A blocking Markov chain Monte Carlo method for inverse stochastic hydrogeological modeling [1]

Jianlin Fu [2]
J. Jaime Gómez-Hernández [3]

Department of Hydraulic and Environmental Engineering
Universitat Politècnica de València
Camino de Vera, s/n, 46022, Valencia, Spain
Tel. (34) 963879615
Fax. (34) 963877618

Appear in
Mathematical Geosciences, 41(2), 105-128, 2009.
DOI: 10.1007/s11004-008-9206-0

[1]The final version for personal research use is available from fu_jianlin_ac@yahoo.com.

[2]Corresponding author. Currently with Department of Energy Resources Engineering, Stanford University, 367 Panama Street, Stanford, CA 94305, USA. E-mail: fu_jianlin_ac@yahoo.com or jianfu@dihma.upv.es

[3]E-mail: jaime@dihma.upv.es

Contents

Abstract

An adequate representation of the detailed spatial variation of subsurface parameters for underground flow and mass transport simulation entails heterogeneous models. Uncertainty characterization generally calls for a Monte Carlo analysis of many equally likely realizations that honor both direct information (e.g., conductivity data) and information about the state of the system (e.g., piezometric head or concentration data). Thus, the problems faced is how to generate multiple realizations conditioned to parameter data, and inverse-conditioned to dependent state data. We propose to use a Markov chain Monte Carlo approach (McMC) with block updating and combined with upscaling to achieve this purpose. Our proposal presents an alternative block updating scheme that permits the application of McMC to inverse stochastic simulation of heterogeneous fields and incorporates upscaling in a multigrid approach to speed up the generation of the realizations. The main advantage of McMC as compared with other methods capable of generating inverse-conditioned realizations (such as the self-calibrating or the pilot point methods) is that it does not require the solution of a complex optimization inverse problem, although it requires the solution of the direct problem many times.

Keywords: Geostatistics, Inverse problem, Model calibration, History matching, Spatial structure, MCMC, Reservoir modeling, Conditional simulation

1 Introduction

Direct and indirect measurements on the property parameters of reservoirs or aquifers provide limited but indispensable knowledge on the subsurface reality. This limited information implies prediction uncertainty, the evaluation of which generally calls for a Monte Carlo analysis involving the generation of multiple realizations that should honor as much information as possible, whether it is parameter data (direct conditioning) or it is state dependent data (inverse conditioning). The problem of direct conditioning was long solved in the geostatistical literature by Journel (1974), however, it was not until the 1990's that the problem of inverse conditioning, in a stochastic modeling framework, was tackled (Sahuquillo et al. 1992; RamaRao et al. 1995; Gómez-Hernández et al. 1997; Zimmerman et al. 1998).

Commonly used inverse-conditioning algorithms include: the maximum likelihood method (Carrera and Neuman 1986), the pilot point method (RamaRao et al. 1995), and the sequential self-calibration method (Gómez-Hernández et al. 1997). Some other methods along with a detailed comparison among them can be found in Zimmerman et al. (1998). All of these approaches take the same route: (i) build a model of spatial variability of the parameter of interest, generally hydraulic conductivity, (ii) generate a realization directly conditioned to the parameter data, (iii) alter this realization so that, without perturbing the conditioning on the parameter data, the solution of the state equation honors the state data at measurement locations. The crux is in the last step, which involves a complex, ill-posed and unstable inverse problem (e.g., Carrera and Neuman 1986). The difference between the various alternatives relies on how to attack this inverse optimization problem in the most efficient and effective way.

Markov chain Monte Carlo (Oliver et al. 1997; Robert and Casella 1999) can be used to achieve the same result replacing the optimization step above by a sampling procedure. The resulting realizations are sampled from the multivariate probability distribution of the parameter of interest conditioned to both direct and inverse data. In addition, given the Bayesian substrate of the procedure, the prior information is particularly relevant, resulting in realizations that are not only conditioned to all conditioning data but also consistently built from the prior model of spatial variability. On the contrary, the approaches mentioned above (self-calibration, pilot points, etc.) cannot force this consistency with the prior model of spatial variability during the optimization step.

Markov chain Monte Carlo has not succeeded in the field of stochastic hydrogeology or petroleum engineering, because the few implementations described in the literature rely on the LU decomposition of the parameter covariance matrix (something unfeasible for middle-sized discretizations) and on an inefficient sampling procedure to build the Markov chain. In this paper we propose a block sampling procedure that does not require the LU decomposition of the covariance matrix over the entire field. We also mention how to use a multi-grid approach based on upscaling for the quick evaluation of the forward flow and/or transport problem in order to speed the Markov chain convergence.

The paper is organized as follows: in the next section the implementation details of the

blocking Markov chain Monte Carlo method (BMcMC) are given, following by a sensitivity analysis, then the coupling of block updating and upscaling is shown to improve convergence velocity, and finally some conclusions summarize the paper.

2 Blocking McMC Method

In McMC conditioning to all types of data is achieved by building a Markov Chain of realizations, in which each realization is generated based on the previously generated one. This chain is the multivariate version of standard scalar Markov chains, in which each member is drawn from a probability distribution which depends only on the last member of the chain. There are two difficulties to consider. The first one is how to establish the multivariate conditional distribution from which to draw the realizations, and the second one is to establish a procedure to sample from it. The first difficulty is solved through a Bayesian formulation, and the second one is tackled with McMC. It is in solving this second difficulty where the contribution of the paper resides.

Consider a random function (RF) discretized at n grid nodes. Suppose there are m hard data (e.g., conductivity) and k nonlinear state data where the term "nonlinear" means that the dependent state data (e.g., piezometric head, concentration data, travel time, etc.) are a non-linear function of the model parameters (the implicit expression of this non-linear function is given by the flow and transport partial differential equations). Specifically, let $\boldsymbol{x} = (x_0, x_1, ..., x_{n-1})^T \subset R^n$ denote the RF, $\boldsymbol{x}_1 = \boldsymbol{x}_{obs} = (x'_0, x'_1, ..., x'_{m-1})^T \subset R^m$ denote the m hard data, and $\boldsymbol{y} = \boldsymbol{y}_{obs} = (y_0, y_1, ..., y_{k-1})^T \subset R^k$ denote the k dependent-state data.

In a stochastic modeling framework, the objective is to generate realizations of $\boldsymbol{x}$, conditioned to $\boldsymbol{x}_1$ and to $\boldsymbol{y}$. If we adopt a multi-Gaussian distribution to characterize the RF, then $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C}_{\boldsymbol{x}})$, where $\boldsymbol{\mu} \subset R^n$ is the prior mean of the RF and $\boldsymbol{C}_{\boldsymbol{x}} \subset R^{n \times n}$ is the two-point covariance. After conditioning to the hard data, the distribution remains multi-Gaussian, $\boldsymbol{x}|\boldsymbol{x}_1 \sim N(\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}_1}, \boldsymbol{C}_{\boldsymbol{x}|\boldsymbol{x}_1})$, where $\boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}_1} \subset R^n$ is the conditional mean of the RF and $\boldsymbol{C}_{\boldsymbol{x}|\boldsymbol{x}_1} \subset R^{n \times n}$ is the conditional covariance, which correspond, respectively, to the simple kriging estimate and the simple kriging covariance commonly used in geostatistics. (We assume that any observation errors in $\boldsymbol{x}_{obs}$ are embedded into the prior covariance model.)

The joint prior (in the sense that it has not yet been conditioned to state data) probability density function (pdf) of the multi-Gaussian random field $\boldsymbol{x}$ conditioned to the hard data is (i.e., Emery 2005),

$$\pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{\theta}) = (2\pi)^{-\frac{n}{2}} \|\boldsymbol{C}_{\boldsymbol{x}|\boldsymbol{x}_1}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}_1})^T \boldsymbol{C}^{-1}_{\boldsymbol{x}|\boldsymbol{x}_1} (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}_1})\right\}. \quad (1)$$

Parameter $\boldsymbol{\theta}$ represents the prior information about the random function, given by the prior mean $\boldsymbol{\mu}$ and the prior covariance $\boldsymbol{C}_{\boldsymbol{x}}$, necessary to compute the simple kriging estimate and the simple kriging covariance. These prior values have to be either adopted as *a priori* subjective estimates or modeled after some measurements.

Assuming a multi-Gaussian error for the discrepancy between the observed state $\boldsymbol{y}$ and the state resulting from the approximate solution of the state equations $\boldsymbol{y}_{sim} = g(\boldsymbol{x})$, $\boldsymbol{y}_{sim}|\boldsymbol{x} \sim N(g(\boldsymbol{x}), \boldsymbol{C}_{\boldsymbol{y}})$, where $\boldsymbol{C}_{\boldsymbol{y}} \subset R^{k \times k}$ is a covariance matrix describing the degree of discrepancy between the transfer function $g(\boldsymbol{x})$ and the true but error-prone observation $\boldsymbol{y}$. The joint pdf of $\boldsymbol{y}$ conditioned to a realization of the parameters $\boldsymbol{x}$ is given by,

$$\pi(\boldsymbol{y}|\boldsymbol{x}) = (2\pi)^{-\frac{k}{2}} \|\boldsymbol{C}_{\boldsymbol{y}}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{y} - g(\boldsymbol{x}))^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(\boldsymbol{y} - g(\boldsymbol{x}))\right\}. \tag{2}$$

The error covariance matrix $\boldsymbol{C}_{\boldsymbol{y}}$ is generally assumed diagonal, see Carrera and Neuman (1986) for a good justification of this choice.

Using Bayes' theorem, we can derive the posterior distribution of $\boldsymbol{x}$ given the observations $\boldsymbol{x}_1$ and $\boldsymbol{y}$, and the prior model $\boldsymbol{\theta}$,

$$\pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{\theta}) = \frac{1}{c} \times \pi(\boldsymbol{y}|\boldsymbol{x}) \times \pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{\theta}), \tag{3}$$

with $c = \int \pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{\theta})d\boldsymbol{x}$ being a normalization factor. Dropping the constant c we may write the posterior pdf as (i.e. Carrera and Neuman 1986; Oliver et al. 1997),

$$\pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}_1})^T \boldsymbol{C}_{\boldsymbol{x}|\boldsymbol{x}_1}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{x}|\boldsymbol{x}_1}) - \frac{1}{2}(\boldsymbol{y} - g(\boldsymbol{x}))^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(\boldsymbol{y} - g(\boldsymbol{x}))\right\}. \tag{4}$$

Now that we have an expression for the conditional multivariate probability distribution function, the task is how to draw independent identically distributed (*i.i.d.*) multivariate samples from this posterior distribution. For this purpose, we will use McMC.

In the sequel the explicit dependency of $\boldsymbol{x}$ on the prior model $\boldsymbol{\theta}$ and on the conditioning data $\boldsymbol{x}_1$ is dropped out to lighten the notation, thus $\pi(\boldsymbol{x}) \equiv \pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{\theta})$ and $\pi(\boldsymbol{x}|\boldsymbol{y}) \equiv \pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{\theta})$.

2.1 Markov chain Monte Carlo method

Due to the complexity of the posterior probability model $\pi(\boldsymbol{x}|\boldsymbol{y})$, it is impossible to sample directly from it. Markov chain Monte Carlo (Metropolis et al. 1953; Hastings 1970; Geman and Geman 1984) is especially suitable for drawing realizations from complex conditional distributions. Metropolis and Hastings demonstrated that using the following procedure, and after some initial iterations, the realizations generated follow the desired posterior distribution:

(1) Initialize a realization $\boldsymbol{x}$;

(2) Update $\boldsymbol{x}$ according to the Metropolis-Hastings rule:

- propose $\boldsymbol{x}^*$ by drawing from the distribution $\boldsymbol{x}^* \sim q(\boldsymbol{x}^*|\boldsymbol{x})$;

- accept $\boldsymbol{x}^*$ with probability $\min\{1, \alpha\}$, where,

$$\alpha = \frac{\pi(\boldsymbol{x}^*|\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x}|\boldsymbol{y})q(\boldsymbol{x}^*|\boldsymbol{x})}; \tag{5}$$

(3) Go to (2) for the next step of the chain.

In step (1), the initialization could be any; however, the closer the first member of the chain is to a suitable candidate, the faster will the chain converge to the sought distribution. In step (2), $q(\boldsymbol{x}^*|\boldsymbol{x})$ is the proposal distribution that permits drawing an alternative realization based on the last realization generated and only based on the last realization generated, it is this proposal distribution that makes the sequence of realizations a Markov chain. The proposal distribution is a multivariate distribution conditioned on the last generated realization, it has to be chosen so that the drawing step is easy to perform and, also, so that the computing of the acceptance probability is feasible. The crucial point in Markov chain Monte Carlo is the selection of this proposal distribution: the closer the proposal is to the sought posterior probability distribution, the faster the Markov chain will converge to the generation of realizations from the desired probability distribution. The sequence of realizations will eventually converge to become a sequence of equally likely realizations drawn from the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$. The largest computation load is on computing the acceptance rate α.

The first attempt to use McMC in the hydrogeology context was made by Oliver et al. (1997) using a single-component updating to build the Markov chain. The transition proposal only modifies a single value of the entire realization to move from member $\boldsymbol{x}$ to member $\boldsymbol{x}^*$: a location j is selected at random out of the n grid nodes and its value is modified according to $x_j^* \sim N(\mu_{x_j|\boldsymbol{x}_1}, \sigma^2_{x_j|\boldsymbol{x}_1})$, with $\mu_{x_j|\boldsymbol{x}_1}$ and $\sigma^2_{x_j|\boldsymbol{x}_1}$ being the simple kriging estimate and simple kriging variance at the location j, respectively. Trying to generate realizations conditioned to state data by modifying a single cell at a time and running a forward simulation model for each modification proved effective but highly inefficient.

2.2 Blocking McMC schemes

What makes our method novel is the proposal distribution we use. The proposed member $\boldsymbol{x}^*$ is built from the previous member $\boldsymbol{x}$ by modifying an entire block of grid nodes. Fig. 1 shows the block being updated in relation to the realization: $\boldsymbol{x}$ denotes the entire field; $\dot{\boldsymbol{x}}$ denotes a subset of the entire field centered at the block to be updated for the purpose of approximating the prior multivariate pdf, $\pi(\dot{\boldsymbol{x}}) \approx \pi(\boldsymbol{x})$; $\hat{\boldsymbol{x}}$ is the block that will be modified in the updating from realization $\boldsymbol{x}$ to realization $\boldsymbol{x}^*$; $\check{\boldsymbol{x}}$ is a skin of grid cells around the block; and $\ddot{\boldsymbol{x}}$ represents an "extended block" made up of the updating block $\hat{\boldsymbol{x}}$ plus the skin $\check{\boldsymbol{x}}$.

The proposal distribution is a multivariate Gaussian distribution of the cells within the block being updated conditioned to the cell values along the skin and whatever prior $\boldsymbol{x}_1$ data

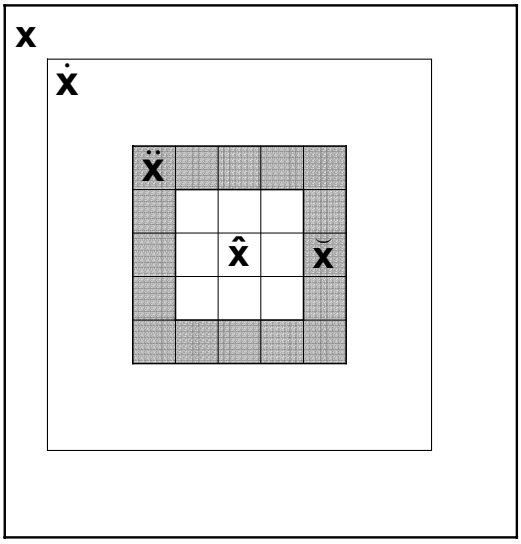

Figure 1: A superblock template $\ddot{\boldsymbol{x}}$ ($\subseteq \dot{\boldsymbol{x}} \subseteq \boldsymbol{x}$) consists of the updating block $\hat{\boldsymbol{x}}$ and its neighborhood $\breve{\boldsymbol{x}}$. Note that $\dot{\boldsymbol{x}}$, which is a subset of $\boldsymbol{x}$, is used to approximate the prior density for a high-dimensional case, i.e., $\pi(\dot{\boldsymbol{x}}) \approx \pi(\boldsymbol{x})$.

there are within the block. This conditional multiGaussian distribution is also multiGaussian with mean and covariance given by the solution of the normal equations (which are the same as the simple kriging equations, Emery 2005).

A block is randomly chosen: $\hat{\boldsymbol{x}}$. The skin values ($\breve{\boldsymbol{x}}$) plus the prior conditioning data within the block are retained as conditioning data to build a multiGaussian distribution $N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{C}}_x)$, where $\hat{\boldsymbol{\mu}}$ is the conditional mean and $\hat{\boldsymbol{C}}_x$ is the conditional covariance, both obtained by simple kriging. A realization is drawn from this distribution just for the block values, which replaces the initial block. The drawing is performed by standard geostatistical techniques for multiGaussian conditional simulations. At the end, the proposal realization $\boldsymbol{x}^*$ is identical to $\boldsymbol{x}$ except for the updating block.

The acceptance rate in (5) can be rewritten as,

$$\alpha = \frac{\pi(\boldsymbol{x}^*)}{\pi(\boldsymbol{x})} \frac{\pi(\boldsymbol{y}|\boldsymbol{x}^*)}{\pi(\boldsymbol{y}|\boldsymbol{x})} \frac{q(\boldsymbol{x}|\boldsymbol{x}^*)}{q(\boldsymbol{x}^*|\boldsymbol{x})}. \tag{6}$$

Take the logarithm,

$$\ln \alpha = \ln \pi(\boldsymbol{x}^*) - \ln \pi(\boldsymbol{x}) + \ln q(\boldsymbol{x}|\boldsymbol{x}^*) - \ln q(\boldsymbol{x}^*|\boldsymbol{x}) + \ln \pi(\boldsymbol{y}|\boldsymbol{x}^*) - \ln \pi(\boldsymbol{y}|\boldsymbol{x}). \tag{7}$$

In this expression (7) we can evaluate the different terms as follows:

$$\ln \pi(\boldsymbol{x}^*) = -\frac{1}{2}(\boldsymbol{x}^* - \boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}^* - \boldsymbol{\mu}) + c_1, \tag{8}$$

$$\ln \pi(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + c_1, \tag{9}$$

$$\ln q(\boldsymbol{x}^*|\boldsymbol{x}) = -\frac{1}{2}(\hat{\boldsymbol{x}}^* - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{C}}_{\boldsymbol{x}}{}^{-1}(\hat{\boldsymbol{x}}^* - \hat{\boldsymbol{\mu}}) + c_2, \tag{10}$$

$$\ln q(\boldsymbol{x}|\boldsymbol{x}^*) = -\frac{1}{2}(\hat{\boldsymbol{x}} - \hat{\boldsymbol{\mu}})^T \hat{\boldsymbol{C}}_{\boldsymbol{x}}{}^{-1}(\hat{\boldsymbol{x}} - \hat{\boldsymbol{\mu}}) + c_2, \tag{11}$$

$$\ln \pi(\boldsymbol{y}|\boldsymbol{x}^*) = -\frac{1}{2}(\boldsymbol{y} - g(\boldsymbol{x}^*))^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(\boldsymbol{y} - g(\boldsymbol{x}^*)) + c_3, \tag{12}$$

$$\ln \pi(\boldsymbol{y}|\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{y} - g(\boldsymbol{x}))^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(\boldsymbol{y} - g(\boldsymbol{x})) + c_3. \tag{13}$$

where c_1, c_2 and c_3 are normalizing constants that cancel out in (7).

Evaluation of the different terms leads to different schemes to compute the acceptance rate.

2.2.1 Scheme #1

Assume the field is small enough such that the inverse of the simple kriging covariance matrix $\boldsymbol{C}_{\boldsymbol{x}}$ can be LU-decomposed, $\boldsymbol{C}_{\boldsymbol{x}} = \boldsymbol{L}\boldsymbol{U}$, in which case (8) is given by,

$$\ln \pi(\boldsymbol{x}^*) = -\frac{1}{2}\boldsymbol{u}^T\boldsymbol{u} + c_1, \tag{14}$$

where $\boldsymbol{u}$ is the solution of $\boldsymbol{L}\boldsymbol{u} = \boldsymbol{x}^* - \boldsymbol{\mu}$; and similarly (9) but for $\boldsymbol{x}$ instead of $\boldsymbol{x}^*$. (Constant c_1 will cancel out in computing $\ln \alpha$ (7)). Notice that the LU decomposition is performed only once at the beginning of the Markov chain and $\boldsymbol{u}$ is simply solved by a quick back-substitution.

To evaluate terms (10) and (11), we recall that they involve a small conditional multi-Gaussian simulation constrained to the superblock $\ddot{\boldsymbol{x}}$; and we propose to use the method described by Davis (1987) and Alabert (1987) for the fast generation of conditional realizations via the LU decomposition, because this approach not only provides the values of the updating block but also the conditional covariance needed in (10) and (11). See also Dimitrakopoulos and Luo (2004) for improving the computational efficiency by a "generalized sequential simulation" method.

Finally, to compute terms (12) and (13) we need to evaluate the forward model $g(\boldsymbol{x})$ (groundwater flow and/or mass transport numerical models), which can be quite expensive to perform if the field is large or the model is complex.

2.2.2 Scheme #2

If the field is too large such that the inverse of the simple kriging covariance matrix cannot be performed, we propose an approximation for (8) and (9) by reducing the extent of the

field to a smaller area $\dot{\boldsymbol{x}}$ centered in the block being updated (Fig. 1), for which the LU decomposition of the covariance matrix can be performed, i.e.,

$$\pi(\boldsymbol{x}^*) \approx \pi(\dot{\boldsymbol{x}}^*) \tag{15}$$

$$\pi(\boldsymbol{x}) \approx \pi(\dot{\boldsymbol{x}}) \tag{16}$$

2.2.3 Scheme #3

Finally, if the field is very large and the block being updated is also very large, we take the decision to use a different proposal distribution: an independent proposal distribution in which the updating block is not made conditional on the skin values but only on the prior conditioning data $\boldsymbol{x}_1$. Recall, that the proposal distribution used in step 2 of the McMC implementation could be any, its choice will influence the convergence speed of the chain to a set of realizations from the conditional distribution, and the ability of the chain of exploring the full domain of this distribution. In such case, $q(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*)$ and the acceptance rate (5) simplifies to,

$$\alpha = \frac{\pi(\boldsymbol{y}|\boldsymbol{x}^*)}{\pi(\boldsymbol{y}|\boldsymbol{x})}. \tag{17}$$

2.3 Multi-grid computation of the likelihood

The expense of running forward simulations with a complete candidate model is an obvious shortcoming in evaluating (7) since numerous candidates should be tested and the forward simulator $g(\boldsymbol{x})$ is generally computational demanding. An alternative way is to use a coarse version of candidates with the aid of upscaling to speedup the forward evaluation. Only those candidates with potentially a high acceptance probability are subject to complete forward simulations. This approach was proposed by Efendiev et al. (2005).

A possible implementation flowchart of the entire multi-grid BMcMC is summarized in Fig. 2. When the average acceptance rate is less than a specified threshold, say 25%, the multi-grid scheme is invoked. We use an economic upscaling scheme for conductivity, e.g., the geometric mean, to yield the upscaled field $\bar{\boldsymbol{x}}$, then a flexible-grid full-tensor finite-difference flow simulator (Fu 2008) and a fast advective-only mass transport simulator, the constant-displacement random-walk particle-tracking approach (Wen and Gómez-Hernández 1996a; Fu 2008), are employed to solve the forward problems. Alternative upscaling schemes were described in Wen and Gómez-Hernández (1996b) and Renard and de Marsily (1997).

The acceptance rate for the coarse model is evaluated as in (7) with the approximation that $\pi(\boldsymbol{y}|\bar{\boldsymbol{x}})$ is the same as in (2), i.e., the error covariance matrix for the predictions based on the upscaled model is the same as that for the fine scale model. Then, if the coarse model is accepted, the fine model is run and the acceptance rate of the fine scale is computed. It

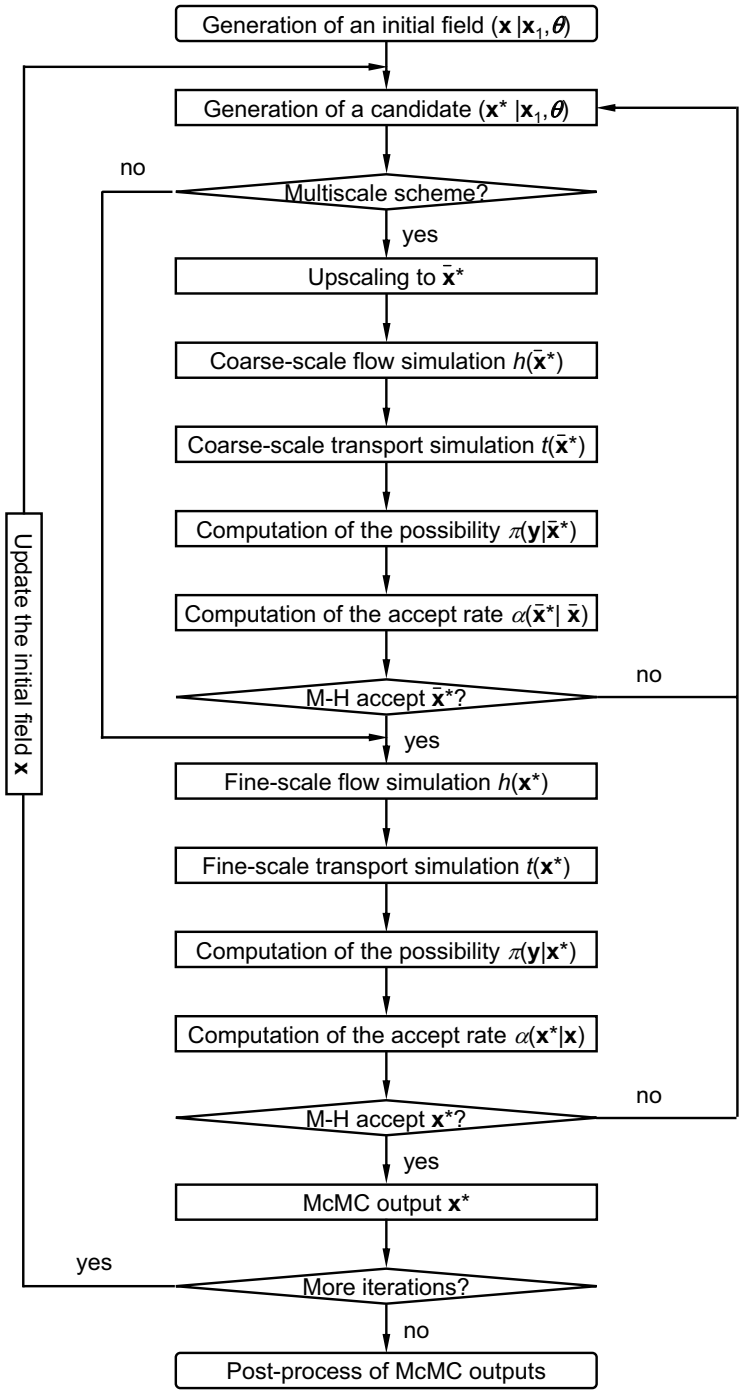

Figure 2: Flowchart of the multi-grid BMcMC scheme

has been shown by Efendiev et al. (2006) that this acceptance rate can be simply computed by,

$$\alpha = \frac{\pi(\boldsymbol{y}|\boldsymbol{x}^*)\pi(\boldsymbol{y}|\bar{\boldsymbol{x}})}{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{y}|\bar{\boldsymbol{x}}^*)}, \tag{18}$$

where $\bar{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}^*$ are the coarse models upscaled from $\boldsymbol{x}$ and $\boldsymbol{x}^*$, respectively. If the coarse model is rejected, then a new proposal at the fine scale must be generated. Note that if the acceptance rate is above the limiting threshold, say 25%, the multi-grid scheme is discarded and only the fine scale simulations are run.

3 Sensitivity Analysis

We now test the three schemes for the computation of the acceptance rate with three objectives: (i) to verify that the proposed approach gives realizations which are conditional to the state data, (ii) to verify that the Markov chain converges to a sequence of realizations drawn from the target conditional distribution, and (iii) to verify that the Markov chain samples the entire model parameter space that the posterior distribution has. The usefulness of the multi-grid scheme is also demonstrated through a numerical experiment.

3.1 A synthetic example

Consider a 2D transient single-phase flow test on a confined aquifer. All parameters have consistent length and time units, the absolute magnitude of which is irrelevant for the purpose of demonstrating the efficiency of the proposed method, only relative magnitudes are mentioned when necessary. The aquifer is discretized in 32×32 cells as shown in Fig. 3. The reference $\ln K$ field is generated using GCOSIM3D (Gómez-Hernández and Journel 1993) with a prior distribution $\ln K \sim N(0, 1)$ and an exponential isotropic variogram, i.e.,

$$\gamma_{\boldsymbol{x}}(r) = \sigma_{\boldsymbol{x}}^2 \left\{ 1 - \exp\left[-\frac{3r}{\lambda_{\boldsymbol{x}}} \right] \right\}, \tag{19}$$

where r is the separation distance, $\sigma_{\boldsymbol{x}}^2$ is the variance, and $\lambda_{\boldsymbol{x}}$ is the range, which is set to 16 cells.

The four boundaries are set to be non-flow. The initial head field is assumed to be zero everywhere in the aquifer. The time discretization for flow simulations employs the so-called time multiplier scheme which assumes that the time increment for each step is multiplied by a constant time-step coefficient α, i.e., $\Delta t_i = \alpha \Delta t_{i-1}, i \in (0, n_t)$. The simulation time of total 500 time units ($t_0 = 0$ and $t_e = 500$) is discretized into 100 steps, i.e., $n_t = 100$, with α equal to 1.05. The advantage of this scheme is that it allows for an adequate time discretization at the early stage of simulation such that the simulated transient head distribution is to the least degree influenced by the time discretization.

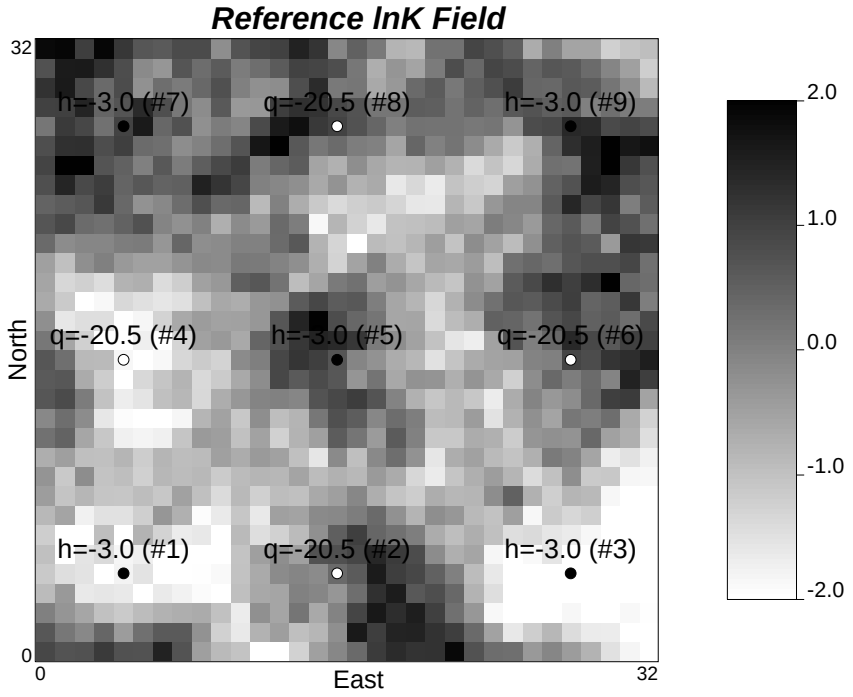

Figure 3: Reference $\ln K$ field: five constant piezometric head extraction wells (in bullet) and four constant flow-rate injection wells (in circle).

Nine wells are drilled throughout this confined aquifer (Fig. 3): four of them are injection wells with a constant flow-rate ($q = 20.5$) and the other five are production wells with a constant pressure 3 length units below the initial state. The flow-rate at the four injection wells and the hydraulic head at the five extraction wells are continuously collected during the first 50 time steps (approximately until after 40 time units).

The inverse stochastic modeling problem, therefore, is to generate hydraulic conductivity fields according to the observed flow-rate and hydraulic head data at the well-bores. As a consequence of the method used for the generation of the conductivity fields, the simulated fields should also preserve the prior distribution statistics, i.e., $\ln K \sim N(0,1)$, and the exponential covariance with the range equal to 16.

Notice that, in this example, we have chosen not to condition to any hydraulic conductivity data, simply to alleviate some of the computations and to focus the analysis on the conditioning to the state data, i.e., the difficult part of inverse stochastic modeling. The methodology has been described and tested elsewhere to account for conditioning to hydraulic conductivity data, too (Fu and Gómez-Hernández 2008; Fu 2008).

Measurement errors are assumed to have a normalized error variance of 16% at all locations and the error covariance matrix is diagonal.

3.2 Convergence velocity

In Markov chain Monte Carlo the starting member of the process could be any, then the sequence of proposal, acceptance/rejection is started; and only after a sufficiently large number of accepted candidates the chain converges to a series of realizations drawn from the target conditional multivariate probability distribution. It is therefore important to choose a starting member as close as possible to a potential draw from the target distribution; and then it is important to accelerate the convergence of the Markov chain to the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$. Notice that the forward model $g(\boldsymbol{x})$ has to be evaluated for each candidate proposed; and this evaluation is usually computationally demanding.

The starting member in this illustrative example is always the same realization drawn from $\pi(\boldsymbol{x})$. It is conditioned to any prior hydraulic conductivity data but not yet to state data.

3.2.1 Effect of block sizes

It is well known in the McMC literature that blocking helps improve the McMC convergence. We have carried out several numerical experiments to relate the blocking size to the convergence velocity. Figure 4 shows the evolution of the mismatch between $\boldsymbol{y}$ measurements and model predictions with respect to the number of iterations needed for the Markov chain to produce a series of realizations with a small mismatch between measurements and simulations (measured by the sum of all square deviations) for different blocking sizes. In all cases, scheme #1 for computing the acceptance rate is used. When the block size is 1×1, that is, no blocking is used (as in Oliver et al. 1997), or it is 2×2, 10^5 iterations are needed

for convergence; whereas for a block size equal to half the range, only a few thousands are necessary. Figure 4 also shows that for any block size above half the range the convergence velocity decreases, concluding that half the range is the optimal block for this particular case.

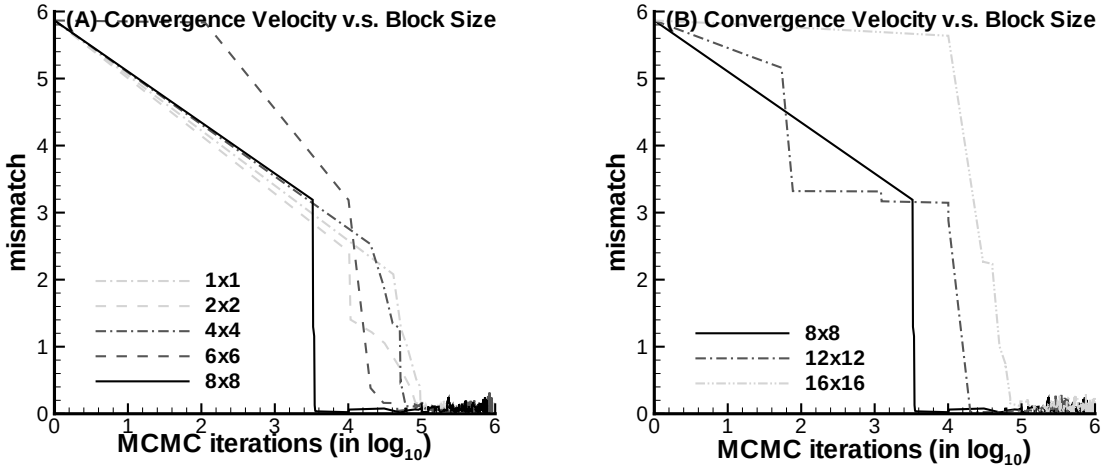

Figure 4: Effect of the block size on convergence velocity of BMcMC: (A) block size up to half the range and (B) block size from half the range up. Note that each member proposal requires a run of the forward simulator.

3.2.2 Effect of BMcMC schemes

Figure 5 displays the evolution of the mismatch between measurements and simulations for the three schemes described earlier for computing the acceptance rate α. For all three cases the blocking size is 4×4. In scheme #1, as explained, the whole domain is used to compute $\pi(\boldsymbol{x})$; whereas in scheme #2, $\pi(\boldsymbol{x})$ is approximated by a subdomain included within an area of 16×16 cells centered at the block being updated.

Schemes #1 and #2 perform similarly, indicating that the approximation introduced in #2 is effective and efficient. However, the best convergence speed is achieved with scheme #3. Indeed, in scheme #3, the acceptance rate is driven solely by the mismatch between measurements and simulations, and therefore, the chain converges faster.

3.3 BMcMC estimation performance

After the chain converges to the target distribution, two types of problems deserve attention: one is the estimation performance of the realizations generated and the other is the mixing

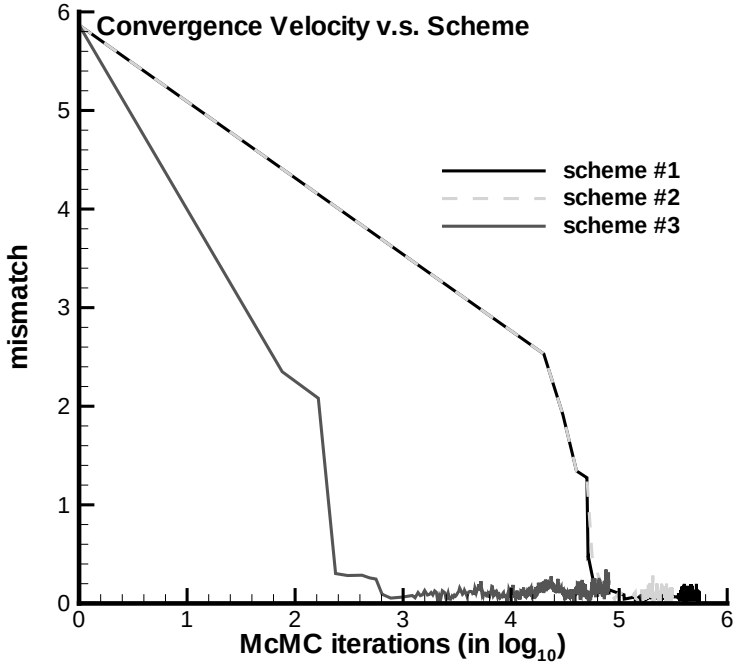

Figure 5: Effect of schemes for computing the acceptance rate α on convergence velocity of BMcMC

speed of the Markov chain, i.e., its efficiency in exploring the space of potential realizations from $\pi(\boldsymbol{x}|\boldsymbol{y})$.

3.3.1 Convergence performance measures

The CUSUM (cumulative sums) plot is a graphical evaluation of convergence of the McMC, which was proposed by Yu and Mykland (1998) and was extended by Brooks (1998). It gives both qualitative and quantitative evaluation of the mixing speed of the chain, i.e., how quickly the sample is moving around in the sample space. Given a set of output realizations (after convergence), $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_{n_r-1})^T$, and an evaluation function, $\eta(\boldsymbol{x}) = (\eta(\boldsymbol{x}_0), \eta(\boldsymbol{x}_1), ..., \eta(\boldsymbol{x}_{n_r-1}))^T$, one can construct CUSUM path plots of scalar summary statistic as follows,

(1) Calculate the mean of the evaluation function,

$$\bar{\eta} = \frac{1}{n_r} \sum_{i=0}^{n_r-1} \eta(\boldsymbol{x}_i),$$

where n_r is the number of realizations generated;

(2) Calculate the CUSUM,

$$\sigma_t = \sum_{i=0}^{t} (\eta(\boldsymbol{x}_i) - \bar{\eta}), \tag{20}$$

for $t = 0, 1, ..., n_r - 1$;

(3) Define a delta function as,

$$\delta_i = \begin{cases} 1 & \text{if } (\sigma_{i-1} - \sigma_i)(\sigma_i - \sigma_{i+1}) < 0 \\ 0 & \text{else} \end{cases}$$

for $i = 1, 2, ..., n_r - 2$;

(4) Calculate the hairiness indices,

$$\Sigma_t = \frac{1}{t-1} \sum_{i=1}^{t-1} \delta_i, \tag{21}$$

for $t = 2, 3, ..., n_r - 2$.

The key outputs are two sequential plots: $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{n_r})^T$ and $\Sigma = (\Sigma_2, \Sigma_3, ..., \Sigma_{n_r-2})^T$. The CUSUM, σ_t, gives a subjective evaluation of the convergence performance of the chain since the mixing rate is reflected by the variance of σ_t (Lin 1992; Brooks 1998). A slowly mixing sequence will lead to a high variance for σ_t and relatively large excursions around 0. When the mixing of the chain is high, the graph of σ is highly irregular (oscillatory) and concentrates around 0. The hairiness index, Σ_t, presents a quantitative measure of the smoothness of the σ plot. An ideal convergence sequence will have a hairiness around 0.5.

In our case, the scalar function used to evaluate the mixing of the BMcMC is the sum of square deviations between measurements and model predictions.

3.3.2 Effect of block sizes

This part presents a numerical experiment to compare the efficiency of three blocking sizes in exploring the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$. Figure 6 plots the mismatch between the measurements and model predictions after the chain converges. Figure 6A, 6B and 6C show that a smaller block produces a smaller mismatch. Note that the average mismatch μ for 2×2 blocking is 0.08 which is slightly less than those for 4×4 blocking ($\mu = 0.09$) and 8×8 blocking ($\mu = 0.10$). On the other hand, the 4×4 block has a faster mixing speed than both the 2×2 block and the 8×8 block as indicated by the higher value of the hairiness index (see Fig. 6D).

3.3.3 Effect of BMcMC schemes

Figure 7 compares the influence of three schemes for computing the acceptance rate on the BMcMC exploration performance and the mixing speed of the chain. One can easily find that scheme #2 almost produces the same results as scheme #1. Their basic statistics are almost the same (Fig. 7A and 7B). But the original version (i.e., scheme #1) does have a faster mixing speed than its alternative (i.e., scheme #2). However, scheme #3 yields worse results. The simulated average mismatch ($\mu = 0.14$) for scheme #3 is larger than those of the other two schemes ($\mu = 0.09$), and its mixing speed is slower than the other two schemes (see Fig. 7D).

This results confirm that scheme #2 provides a good approximation to the "exact" scheme #1 and that since scheme #3 focuses on the measurement reproduction it will not control so much how the proposed realizations relate to the target distribution, thus producing a worse mixing than the other two schemes.

3.4 Efficiency of the multi-grid scheme

Figure 8 displays the average number of iterations needed to generate an acceptable realization. As can be seen, the acceptance rate is low at the early stages of the Markov chain, and becomes higher as the chain converges and stays around one of the modes of the posterior conditional distribution. Thus, it seems appropriate that the multi-grid procedure be introduced in the generation process, running a coarse model at the early stages to evaluate whether to accept or not a realization. Figure 9 compares the evolution of the mismatch between measurements and predictions as a function of the number of iterations in two cases: one is that always the fine scale model is used, and the other is that a coarse model is used as a filter when the acceptance rate is below 25%. Both chains converge towards realizations of the same quality; however, the computational burden of the method that uses a single fine

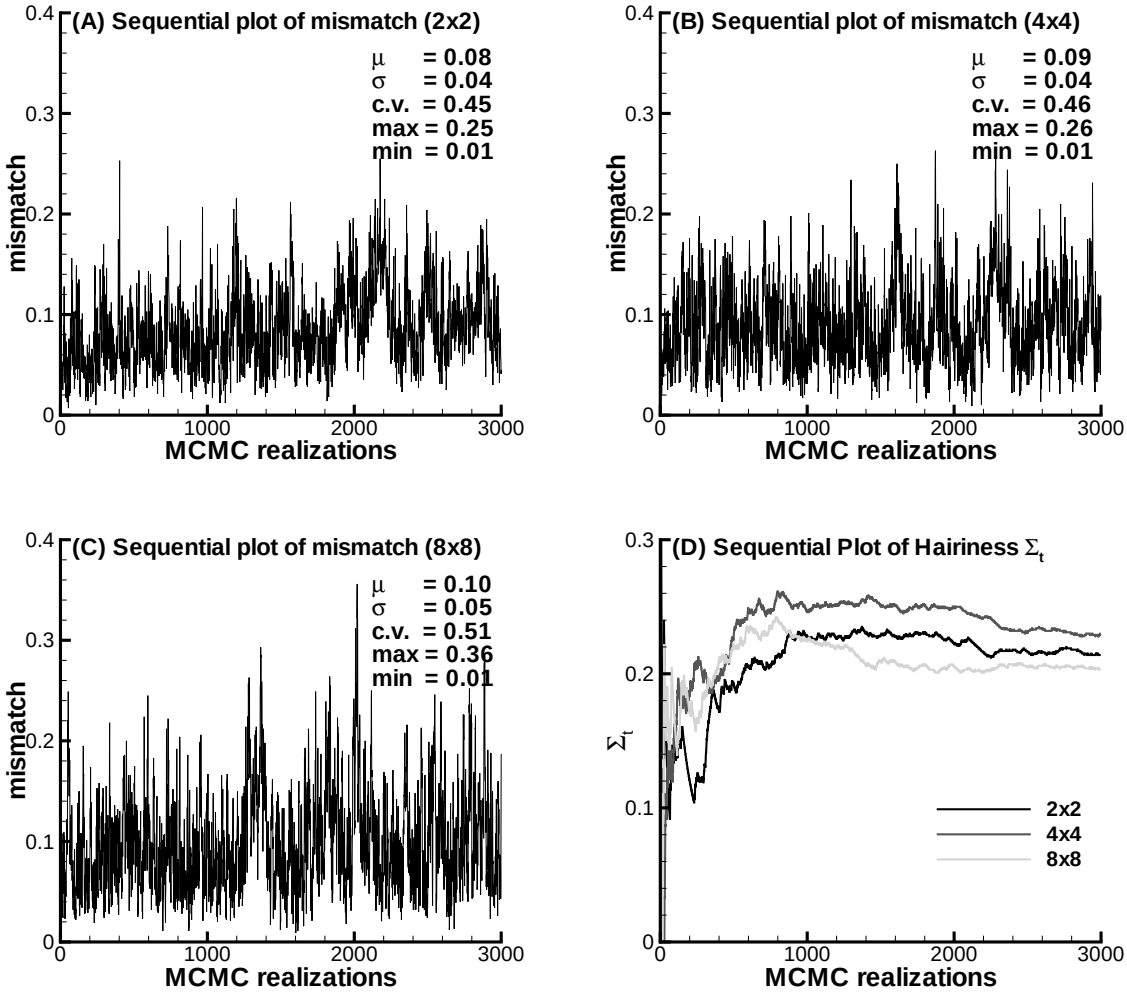

Figure 6: Effect of the updating block size on the BMcMC performance: (A) 2×2, (B) 4×4, (C) 8×8, and (D) hairiness indices of (A), (B) and (C).

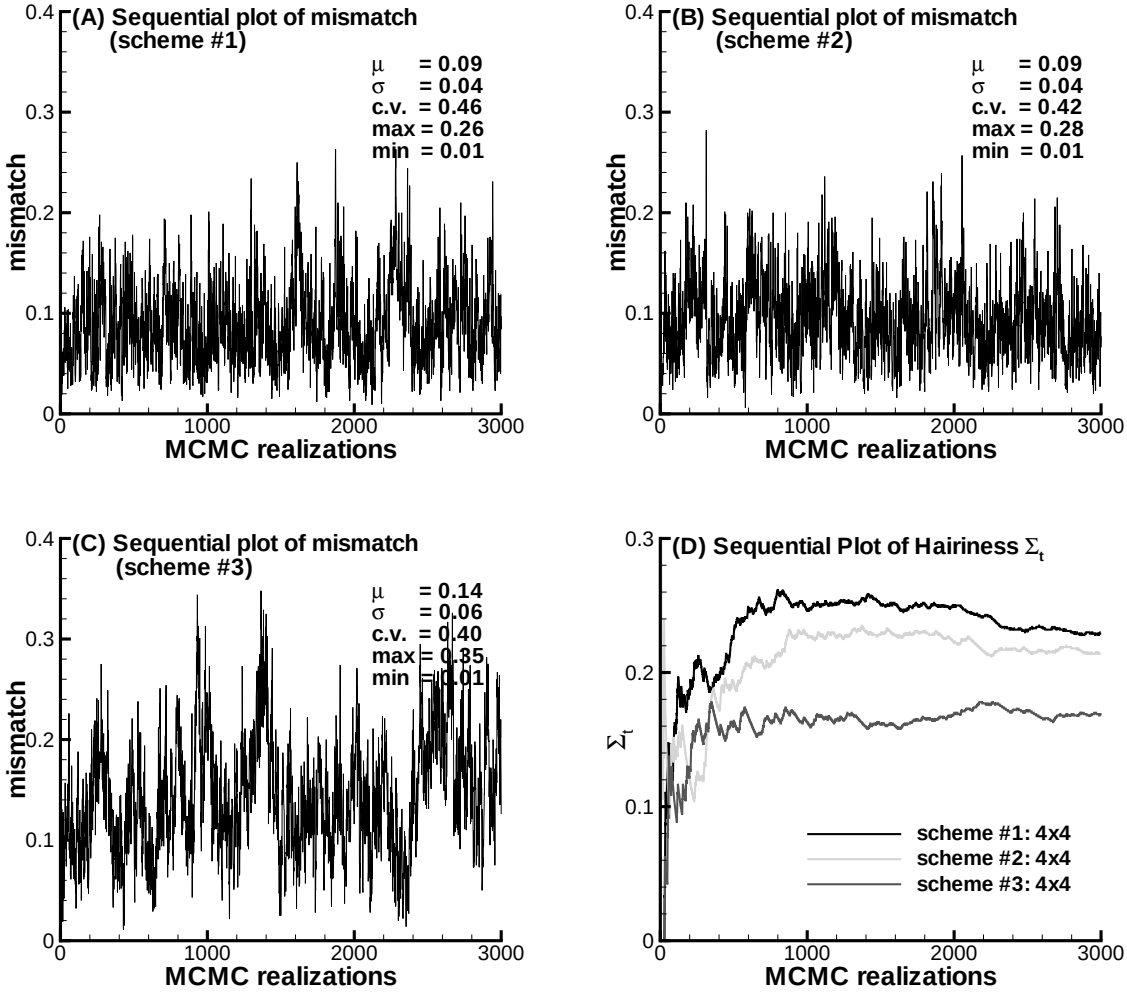

Figure 7: Effect of three schemes for computing the acceptance rate on the BMcMC performance: (A) sequential plot for scheme #1, (B) sequential plot for scheme #2, (C) sequential plot for scheme #3, and (D) hairiness indices of (A), (B) and (C).

grid is considerably higher than that of the method that uses the coarse as a filter before passing the realization for consideration to the fine scale simulator.

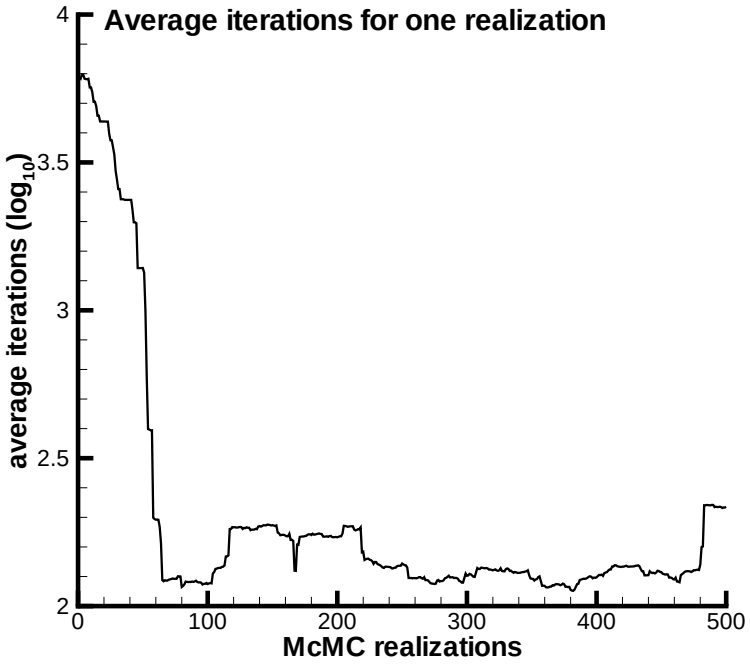

Figure 8: Average number of iterations needed to get an acceptable realization.

4 Coupling BMcMC schemes

4.1 Scheme #4

From the previous results, it appears that a large updating block (of size half the range) improves the convergence velocity of the Markov chain. It is also apparent that for the same blocking size, scheme #3 is the faster to converge. On the other hand, once the chain has converged, scheme #2 and a smaller updating block gives a better mixing of the chain. Thus, we propose a mixed scheme in which a large updating block and scheme #3 in combination with the multi-grid scheme are used to drive the chain into the region of convergence, then, after a "burn-in" period after convergence, the proposal scheme is switched to a smaller updating block (one fourth the range) and scheme #2 to compute the acceptance rate.

Figure 10 shows the number of iterations needed to generate 1000 independent realizations with the mixed scheme compared to the schemes previously analyzed. The mixed

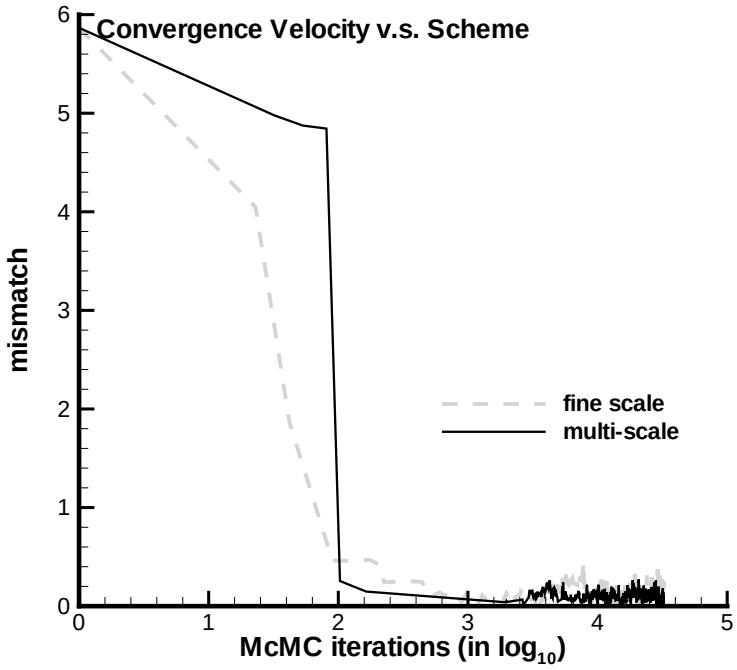

Figure 9: Convergence velocity for the fine scale BMcMC and the multiple scale one. Convergence is similar while computation needs are much smaller for the multiple scale approach.

scheme uses scheme #3 with an updating block of 8×8, then after 50 accepted realizations after convergence (burn-in period) it switches to scheme #2 with a block of 4×4.

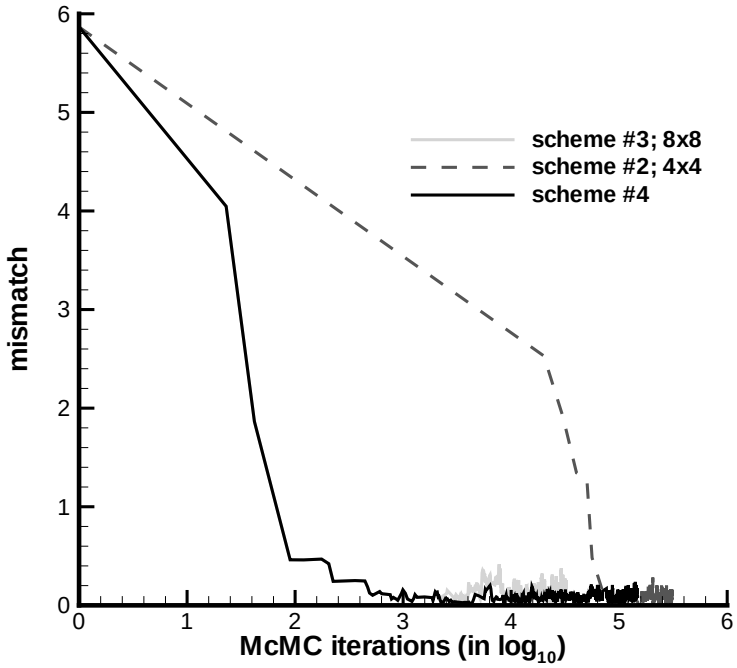

Figure 10: Improving convergence velocity with scheme #4. (Schemes #3 and #4 are indistinguishable at the beginning by construction of scheme #4)

Figure 11 shows the mismatch between measurements and simulations for the three schemes analyzed. It is clearly seen that scheme #4 shares the best of schemes #2 and #3.

4.2 Scheme #5

The scheme #4 can be modified to further improve mixing. Consider two separate chains that evolve in parallel: one is constructed by a large updating block and scheme #3, and the other with a small updating block and scheme #2. Every certain iterations, the two chains exchange accepted members to form coupled Markov chains, that, in the end, should benefit from the strengths of each chain. This coupled Markov chain concept leads us to propose an BMcMC similar to scheme #4 in which we switch back and forth between the two contributing chains. First, we run scheme #3, then we switch to scheme #2 but after generating a number of realizations, we switch back to scheme #3 to locate a new mode of the posterior distribution, and then switch to scheme #2 to generate realizations around this

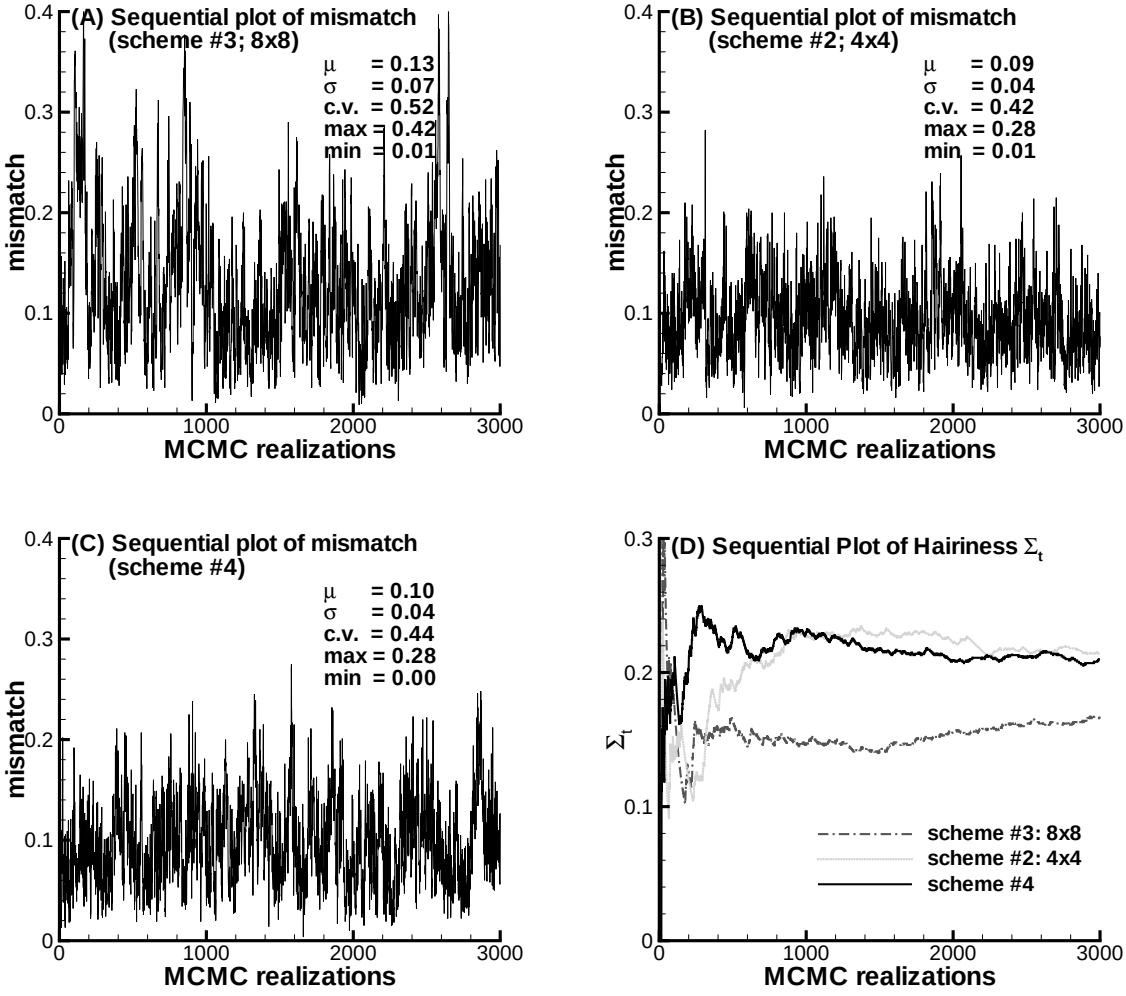

Figure 11: Improving convergence velocity with scheme #4: (A) sequential plot for scheme #3, (B) sequential plot for scheme #2, (C) sequential plot for scheme #4, and (D) hairiness indices of (A), (B) and (C).

new mode, and so on, and so forth. This combination should produce a better and faster mixing, as it is shown in Fig. 12.

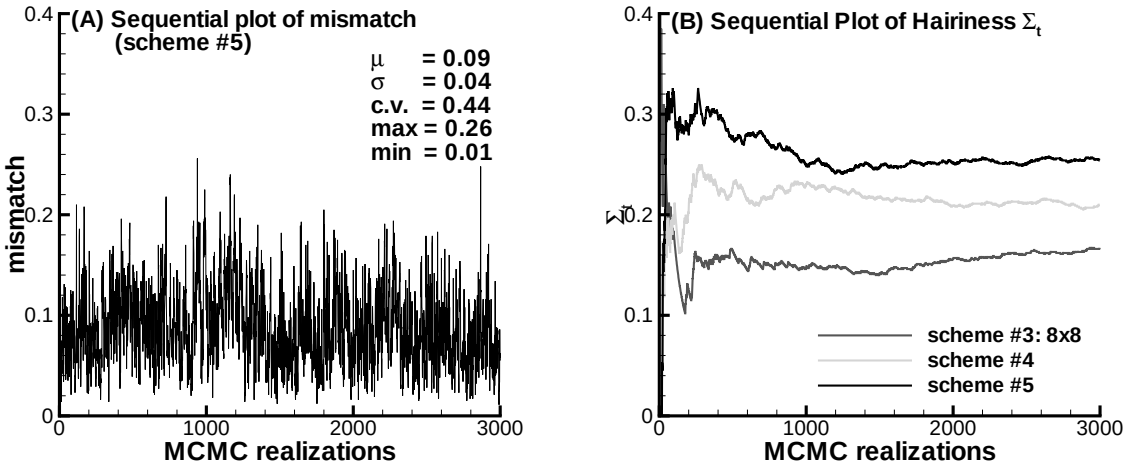

Figure 12: Improving performance with scheme #5: (A) sequential plot of scheme #5 and (B) hairiness indices of scheme #3, scheme #4, and scheme #5.

4.3 A simulation result

One hundred unconditional realizations of $\ln K$ are generated following a multi-Gaussian distribution with zero mean, unit variance and variogram (19). Also, one hundred realizations are generated inverse conditional to flow-rate and piezometric head as in the synthetic example by BMcMC using scheme #5. On the left column of Fig. 13 the measurements and predictions at three selected wells from the 100 unconditional realizations are displayed; on the right column of the same figures, the same results for the conditional realizations are shown. For the unconditional case, in which the piezometric head data was not accounted for in the generation process, the reproduction of the piezometric head measurements is poor, as expected. Some unconditional realizations are able to get close in reproducing the piezometric head data, but most do not, resulting in a large uncertainty about piezometric head predictions for the unconditional realizations. On the contrary, while the reproduction of the piezometric head measurements by the conditional realizations is not perfect (do not forget that the measurement error covariance matrix $\boldsymbol{C_y}$ is not null and a fluctuation of the model predictions about the measurements is expected), the uncertainty reduction is important. It is interesting to note that the "real" measurements are included in the cloud of prediction curves for both the unconditional and conditional cases. What conditioning does is reducing the uncertainty on model predictions.

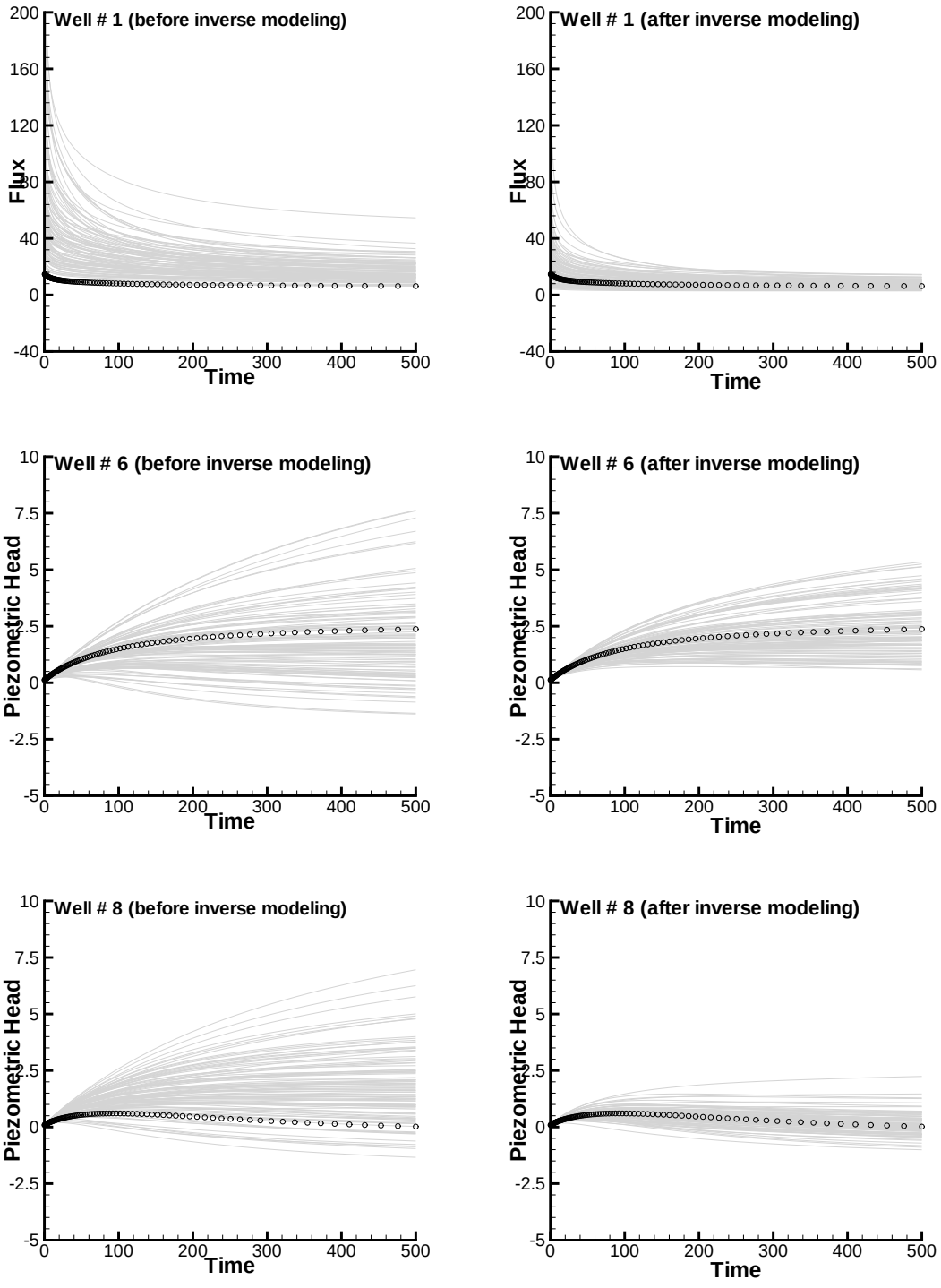

Figure 13: Three wells' performance reproduction (before 40.1 days) and prediction (after 40.1 days) by scheme #5: The left column shows the results of the unconditional realizations. The right column plots the results after inverse conditioning to wells' data by BMcMC.

In Fig. 14, one of the inverse conditional realizations is shown. From a qualitative point of view, we can appreciate that this realization displays the same patterns of variability at large scale and also the same degree of short scale variability as the reference field in Fig. 3.

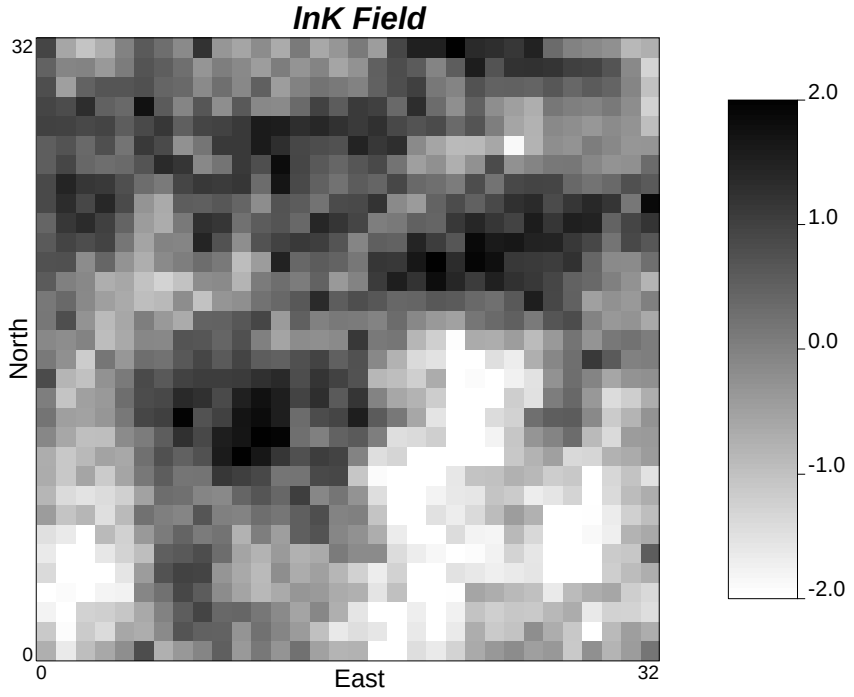

Figure 14: A typical realization of a $\ln K$ field generated by BMcMC

From a quantitative standpoint, the statistics (histogram and variogram, Fig. 15 and 16) of the 100 inverse conditional realizations stand in between those of the prior model (mean of 0, and standard deviation of 1 and the statistics of the reference field (mean of -0.33 and standard deviation of 1.07). The prior model drives the McMC simulations since it is used in each step of the simulation process at the time of computing the acceptance probability, while the reference realization exerts its control through the conditioning data. The Bayesian character of the generation procedure tends to produce realizations respecting the prior model, however, the information contained in the conditioning data about the spatial structure of the reference field clashes against the prior model, resulting in realizations with a spatial structure somewhere in between that of the reference and the prior model. This reveals a limited but important self-adaptation in the method to adjust spatial statistics and model structure in order to fit the real case.

We wish to emphasize that both the prior model and the conditioning data play a role in the statistical structure of the final ensemble of realizations. Both prior model and conditioning data are used in each step of the generation process, in each iteration of the proposal-acceptance/rejection the prior model drives the construction of the new candidate,

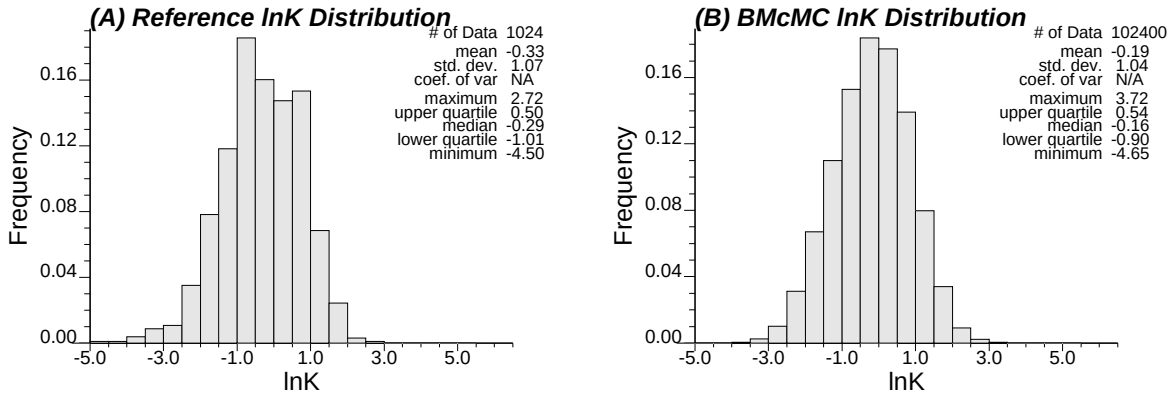

Figure 15: Histograms of the reference (A) and all 100 simulations by scheme #5 (B)

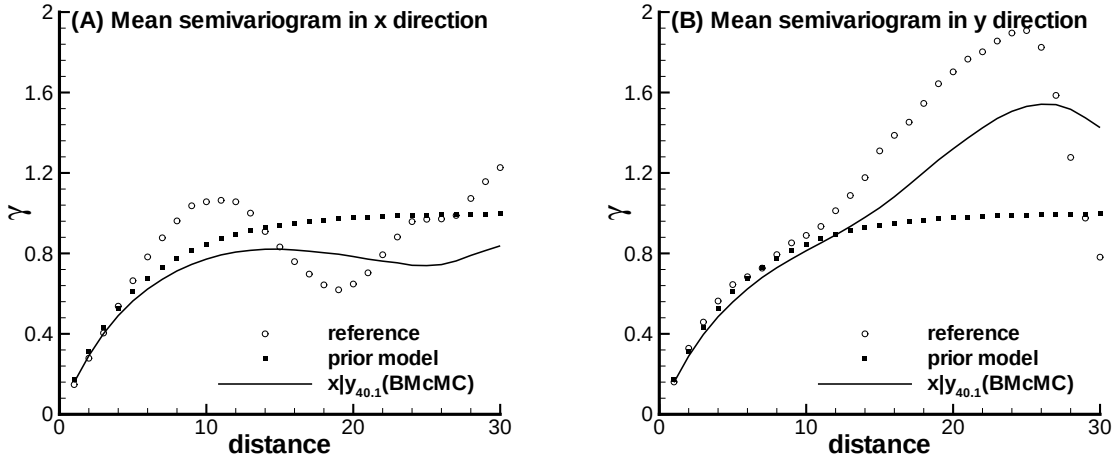

Figure 16: Variograms of the reference (A) and all 100 simulations by scheme #5 (B): The unit of the lag distance is in [$cells$].

and the conditioning data influences the probability of acceptance. This is different from other approaches such as self-calibration, pilot points, or even ensemble Kalman filtering, in which the prior model is used only for the generation of the seed realizations, and, then the conditioning data (and only the conditioning data) drives the optimization process in order to reach the reproduction of the state data.

5 Conclusions

A blocking McMC method is presented to generate independent, identically distributed realizations of hydraulic conductivity that honor both hydraulic conductivity measurements and piezometric head observations. The procedure is general enough to be applied to other fields, in which a spatial distribution of a parameter or parameters is necessary, and there are sparse data for both the parameter and some state variable (related to the parameter through a state equation). The method presented falls into the category of stochastic inverse methods, since it seeks the identification of multiple representations of conductivity, in each of which, the solution of the groundwater flow equation matches the piezometric head at data locations. As opposed to standard inverse methods that seek some "optimal" unique representation, which generally results in an oversmooth representation instead of in a likely representation of the spatial distribution.

McMC has been little explored in the communities of hydrogeology or petroleum engineering, mostly because an efficient implementation of the methodology was not found. McMC requires multiple evaluations of the state of the system to evaluate the departure between the measured and predicted states, each of such evaluation calls for the time consuming run of a complex flow simulator. In the seminal work of Oliver et al. (1997), they modified a single conductivity cell value at a time, and then run the flow simulator. Although their approach worked, the algorithm proved very time consuming. We go one step further by proposing a blocking McMC approach, in which each proposed member of the chain is different from the previous one over an entire, relatively large, block of cells. This blocking scheme, indeed, accelerates the creation of the chain of realizations.

Although all conclusions are particular to the specific example analyzed, we found a couple of rules that could be used as a reference for new applications. While proposing the candidates we should alternate between two schemes: (i) a proposal block of size about half the range of the prior statistical model and an independent proposal distribution that makes the acceptance probability only depend on the deviation between state variable measurements and model predictions; and (ii) a proposal block of a quarter of the range or less and a proposal distribution as in scheme #2. The McMC starts with scheme (i) until the deviation between measurements and predictions is small, and then switches to scheme (ii).

The method still requires a significant number of computations of the system state. For this reason, it is recommended to use some simple upscaling rules to make a quick numerical evaluation of the system state and discard those realizations that are too far from the state measurements. In other word, the multi-grid scheme can be used to reduce efficiently the

computation burden.

Other two critical problems of McMC are the convergence speed and the mixing of the chain. Both problems have been extensively studied in the specialized McMC literature, without foolproof solutions. By convergence we mean that the chain, since it starts from a realization that may be far from being inverse conditioned, may take some time until it becomes a chain of inverse conditional realizations (as it can be appreciated in Fig. 4 or Fig. 5). By mixing we mean that once the chain reaches convergence, it is desirable that the space of possible realizations is explored extensively. To accelerate the convergence we propose to use an independent sampler with a relatively larger updating block. To improve the mixing, we propose to run two chains in parallel using different proposal alternatives and to swap chain members after a certain number of iterations which reactivates the search in another area of the space of variability. This procedure is what we have called scheme #5.

BMcMC does not solve any optimization problem, therefore it is remarkably stable and well-behaved for inverse conditioning. The proposed method is easy to implement and could be extended to alternative prior stochastic models, different from the multi-Gaussian, as long as the different probability distribution involved in computing the acceptance rate can be estimated.

The disadvantages of the proposed method are related to the computation demands. The forward model has to be run numerous times, which can yield it unfeasible for very large domains. Also, the convergence speed of the Markov chain and its efficiency to explore the parameter space is not easy to quantify, as is common in McMC applications to other fields.

Acknowledgments

We thank the Spanish Ministry of Education for the financial support through grant CGL2004-02008. A doctoral fellowship and an extra travel grant awarded to the the first author by the Universitat Politècnica de València, Spain, is gratefully acknowledged. Parts of the work were carried out while the second author was on sabbatical leave. *Junfa Liu* and *Jugang Peng* kindly provided technical support in McMC and code development. We thank two anonymous reviewers for very constructive remarks to the first version of this manuscript that greatly helped us improve its final version.

References

[1] Alabert F (1987) The practice of fast conditional simulations through the LU decomposition of the covariance matrix. Math Geol, 19(5), 369-386

[2] Brooks SP (1998) Quantitative convergence assessment for Markov chain Monte Carlo via cusums. Stat Comput, 8(3), 267-274

[3] Carrera J, Neuman SP, (1986) Estimation of aquifer parameters under transient and steady state conditions, 1. Maximum likelihood method incorporating prior information. Water Resou Res, 22(2), 199-210

[4] Davis MW (1987) Production of conditional simulations via the LU triangular decomposition of the covariance matrix. Math Geol, 19(2), 91-98

[5] Dimitrakopoulos R, Luo X (2004) Generalized sequential Gaussian simulation on Group size ν and screen-effect approximations for large field simulations. Math Geol, 36(5), 567-591

[6] Efendiev Y, Datta-Gupta A, Ginting V, Ma X, Mallick B (2005) An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. Water Resou Res, 41, W12423

[7] Efendiev Y, Hou T, Luo W (2006) Preconditioning Markov chain Monte Carlo simulations using coarse-scale methods. Siam J Sci Comput, 28(2), 776-803

[8] Emery X (2005) Simple and ordinary multiGaussian kriging for estimating recoverable reserves. Math Geol, 37(3), 295-319

[9] Fu J (2008) *A Markov chain Monte Carlo method for inverse stochastic modeling and uncertainty assessment.* Unpublished doctoral dissertation, Technical University of Valencia, Spain, 158pp

[10] Fu J, Gómez-Hernández JJ (2008) Preserving spatial structure for inverse stochastic simulation using blocking Markov chain Monte Carlo method. Inverse Problem in Sciences and Engineering (in press)

[11] Geman S, Geman D (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. IEEE T Pattern Anal, 6(6), 721-741

[12] Gómez-Hernández JJ, Journel AG (1993) Joint simulation of multiGaussian random variables. In *Geostatistics Troia'92*, volume 1, edited by A. Soares, Springer, New York, pp. 85-94

[13] Gómez-Hernández JJ, Sahuquillo A, Capilla JE (1997) Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: I. Theory. J Hydrol, 203, 162-174

[14] Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their application. Biometrika, 57(1), 97-109

[15] Journel AG (1974) Geostatistics for conditional simulation of ore bodies. Economic Geology, 69(5):673-687

[16] Lin ZY (1992) On the increments of partial sums of a -mixing sequence. Theoret Probab Appli, 36, 316-326

[17] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. J Chem Phy, 21(3), 1087-1092

[18] Oliver DS, Cunha LB, Reynolds AC (1997) Markov chain Monte Carlo methods for conditioning a log-permeability field to pressure data. Math Geol, 29(1), 61-91

[19] RamaRao BS, LaVenue AM, de Marsily G, Marietta MG (1995) Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments. Water Resou Res, 31(3), 475-493

[20] Renard P, de Marsily G (1997) Calculating equivalent permeability: a review. Adv Water Resour, 20(5-6), 253-278

[21] Robert CP, Casella G, 1999. *Monte Carlo Statistical Methods*, Springer-Verlag, New York, pp. 507

[22] Sahuquillo A, Capilla JE, Gómez-Hernández JJ, Andreu J (1992) Conditional simulation of transmissivity fields honouring piezometric head data. In *Hydraulic Engineering Software IV, "Fluid Flow Modeling"*, volume II, edited by W.R. Blair and E. Cabrera, London, UK, Elsevier Applied Science, pp. 201-214

[23] Wen XH, Gómez-Hernández JJ (1996a) The constant displacement scheme for tracking particles in heterogeneous aquifers. Ground Water, 34(1), 135-142

[24] Wen XH, Gómez-Hernández JJ (1996b) Upscaling hydraulic conductivities in heterogeneous media: An overview. J Hydrol, 183, ix-xxxii

[25] Yu B, Mykland P (1998) Looking at Markov samplers through cusum path plots: a simple diagnostic idea. Stat Comput, 8(3), 275-286

[26] Zimmerman DA, de Marsily, Gotway CA, Marietta MG, Axness CL, Beauheim R, Bras R, Carrera J, Dagan G, Davies PB, Gallegos D, Galli A, Gómez-Hernández J J, Grindrod P, Gutjahr AL, Kitanidis P, LaVenue AM, McLaughlin D, Neuman SP, RamaRao BS, Ravenne C, and Rubin Y (1998) A comparison of seven geostatistically-based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow. Water Resou Res, 34(6):1373-1413