

Preserving spatial structure for inverse stochastic simulation using blocking Markov chain Monte Carlo method 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

Inverse Problems in Science and Engineering, 16(7), 865-884, 2008.

DOI: 10.1080/17415970802015781

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

²Corresponding 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

³E-mail: jaime@dihma.upv.es

Contents

1	Introduction	1
2	Methodology	3
2.1	Bayesian formulation	3
2.2	Blocking McMC scheme	4
2.3	Performance assessment of BMcMC	7
3	An Illustrative Example	8
3.1	Reference models	8
3.2	An inverse-conditioning experiment	8
3.3	Factors that affect the performance of BMcMC	10
4	Uncertainty Assessment	15
4.1	Model uncertainty	15
4.2	Uncertainty of model responses	16
4.3	Results	17
5	Summary	18
6	An LU-decomposition-based Sampler	22
6.1	Unconditional sampler	22
6.2	Conditional sampler	23

Abstract

State data (piezometric head) provide valuable information for identifying the spatial pattern of aquifer parameters (hydraulic conductivity) and to reduce the uncertainty of aquifer models. To extract such spatial information from the measurements and accurately quantify the uncertainty, a Monte Carlo method typically calls for a large number of realizations that are conditional to hard data and inverse-conditional to state data. However, inverse stochastic simulation is extremely computationally intensive since a nonlinear inverse problem is involved. In contrast to some classical nonlinear optimizers, a blocking Markov chain Monte Carlo (BMcMC) scheme is presented to generate independent, identically distributed (*i.i.d*) realizations by sampling directly from a posterior distribution that incorporates *a priori* information and *a posteriori* observations in a Bayesian framework. The realizations are not only conditioned to hard data and inverse-conditioned to state data but also preserve expected spatial structures. A synthetic example demonstrates the effectiveness of the proposed method.

Keywords: Inverse conditional simulation, Model calibration, Model structure, Spatial statistics, McMC

1 Introduction

The model parameters of interest in groundwater and petroleum engineering include hydraulic conductivity or permeability, storage coefficient or porosity, dispersivity, retardation factor, mass transfer rate, etc. Amongst them, the spatial variation of hydraulic conductivity or permeability predominantly controls flow and transport of solutes. It is also recognized that this key variable can vary significantly over a short distance showing strong heterogeneity and a log-normal distribution in space.

Unfortunately, the physical model of parameters of interest is almost impossible to obtain fully and directly from the underground. On the basis of some direct but sparse observations from well-bore and some indirect measurements by means of geophysical tools, two types of mathematical tools are developed to compensate for this under-determination or unavailability of the underground reality: one is the estimation algorithm and the other is the simulation algorithm. The estimation algorithm focuses on seeking a single estimate, e.g. the maximum likelihood (ML) estimate (e.g. *Carrera and Neuman, 1986; RamaRao et al., 1995*) and the maximum *a posteriori* (MAP) estimate (e.g. *McLaughlin and Townley, 1996; Kitanidis, 1996*), of the unknown parameters. The simulation algorithm, on the other hand, aims at constructing a series of “equally likely” models of unknown parameters (*Gomez-Hernandez et al., 1997*). One of the main advantages of the simulation over the estimation lies in that the former generates a set of alternative models as input to the complex functions to appraise the uncertainty of the spatially variable parameters of interest. Indeed, there are a large number of realizations consistent with the observations since the parameter measurement is limited in comparison with the unknown parameter field of interest.

However, this set of “equally likely” realizations should be constrained to all information available in order to further reduce the uncertainties of underground reservoirs and aquifers. This information may include some prior concepts, e.g. those from the experts’ subjective imagination, and some posterior observations, e.g. those from the measurements in the field. It is possible that those observations are linearly related to the parameters of interest, e.g. those space-dependent static hard and soft data, or even nonlinearly related, e.g. those time-dependent dynamic observations known as state data. Basically, three types of quantitative data might be drawn from this information: the hard data \mathbf{x}_1 , the state data \mathbf{y} , and the hyperparameters $\boldsymbol{\theta}$ on model statistics and spatial structure. The objective of conditional simulation is to draw *i.i.d* samples $\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \boldsymbol{\theta}$, with \mathbf{x} being the parameter of interest in K .

Amongst them, constraining to nonlinear state data $\mathbf{x}|\mathbf{y}$ is very computationally challenging since a highly nonlinear inverse problem is involved. Mathematically, the forward problem can be simplified as,

$$\mathbf{y} = g(\mathbf{x}) + \boldsymbol{\epsilon}, \quad (1)$$

where $\mathbf{x} \in R^n$ is the parameterized physical model of interest; $\mathbf{y} = \mathbf{y}_{obs} \in R^k$ is the error-prone dynamic observations; $g(\mathbf{x}) \in R^k$ is the transfer operator that maps the parameters of interest \mathbf{x} to the observation attribute \mathbf{y} ; and $\boldsymbol{\epsilon} \in R^k$ is the system error $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$

where Σ is a covariance matrix. The objective of the conditional simulation is to make inference to the physical model \mathbf{x} given the observation \mathbf{y} by assuming some prior error ϵ . In this sense, the conditional simulation $\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \theta$ involves a complex inverse problem; the procedure of conditioning to the nonlinear data $\mathbf{x}|\mathbf{y}$ is termed as “inverse-conditioning problem” just because of the existence of such inverse problem, i.e. $g^{-1}(\mathbf{y})$. Note that the forward operator $g(\mathbf{x})$ is generally analytically untractable. Most often, a complex numerical simulator is employed to solve the forward problem.

Traditional optimization-based inverse approaches to generate conditional and inverse-conditional realizations typically consists of three stages (e.g. *Zimmerman et al.*, 1998): (1) identify model statistics and spatial structure θ ; (2) generate conditional realizations $\mathbf{x}|\mathbf{x}_1, \theta$; and (3) constrain these realizations to state data $\mathbf{x}|\mathbf{x}_1, \mathbf{y}$. However, a problem arises: most often, the final models do not preserve the expected spatial structure, i.e. $\mathbf{x}|\mathbf{x}_1, \mathbf{y} \neq \mathbf{x}|\mathbf{x}_1, \mathbf{y}, \theta$, since the hyperparameters θ are readily modified during the model calibration. One of the objectives of this study, therefore, is to generate *i.i.d* realizations that preserve the spatial structure, i.e. $\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \theta$.

The construction of physical models honoring the prior information, the linear data and the nonlinear data is only one aspect of the geostatistically-based conditional and inverse-conditional simulation. Of equal importance is to perform uncertainty analysis, e.g. to quantify the reliability of those models, to identify key uncertainty resources, and to assess the resolutions and confidence regions of the conditional realizations so as to measure how much the property parameters can depart from the conditional realizations. Note that the purpose of the quantitative uncertainty analysis is not to reduce uncertainty which can only be achieved by collecting additional effective information. However, not all conditioning algorithms can detect such uncertainty reduction introduced by additional effective information. Therefore, quantification of uncertainty can reflect the efficiency of a conditioning algorithm.

Uncertainty assessment and ranking of uncertainty sources are useful in practice, e.g. for experimental network design. Once the relative importance of the various error sources to the prediction of aquifer responses has been established, one can rank the sources of uncertainty, i.e. to rank the contributions to the error of a response from different sources, e.g. the model structure, the parameter estimation, and the inherent variation of aquifers. Inevitably, the relative significance of different sources is problem specific and it is not expected that a general conclusion can be drawn from an individual case study.

To these two ends, a blocking Markov chain Monte Carlo (BMcMC) method is presented to perform Monte Carlo stochastic simulation. This article is organized as follows. The second section presents the BMcMC sampling method and the measures of the performance of the BMcMC. The third section displays the efficiency of the proposed BMcMC method for inverse-conditional simulation and some of the influential factors via a synthetic aquifer. The fourth section quantifies the model uncertainty and the response uncertainty and points out the main sources of error. Finally, several conclusions and further research directions are summarized.

2 Methodology

Consider a stochastic simulation at n grid nodes conditional to m hard data and k non-linear state data. Specifically, let $\mathbf{x} = (x_0, x_1, \dots, x_{n-1})^T \in R^n$ denote a realization conditional to m hard data $\mathbf{x}_1 = \mathbf{x}_{obs} = (x'_0, x'_1, \dots, x'_{m-1})^T \in R^m$ and k state data $\mathbf{y} = \mathbf{y}_{obs} = (y_0, y_1, \dots, y_{k-1})^T \in R^k$. Assuming a multi-Gaussian process, the spatial distribution of \mathbf{x} follows, $\mathbf{x}|\boldsymbol{\theta} \sim N(\boldsymbol{\mu}, \mathbf{C}_x)$, where $\boldsymbol{\theta}$ denotes the hyperparameters, $\boldsymbol{\mu}$, the prior mean of the Gaussian process, and \mathbf{C}_x its covariance. The observation errors of \mathbf{x}_{obs} are assumed to be assimilated into the prior statistical model. Assuming a multi-normal error, the simulated state \mathbf{y}_{sim} for a given sample \mathbf{x} can be expressed as, $\mathbf{y}_{sim}|\mathbf{x} \sim N(g(\mathbf{x}), \mathbf{C}_y)$, where \mathbf{C}_y describes the degree of discrepancy between the transfer function $g(\mathbf{x})$ and the true but error-prone observation \mathbf{y} . The transfer function $g(\mathbf{x})$ is error-prone since most often an analytical expression is not available. One generally has to resort to some complex computer models to simulate the physical process. The accuracy depends also on the ability of the mathematical model to grab the physics of undergoing phenomena. As the dimension of parameterization grows, the transfer function becomes more accurate at the expense of the computational efforts. Also, there may exist some observation errors of \mathbf{y} that can be included in this statistical model. In this sense, \mathbf{C}_y measures both the modeling errors and the measurement errors.

In summary, the objective of the stochastic inverse-conditional simulation is to infer \mathbf{x} from \mathbf{y} by assuming some spatial statistical structures and other hyperparameters $\boldsymbol{\theta}$, where \mathbf{y} is nonlinearly related to \mathbf{x} through a forward operator $g(\mathbf{x})$ and \mathbf{x} may be partly observed. The most challenging part of the conditional simulation is basically an inverse problem since an inverse operator $g^{-1}(\mathbf{y})$ is applied to the conditioning procedure.

2.1 Bayesian formulation

Assuming a multi-Gaussian distribution $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{C}_x)$, the joint prior density function (pdf) of the random field is,

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

where $\pi(\mathbf{x}|\mathbf{x}_1, \boldsymbol{\theta})$ denotes the prior pdf of $\mathbf{x} \in R^n$; n is the length of the vector \mathbf{x} ; $\boldsymbol{\mu} \in R^n$ is the prior mean of the random field; and $\mathbf{C}_x \in R^{n \times n}$ is the positive-definite covariance matrix of the vector \mathbf{x} . Note that \mathbf{x} may be partly observed, say, $\mathbf{x}_1 = \mathbf{x}_{obs} \in R^m$, but not fully known, i.e. $m < n$. The prior pdf represents some prior knowledge about the parameterization of a physical model \mathbf{x} through the configuration of $\boldsymbol{\mu}$ and \mathbf{C}_x which, together with other parameters, boil down to a hyperparameter set $\boldsymbol{\theta}$. The hyperparameters $\boldsymbol{\theta}$ are inferred from both the *a posteriori* measurements and the *a priori* subjective imagination. It should allow for the greatest uncertainty while obeying the constraints imposed by the prior information.

Assuming that the observation and modeling errors are normally distributed, the conditional probability for observing \mathbf{y} given the attribute \mathbf{x} , $\pi(\mathbf{y}|\mathbf{x})$, or equivalently, the likelihood model, $L(\mathbf{x}|\mathbf{y})$, is,

$$\pi(\mathbf{y}|\mathbf{x}) = (2\pi)^{-\frac{k}{2}} \|\mathbf{C}_y\|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (g(\mathbf{x}) - \mathbf{y})^T \mathbf{C}_y^{-1} (g(\mathbf{x}) - \mathbf{y}) \right\}, \quad (3)$$

where $\mathbf{y} = \mathbf{y}_{obs} \in R^k$ represents the values of the observations; $g(\mathbf{x})$ is usually a highly complex transfer function of \mathbf{x} , by which \mathbf{x} relates to \mathbf{y} ; and $\mathbf{C}_y \in R^{k \times k}$ is the covariance matrix of the vector \mathbf{y} . Note that if the observation errors of \mathbf{y} are fully independent from each other, then \mathbf{C}_y is a diagonal matrix.

Using the Bayes' theorem, the posterior distribution of \mathbf{x} given \mathbf{x}_1 , \mathbf{y} , and $\boldsymbol{\theta}$ may be written as $\pi(\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \boldsymbol{\theta}) = \pi(\mathbf{y}|\mathbf{x}) \times \pi(\mathbf{x}|\mathbf{x}_1, \boldsymbol{\theta})/c$, with $c = \int \pi(\mathbf{y}|\mathbf{x}) \pi(\mathbf{x}|\mathbf{x}_1, \boldsymbol{\theta}) d\mathbf{x}$ being a normalization factor. Dropping the constant c , we can write the posterior pdf,

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

The posterior pdf measures how well a parameter model \mathbf{x} agrees with the prior information and the observed data \mathbf{y} . The objective of the stochastic conditional and inverse-conditional simulation is then to draw *i.i.d* samples for \mathbf{x} from this posterior distribution $\pi(\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \boldsymbol{\theta})$. For the simplicity of presentation, \mathbf{x}_1 and $\boldsymbol{\theta}$ are dropped out such that $\pi(\mathbf{x}) \equiv \pi(\mathbf{x}|\mathbf{x}_1, \boldsymbol{\theta})$ and $\pi(\mathbf{x}|\mathbf{y}) \equiv \pi(\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \boldsymbol{\theta})$.

2.2 Blocking MCMC scheme

Due to the highly complexity of the log-likelihood model, it is impossible to sample directly from this posterior distribution $\pi(\mathbf{x}|\mathbf{y})$. The Markov chain Monte Carlo method (*Metropolis et al.*, 1953; *Hastings*, 1970; *Oliver et al.*, 1997), however, is especially suitable for exploring the parameter space of such type of complicated posterior distribution. A typical MCMC algorithm employing the Metropolis-Hastings rule to explore the posterior distribution $\pi(\mathbf{x}|\mathbf{y})$ is as follows,

- (1) Initialize the parameters \mathbf{x} ;
- (2) Update \mathbf{x} according to the Metropolis-Hastings rule:
 - propose $\mathbf{x}^* \sim q(\mathbf{x}^*|\mathbf{x})$;
 - accept \mathbf{x}^* with probability $\min\{1, \alpha\}$, where $\alpha = \frac{\pi(\mathbf{x}^*|\mathbf{y})q(\mathbf{x}|\mathbf{x}^*)}{\pi(\mathbf{x}|\mathbf{y})q(\mathbf{x}^*|\mathbf{x})}$;
- (3) Go to (2) for the next step of the chain.

After the chain converges, it will give the realizations of \mathbf{x} with the stationary posterior distribution $\pi(\mathbf{x}|\mathbf{y})$.

One of the most interesting problems in this algorithm is the configuration of the proposal transition kernel $q(\mathbf{x}^*|\mathbf{x})$, which plays a crucial role in the computational efficiency of a Metropolis-Hastings-type MCMC method. Unlike the traditional MCMC methods (*Geman and Geman, 1984; Oliver et al., 1997*) whose proposal distribution is generally single-component, e.g. $x^* \sim N(\mu, \sigma_x^2)$, and only a single grid node is updated each time, the proposed method in this study adopts a blocking updating scheme, $\mathbf{x}^* \sim N(\boldsymbol{\mu}, \mathbf{C}_x)$, which is more efficient since it approximates the posterior distribution more closely, i.e. $q(\mathbf{x}^*|\mathbf{x}) = \pi(\mathbf{x}^*|\mathbf{x})$. Moreover, it is well known that the blocking scheme helps speed up the convergence and improve the mixing of Markov chain (*Liu, 1996; Roberts and Sahu, 1997*). A more interesting feature of the blocking updating scheme is that the generated realizations preserve the spatial structure as specified *a priori* since the hyperparameters $\boldsymbol{\theta}$ remain unchanged during the sampling procedure.

The meaningfulness of “blocking” is twofold: (1) the updating unit is in a block as opposed to the single component and (2) the updating transition kernel is correlated such that it has the prior spatial statistics and structure. Specifically, the proposal kernel $\hat{\mathbf{x}}^*|\check{\mathbf{x}} \sim N(\hat{\boldsymbol{\mu}}, \hat{\mathbf{C}}_x)$, which has the identical spatial distribution as the prior model $\mathbf{x}^*|\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{C}_x)$ except that their dimensions are different, where $\hat{\mathbf{x}}^* \subseteq \mathbf{x}^*$ denotes the new parameters of the updating block and $\check{\mathbf{x}} \subset \mathbf{x}$ represents the old parameters of a limited neighbor around the updating block (see Figure 1). The superblock that consists of the updating block $\hat{\mathbf{x}}$ and its neighbor $\check{\mathbf{x}}$ can be defined as a template over which the proposal scheme generally works, i.e. $\check{\mathbf{x}} = (\hat{\mathbf{x}}^*, \check{\mathbf{x}})^T \subseteq \mathbf{x}$. As the Markov chain evolves, the models are continuously updated as the template $\check{\mathbf{x}}$ moves within the entire field \mathbf{x} following a random scanning path.

An important implementation detail is the fast generation of the proposal transition kernel $\hat{\mathbf{x}}^*$ which fully depends on the LU-decomposition of the covariance matrix (*Davis, 1987; Alabert, 1987*). Since the spatial structure of physical model is specified *a priori* and should apply to all candidates, the covariance matrix remains unchanged which makes the LU-decomposition method full of advantages for repetitive generation of candidates because the decomposition operator is applied only once. Appendix A gives an outline on the implementation of the LU-decomposition-based sampler.

The acceptance rate α is calculated by,

$$\alpha = \frac{\pi(\mathbf{x}^*)}{\pi(\mathbf{x})} \frac{\pi(\mathbf{y}|\mathbf{x}^*)}{\pi(\mathbf{y}|\mathbf{x})} \frac{\pi(\hat{\mathbf{x}}|\check{\mathbf{x}})}{\pi(\hat{\mathbf{x}}^*|\check{\mathbf{x}})}. \quad (5)$$

Take its logarithm,

$$\ln \alpha = \ln \pi(\mathbf{x}^*) - \ln \pi(\mathbf{x}) + \ln \pi(\mathbf{y}|\mathbf{x}^*) - \ln \pi(\mathbf{y}|\mathbf{x}) + \ln \pi(\hat{\mathbf{x}}|\check{\mathbf{x}}) - \ln \pi(\hat{\mathbf{x}}^*|\check{\mathbf{x}}). \quad (6)$$

Dropping the constants, the computations of the six items are listed as follows,

$$\ln \pi(\mathbf{x}^*) \propto -\frac{1}{2}(\mathbf{x}^* - \boldsymbol{\mu})^T \mathbf{C}_x^{-1}(\mathbf{x}^* - \boldsymbol{\mu}), \quad (7)$$

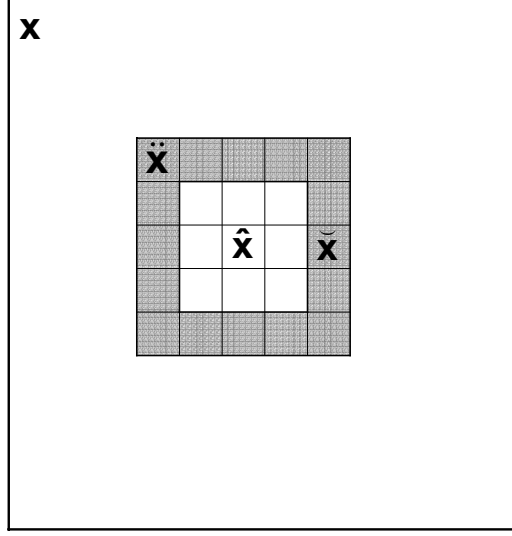


Figure 1: A superblock template $\check{\mathbf{x}} \subseteq \mathbf{x}$ consists of the updating block $\hat{\mathbf{x}}$ and its neighbor $\check{\mathbf{x}}$.

$$\ln \pi(\mathbf{x}) \propto -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}_{\mathbf{x}}^{-1}(\mathbf{x} - \boldsymbol{\mu}), \quad (8)$$

$$\ln \pi(\mathbf{y}|\mathbf{x}^*) \propto -\frac{1}{2}(g(\mathbf{x}^*) - \mathbf{y})^T \mathbf{C}_{\mathbf{y}}^{-1}(g(\mathbf{x}^*) - \mathbf{y}), \quad (9)$$

$$\ln \pi(\mathbf{y}|\mathbf{x}) \propto -\frac{1}{2}(g(\mathbf{x}) - \mathbf{y})^T \mathbf{C}_{\mathbf{y}}^{-1}(g(\mathbf{x}) - \mathbf{y}), \quad (10)$$

$$\ln \pi(\hat{\mathbf{x}}^*|\check{\mathbf{x}}) \propto -\frac{1}{2}(\check{\mathbf{x}}^* - \check{\boldsymbol{\mu}}_{\check{\mathbf{x}}})^T \check{\mathbf{C}}_{\check{\mathbf{x}}}^{-1}(\check{\mathbf{x}}^* - \check{\boldsymbol{\mu}}_{\check{\mathbf{x}}}), \quad (11)$$

$$\ln \pi(\hat{\mathbf{x}}|\check{\mathbf{x}}) \propto -\frac{1}{2}(\check{\mathbf{x}} - \check{\boldsymbol{\mu}}_{\check{\mathbf{x}}})^T \check{\mathbf{C}}_{\check{\mathbf{x}}}^{-1}(\check{\mathbf{x}} - \check{\boldsymbol{\mu}}_{\check{\mathbf{x}}}), \quad (12)$$

where the superblock $\check{\mathbf{x}}^* = (\hat{\mathbf{x}}^*, \check{\mathbf{x}})^T \subseteq \mathbf{x}^*$; $\check{\mathbf{x}} = (\hat{\mathbf{x}}, \check{\mathbf{x}})^T \subseteq \mathbf{x}$; $\check{\boldsymbol{\mu}}_{\check{\mathbf{x}}}$ is the kriging estimate for the superblock from the neighbor $\check{\mathbf{x}}$; and $\check{\mathbf{C}}$ is the covariance matrix of the superblock. The computational burden of Equations (7) and (8) is relatively small since $\mathbf{C}_{\mathbf{x}}$ is LU-decomposed once for all. The computations of Equations (9) and (10) are straightforward since $\mathbf{C}_{\mathbf{y}}$ is generally a diagonal matrix and the forward simulator $g(\mathbf{x})$ is called in a black-box way.

For the blocking MCMC method, the proposal kernel $\hat{\mathbf{x}}^*|\check{\mathbf{x}} \sim N(\hat{\boldsymbol{\mu}}, \hat{\mathbf{C}}_{\mathbf{x}})$, which entails that the kriging estimates and the kriging covariances are needed to compute firstly in that they fully depend on the current state of the chain. This is quite computationally demanding. An economical alternative can be found to compute $\ln \pi(\hat{\mathbf{x}}^*|\check{\mathbf{x}})$, i.e.

$$\ln \pi(\hat{\mathbf{x}}^* | \check{\mathbf{x}}) \propto -\frac{1}{2} \mathbf{z}_{\hat{\mathbf{x}}^*}^T \mathbf{z}_{\hat{\mathbf{x}}^*}, \quad (13)$$

where $\mathbf{z}_{\hat{\mathbf{x}}^*} \sim N(\mathbf{0}, \mathbf{1})$ that yields the random realization for the updating block, since,

$$-\frac{1}{2}(\hat{\mathbf{x}}^* - \hat{\boldsymbol{\mu}}_{\check{\mathbf{x}}})^T \mathbf{C}_{\hat{\mathbf{x}}^*}^{-1}(\hat{\mathbf{x}}^* - \hat{\boldsymbol{\mu}}_{\check{\mathbf{x}}}) = -\frac{1}{2}(\mathbf{L}_{\hat{\mathbf{x}}^* \hat{\mathbf{x}}^*} \mathbf{z}_{\hat{\mathbf{x}}^*})^T \mathbf{C}_{\hat{\mathbf{x}}^*}^{-1}(\mathbf{L}_{\hat{\mathbf{x}}^* \hat{\mathbf{x}}^*} \mathbf{z}_{\hat{\mathbf{x}}^*}) = -\frac{1}{2} \mathbf{z}_{\hat{\mathbf{x}}^*}^T \mathbf{z}_{\hat{\mathbf{x}}^*}. \quad (14)$$

where $\hat{\boldsymbol{\mu}}_{\check{\mathbf{x}}}$ is the kriging estimate for the updating block from its neighbor $\check{\mathbf{x}}$. And,

$$\ln \pi(\hat{\mathbf{x}} | \check{\mathbf{x}}) \propto -\frac{1}{2} \mathbf{u}_{\hat{\mathbf{x}}}^T \mathbf{u}_{\hat{\mathbf{x}}}, \quad (15)$$

where $\mathbf{u}_{\hat{\mathbf{x}}}$ satisfies $\mathbf{L}_{\hat{\mathbf{x}} \hat{\mathbf{x}}} \mathbf{u}_{\hat{\mathbf{x}}} = \hat{\mathbf{x}} - \hat{\boldsymbol{\mu}}_{\check{\mathbf{x}}}$. Note that the computation of $\mathbf{L}_{\hat{\mathbf{x}} \hat{\mathbf{x}}}$ is quite expensive. But the computational burden can be reduced by narrowing down the neighbor size of the updating block, i.e. reducing the number of conditioning data for the kriging estimate.

2.3 Performance assessment of BMcMC

An empirical approach to convergence control is to draw pictures of the output of a chain in order to detect deviant or nonstationary behaviors (*Robert and Casella, 1999*). The key output of this method is a sequential plot, $\eta(\mathbf{x}) = (\eta(\mathbf{x}_0), \eta(\mathbf{x}_1), \dots, \eta(\mathbf{x}_{n_r-1}))^T$, given a set of output realizations $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n_r-1})^T$ and an evaluation function $\eta(\cdot)$.

Based solely on a single replication, the cumulative sums (CUSUM) plot is a graphical evaluation of convergence of the McMC, which was proposed by *Yu and Mykland (1998)* and 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), $\mathbf{x} = (\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_{n_r-1})^T$, and an evaluation function, $\eta(\mathbf{x}) = (\eta(\mathbf{x}_0), \eta(\mathbf{x}_1), \dots, \eta(\mathbf{x}_{n_r-1}))^T$, one can construct CUSUM path plots of scalar summary statistic as follows,

- (1) Calculate the mean of the evaluation function $\bar{\eta} = \frac{1}{n_r} \sum_{i=0}^{n_r-1} \eta(\mathbf{x}_i)$;
- (2) Calculate the CUSUM $\sigma_t = \sum_{i=0}^t (\eta(\mathbf{x}_i) - \bar{\eta})$, $t \in [0, n_r)$, and $\sigma_{n_r} = 0$;
- (3) Define a delta function δ_i , $i \in [1, n_r)$ as follows: if $(\sigma_{i-1} - \sigma_i)(\sigma_i - \sigma_{i+1}) < 0$, then $\delta_i = 1$; else $\delta_i = 0$;
- (4) Calculate the hairiness indices $\Sigma_t = \frac{1}{t-1} \sum_{i=1}^{t-1} \delta_i$, $t \in [2, n_r)$.

The key outputs are two sequential plots: $\boldsymbol{\sigma} = (\sigma_0, \sigma_1, \dots, \sigma_{n_r})^T$ and $\boldsymbol{\Sigma} = (\Sigma_0, \Sigma_1, \dots, \Sigma_{n_r})^T$. The CUSUM, σ_t , gives a subjective evaluation of convergence performance of the chain since the mixing rate is reflected by the variance CUSUMs over blocks of the sequence (*Lin, 1992; Brooks, 1998*). A slowly mixing sequence will lead to a high variance for σ_t and a relatively

large excursion size before returning to 0 at n_r . When the mixing of the chain is high, the graph of σ is highly irregular (oscillatory or “fractal”) and concentrates around 0. When the mixing is slow, the CUSUM path is smooth and has a bigger excursion size. The hairiness index, Σ_t , presents a quantitative measure of smoothness to evaluate the convergence performance of a chain. An ideal convergence sequence will be centered at around 0.5.

3 An Illustrative Example

3.1 Reference models

A synthetic two-dimensional aquifer is presented to demonstrate the efficiency and effectiveness of the BMcMC method. Figure 2(A) shows the spatial distribution of the reference 16×16 [cells] $\ln K$ field generated by LUSIM (*Deutsch and Journel, 1998*), having a multi-Gaussian characteristic. The mean and standard deviation of the $\ln K$ are set to zero and one [md^{-1}], respectively. The spatial correlation structure is specified as an exponential function with a correlation length $\lambda_{\mathbf{x}}$ equal to 16 [cells], which means that the field is highly correlated. An obvious pattern of the $\ln K$ distribution is that the lower values, mainly located near the left-lower corner, are spread along the north-west direction while the higher values are mostly located around the right-lower corner and the left-upper corner.

An assumed steady flow experiment is performed on this aquifer. The left and right boundary conditions are set as a prescribed head with fixed values equal to ten and zero [m], respectively. Both the upper and the lower bounds are set as non-flow. A block-centered finite-difference method is employed to solve the flow problem and the resulting head field is shown in the Figure 2(C). Nine head observations as the conditioning data set of state variables are sampled from the head field as shown in Figure 2(D). Nine $\ln K$ observations as hard data to be conditioned are also sampled from the $\ln K$ field at the same locations as head data, which are shown in Figure 2(B).

3.2 An inverse-conditioning experiment

The objective of the inverse-conditional simulation is to generate *i.i.d* realizations inverse-conditional to the nine head samples using the proposed BMcMC method. Assume that the flow boundary conditions and the mean and covariance of $\ln K$ are correctly observed, i.e. $\mu = 0$, $\sigma_x^2 = 1.0$, and $\lambda_{\mathbf{x}} = \lambda_x = \lambda_y = 16$ [cells]. The relative measurement error of state variables is set as 0.1, i.e. $\sigma_y^2 = 0.1$. Figure 3 shows a typical BMcMC realization of $\ln K$ and the corresponding head field. Both fields reveal some of main patterns of reference models.

Figure 4 shows the most probable estimate which is obtained by averaging over 3000 BMcMC realizations. The estimated $\ln K$ distribution shown in Figure 4(A) well matches the distribution of the reference $\ln K$ field: the lower values, mainly locating at the left-lower corner, spread along the north-west direction while the higher values locate mainly at the right-lower corner, although the reproduction of the left-upper corner is not that so clear.

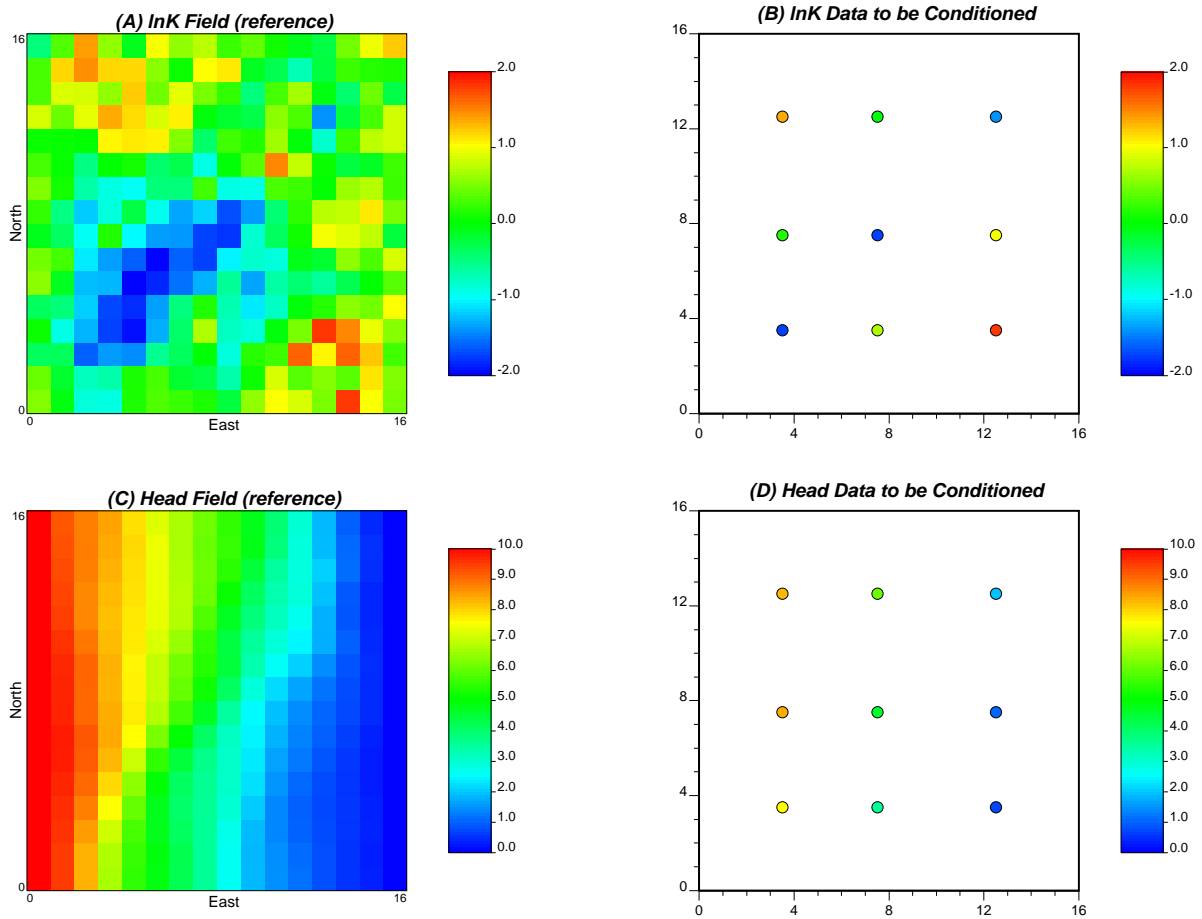


Figure 2: Conditioning data sets and reference fields: (A) reference $\ln K$ field, (B) nine hard data ($\ln K$), (C) reference head field, and (D) nine head data. The unit of hydraulic conductivity and head is in $[md^{-1}]$ and $[m]$, respectively.

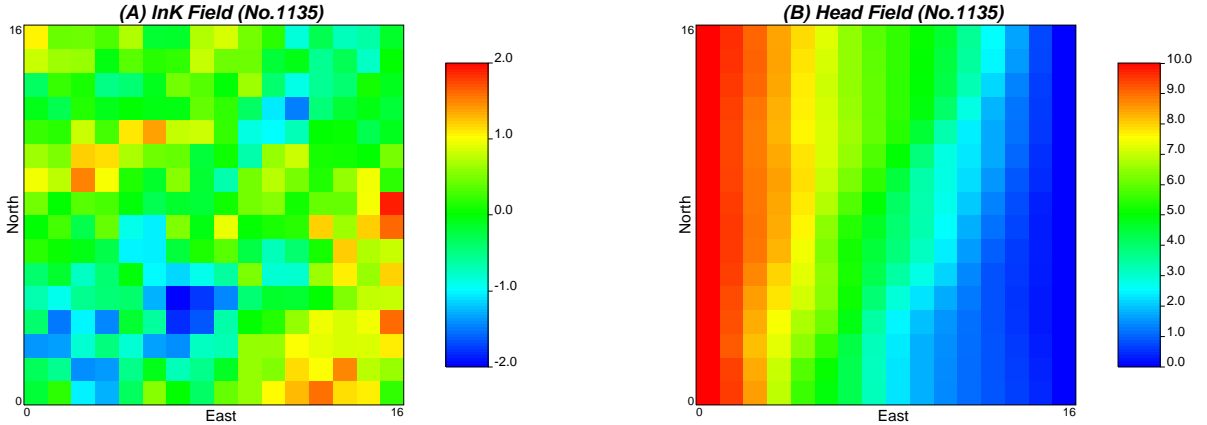


Figure 3: A typical BMcMC realization (A) and the simulated head field (B) from (A): The unit of hydraulic conductivity and head is in $[md^{-1}]$ and $[m]$, respectively.

The precision of such estimate is plotted in Figure 4(B). One can find that the regions close to the non-flow boundaries have the most uncertainties. The mean head field shown in Figure 4(C) reasonably reflects the main characteristics of the reference head field (Figure 2(C)). The distribution of head uncertainty as plotted in Figure 4(D) also demonstrates that the regions near the non-flow boundaries have the most uncertainties.

The realizations generated closely follow the prior configuration for model parameters. For example, the experimental mean and variance of the $\ln K$ are -0.13 and 0.98 , respectively (Figure 5(A)), which approximate to the corresponding prior specifications, i.e. $\mu = 0$ and $\sigma_x^2 = 1$, and those of the reference field, i.e. $\mu = -0.05$ and $\sigma_x^2 = 0.84$. Figure 5 shows that the $\ln K$ distribution obeys on the Gaussian assumption and the squared head mismatch follows the log-Gaussian assumption as implied in Equation (2)-(3).

In summary, head observations may contain important information for the identification of conductivity distribution. The BMcMC scheme is effective in performing stochastic simulation inverse-conditioning to head observations.

3.3 Factors that affect the performance of BMcMC

Two important factors are considered in this section to examine their influences on the performance of the BMcMC method: the correlation length (λ_x) and whether or not conditioning on the hard data ($\mathbf{x}|\mathbf{x}_1$). Several performance measures considered in this study include the reproduction of reference models, the acceptance rate, and the mixing of Markov chain.

Four scenarios of BMcMC simulations with distinct configurations are run on the basis of the same conditioning data set as above. The most probable estimates are plotted in Figure 6. Compared to the larger prior correlation length (Figure 6(A) and (C)), the magnitude of

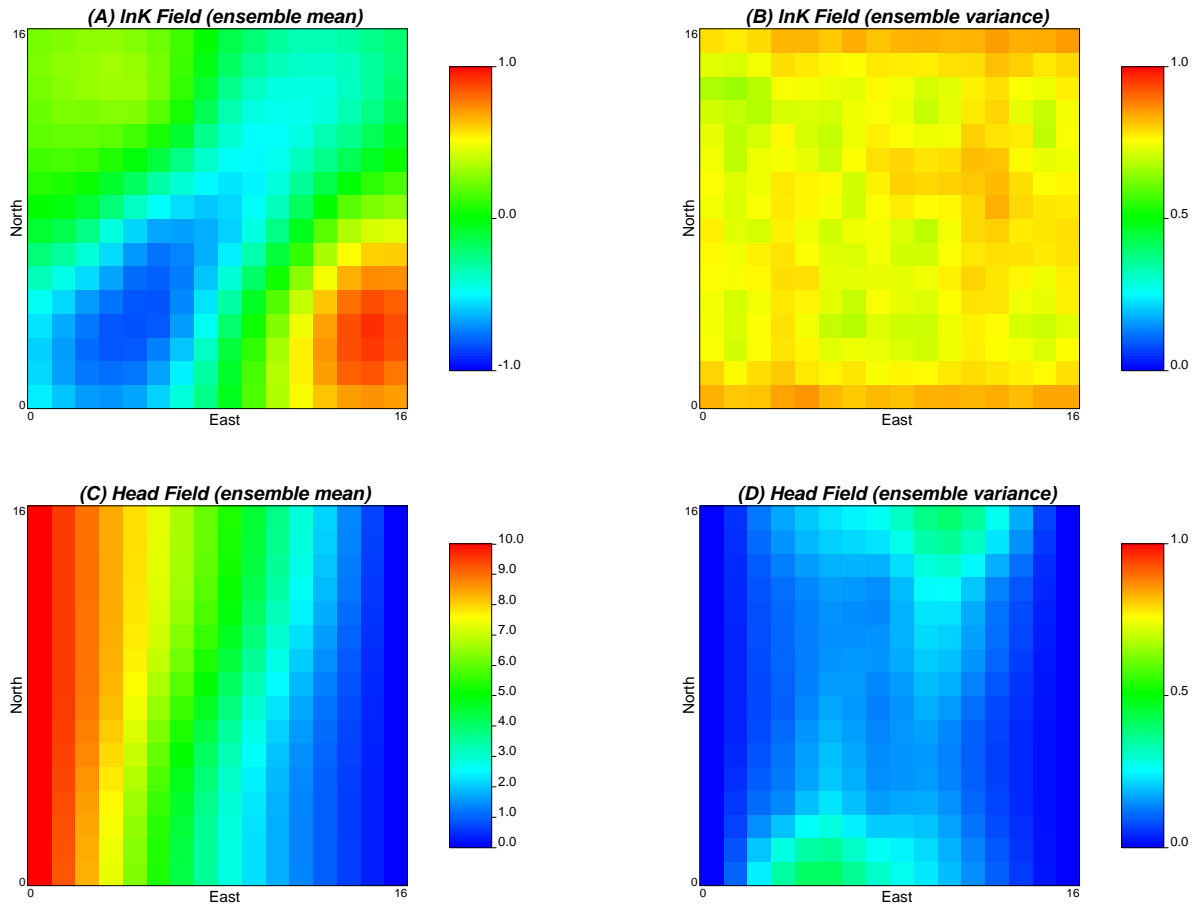


Figure 4: A BMcMC estimate by inverse-conditioning to head observations: the simulated mean $\ln K$ field (A) and its variance (B), the simulated mean head field (C) and its variance (D). The unit of hydraulic conductivity and head is in $[md^{-1}]$ and $[m]$, respectively.

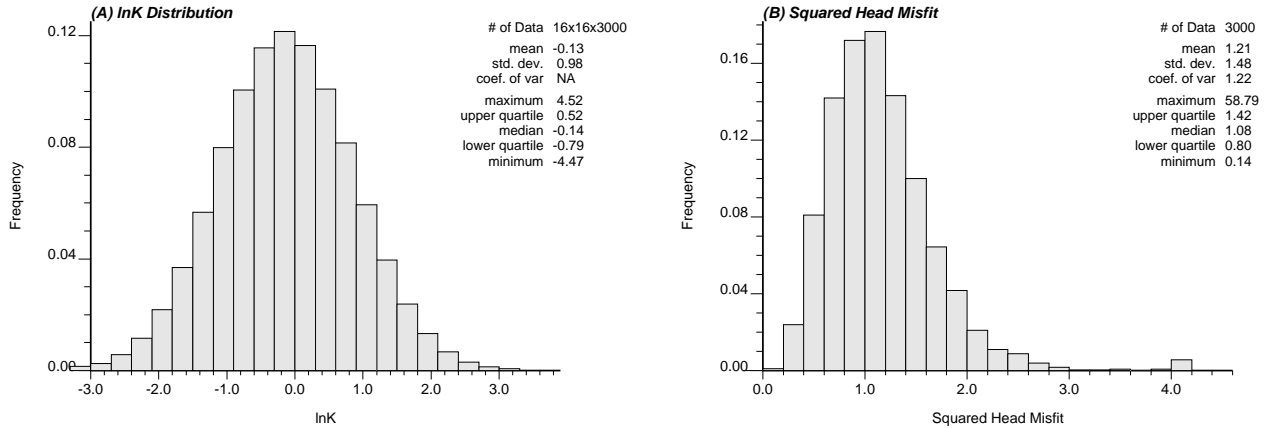


Figure 5: $\ln K$ [md^{-1}] distribution (A) and squared head [m] misfit (B)

$\ln K$ is underestimated by the smaller prior correlation length for both the unconditional and the conditional cases as plotted in Figure 6(B) and (D), respectively. The identified mean $\ln K$ field for the conditional case does contain the contributions from both the kriging and inverse-conditional estimate.

Figure 7 compares the effects of various BMcMC configurations on the performance of Markov chain: (A)-(D) display the sequential plots of four chains with different BMcMC configurations after the chains are convergent; (E) and (F) compare the CUSUM plots of these four chains and their hairiness indices, respectively. The BMcMC parameter configurations are marked in each figures. The mean squared head mismatch in Figure 7(A)-(D) is calculated by averaging the 3000 independent realizations. One can find that the effect of the prior correlation length is minor: in the unconditional case, the correct correlation length ($\lambda_x = 16$) yields a slightly better mean squared head mismatch (as shown in Figure 7(A) and (B)) but a slightly worse mixing of chain (a larger excursion size in Figure 7(E) but better hairiness indices in Figure 7(F)); in the conditional case, the shorter correlation length produces better results (from Figure 7(C) through (F)). However, conditioning to hard data has considerable influence on the performance of Markov chain. The error of mean squared head misfit obviously decreases, for example, by comparing Figure 7(C) to (A), and the chain mixes more rapidly, i.e. a smaller excursion size and a larger hairiness indices.

In addition, the acceptance rate is quite sensitive to the prior correlation length and whether or not conditioning on hard data. Table 1 lists 14 scenarios of stochastic simulations and their BMcMC configurations. One can easily find that conditioning to hard data significantly enhances the acceptance rate (e.g. by comparing Scenario 9-14 to Scenario 3-8). The effect of correlation length is quite complicated: for unconditional case, a correct correlation length helps increase the acceptance rate; but for the conditional case, the correct correlation length decreases the acceptance rate.

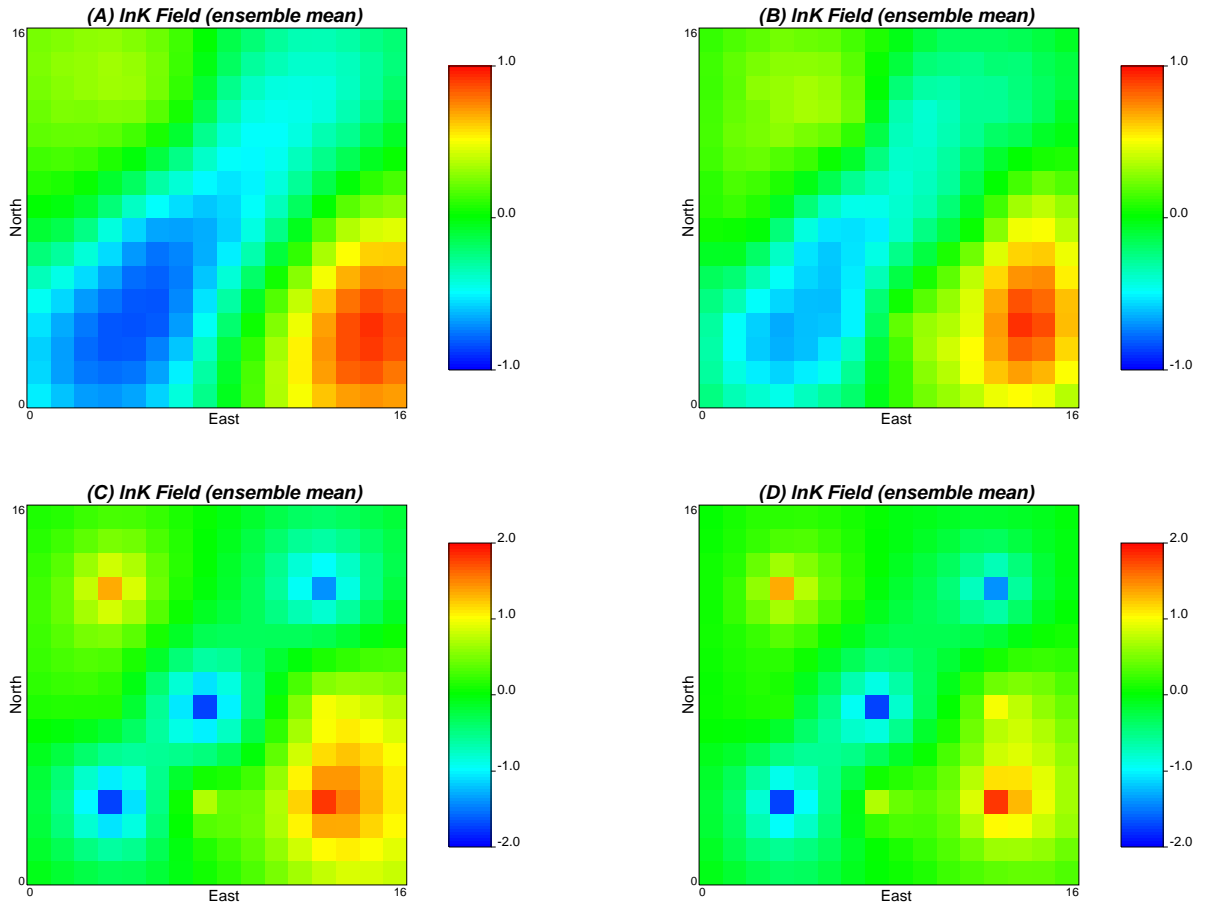


Figure 6: Factors affecting the reproduction of reference models: (A) $\mathbf{x}|\mathbf{y}, \lambda_{\mathbf{x}} = 16$, (B) $\mathbf{x}|\mathbf{y}, \lambda_{\mathbf{x}} = 4$, (C) $\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \lambda_{\mathbf{x}} = 16$, and (D) $\mathbf{x}|\mathbf{x}_1, \mathbf{y}, \lambda_{\mathbf{x}} = 4$

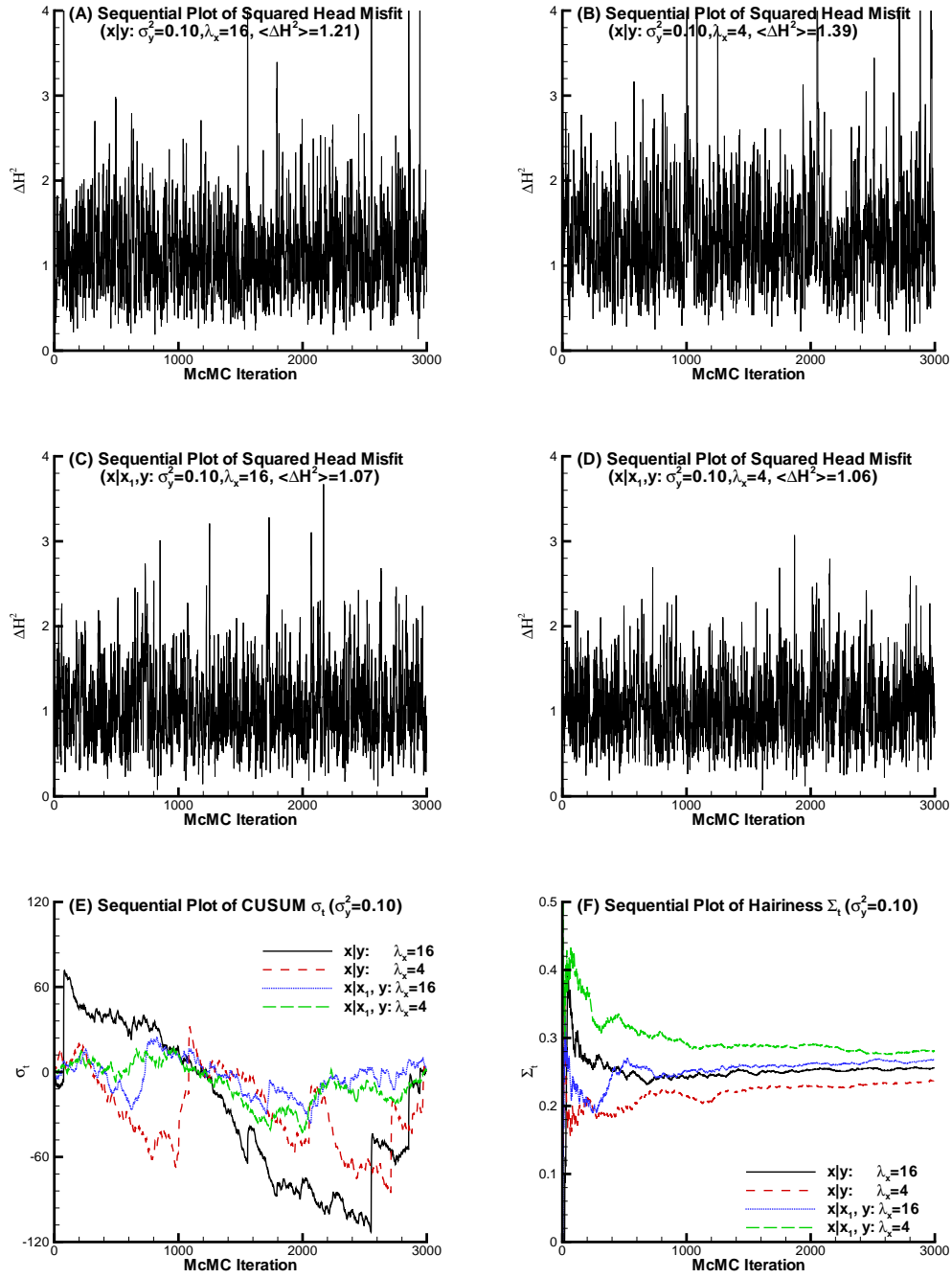


Figure 7: Factors affecting the performance of Markov chain: (A)-(D) show the sequential plots of four chains ((A) $x|y, \lambda_x = 16$, (B) $x|y, \lambda_x = 4$, (C) $x|x_1, y, \lambda_x = 16$, and (D) $x|x_1, y, \lambda_x = 4$); (E) and (F) display the CUSUM plots of these four chains and their hairiness indices, respectively.

Table 1: Summary of stochastic simulations, BMcMC configurations and performance indices

Scenario	\mathbf{x}	$\lambda_{\mathbf{x}}$	σ_y^2	$\alpha(\%)$	$\langle \Delta y^2 \rangle$	Σ_t
1	$\mathbf{x} -$	16				
2	$\mathbf{x} \mathbf{x}_1$	16				
3	$\mathbf{x} \mathbf{y}$	16	0.10	0.1273	1.21	0.256
4	$\mathbf{x} \mathbf{y}$	16	0.05	0.0175	1.24	0.047
5	$\mathbf{x} \mathbf{y}$	16	0.02	0.0030	2.75	0.101
6	$\mathbf{x} \mathbf{y}$	4	0.10	0.0903	1.39	0.238
7	$\mathbf{x} \mathbf{y}$	4	0.05	0.0106	1.34	0.072
8	$\mathbf{x} \mathbf{y}$	4	0.02	0.0028	2.38	0.108
9	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	16	0.10	0.8381	1.07	0.268
10	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	16	0.05	0.1128	0.63	0.247
11	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	16	0.02	0.0082	0.64	0.049
12	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	4	0.10	1.4622	1.06	0.281
13	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	4	0.05	0.1985	0.65	0.264
14	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	4	0.02	0.0109	0.52	0.059

4 Uncertainty Assessment

On the basis of the synthetic example presented above, the uncertainty reduction obtained by conditioning and inverse-conditioning is assessed quantitatively in terms of several measures. Two types of uncertainties are considered in this section: model uncertainty and response uncertainty. The model uncertainty is important at the spatiotemporal scale: not only because the models generated form the basis for future performance prediction at the existing wells but also because they are helpful for risk evaluation in locating new wells. The response uncertainty directly measures the prediction ability of models at the time scale.

4.1 Model uncertainty

Although the reference model is well defined and observable in this study, we generally do not know what it is ahead in practice. A practical way is to use the ensemble average of simulated outputs instead of the real model. Two parameters are computed as the metrics of performance measure to this end, the ensemble average error ($I(\mathbf{x})_1$) and the standard deviation of the ensemble average error ($I(\mathbf{x})_2$), which are defined as the L_1 -norm and L_2 -norm between the simulated models and the mean models, respectively,

$$I(\mathbf{x})_1 = \|\mathbf{x}_{sim} - \bar{\mathbf{x}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r} - \bar{x}_i|, \quad (16)$$

$$I(\mathbf{x})_2^2 = \|\mathbf{x}_{sim} - \bar{\mathbf{x}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r} - \bar{x}_i)^2, \quad (17)$$

where n_r is the number of realizations, n_{xyz} is the number of grid cells, \mathbf{x}_{sim} is the vector of simulated attribute values, and $\bar{\mathbf{x}}_{sim}$ is the ensemble average vector of simulated attribute values.

In the synthetic example considered in this study, however, the model uncertainty can be measured by the simulated errors to validate the efficiency of the proposed method since the real model is available (e.g. *Deng et al.*, 1993). In such case, the L_1 -norm and L_2 -norm between the simulated models and the real models are defined as,

$$I(\mathbf{x})_3 = \|\mathbf{x}_{sim} - \mathbf{x}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r}^{sim} - x_i^{ref}|, \quad (18)$$

$$I(\mathbf{x})_4^2 = \|\mathbf{x}_{sim} - \mathbf{x}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r}^{sim} - x_i^{ref})^2, \quad (19)$$

respectively. Note that \mathbf{x}_{ref} is the vector of reference attribute values. Obviously, the smaller $I(\mathbf{x})_3$ and $I(\mathbf{x})_4$ are, the closer to the real model the generated realizations are.

4.2 Uncertainty of model responses

A method to examining the effect of conditioning to head data on the uncertainty reduction of the spatial distribution of hydraulic conductivity is to examine the decrease of the L_1 -norm and L_2 -norm of the predicted model responses (*Hoeksema and Kitanidis*, 1984; *Kitanidis*, 1986). The four metrics for the model responses are defined,

$$I(\mathbf{y})_1 = \|\mathbf{y}_{sim} - \bar{\mathbf{y}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r} - \bar{y}_i|, \quad (20)$$

$$I(\mathbf{y})_2^2 = \|\mathbf{y}_{sim} - \bar{\mathbf{y}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r} - \bar{y}_i)^2, \quad (21)$$

$$I(\mathbf{y})_3 = \|\mathbf{y}_{sim} - \mathbf{y}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r}^{sim} - y_i^{ref}|, \quad (22)$$

$$I(\mathbf{y})_4^2 = \|\mathbf{y}_{sim} - \mathbf{y}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r}^{sim} - y_i^{ref})^2. \quad (23)$$

In essence, $I(\mathbf{x})_1$, $I(\mathbf{x})_2$, $I(\mathbf{y})_1$, and $I(\mathbf{y})_2$ measure the degree of precision that the MCMC simulations could render, that is, how narrow the confidence interval of MCMC simulations is.

$I(\mathbf{x})_3$, $I(\mathbf{x})_4$, $I(\mathbf{y})_3$, and $I(\mathbf{y})_4$ measure the degree of accuracy that the MCMC simulations may attain, that is, how they are close to the true model and its response. From the standpoint of estimate and uncertainty, $I(\mathbf{x})_1$, $I(\mathbf{x})_3$, $I(\mathbf{y})_1$, and $I(\mathbf{y})_3$ measure the reliability of the estimated models and their responses while $I(\mathbf{x})_2$, $I(\mathbf{x})_4$, $I(\mathbf{y})_2$, and $I(\mathbf{y})_4$ measure the uncertainty of the estimates and their responses.

4.3 Results

Totally fourteen scenarios of numerical experiments (see Table 1) are carried out to systematically evaluate the uncertainty reduction obtained by conditioning the hydraulic conductivity field on $\ln K$ (hard) data and head data.

Table 2 summarizes the model uncertainty reduction in terms of four model metrics, $I(\mathbf{x})_1$, $I(\mathbf{x})_2$, $I(\mathbf{x})_3$, and $I(\mathbf{x})_4$. First, one can observe that the real uncertainties indicated by $I(\mathbf{x})_1$ and $I(\mathbf{x})_2$ are uniformly larger than the estimated uncertainties represented by $I(\mathbf{x})_3$ and $I(\mathbf{x})_4$. This indicates that the model uncertainty is always underestimated by the BMCMC simulation. However, the latter two metrics can be used as a substitute of the former two for uncertainty assessment since they completely reflect the common rules that the former two reveal. Second, as the measurement errors of state variables decrease, the model accuracy does not increase much.

Table 2: Mean absolute error and mean variance of $\ln K$

		\mathbf{x}	$\mathbf{x} \mathbf{x}_1$	$\mathbf{x} \mathbf{y}$	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	$\mathbf{x} \mathbf{y}$	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$
σ_y^2	$I(\mathbf{x})$	$\lambda_{\mathbf{x}} = 16$	$\lambda_{\mathbf{x}} = 16$	$\lambda_{\mathbf{x}} = 16$	$\lambda_{\mathbf{x}} = 16$	$\lambda_{\mathbf{x}} = 4$	$\lambda_{\mathbf{x}} = 4$
0.10	$I(\mathbf{x})_1$	1.0454	0.8089	0.8686	0.7287	0.9259	0.8684
0.10	$I(\mathbf{x})_2^2$	1.3072	1.0120	1.0839	0.9216	1.1570	1.1048
0.10	$I(\mathbf{x})_3$	0.7978	0.6205	0.6930	0.4838	0.7525	0.7026
0.10	$I(\mathbf{x})_4^2$	0.9992	0.7929	0.8690	0.6238	0.9439	0.8977
0.05	$I(\mathbf{x})_1$	1.0454	0.8089	0.8623	0.7174	0.9222	0.8631
0.05	$I(\mathbf{x})_2^2$	1.3072	1.0120	1.0757	0.9072	1.1531	1.0985
0.05	$I(\mathbf{x})_3$	0.7978	0.6205	0.6871	0.4743	0.7517	0.6961
0.05	$I(\mathbf{x})_4^2$	0.9992	0.7929	0.8615	0.6119	0.9437	0.8900
0.02	$I(\mathbf{x})_1$	1.0454	0.8089	0.8766	0.7156	0.9375	0.8583
0.02	$I(\mathbf{x})_2^2$	1.3072	1.0120	1.0954	0.9056	1.1726	1.0924
0.02	$I(\mathbf{x})_3$	0.7978	0.6205	0.7040	0.4718	0.7648	0.6925
0.02	$I(\mathbf{x})_4^2$	0.9992	0.7929	0.8850	0.6095	0.9602	0.8865
	Ranking	6	2	3	1	5	4

Third, the bottom line in Table 2 ranks the importance of BMCMC configurations on the model uncertainty reduction. Compared to the unconditional case, conditional simulations do reduce the model uncertainty. The case with a correct specification of model structure,

e.g. $\lambda_{\mathbf{x}} = 16$, conditioning to $\ln K$ and inverse-conditioning to head data reduces the model uncertainty to the largest extent. Conditioning to $\ln K$ reduces the uncertainty more efficiently than inverse-conditioning to head data, which reveals that direct measurements of $\ln K$ are more effective for the model parameter estimation than indirect measurements of dependent state variables. However, a correct configuration of the spatial model, e.g. the correlation length $\lambda_{\mathbf{x}}$, plays the most important role in the model uncertainty reduction. Due to the wrong specification of the correlation length, i.e. $\lambda_{\mathbf{x}} = 4$, the conditional and inverse-conditional result only ranks the fourth, while the correct one ranks the first. In summary, the correct configuration for model structure plays a crucial role in model estimation. The direct measurement of parameters plays a more important role than the indirect measurement of dependent state data for the model parameter estimation.

Table 3 summarizes the uncertainty reduction of model response in terms of four response metrics, $I(\mathbf{y})_1$, $I(\mathbf{y})_2$, $I(\mathbf{y})_3$, and $I(\mathbf{y})_4$. First, just as the observation for the model uncertainty, $I(\mathbf{y})_1$ and $I(\mathbf{y})_2$ are generally larger than $I(\mathbf{y})_3$ and $I(\mathbf{y})_4$, which also shows that the response uncertainty is always underestimated by the BMcMC simulation. Second, as the measurement errors of state variables decrease, the model accuracy does not increase as much. Third, the bottom line in Table 3 ranks the importance of BMcMC configurations on uncertainty reduction of model response. Again, conditional and/or inverse-conditional simulations reduce the response uncertainty. The case that has a correct specification for model structure, conditioning to $\ln K$ and inverse-conditioning to head data reduces the response uncertainty to the largest extent. However, different from the model uncertainty, inverse-conditioning to head data reduces the uncertainty more efficiently than conditioning to $\ln K$, which indicates that indirect measurements of dependent state variables are more helpful to reduce the response uncertainty than direct measurements of $\ln K$. Moreover, a correct configuration of the spatial covariance plays a less important role in reducing the response uncertainty. In summary, the response uncertainty is more sensitive to the measurements of model response than those of model parameters.

It is worth mentioning that, even though the configurations for model parameter are not correct, the realizations generated well reproduce the crucial patterns of reference models (e.g. Figure 6 (B) and (D)) and both the model uncertainty and the response uncertainty are reduced by inverse-conditioning and/or conditioning (Table 2 and Table 3), which proves the robustness of the proposed BMcMC method.

5 Summary

Aiming at preserving spatial structure for inverse stochastic simulation, a blocking Markov chain Monte Carlo method is presented to generate *i.i.d* realizations which are conditional to hard data and inverse-conditional to state data. The proposal kernel for the construction of Markov chain is an appropriate (blocking) approximation to the target posterior distribution in the Bayesian framework. The generation of candidate realizations is very fast on the basis of the LU-decomposition of the covariance matrix. A numerical experiment on a synthetic

Table 3: Mean absolute error and mean variance of predicted head

σ_y^2	\mathbf{x}	$\mathbf{x} \mathbf{x}_1$	$\mathbf{x} \mathbf{y}$	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	$\mathbf{x} \mathbf{y}$	$\mathbf{x} \mathbf{x}_1, \mathbf{y}$	
$I(\mathbf{y})$	$\lambda_x = 16$	$\lambda_x = 16$	$\lambda_x = 16$	$\lambda_x = 16$	$\lambda_x = 4$	$\lambda_x = 4$	
0.10	$I(\mathbf{y})_1$	0.9220	0.4943	0.3072	0.2725	0.3583	0.3049
0.10	$I(\mathbf{y})_2^2$	1.3217	0.7192	0.4347	0.3815	0.5053	0.4289
0.10	$I(\mathbf{y})_3$	0.6806	0.3836	0.2524	0.2252	0.2834	0.2624
0.10	$I(\mathbf{y})_4^2$	0.9516	0.5525	0.3580	0.3160	0.3992	0.3692
0.05	$I(\mathbf{y})_1$	0.9220	0.4943	0.2946	0.2383	0.3446	0.2771
0.05	$I(\mathbf{y})_2^2$	1.3217	0.7192	0.4453	0.3363	0.5054	0.3951
0.05	$I(\mathbf{y})_3$	0.6806	0.3836	0.2436	0.1983	0.2825	0.2393
0.05	$I(\mathbf{y})_4^2$	0.9516	0.5525	0.3749	0.2797	0.4121	0.3400
0.02	$I(\mathbf{y})_1$	0.9220	0.4943	0.3616	0.2317	0.3945	0.2621
0.02	$I(\mathbf{y})_2^2$	1.3217	0.7192	0.5885	0.3425	0.5957	0.3846
0.02	$I(\mathbf{y})_3$	0.6806	0.3836	0.3146	0.1923	0.3265	0.2280
0.02	$I(\mathbf{y})_4^2$	0.9516	0.5525	0.5103	0.2881	0.4838	0.3327
	Ranking	6	5	3	1	4	2

aquifer is carried out to demonstrate the efficiency of the proposed method in performing inverse-conditional simulation. The performance of this method is also widely evaluated based on the synthetic aquifer. Numerical experiments show that the degree of heterogeneity of aquifer, e.g. correlation length (λ_x), and whether or not conditioning on the hard data ($\mathbf{x}|\mathbf{x}_1$) may play important roles in the results of inverse-conditional simulations and also affect the performance of BMcMC itself.

The model uncertainty and the response uncertainty are also assessed. Both types of uncertainties are reduced due to conditioning on hard data and/or inverse-conditioning on state data. However, their influences on the uncertainty reduction may have different weights. As for the effect of the piezometric head (state data) and the hydraulic conductivity (hard data) upon the model uncertainty of $\ln K$, the measurement of $\ln K$ plays a major role in reducing such uncertainty compared to the head data. Conditioning to head data, however, does improve the model estimation of $\ln K$ compared to conditioning to $\ln K$ solely. This conclusion is completely consistent with the result of *Dagan* (1985). The measurement of head is informative on the large-scale configuration of $\ln K$. As for the effect of the piezometric head and the hydraulic conductivity upon the prediction uncertainty of head distribution, the measurement of head plays a major role in reducing such uncertainty compared to that of $\ln K$. Although the prediction of head is quite insensitive to local $\ln K$ values, the joint conditioning to $\ln K$ does improve the prediction of head compared to inverse-conditioning to head data solely. The local measurement of $\ln K$ does not carry too much information about the spatial trend of $\ln K$.

However, the efficiency and applicability of the proposed BMcMC method for conditional

and inverse-conditional simulation suffers from several shortcomings. First, the fast generation of proposal kernel is totally based on the LU-decomposition of the covariance matrix which makes it very limited in dealing with the high-resolution cases. Second, the computation of the likelihood is extremely time-consuming since the forward simulation $g(\mathbf{x})$ is CPU expensive. These two bottlenecks limit this method to the extensive application in practice for the conditional and inverse-conditional simulation. In addition, this study only shows the capability of the proposed BMcMC method conditioning on head data. Actually, the BMcMC method has a great flexibility in incorporating various data from different sources, e.g. the concentration data. An improved version of BMcMC method is expected to handle with the high-resolution case for the integration of various data.

Acknowledgements

A doctoral fellowship and an extra travel grant to the first author by the Universitat Politècnica de València (UPV), Spain, is gratefully acknowledged.

References

- [1] Alabert, F., 1987. The practice of fast conditional simulations through the LU decomposition of the covariance matrix, *Mathematical Geology*, **19**(5), 369-386.
- [2] Brooks, S.P., 1998. Quantitative convergence assessment for Markov chain Monte Carlo via cusums, *Statistics and Computing*, **8**(3), 267-274.
- [3] Carrera, J., and S.P. Neuman, 1986. Estimation of aquifer parameters under transient and steady state conditions, 1. Maximum likelihood method incorporating prior information, *Water Resources Research*, **22**(2), 199-210.
- [4] Dagan, G., 1985. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem, *Water Resources Research*, **21**(1), 65-72.
- [5] Deng, F.W., J.H. Cushman, and J.W. Delleur, 1993. Adaptive estimation of the log fluctuating conductivity from tracer data at the Cape Cod site, *Water Resources Research*, **29**(12), 4011-4018.
- [6] Deutsch, C.V., and A.G. Journel, 1998. *GSLIB: Geostatistical software library and user's guide*, second edition, Oxford University Press, pp369.
- [7] Davis, M.W., 1987. Production of conditional simulations via the LU triangular decomposition of the covariance matrix, *Mathematical Geology*, **19**(2), 91-98.
- [8] Geman, S., and D. Geman, 1984. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE transactions on Pattern Analysis and Machine Intelligence*, **6**(6), 721-741.

- [9] Gomez-Hernandez, J.J., A. Sahuquillo, and J.E. Capilla, 1997. Stochastic simulation of transmissivity fields conditional to both transmissivity and piezometric data: I. Theory, *Journal of Hydrology*, **203**, 162-174.
- [10] Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their application, *Biometrika*, **57**(1), 97-109.
- [11] Hoeksema, R.J., and P.K. Kitanidis, 1984. An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resources Research*, **20**(7), 1003-1020.
- [12] Kitanidis, P.K., 1986. Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resources Research*, **22**(4), 499-507.
- [13] Kitanidis, P.K., 1996. On the geostatistical approach to the inverse problem, *Advances in Water Resources*, **19**(6), 333-342.
- [14] Lin, Z.Y., 1992. On the increments of partial sums of a α -mixing sequence, *Theoretical Probability and its Applications*, **36**, 316-326.
- [15] Liu, J.S., 1996. Metropolized independent sampling with comparisons to rejection sampling and importance sampling, *Statistics and Computing*, **6**(2), 113-119.
- [16] McLaughlin, D., and L.R. Townley, 1996. A reassessment of the groundwater inverse problem, *Water Resources Research*, **32**(5), 1131-1161.
- [17] Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, 1953. Equations of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**(3), 1087-1092.
- [18] Oliver, D.S., L.B. Cunha, and A.C. Reynolds, 1997. Markov chain Monte Carlo methods for conditioning a log-permeability field to pressure data, *Mathematical Geology*, **29**(1), 61-91.
- [19] RamaRao, B.S., A.M. LaVenue, G. de Marsily, and M.G. Marietta, 1995. Pilot point methodology for automated calibration of an ensemble of conditionally simulated transmissivity fields: 1. Theory and computational experiments, *Water Resources Research*, **31**(3), 475-493.
- [20] Robert, C.P., and G. Casella, 1999. *Monte Carlo Statistical Methods*, Springer-Verlag, New York, pp. 507.
- [21] Roberts, G.O., and S.K. Sahu, 1997. Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler, *Journal of Royal Statistical Society B*, **59**(2), 291-317.

- [22] Yu, B., and P. Mykland, 1998. Looking at Markov samplers through cusum path plots: a simple diagnostic idea, *Statistics and Computing*, **8**(3), 275-286.
- [23] Zimmerman, D.A., G. de Marsily, C.A. Gotway, M.G. Marietta, C.L. Axness, R.L. Beauheim, R.L. Bras, J. Carrera, G. Dagan, P.B. Davies, D.P. Gallegos, A. Galli, J. Gomez-Hernandez, P. Grindrod, A.L. Gutjahr, P.K. Kitanidis, A.M. Lavenue, D. McLaughlin, S.P. Neuman, B.S. RamaRao, C. Ravenne, and Y. Rubin, 1998. A comparison of seven geostatistically based inverse approaches to estimate transmissivities for modeling advective transport by groundwater flow, *Water Resources Research*, **34**(6), 1373-1413.

6 An LU-decomposition-based Sampler

The BMcMC computation needs a fast sampler to generate a large number of candidate realizations. The joint prior density of a multi-Gaussian random field is,

$$\pi(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} \|\mathbf{C}_x\|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{C}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\},$$

where $\pi(\mathbf{x})$ denotes the prior pdf of $\mathbf{x} \in R^n$; n is the length of the vector \mathbf{x} ; $\boldsymbol{\mu} \in R^n$ is the prior mean of the random field; and $\mathbf{C}_x \in R^{n \times n}$ is the positive-definite covariance matrix of the vector \mathbf{x} . Note that \mathbf{x} may be partly observed, say, $\mathbf{x}_{obs} \in R^m$, but seldom fully known, i.e. $m < n$. In such case, the sample is called a conditional simulation on linear hard data, i.e. $\mathbf{x}_2 | \mathbf{x}_1$, where, $\mathbf{x}_1 = \mathbf{x}_{obs} \in R^m$, $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$, and $\mathbf{x}_2 \in R^{n-m}$.

The objective is to draw randomly *i.i.d* realizations from the distribution $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{C}_x)$. The LU-decomposition algorithm is quite efficient and effective in generating a large number of realizations as required by the BMcMC computation since the LU-decomposition of the covariance matrix can be done once for all (Davis, 1987; Alabert, 1987). The simulated results are rather more precise and accurate than some of others, e.g. the sequential simulation algorithm.

6.1 Unconditional sampler

Sample Algorithm 1. Unconditional sample $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{C}_x)$, where $\mathbf{x}, \boldsymbol{\mu} \in R^n$, and $\mathbf{C}_x \in R^{n \times n}$:

- (1) Cholesky decompose $\mathbf{C}_x = \mathbf{L}\mathbf{L}^T$, where $\mathbf{L} \in R^{n \times n}$;
- (2) Randomly draw $\mathbf{z} \sim N(\mathbf{0}, \mathbf{1}) \in R^n$;
- (3) Calculate $\mathbf{v} = \mathbf{L}\mathbf{z} \in R^n$;
- (4) Generate an unconditional sample $\mathbf{x} = \boldsymbol{\mu} + \mathbf{v}$.

Note that the step 1 is only needed to be done once, which takes most of computational time. More realizations can be obtained by repeating from the step 2 to the step 4. The computational effort lies in the matrix-vector multiplication, i.e. \mathbf{Lz} , as in the step 3.

6.2 Conditional sampler

The joint distribution of $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T$ is,

$$\begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \right),$$

where $\mathbf{x}_1 \subset R^m$ is the (normalized) conditioning dataset; $\mathbf{x}_2 \subset R^{n-m}$ is the (normalized) conditional simulated values; $\mathbf{C}_{11} \subset R^{m \times m}$ is the data-to-data covariance matrix; $\mathbf{C}_{22} \subset R^{(n-m) \times (n-m)}$ is the unknowns-to-unknowns covariance matrix; and $\mathbf{C}_{21} = \mathbf{C}_{12}^T$ is the unknowns-to-data covariance matrix, $\mathbf{C}_{21} \subset R^{(n-m) \times m}$. It can be shown that the expected value of \mathbf{x}_2 is $\boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$, which is known as the simple kriging estimate, and the covariance matrix of \mathbf{x}_2 is $\mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$. Therefore, the conditional realizations can be drawn from $\mathbf{x}_2 \sim N(\boldsymbol{\mu}^*, \mathbf{C}^*)$, where $\boldsymbol{\mu}^* = \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ and $\mathbf{C}^* = \mathbf{C}_{22} - \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}$.

The covariance matrix for all n grid nodes including m conditioning data can be decomposed into the product of a lower triangular matrix and an upper one,

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} = \mathbf{LU} = \begin{bmatrix} \mathbf{L}_{11}\mathbf{U}_{11} & \mathbf{L}_{11}\mathbf{U}_{12} \\ \mathbf{L}_{21}\mathbf{U}_{11} & \mathbf{L}_{21}\mathbf{U}_{12} + \mathbf{L}_{22}\mathbf{U}_{22} \end{bmatrix}.$$

Therefore, \mathbf{L}_{11} , \mathbf{L}_{21} , and \mathbf{L}_{22} can be obtained by $\mathbf{C}_{11} = \mathbf{L}_{11}\mathbf{U}_{11}$, $\mathbf{C}_{21} = \mathbf{L}_{21}\mathbf{U}_{11}$, and $\mathbf{C}_{22} = \mathbf{L}_{21}\mathbf{U}_{12} + \mathbf{L}_{22}\mathbf{U}_{22}$, respectively. Note that the matrix multiplication, the matrix minus and the LU-decomposition are involved in the procedure of calculating the lower triangle matrices.

A conditional realization \mathbf{x} is obtained by the multiplication of \mathbf{L} with a column vector $\mathbf{z} \subset R^n$,

$$\begin{pmatrix} \mathbf{x}_1 - \boldsymbol{\mu}_1 \\ \mathbf{x}_2 - \boldsymbol{\mu}_2 \end{pmatrix} = \mathbf{Lz} = \begin{bmatrix} \mathbf{L}_{11} & \mathbf{0} \\ \mathbf{L}_{21} & \mathbf{L}_{22} \end{bmatrix} \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{pmatrix},$$

where the sub-vector $\mathbf{z}_1 \subset R^m$ is set as $\mathbf{z}_1 = \mathbf{L}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ and the sub-vector $\mathbf{z}_2 \subset R^{n-m}$ consists of the $n - m$ independent standard normal deviates. Therefore, a conditional realization can be obtained by,

$$\mathbf{x}_2 = \boldsymbol{\mu}_2 + \mathbf{L}_{21}\mathbf{z}_1 + \mathbf{L}_{22}\mathbf{z}_2 = \boldsymbol{\mu}_2 + \mathbf{L}_{21}\mathbf{L}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + \mathbf{L}_{22}\mathbf{z}_2.$$

Sample Algorithm 2. Conditional sample $\mathbf{x}_2|\mathbf{x}_1$, where the unknowns $\mathbf{x}_2 \subset R^{n-m}$ and the hard data $\mathbf{x}_1 \subset R^m$, and $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{C}_x)$, in which $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)^T \subset R^n$, $\boldsymbol{\mu} \subset R^n$, and $\mathbf{C}_x \subset R^{n \times n}$:

- (1) Calculate $\mathbf{L}_{11} \subset R^{m \times m}$ from $\mathbf{C}_{11} = \mathbf{L}_{11}\mathbf{L}_{11}^T$;

- (2) Calculate $\mathbf{L}_{21} \in R^{(n-m) \times m}$ from $\mathbf{C}_{21} = \mathbf{L}_{21} \mathbf{L}_{11}^T$;
- (3) Calculate $\mathbf{L}_{22} \in R^{(n-m) \times (n-m)}$ from $\mathbf{C}_{22} = \mathbf{L}_{21} \mathbf{U}_{12} + \mathbf{L}_{22} \mathbf{L}_{22}^T$;
- (4) Calculate the simple kriging field $\mathbf{v}_1 = \mathbf{L}_{21} \mathbf{L}_{11}^{-1} (\mathbf{x}_1 - \boldsymbol{\mu}_1) \in R^{n-m}$;
- (5) Randomly draw $\mathbf{z}_2 \sim N(\mathbf{0}, \mathbf{1}) \in R^{n-m}$;
- (6) Calculate $\mathbf{v}_2 = \mathbf{L}_{22} \mathbf{z}_2 \in R^{n-m}$;
- (7) Generate a conditional sample $\mathbf{x}_2 = \boldsymbol{\mu}_2 + \mathbf{v}_1 + \mathbf{v}_2$.

Note that the step 1 through the step 4 are only needed to be done once which consumes the largest part of the computational efforts of this algorithm. More realizations can be obtained by repeating from the step 5 to the step 7. The computational effort focuses on the matrix-vector multiplication, i.e. $\mathbf{L}_{22} \mathbf{z}_2$, as in the step 6.