the region within the fine scale exterior is then refined, conditioned on matching its coarse values and conditioned on the fine scale pixels neighboring the

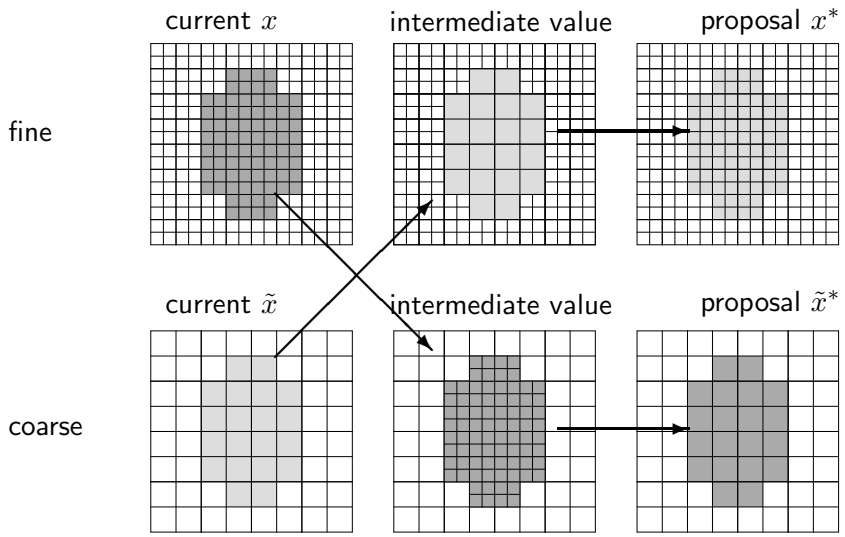

Fig. 5. A proposal that swaps only a piece of the image between the coarse and fine scales. Given the current values for x and $\tilde{x}$, the shaded regions of the two images are exchanged giving the intermediate values. The coarse shaded piece is refined to give a fine proposal x^* and the fine shaded piece is coarsened to give a coarse proposal $\tilde{x}^*$. The stochastic refining of the coarse shaded piece conditions on its previous coarse value as well as its neighboring fine scale pixels.

region. This gives the proposal a better chance of being accepted – about one in ten swap proposals are accepted. Fig. 4 shows an accepted swap along with coarse and fine scale posterior mean images.

B. Application in hydrology

An important problem in engineering is the study of the flow of liquids, particularly groundwater, through porous media, such as soil. Applications include contaminant cleanup and oil production. Reliable solutions exist for the forward problem of determining how water flows when the physical characteristics (e.g., permeability and porosity) of the aquifer are known. Of interest to statisticians and engineers alike is the inverse problem of using flow data to infer the permeability structure of the aquifer. In this section, we apply the methods of this paper to the inverse problem. Additional description of the inverse problem, as well as additional references, can be found in Lee et al. [19].

The data in the section are from a flow experiment run at Hill Air Force Base in Utah, and are part of a larger study [20], [21]. The ground at this site holds a number of contaminants. We focus on data from an experiment using a conservative tracer which does not appreciably interact with the contaminants, so that the flow data reflects the permeability structure of the aquifer. The test site measures 14 feet by 11 feet. There are four injection wells along one short edge where water is pumped into the ground, and at a particular point in time, a tracer is added so that its concentration can be monitored at other sites. Three production wells along the opposite edge extract water from the ground, forcing water to flow across the site. There are five sampling wells in the middle where the concentration of the tracer is measured (see the upper left plot of Figure 6). Water flows more easily through areas of high permeability (shown as lighter regions in the figure), so by comparing how long it takes the tracer to reach each of the sampling wells (called the *breakthrough*

times), one can make inferences about the underlying permeability structure.

We model the unknown permeabilities on a 32 by 32 grid and use an MRF prior. Note that as permeabilities are log-normally distributed, we do all operations on the log scale. Conditional on a permeability field specified on a grid, the breakthrough times are found from the solution of differential equations given by physical laws, i.e., conservation of mass, Darcy's Law, and Fick's Law. For our forward simulator, we use the S3D streamtube code of Datta-Gupta [22] to find the fitted breakthrough times, $\hat{y}_h$, and use a Gaussian likelihood on these breakthrough times centered at the observed times y_h. Hence the breakthrough times are just a non-linear transformation of the permeabilities. Note that in this example $y = \tilde{y}$, i.e., the data are the same for the coarse and fine scales. Combining the models on the fine and coarse scales gives the full likelihood:

$$p(y|x,\theta) \propto \exp\left\{-\frac{1}{2}\sum_{h=1}^{5}\left[\theta_y(\hat{y}_h - y_h)^2 + \tilde{\theta}_y(\hat{\tilde{y}}_h - y_h)^2\right]\right\}. \tag{9}$$

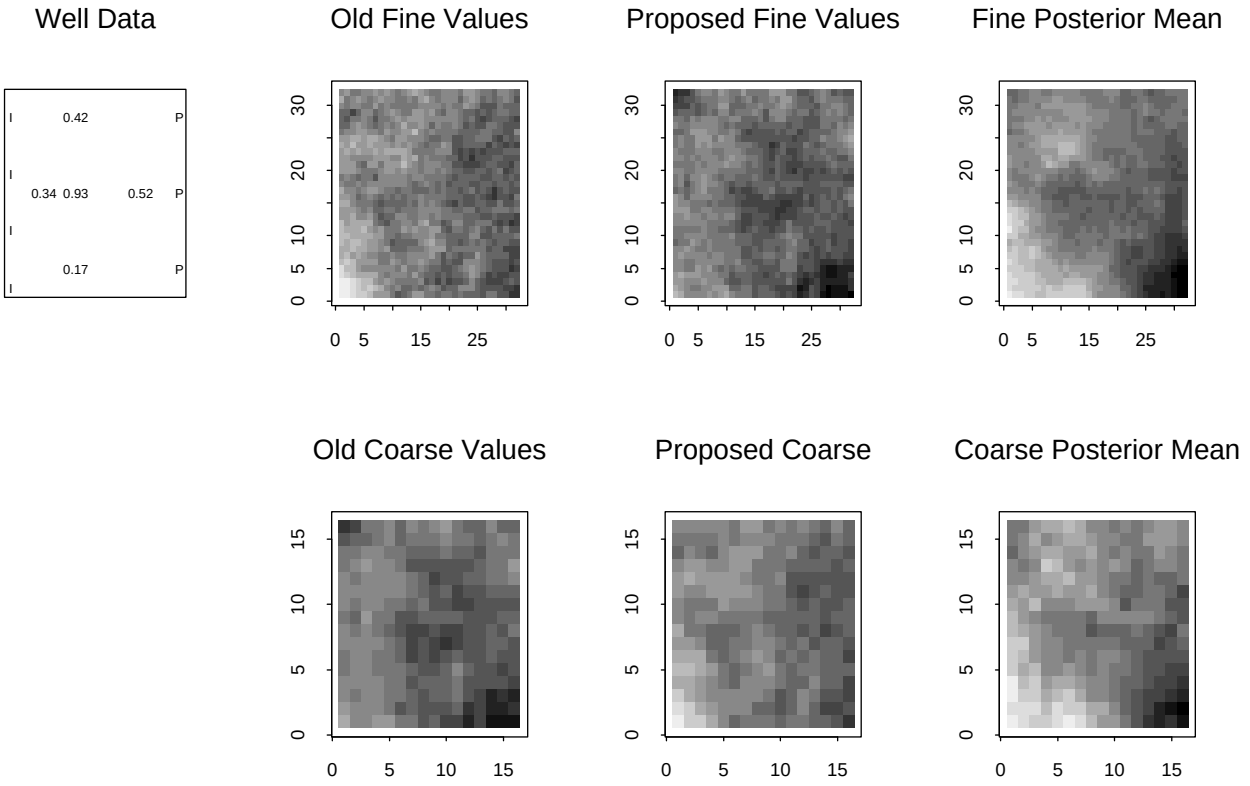

Fig. 6. Layout of wells, a sample swap, and posterior means for the Hill Air Force Base data. In the upper left plot, the wells are labeled "I" for injectors, "P" for producers, and the samplers are shown with numbers where the value is the breakthrough time (in days) for each well. For the permeability plots, all values are on the log scale, and lighter regions correspond to higher permeability values.

We put a first-order symmetric MRF prior on the coarse and fine scales

$$\pi(x, \tilde{x}|\theta_x, \tilde{\theta}_{\tilde{x}}) \propto \exp\{-0.5\theta_x x^T W x\} \exp\{-0.5\tilde{\theta}_{\tilde{x}}\tilde{x}^T \tilde{W}\tilde{x}\} \tag{10}$$

where W and $\tilde{W}$ are precision matrices with diagonal elements W_{ii} equal to the number of neighbors of pixel i, $W_{ij} = -1$ if pixels i and j are adjacent, and $W_{ij} = 0$ otherwise. Finally we specify informative gamma priors on θ_y, $\tilde{\theta}_y$, θ_x, and $\tilde{\theta}_{\tilde{x}}$.

Figure 6 shows a sample swap between the coarse and fine scales, as well as posterior means in the rightmost column. Note that the lower and central sampling wells have the earliest and latest breakthrough times respectively. In order to be consistent with this data, the posterior permeability realizations must show a high permeability path from the injectors to the lower sampling well, and a fairly abrupt decrease near the central sampling well. With this structure in the data, the fine scale model can have difficulties escaping from local modes, so the combining of information across scales allows the coarse scale model to better explore the model space and to pass this information on to the finer scale. The coupled chain appears to be about twice as efficient in computational effort as compared to a single chain run on the fine scale. We expect a parallel implementation to give a many-fold increase in speed over the single chain.

IV. Discussion

The results of this paper show that augmenting the original formulation of an inverse problem with a coarsened version can greatly reduce the amount of time it takes to explore the posterior distribution via MCMC. This improvement is due largely to the faster simulations under the coarse formulation, although the MCMC behavior is typically somewhat better under the coarse formulation as well. In this paper, we restrict ourselves to considering only an additional coarse scale formulation. As we develop parallel implementations of this scheme, we will experiment with using many additional Markov chains to explore posteriors, perhaps with even more coarsened formulations.

The methodology of this paper applies equally well when data has actually been collected on two separate scales. Such data arises naturally in a number of different fields including hydrology (flow data on a coarser scale, core data on a fine scale) and finance (daily returns and aggregated weekly or monthly returns). The coupled chains provide a mechanism for fitting a model jointly using data on multiple scales.

Acknowledgments

We would like to thank Akhil Datta-Gupta for helpful suggestions and assistance with implementing the S3D flow simulator. We would also like to thank John Trangenstein, Mike West, and Marco Ferreira for many helpful comments. This work was partially supported by National Science Foundation grant DMS 9873275.

Appendices

A. Notation

$\tilde{x}_i$	the ith coarse cell
$x_{i,j}$	the jth fine cell corresponding to the ith coarse cell
$\tilde{x}_i^*$	the proposed ith coarse cell
$x_{i,j}^*$	the proposed jth fine cell corresponding to the ith coarse cell
θ_x	the precision parameter for the fine scale MRF prior
$\tilde{\theta}_{\tilde{x}}$	the precision parameter for the coarse MRF prior
$\tilde{\theta}_y$	the precision in the coarse scale likelihood
θ_y	the precision in the fine scale likelihood
$\tilde{\theta}_y^*$	the proposed precision in the coarse scale likelihood
θ_y^*	the proposed precision in the fine scale likelihood
m	the number grid cells at the fine level
$\tilde{m}$	the number grid cells at the coarse level
k	the number of columns in the coarse grid
n	the number of data points
y_h	the hth observed data point (e.g., photon count, breakthrough time at a well)
$\tilde{y}_h$	the hth observed coarse scale data point
$\hat{y}_h$	the hth fine scale fitted value
$\hat{y}_h^*$	the hth fitted value using the current fine proposal
$\hat{\tilde{y}}_h$	the hth coarse scale fitted value
$\hat{\tilde{y}}_h^*$	the hth fitted value using the current coarse proposal
C	coarsening matrix
$\breve{C}$	augmented coarsening matrix (full-rank)
W	prior precision matrix for the fine scale
z	fine cells for full-rank multivariate joint proposal
M	conditional mean matrix, $E[z\|\tilde{x}] = M\tilde{x}$
V	conditional variance matrix, $Var(z\|\tilde{x}) = V$
a, b, $\tilde{a}$, $\tilde{b}$	hyperparameters

B. Joint coarse to fine update

Here we use a multivariate draw to create a fine-scale proposal from a single multivariate normal draw. We want to create x^* from $\tilde{x}$. We sample from the prior, but restricted so that the mean of the four fine cells corresponding to a coarse cell must be equal to the old coarse cell value. First, let z be the subset of x containing all but one of the fine cells corresponding to each coarse cell. For example, in a two-dimensional problem, z contains three of the four cells of each group corresponding to a single coarse cell (e.g., the upper left, upper right, and lower left, but not the lower right). Since we condition on the mean of each group, the

value of the last cell of each group is determined by the values of the others, so we only need to draw $z|\tilde{x}$.

Let W be the prior precision as described after Equations (4), (8), or (10). Let C be the coarsening matrix, each row representing a coarse pixel, each column a fine pixel; for the imaging example, $C_{ij} = 1$ if fine pixel j is aggregated into coarse pixel i (a summation operation); for the hydrology example, $C_{ij} = 0.25$ if fine pixel j is aggregated into coarse pixel i (an averaging operation); $C_{ij} = 0$ otherwise. Thus $\tilde{x}^* = Cx$. Augment C to get a full-rank square matrix, $\breve{C}$, by adding rows of zeroes with a one in a single column, that column corresponding to one of the fine pixels in a group except for the last one. For example, in the hydrology example, there are four fine pixels corresponding to a single coarse pixel; for each coarse pixel there are three added rows in $\breve{C}$, one for the upper left fine pixel, one for the upper right, and one for the lower left; combined with the row of C for the coarse pixel, one can recover the fine pixels from these four rows. Furthermore, we now have that $\begin{pmatrix} \tilde{x} \\ z \end{pmatrix} = \breve{C}x^*$. In drawing from the prior, x (or x^*) is multivariate normal and $\breve{C}$ is a linear transformation matrix, so $\tilde{x}$ and z are jointly multivariate normal. $\breve{C}x^*$ has precision matrix $\Lambda = (\breve{C}^T)^{-1}W\breve{C}^{-1}$ which we partition into parts corresponding to $\tilde{x}$ and z, $\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}$. Simple algebra shows that $z|\tilde{x} \sim N_{m-\tilde{m}}(M\tilde{x}, V)$ where $M = -\Lambda_{22}^{-1}\Lambda_{21}$ and $V = \Lambda_{22}^{-1}$. We can then easily draw $z|\tilde{x}$ by simulating a vector of standard normals, multiplying by the Cholesky factor of V (and by $1/\sqrt{\theta_x}$ if that is not included in W), and adding $M\tilde{x}$. The complete draw x^* is then constructed by augmenting (and re-ordering) z with the value of the last fine cell for each group such that the group mean equals the old coarse value. We need only compute M and the Cholesky decomposition of V once at the beginning, allowing for relatively easy proposal generation at each MCMC step. Since this draw is not directly from the prior, the full form of the Metropolis-Hastings algorithm must be used to compute the probability of acceptance. θ_x and/or θ_y can also be updated as part of this proposal, as is done in more detail in Appendix C (see Equations (18)-(19) and (21)).

C. Sequential coarse to fine update

An outline of our sequential algorithm for creating a swap proposal is as follows, with details explained afterwards:

1. Create a coarse proposal by summing or averaging over the corresponding cells in the old fine chain.
2. Create the first stage of the fine proposal by setting all fine cells corresponding to a coarse cell equal to the value of the old coarse cell.
3. For each coarse cell (in any arbitrary, but fixed order) draw a joint sample of corresponding fine cells from the multivariate normal distribution defined by the prior MRF conditional on the values of the other fine cells at this stage, and also conditional on their mean equaling the value of the old coarse cell.
4. Sample a proposed θ_y^* and $\tilde{\theta}_y^*$ independently for the fine and coarse scales, sampled from their complete conditional distributions. One can also easily update θ_x and $\tilde{\theta}_x$ here, although that is not done in this example.

In the equations that follow, we use a fine grid with twice as many elements in each dimension as the coarse grid, although the equations easily generalize to other scaling factors. The equations are illustrated for the two-dimensional examples, so that there are four fine cells corresponding to a single coarse cell. In the one-dimensional example, there would only be two fine cells per coarse cell. Step 1 creates a proposed coarse chain by deterministically averaging the corresponding fine cells for each coarse cell. For $i = 1, \ldots, n$,

$$\tilde{x}_i^* = \frac{1}{4} \sum_{j=1}^{4} x_{i,j} \,. \tag{11}$$

Step 2 creates a base fine proposal which exactly matches the old coarse field:

$$x_{i,j}^* = \tilde{x}_i \tag{12}$$

for $i = 1, \ldots, n$, and $j = 1, 2, 3, 4$. Step 3 modifies this base proposal to smooth it out so that it has a larger probability of being accepted. For each $i = 1, \ldots, n$, in order, we draw $\{x_{i,1}^*, x_{i,2}^*, x_{i,3}^*, x_{i,4}^*\}$ from the multivariate normal distribution defined by the MRF prior using the current value of precision at the fine scale, i.e., $\theta^\dagger = \theta_x$, and such that the four are constrained to the have an average equal to the old coarse cell, $\tilde{x}_i$ (for reversibility). Thus for an interior coarse cell (one with four neighbors), we draw

$$\begin{pmatrix} x_{i,1}^* \\ x_{i,2}^* \\ x_{i,3}^* \end{pmatrix} \sim N_3 \left(\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \frac{1}{\theta_x} \begin{bmatrix} \frac{1}{6} & -\frac{1}{24} & -\frac{1}{24} \\ -\frac{1}{24} & \frac{1}{6} & -\frac{1}{12} \\ -\frac{1}{24} & -\frac{1}{12} & \frac{1}{6} \end{bmatrix} \right) \tag{13}$$

where

$$\mu_1 = \frac{1}{24} \left(4x_{i-1,2}^* + 4x_{i-k,3}^* - 2x_{i+1,3}^* - 2x_{i+k,2}^* - x_{i-k,4}^* - x_{i+1,1}^* - x_{i-1,4}^* - x_{x+k,1}^* + 24\tilde{x}_i \right) \tag{14}$$

$$\mu_2 = \frac{1}{24} \left(-x_{i-1,2}^* - x_{i-k,3}^* - x_{i+1,3}^* - x_{i+k,2}^* + 4x_{i-k,4}^* + 4x_{i+1,1}^* - 2x_{i-1,4}^* - 2x_{x+k,1}^* + 24\tilde{x}_i \right) \tag{15}$$

$$\mu_3 = \frac{1}{24} \left(x_{i-1,2}^* - x_{i-k,3}^* - x_{i+1,3}^* - x_{i+k,2}^* - 2x_{i-k,4}^* - 2x_{i+1,1}^* + 4x_{i-1,4}^* + 4x_{x+k,1}^* + 24\tilde{x}_i \right) \tag{16}$$

and set

$$x_{i,4}^* = 4\tilde{x}_i - x_{i,1}^* - x_{i,2}^* - x_{i,3}^* \,. \tag{17}$$

For an edge cell, there are corresponding equations, or one can approximate by extending the grid.

Finally, in Step 4, we generate new proposals for the precision parameters in the likelihoods at both scales. We use priors with fixed hyperparameters, $\tilde{\theta}_y^* \sim \Gamma(\tilde{a}, \tilde{b})$ and $\theta_y^* \sim \Gamma(a, b)$. Denote by $\hat{\tilde{y}}_h^*$ the fitted observation at point h (e.g., breakthrough time at a well) using the proposed coarse permeabilities $\tilde{x}_i^*$, and by $\hat{y}_h^*$ the fitted breakthrough times using the proposed fine permeabilities, $x_{i,j}^*$.

$$\tilde{\theta}_y^* = \Gamma \left(\tilde{a} + \frac{n}{2}, \tilde{b} + \frac{1}{2} \sum_{h=1}^{n} (\hat{\tilde{y}}_h^* - y_h)^2 \right) \tag{18}$$

$$\theta_y^* = \Gamma \left(a + \frac{n}{2}, b + \frac{1}{2} \sum_{h=1}^{n} (\hat{y}_h^* - y_h)^2 \right) \tag{19}$$

This joint proposal for all the $x^*_{i,j}, \tilde{x}^*_i, \tilde{\theta}^*_y$, and θ^*_y is accepted or rejected according to the Metropolis-Hastings algorithm, using the full form of this algorithm, as the proposal generation mechanism is not symmetric. The probability of accepting the candidate is

$$\min\left(1, \frac{L(y^*, \tilde{y}^*|\theta^*, x^*, \tilde{\theta}^*, \tilde{x}^*)\pi(\theta^*, x^*, \tilde{\theta}^*, \tilde{x}^*)q(\theta, x, \tilde{\theta}, \tilde{x}|\theta^*, x^*, \tilde{\theta}^*, \tilde{x}^*)}{L(y, \tilde{y}|\theta, x, \tilde{\theta}, \tilde{x})\pi(\theta, x, \tilde{\theta}, \tilde{x})q(\theta^*, x^*, \tilde{\theta}^*, \tilde{x}^*|\theta, x, \tilde{\theta}, \tilde{x})}\right) \tag{20}$$

where the likelihood $L(y, \tilde{y}|\theta, x, \tilde{\theta}, \tilde{x})$ is given in Equation (3) and its coarse counterpart for the 1-d example, or Equation (9) for the hydrology example; the prior is given in Equations (4), (5) and their coarse counterparts, or Equation (10); the transition kernel density q is composed of two parts, one for the joint permeability draw, one for the precisions:

$$q(\theta^*, x^*, \tilde{\theta}^*, \tilde{x}^*|\theta, x, \tilde{\theta}, \tilde{x}) = q_1(x^*|\tilde{x}, \theta_x)q_2(\tilde{\theta}^*_y, \theta^*_y|\hat{\tilde{y}}^*_h, \hat{y}^*_h)\,.$$

q_1 depends only on the proposed fine scale permeabilities and the old coarse scale, because the proposed coarse scale is a deterministic function of the old fine scale. q_1 is a product of the sequential multivariate normal kernels:

$$\begin{aligned}
q_1(x^*|\tilde{x}, \theta_x) &= \prod_{i=1}^{m} q_{1,i}(x^*_{i,1}, \ldots, x^*_{i,4}|x^*_{(-i)}, \tilde{x}, \theta_x) \\
q_{1,i}(x^*_i|x^*_{(-i)}, \tilde{x}, \theta_x) &= \exp\big\{-\theta_x\big(8(x^*_{i,1} - \mu_{1,i})^2 + 10(x^*_{i,2} - \mu_{2,i})^2 \\
&\quad + 10(x^*_{i,3} - \mu_{3,i})^2 + 8(x^*_{i,1} - \mu_{1,i}) * (x^*_{i,2} - \mu_{2,i}) \\
&\quad + 8(x^*_{i,1} - \mu_{1,i}) * (x^*_{i,3} - \mu_{3,i}) + 12(x^*_{i,2} - \mu_{2,i}) * (x^*_{i,3} - \mu_{3,i})\big)\big\}
\end{aligned}$$

where $q_{1,i}$ is just the evaluation of the density of Equation (13), $x^*_{(-i)}$ are the current proposed values of x excluding $x^*_{i,j}$ for $j = 1, \ldots, 4$ (note that $x^*_{(-i)}$ will be updated sequentially for each i), and $\mu_{1,i}$, $\mu_{2,i}$, $\mu_{3,i}$, x^* are as defined in Equations (14)–(16). q_2 is a product of terms for $\tilde{\theta}^*_y$ and θ^*_y:

$$\begin{aligned}
&q_2(\tilde{\theta}^*_y, \theta^*_y|\hat{\tilde{y}}^*_h, \hat{y}^*_h) = \\
&\quad\left(\tilde{b} + \frac{1}{2}\sum_{h=1}^{n}(\hat{\tilde{y}}^*_h - y_h)^2\right)^{(\tilde{a}+n/2)} \left(\tilde{\theta}^*_y\right)^{(\tilde{a}+n/2-1)} \exp\left\{-\left(\tilde{b} + \frac{1}{2}\sum_{h=1}^{n}(\hat{\tilde{y}}^*_h - y_h)^2\right)\tilde{\theta}^*_y\right\} \\
&\quad * \left(b + \frac{1}{2}\sum_{h=1}^{n}(\hat{y}^*_h - y_h)^2\right)^{(a+n/2)} \left(\theta^*_y\right)^{(a+n/2-1)} \exp\left\{-\left(b + \frac{1}{2}\sum_{h=1}^{n}(\hat{y}^*_h - y_h)^2\right)\theta^*_y\right\}\,.
\end{aligned} \tag{21}$$

Note that q_2 could easily be modified to update θ_x instead of, or in addition to θ_y. We omit from q_1 and q_2 normalizing constants and other terms which cancel out in Equation (20).

References

[1] Frans J. T. Floris, Mike D. Bush, Maarten Cuypers, Frederic Roggero, and Anne-Randi Syversveen, "Comparison of production forecast uncertainty quantification methods — an integrated study," 1st Symposium on Petroleum Geostatistics, Toulouse, 20–23 April 1999, 1999.

[2] John W. Barker, Maarten Cuypers, and Lars Holden, "Quantifying uncertainty in production forecasts: Another look at the PUNQ-S3 problem," Society of Petroleum Engineers 2000 Annual Technical Conference, SPE 62925, 2000.

[3] Charles J. Geyer, "Markov chain Monte Carlo maximum likelihood," in *Computing Science and Statistics. Proceedings of the 23rd Symposium on the Interface*. 1991, pp. 156–163, Interface Foundation of North America (Fairfax Station, VA).

[4] Peter J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, pp. 711–732, 1995.

[5] Valen E. Johnson, "A coupling-regeneration scheme for diagnosing convergence in Markov chain Monte Carlo algorithms," *Journal of the American Statistical Association*, vol. 93, pp. 238–248, 1998.

[6] J. G. Propp and D. B. Wilson, "Exact sampling with coupled Markov chains and applications to statistical mechanics," *Random Structures and Algorithms*, vol. 9, pp. 223–252, 1996.

[7] P. Barone, G. Sebastiani, and J. Stander, "Over-relaxation methods and coupled Markov chains for Monte Carlo simulation," *to appear in Statistics and Computing*, 2001.

[8] P. Barone, G. Sebastiani, and J. Stander, "General over-relaxation Markov chain Monte Carlo algorithms for Gaussian densities," *to appear in Statistics and Probability Letters*, 2001.

[9] J. Besag, P. J. Green, D. M. Higdon, and K. Mengersen, "Bayesian computation and stochastic systems (with discussion)," *Statistical Science*, vol. 10, pp. 3–66, 1995.

[10] W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, vol. 57, pp. 97–109, 1970.

[11] Michael Evans and Tim Swartz, "Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems (disc: V11 p54-64)," *Statistical Science*, vol. 10, pp. 254–272, 1995.

[12] Bradley P. Carlin and Siddhartha Chib, "Bayesian model choice via Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society, Series B, Methodological*, vol. 57, pp. 473–484, 1995.

[13] Stephen P. Brooks, "Markov chain Monte Carlo method and its application," *The Statistician*, vol. 47, pp. 69–100, 1998.

[14] Sylvia Richardson and Peter J. Green, "On Bayesian analysis of mixtures with an unknown number of components (disc: P758-792) (corr: 1998v60 p661)," *Journal of the Royal Statistical Society, Series B, Methodological*, vol. 59, pp. 731–758, 1997.

[15] J. Besag, "Markov chain Monte Carlo for statistical inference," Tech. Rep. 9, University of Washington, Center for Statistics and the Social Sciences, 2000.

[16] Y. Vardi, L. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *Journal of the American Statistical Association*, vol. 80, pp. 8–25, 1985.

[17] I. Weir, "Fully Bayesian reconstructions from single photon emission computed tomography," *Journal of the American Statistical Association*, vol. 92, pp. 49–60, 1997.

[18] D. M. Higdon, V. E. Johnson, J. E. Bowsher, T. G. Turkington, D. R. Gilland, and R. J. Jaszczack, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Transactions on Medical Imaging*, vol. 16, pp. 516–526, 1997.

[19] Herbert Lee, David Higdon, Zhuoxin Bi, Marco Ferreira, and Mike West, "Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media," Tech. Rep. 00-35, Duke University, ISDS, 2000.

[20] Michael D. Annable, P. S. C. Rao, Kirk Hatfield, Wendy D. Graham, A. L. Wood, and C. G. Enfield, "Partitioning tracers for measuring residual NAPL: Field-scale test results," *Journal of Environmental Engineering*, vol. 124, pp. 498–503, 1998.

[21] Seongsik Yoon, *Dynamic Data Integration Into High Resolution Reservoir Models Using Streamline-Based Inversion*, Ph.D. thesis, Texas A&M University, Department of Petroleum Engineering, 2000.

[22] Michael J King and Akhil Datta-Gupta, "Streamline simulation: A current perspective," *In Situ*, vol. 22, no. 1, pp. 91–140, 1998.