A Markov Chain Monte Carlo Method for Inverse Stochastic Modeling and Uncertainty Assessment

Doctoral Thesis submitted by
Jianlin Fu

Advisor:
J. Jaime Gómez-Hernández

A Markov Chain Monte Carlo Method for Inverse Stochastic Modeling and Uncertainty Assessment

PhD Thesis submitted by
Jianlin Fu

Advisor:
J. Jaime Gómez-Hernández

Departamento de Ingeniería Hidráulica y Medio Ambiente
Universidad Politécnica de Valencia
Valencia, Spain

December 2007

Life is like a box of chocolates.
You never know what you gonna get.
$\langle\langle$*Forest Gump*$\rangle\rangle$

Life is like a Markov chain.
You never know where you gonna go.
(?)

Abstract

Unlike the traditional two-stage methods, a conditional and inverse-conditional simulation approach may directly generate independent, identically distributed ($i.i.d$) realizations to honor both static data and state data in one step. The Markov chain Monte Carlo (McMC) method was proved a powerful tool to perform such type of stochastic simulation. One of the main advantages of the McMC over the traditional sensitivity-based optimization methods to inverse problems is its power, flexibility and well-posedness in incorporating observation data from different sources. In this work, an improved version of the McMC method is presented to perform the stochastic simulation of reservoirs and aquifers in the framework of multi-Gaussian geostatistics.

First, a blocking scheme is proposed to overcome the limitations of the classic single-component Metropolis-Hastings-type McMC. One of the main characteristics of the blocking McMC (BMcMC) scheme is that, depending on the inconsistence between the prior model and the reality, it can preserve the prior spatial structure and statistics as users specified. At the same time, it improves the mixing of the Markov chain and hence enhances the computational efficiency of the McMC. Furthermore, the exploration ability and the mixing speed of McMC are efficiently improved by coupling the multiscale proposals, i.e., the coupled multiscale McMC method. In order to make the BMcMC method capable of dealing with the high-dimensional cases, a multiscale scheme is introduced to accelerate the computation of the likelihood which greatly improves the computational efficiency of the McMC due to the fact that most of the computational efforts are spent on the forward simulations. To this end, a flexible-grid full-tensor finite-difference simulator, which is widely compatible with the outputs from various upscaling subroutines, is developed to solve the flow equations and a constant-displacement random-walk particle-tracking method, which enhances the computational efficiency at different scales, is employed to solve the transport problems.

Second, the usefulness and efficiency of the proposed method are validated by a synthetic example. The uncertainty reduction due to conditioning on various types of data from different sources is assessed with the aid of the synthetic example. One of the novel achievements in this work is that the physical models are constrained to the temporal moments of BTCs that are

more easily accessible than the concentration data which are only sparsely distributed in space. The worth on uncertainty reduction is evaluated by comparing to other data sources.

Third, by comparing the BMcMC to the ensemble Kalman filtering (EnKF), the importance of honoring the prior information for inverse stochastic modeling is ascertained in two synthetic examples. Numerical simulations show that, even though the EnKF method may efficiently provide a better reproduction of observed dynamic data than the BMcMC method, the preservation of spatial statistics and model structure makes the BMcMC simulations competitive for some cases in predicting accurately and reliably the future performance of reservoirs particularly at new well locations. This is because the spatial structure and statistics of models may be one of the most important error sources to the prediction of the future performance of reservoirs and aquifers, it should be consistent with the given information just as conditioning to linear data and inverse-conditioning to nonlinear data. In other words, the realizations generated should preserve the given spatial structure and statistics during the procedure of conditioning and inverse-conditioning.

Resumen

La recopilación de observaciones de la altura piezométrica y las medidas de la conductividad hidráulica local (o transmisividad) proporcionan una inestimable información para identificar el patrón espacial de los parámetros en acuíferos, incluso los caminos de flujo o barreras de flujo, y para reducir la incertidumbre de los modelos de acuíferos. Para obtener dicha información de conectividad a partir de las medidas y cuantificar la incertidumbre con exactitud, el método Monte Carlo es normalmente utilizado para generar un gran número de realizaciones de los parámetros de acuíferos condicionados a datos duros (conductividad) e inversamente condicionados a los datos de estado (altura piezométrica). No obstante, la simulación inversa condicionada de los parámetros de acuíferos es computacionalmente muy pesada, ya que implica una optimización no lineal del problema para generar cada una de las realizaciones inversas condicionadas. En contraste con algunos de los optimizadores no lineales clásicos y buscadores de algoritmos, en este estudio se presenta un esquema de cadena Markov Monte Carlo (McMC) para generar realizaciones condicionadas multi-Gaussianas, muestreando directamente de una distribución posterior que incorpora información a priori y observaciones a posteriori en un esquema Bayesiano. Lo que hace de este método bastante eficiente en la exploración del espacio de los parámetros del modelo es que el núcleo propuesto es una aproximación apropiada a la distribución del objetivo posterior y que la generación de realizaciones de candidatos es muy rápida debido a la descomposición LU de la matriz de covarianza. Las realizaciones generadas de esta forma no están únicamente condicionadas por los datos duros sino que también tienen la estructura espacial esperada. El funcionamiento del esquema McMC propuesto es ampliamente evaluado mediante un ejemplo sintético que simula el caso de flujo por gradiente natural. La propagación de incertidumbre debida al mapeado condicionado e inverso condicionado de los modelos para acuíferos es entonces cuantificada en términos estadísticos de tiempos de llegada, resolviendo los estados estacionarios de flujo asumidos y los problemas de transporte conservativo ideal. La reducción en la incertidumbre de la predicción, implica no sólo el valor de la altura piezométrica, sino también el significado de los momentos temporales y las estadísticas de conectividad en el mapeado de parámetros de acuíferos.

Una representación adecuada de la variación espacial detallada de los parámetros superficiales requiere modelos de acuíferos de alta resolución. La caracterización precisa de estos modelos a gran escala en un método Monte Carlo recurre típicamente a una simulación estocástica capaz de condicionar los datos duros (ej. conductividad) y los datos de estado dependientes (ej. altura piezométrica, concentración, etc.), conocida como modelación condicionada e inversa condicionada respectivamente. Se ha comprobado que un esquema de cadena Markov Monte Carlo (McMC) resulta efectivo y eficiente para llevar a cabo este tipo de simulaciones condicionada e inversa condicionada, muestreando directamente en una distribución posterior que incorpora la información previa y las observaciones posteriores en un marco de trabajo Bayesiano. A pesar de esto, la utilidad de los métodos McMC previamente mencionados, se debe a la limitada capacidad de la descomposición LU de la matriz de covarianza en desacuerdo con los casos de alta resolución. En este estudio se presenta un nuevo esquema McMC para generar realizaciones condicionadas multi-Gaussianas de alta resolución. Lo que hace de este método muy eficiente en la exploración de los parámetros espaciales de modelos de elevadas dimensiones, es que el núcleo propuesto es una aproximación apropiada para la distribución posterior del objetivo seleccionado y que la generación de realizaciones de candidatos está basada en la descomposición espectral de la matriz covarianza con el fin de aumentar la velocidad de la transformada de Fourier. Las realizaciones generadas de esta forma, no sólo están condicionadas por el registro de conductividad, la altura piezométrica y los momentos temporales de la concentración de soluto, sino que también tienen la estructura espacial esperada. La propagación de incertidumbre debida al mapeado condicionado e inverso condicionado de los modelos para acuíferos también es cuantificada.

Resum

La recopilació d'observacions de la càrrega piezomètrica i les mesures de la conductivitat hidràulica local (o transmissivitat) proporcionen una inestimable informació per a identificar el patró espacial dels paràmetres en aqüífers, inclús els camins de fluix o barreres de fluix, i per a reduir la incertesa dels models d'aqüífers. Per a obtindre la dita informació de conectivitat a partir de les mesures i quantificar la incertesa amb exactitud, el mètode Monte Carlo és normalment utilitzat per a generar un gran nombre de realitzacions dels paràmetres d'aqüífers condicionats a dades durs (conductivitat) i inversament condicionats a les dades d'estat (càrrega piezomètrica). No obstant això, la simulació inversa condicionada dels paràmetres d'aqüífers és computacionalment molt pesada, ja que implica una optimació no lineal del problema per a generar cada una de les realitzacions inverses condicionades. En contrast amb alguns dels optimadors no lineals clàssics i buscadors d'algoritmes, en este estudi es presenta un esquema de cadena Markov Monte Carlo (McMC) per a generar realitzacions condicionades multi-gaussianes, mostrejant directament d'una distribució posterior que incorpora informació a priori i observacions a posteriori en un esquema Bayesiano. El que fa d'este mètode prou eficient en l'exploració de l'espai dels paràmetres del model és que el nucli proposat és una aproximació apropiada a la distribució de l'objectiu posterior i que la generació de realitzacions de candidats és molt ràpida a causa de la descomposició LU de la matriu de covarianza. Les realitzacions generades d'esta forma no estan únicament condicionades per les dades durs sinó que també tenen l'estructura espacial esperada. El funcionament de l'esquema McMC proposat és àmpliament avaluat per mitjà d'un exemple sintètic que simula el cas de fluix per gradient natural. La propagació d'incertesa deguda al mapeado condicionat i invers condicionat dels models per a aqüífers és llavors quantificada en termes estadístics de temps d'arribada, resolent els estats estacionaris de fluix assumits i els problemes de transport conservatiu ideal. La reducció en la incertesa de la predicció, implica no sols el valor de la càrrega piezomètrica, sinó també el significat dels moments temporals i les estadístiques de conectivitat en el mapeado de paràmetres d'aqüífers.

Una representació adequada de la variació espacial detallada dels paràmetres superficials requerix models d'aqüífers d'alta resolució. La caracterització pre-

cisa d'estos models a gran escala en un mètode Monte Carlo recorre típicament a una simulació estocàstica capaç de condicionar les dades durs (ex. conductivitat) i les dades d'estat dependents (ex. càrrega piezomètrica, concentració, etc.), coneguda com modelació condicionada i inversa condicionada respectivament. S'ha comprovat que un esquema de cadena Markov Monte Carlo (McMC) resulta efectiu i eficient per a portar a terme este tipus de simulacions condicionada i inversa condicionada, mostrejant directament en una distribució posterior que incorpora la informació prèvia i les observacions posteriors en un marc de treball Bayesiano. A pesar d'açò, la utilitat dels mètodes McMC prèviament mencionats, es deu a la limitada capacitat de la descomposició LU de la matriu de covarianza en desacord amb els casos d'alta resolució. En este estudi es presenta un nou esquema McMC per a generar realitzacions condicionades multi-gaussianes d'alta resolució. El que fa d'este mètode molt eficient en l'exploració dels paràmetres espacials de models d'elevades dimensions, és que el nucli proposat és una aproximació apropiada per a la distribució posterior de l'objectiu seleccionat i que la generació de realitzacions de candidats està basada en la descomposició espectral de la matriu covarianza a fi d'augmentar la velocitat de la transformada de Fourier. Les realitzacions generades d'esta forma, no sols estan condicionades pel registro de conductivitat, la càrrega piezomètrica i els moments temporals de la concentració de soluto, sinó que també tenen l'estructura espacial esperada. La propagació d'incertesa deguda al mapeado condicionat i invers condicionat dels models per a aqüífers també és quantificada.

Acknowledgements

First, I wish to thank my advisor *J. Jaime Gomez-Hernandez* for accepting me into his research group and providing me the opportunity to carry out this study. I also should thank *Jaime* for his instruction in academia and for finding me the financial support during my stay. A doctoral fellowship and an extra travel grant to the author by the Universidad Politecnica de Valencia (UPV), Spain, is gratefully acknowledged.

I would like to express my thanks and gratitude to *Dr. Daniel Fernandez-Garcia* for valuable inputs and stimulating discussions in the development of the forward simulators. He was so patient in helping everyone deal with various problems that I benefited greatly from his energy and drive. One researcher cannot expect to find a better coworker and advisor than him. *Carl Axness*, *Andres Sahuquillo*, *Jose Capilla* and *Eduardo Cassiraga* also played important roles in leading me into the groundwater field. Especially, *Eduardo* helped me in the presentation of several figures. *Teresa Martinez*, *Javier Moreno* and other secretaries from the department helped me so much in preparing various documents. A special thank should go to *Prof. Henning Omre* and his group for their hospitality during my visit in Trondheim, Norway, in 2006. The EnKF algorithm and the case studies in Chapter 5 were developed during that time. *Henning* carefully read the early manuscript of this part and provided some valuable comments and suggestions.

I would never forget to say thanks to *Carolina Guardiona* who was so kind to me and found me a nice house before I came to Valencia on 2002, although maybe she did not know how much it meant to me at that time when I did not speak Spanish at all. I need to thank *Rafael Aliaga* who picked me up at the airport when I arrived at Valencia the very first day and organized a welcome party for me. He was such a funny actor that always brought laughing into the office and made us full of joys. I really appreciate *Rafa* for his patience in helping me configure the computers and network. I should say thanks to *Carlos*, *Gero*, and the rest of my colleagues in the research group and the department whose names are not listed here but in my heart.

Luis Borras was so patient to help me in preparing several Spanish materials. *Josue Chanona* and *Josep Ribes* also helped me a lot especially in learning Spanish when we lived together. I cannot forget the happy parties

with them and their group. From November 2005, it turned out to be a memorable period to share a house with *Fran* and *Edu* who were so kind and helpful at that time. I was so lucky to meet all of them who made my life in Valencia so wonderful and enjoyable. I always appreciate the friendship and supports that *Junfa Liu* and *Jugang Peng* provided to me. I really do not know how I can pay back to you even though I search so hard.

Finally, I would like to thank my wife *Ying* and our families. Without their support, I really cannot finish my study in Spain. I owe them so much. This thesis is dedicated to them.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Motivation and Objectives

Since the physical parameters in subsurface vary in a highly non-deterministic and unpredictable manner in space, a stochastic simulation approach is often adopted to quantify this kind of uncertainty. As for a complicated system, such uncertainty analysis will have to resort to a numerical simulation which is typically costly in computation. The Monte Carlo method is an indispensable tool for solving this type of difficult computational problem. Suppose there is a process generating a random vector $\boldsymbol{x}$ and we wish to compute $E[f(\boldsymbol{x})]$ and $Var[f(\boldsymbol{x})]$ given a function $f(\boldsymbol{x})$. Suppose further the random vector is associated with a probability density function $p(\boldsymbol{x})$. In this study, the terms "density" and "distribution" are interchangeably used when referring to the mechanism for generating a random process. The expectation and the variance may be written as,

$$E[f(\boldsymbol{x})] = \int f(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}, \tag{1.1}$$

$$Var[f(\boldsymbol{x})] = E[(f(\boldsymbol{x}) - E[f(\boldsymbol{x})])^2], \tag{1.2}$$

respectively.

In the communities of petroleum engineering and groundwater, the main objective of stochastic simulation is to build up a large number of independent, identically distributed $(i.i.d)$ reservoir models $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_{n_r-1})^T$ from a

target distribution $\pi(\boldsymbol{x})$ as input to assessment systems of the uncertainty of reservoir behavior, e.g., $f(\boldsymbol{x}) = (f(\boldsymbol{x}_0), f(\boldsymbol{x}_1), ..., f(\boldsymbol{x}_{n_r-1}))^T$, where n_r is the number of reservoir models or realizations. The average estimate of reservoir behavior can be approximated by,

$$\hat{\mu}_f = \frac{1}{n_r} \sum_{i=0}^{n_r-1} f(\boldsymbol{x}_i), \tag{1.3}$$

which represents the maximum probable estimate and the corresponding variance is,

$$\hat{\sigma}_f^2 = \frac{1}{n_r} \sum_{i=0}^{n_r-1} (f(\boldsymbol{x}_i) - \hat{\mu}_f)^2, \tag{1.4}$$

which measures the error scope of the maximum probable estimate.

However, this set of *i.i.d* realizations should be constrained by all information available, even though incomplete, to exhaust the resources that experiments provide and reduce the uncertainty to the greatest extent so that they can better reflect the underground reality of reservoir and aquifer. These information may include some prior concepts about models (e.g., those from the experts' subjective imagination on the basis of outcrops, sedimentary petrography and other geological and geophysical information) and all posterior objective observations (e.g., those from measurements in the field). It is also possible that the measurement data are available at different scales with different precisions.

Basically in mathematics, there are two kinds of measurements being used to constrain the stochastic models: one is linear and the other is nonlinear. The linear data may include hard data $\boldsymbol{x}_1$ on the measurement scale typically with a higher precision, such as permeability measured at the core scale or derived from well-logging or well-test data, and soft data $\underline{\boldsymbol{x}}$ at a larger scale with a lower precision, such as seismic data which is beyond the scope of this work. There also is another type of special data available, i.e., the hyperparameters $\boldsymbol{\theta}$, which is a combination of the prior and posterior information, such as, variogram (including sill, hole effect, correlation length, anisotropy and principal direction, variogram type...), histogram and/or multi-point statistics, since they depend on both the actual observations and expertise's imagination. These data can be directly incorporated into the models by geostatistical tools currently available. The widely used non-conditioning and linear conditioning algorithms for sampling $\boldsymbol{x} \sim \pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{\theta})$ include: the LU-decomposition algorithm (*Davis*, 1987; *Alabert*, 1987), the sequential Gaussian simulation (*Gomez-Hernandez and Journel*, 1993), the sequential indicator simulation (*Gomez-Hernandez and Srivastava*, 1990), the p-field simulation, the simulated annealing algo-

rithm (*Deutsch and Journel*, 1998), the FFT-based spectral method (*Pardo-Iguzquiza and Chica-Olmo*, 1993; *Robin et al.*, 1993; *Gutjahr et al.*, 1997; *Ruan and McLaughlin*, 1998), etc.

The other type is the nonlinear data $\boldsymbol{y}$ which are typically time-dependent known as dynamic data, such as the pressure measurements and water cut history. In order to integrate nonlinear data into the models, an ill-posed inverse problem arises and typically has to be solved by some complicated optimization methods. In the groundwater community this procedure is called model calibration while in petroleum engineering it is called automatic history matching. In geostatistics it can be named as inverse-conditioning problem since an inverse procedure is involved.

The practical meaningfulness of such inverse-conditioning simulation is obvious. For example, in petroleum engineering, one of the main goals of the numerical modeling is to predict reservoir performance at the spatiotemporal scale in more details and with higher accuracy than is possible with simple techniques such as extrapolation. The heterogeneities of subsurface reservoirs and aquifers are so complex that there are no enough data to predict a future performance with complete confidence and accuracy. However, the validity of a physical model can be evaluated by calculating the past performance and comparing the calculated results to the actual field observations. If the misfit between the model and the observations is unacceptable, then the input model parameters must be adjusted until a satisfied consistence is obtained, which is known as history matching. History matching can be time-consuming, expensive, and frustrating, primarily because reservoir performance can be complex, responding to numerous interactions that, as a whole, may be difficult to comprehend. In general, the parameters to be constructed include permeability, porosity, geologic facies, fault, and boundary conditions. The dynamic observation data to be matched consist of pressure, WOR, GOR, gas/water ratio, fluid contact movement, water and gas arrival times, and fluid saturations measured in cores, well-logs, and chemical tracer tests.

Similarly, this topic is also found in the groundwater community. For example, groundwater flow and contaminant transport modeling has been used at many hazardous waste sites. Models may be used throughout all phases of the site investigation and remediation processes. The ability to reliably predict groundwater flow and contaminant transport is critical in planning, implementing and managing groundwater remediations. Those models should not violate specified constraints imposed on them. One type of important constraint is the collection of model responses in the real field. One can compare the observed model responses in the real system with those predicted by the models. The sought values of model parameters are those that will make the two sets of values of state variables identical. However, because the models are only an approximation of the real system, one should never expect these

two sets of values to be identical. Instead, the "best fit" or "optimal" (i.e., those values that make the predicted values and the measured ones sufficiently close to each other) between them must be sought according to some criterion. Several important model parameters to be identified include hydraulic conductivity, storage coefficients, dispersivity, retardation factor, mass transfer rate, aquifer boundary conditions, etc. The model responses that could be collected in the field are piezometric heads, concentrations of contaminants, travel times of tracers, etc.

The commonly used inverse-conditioning algorithms for sampling $\boldsymbol{x} \sim \pi(\boldsymbol{x}|\boldsymbol{y})$ include: the cokriging method (*Kitanidis and Vomvoris*, 1983), the maximum likelihood method (*Carrera and Neuman*, 1986), the randomized maximum likelihood method (*Oliver et al.*, 1996), the pilot point method (*RamaRao et al.*, 1995), the sequential self-calibration method (*Gomez-Hernandez et al.*, 1997), the simulated annealing method (*Datta-Gupta et al.*, 1995), the gradual deformation method (*Hu*, 2000), the ensemble Kalman filtering method (*Evensen*, 1994, 2003), etc. A detailed review on these algorithms is outside the scope of this work. Readers can refer to *Yeh* (1986), *McLaughlin and Townley* (1996) and *Zimmerman et al.* (1998) for more details. In general, these classical two-stage methods start from generation of a series of independent seed fields and then calibrate those fields to honor nonlinearly dependent state data from different sources.

The objective of this thesis is to develop a stochastic simulation algorithm to generating *i.i.d* realizations which not only honor the static hard and soft data and the dynamic state data but also have the specified spatial structure and statistics for models, i.e., $\boldsymbol{x} \sim \pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{\theta})$. Instead of the two-stage method, an alternative choice is to generate realizations in one step which are both conditioned on the linear data and inverse-conditioned on the nonlinear data. With the aid of the Bayesian theorem, a conditional and inverse-conditional simulation method can draw samples directly from the posterior distribution that incorporates both the prior information and the posterior observations. It should be pointed out that, although some of the two-stage methods are also based on the Bayesian theorem, the ways to use it are essentially different. The classic two-stage methods, e.g., the maximum *a posteriori* (MAP) method, try to build up parameter models by maximizing the posterior probability while the method presented in this thesis only weights the candidate realizations by their posterior densities and just samples those models with the maximum posterior probabilities.

Thanking to the pioneering works by *Omre and Tjelmeland* (1996) and *Oliver et al.* (1997), the Markov chain Monte Carlo (McMC) method has already been introduced into the petroleum engineering community to perform such conditional and inverse-conditional stochastic simulation and uncertainty assessment. The McMC outstands itself from other inverse-conditioning ap-

proaches because it completely circumvents the numerical instability or the ill-posedness problem due to the fact that the inverse-conditioning is a sampling procedure rather than an optimization step. But its efficiency deserves more improving since the McMC is extremely computationally demanding especially for the high-dimensional case. In this study, we propose an improved McMC method, called the blocking McMC (BMcMC), to enhance the computational efficiency. To enable the BMcMC capable of handling the high-resolution case, a multi-scale scheme is introduced to form an extended BMcMC version, the MsBMcMC. Beside the improvement in computational aspects, a more striking characteristic of the proposed MsBMcMC method is that the prior specification on models can be preserved owing to the introduction of the BMcMC scheme.

1.2 Thesis Organization

The thesis is organized as follows. This chapter gives some simple introduction on the motivation and objectives of this dissertation.

Chapter 2 provides a detailed description on the numerical implementation of the forward simulations which form the basis of the following inverse problems. A special emphasis is put on the scale problem which is a key factor to accelerate the forward computation. To this end, a flexible-grid full-tensor finite-difference simulator, which is widely compatible with the outputs from various upscaling subroutines, is developed to solve the flow equation and a constant-displacement random-walk particle-tracking method, which enhances the computational efficiency at different scales, is employed to solve the transport problems.

The following three chapters consist in three separate, self-contained papers which may have some contents in common for the sake of completeness. Chapter 3 introduces a blocking scheme into the classic Metropolis-Hastings-type McMC method in order to overcome the slow mixing of the Markov chain and better preserve the spatial structure of physical models. Aiming at enabling the BMcMC method to deal with the high-resolution cases, a multi-scale BMcMC (MsBMcMC) method is developed to efficiently perform the conditional and inverse-conditional stochastic simulation. The multi-scale scheme greatly accelerates the computation of the likelihood especially for the cases with small measurement errors of state variables. The generation of the proposal kernel is simply based on the FFT-based conditional and unconditional spectral sampler which makes the proposal of candidate realizations also fast even for the high-dimensional case.

Chapter 4 presents a complete assessment on uncertainty reduction owing to conditioning and inverse-conditioning on various types of data from differ-

ent sources by means of the proposed MsBMcMC method. One of the novel achievements is that the physical models are constraint to the temporal moments of chemical tracers. Their worth on uncertainty reduction is evaluated by comparing to other data sources.

In Chapter 5, the BMcMC method has been compared with the ensemble Kalman filtering method (EnKF). Two synthetic examples show that, even though the EnKF method may efficiently provide a better reproduction of observed dynamic data than the BMcMC method, the preservation of spatial statistics and model structure makes the BMcMC simulations competitive for some cases in predicting accurately and reliably the future performance of reservoirs particularly at new well locations. Therefore, including the prior information in the inverse stochastic simulation is of significance for accurate assessment of model uncertainties and response uncertainties if the prior parameter information effectively reflects the underground reality.

Finally, in Chapter 6, several important contributions and conclusions are summarized and some interesting topics are outlined for further investigations.

2
Forward Simulators

Abstract

Since reservoir forward simulations are extremely computationally intensive, an effective scheme is to reduce the dimension of reservoir (e.g., by upscaling) and run the simulations at a coarsen scale, which calls for the forward simulators capable of dealing with the flow and transport problems at various scales. For this purpose, a flexible-grid full-tensor finite-difference flow simulator, which is widely compatible with the outputs from various upscaling subroutines (*Wen and Gomez-Hernandez*, 1996b), is developed for the fast computation of the forward flow problem. This flow simulator has ample abilities to accept the input of aquifer models within irregular grids and the input $\ln K$ field could be a full tensor. A constant-displacement random-walk particle-tracking approach is employed to solve the transport equations quickly and accurately. In contrast to the scale-dependent constant-time-step scheme, the scale-independent constant-displacement scheme is implemented to calculate the travel time of particles aiming at the different scale's transport problems. With this scheme, the computation of travel times of particles is only done in specified steps within one cell, by which numerous computation times are saved in solving transport problems at the coarse scale and, moreover, the results are proved to be quite accurate compared to the constant time scheme (*Wen and Gomez-Hernandez*, 1996a). In summary, these two forward simulators developed in this work, i.e., the multi-scale-oriented flow and transport solvers, pay much attention on the scale problem and hence are especially suitable for the multi-scale McMC computation.

2.1 Flow and Transport Problems

The flow of an incompressible or slightly compressible fluid in saturated porous media is described by,

$$\nabla \cdot (\mathbf{K}\nabla h) + q = s_s \frac{\partial h}{\partial t}, \tag{2.1}$$

where $\mathbf{K} = \mathbf{K}(\mathbf{x})$ is the 3×3 hydraulic conductivity tensor, $[LT^{-1}]$, i.e.,

$$\mathbf{K} = \begin{bmatrix} K_{xx} & K_{xy} & 0 \\ K_{yx} & K_{yy} & 0 \\ 0 & 0 & K_{zz} \end{bmatrix};$$

$h = h(\mathbf{x}, t)$ is the piezometric head, $[L]$; $q = q(\mathbf{x}, t)$ is the source or sink term (positive if fluid is extracted from the reservoir), $[T^{-1}]$; $s_s = s_s(\mathbf{x})$ is the specific storage coefficient, $[L^{-1}]$; t is the time, $[T]$; $\nabla\cdot = (\frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z})$ is the divergence operator of a vector field; and $\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})^T$ is the gradient operator of a scalar field. For a single-phase flow involving both rock and fluid properties, the partial derivative equation can be rewritten as,

$$\nabla \cdot \left(\frac{\rho g}{\mu} \mathbf{k} \nabla h \right) + q = \rho g \phi c_t \frac{\partial h}{\partial t}, \tag{2.2}$$

where $\mathbf{k} = \mathbf{k}(\mathbf{x})$ is the 3×3 intrinsic permeability tensor, $[L^2]$; $h = h(\mathbf{x}, t) = \frac{p}{\rho g} + z$ is the piezometric head, $[L]$; z is the elevation, increasing upward, $[L]$; $c_t = c_b + \phi c_f$ is the total compressibility of the system, $[M^{-1}LT^2]$; c_b is the compressibility of the bulk porous medium, $[M^{-1}LT^2]$; c_f is the compressibility of the single-phase fluid, $[M^{-1}LT^2]$; ρ is the density of single-phase fluid, $[ML^{-3}]$; μ is the viscosity of single-phase fluid, $[ML^{-1}T^{-1}]$; $\phi(\mathbf{x})$ is the porosity, $[dimensionless]$; and g is the acceleration of gravity, $[LT^{-2}]$. Several important relations between those parameters are listed as follows,

$$\begin{aligned} \mathbf{K} &= \frac{\rho g}{\mu} \mathbf{k}, \\ h &= \frac{p}{\rho g} + z, \\ s_s &= \rho g \phi c_t. \end{aligned}$$

Following the Darcy's law, the interblock velocity field is calculated by,

$$\mathbf{q} = \phi \mathbf{v} = -\mathbf{K}\nabla h, \tag{2.3}$$

where $\mathbf{q} = \mathbf{q}(\mathbf{x}, t) = (q_{xx}(\mathbf{x}, t), q_{yy}(\mathbf{x}, t), q_{zz}(\mathbf{x}, t))^T$ is the 3×1 Darcian flux vector, $[LT^{-1}]$; $\phi(\mathbf{x})$ is the porosity of the porous medium, $[dimensionless]$;

and $\mathbf{v} = \mathbf{v}(\mathbf{x}, t) = (v_{xx}(\mathbf{x}, t), v_{yy}(\mathbf{x}, t), v_{zz}(\mathbf{x}, t))^T$ is the 3×1 pore fluid velocity vector, $[LT^{-1}]$.

The governing equation for three-dimensional advective-dispersive contaminant transport in ground water may be written as follows,

$$r\phi\frac{\partial c}{\partial t} + \nabla \cdot (\mathbf{q}c) - \nabla \cdot (\phi \mathbf{D} \nabla c) + \lambda r c = 0, \tag{2.4}$$

where $r = r(\mathbf{x})$ is the retardation factor, [dimensionless]; $c = c(\mathbf{x}, t)$ is the solute concentration, $[ML^{-1}]$ or $[ML^{-3}]$; $\mathbf{D} = \mathbf{D}(\mathbf{x}, t)$ is the 3×3 local hydrodynamic dispersion coefficient tensor, $[L^2T^{-1}]$; and λ is the first-order decay constant, $[T^{-1}]$.

The main workflow for forward problems is that, for a given aquifer with physical parameters, e.g., $\mathbf{K}(\mathbf{x})$, $s_s(\mathbf{x})$, $\phi(\mathbf{x})$, and $r(\mathbf{x})$, the flow problem with a set of assumed boundary conditions is firstly solved to obtain the piezometric head field $h(\mathbf{x}, t)$, then the corresponding velocity field $\mathbf{q}(\mathbf{x}, t)$ (so are $\mathbf{v}(\mathbf{x}, t)$ and $\mathbf{D}(\mathbf{x}, t)$) is established by applying the Darcy's law, and finally the concentration field $c(\mathbf{x}, t)$ is obtained by solving the transport problem.

2.2 Multi-scale-oriented Flow Simulator

2.2.1 Flow difference equation

Consider an unconstructed confined aquifer with a full conductivity tensor, the flow equation can be rewritten as,

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial h}{\partial x} + K_{xy}\frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial y}\left(K_{yx}\frac{\partial h}{\partial x} + K_{yy}\frac{\partial h}{\partial y}\right) + \frac{\partial}{\partial z}\left(K_{zz}\frac{\partial h}{\partial z}\right) + q = s_s\frac{\partial h}{\partial t}. \tag{2.5}$$

If the eleven-point block-centered finite-difference full-tensor scheme is used to solve the flow problem, the flow equation then can be discretized as follows,

$$\begin{aligned}
&\left[\left(K_{xx}\frac{\Delta h}{\Delta x} + K_{xy}\frac{\Delta h}{\Delta y}\right)s|_{i-\frac{1}{2},j,k} - \left(K_{xx}\frac{\Delta h}{\Delta x} + K_{xy}\frac{\Delta h}{\Delta y}\right)s|_{i+\frac{1}{2},j,k}\right] + \\
&\left[\left(K_{yx}\frac{\Delta h}{\Delta x} + K_{yy}\frac{\Delta h}{\Delta y}\right)s|_{i,j-\frac{1}{2},k} - \left(K_{yx}\frac{\Delta h}{\Delta x} + K_{yy}\frac{\Delta h}{\Delta y}\right)s|_{i,j+\frac{1}{2},k}\right] + \\
&\left[K_{zz}\frac{\Delta h}{\Delta z}s|_{i,j,k-\frac{1}{2}} - K_{zz}\frac{\Delta h}{\Delta z}s|_{i,j,k+\frac{1}{2}}\right] + q_{i,j,k} = \\
&s_{s,i,j,k}\frac{\Delta h_t}{\Delta t_t}\Delta x_{i,j,k}\Delta y_{i,j,k}\Delta z_{i,j,k},
\end{aligned} \tag{2.6}$$

where $i \in [0, n_x), j \in [0, n_y), k \in [0, n_z)$, and,

$$\left(K_{xx}\frac{\Delta h}{\Delta x}s\right)|_{i-\frac{1}{2},j,k} = (K_{xx}s)|_{i-\frac{1}{2},j,k}\frac{h_{i-1,j,k}-h_{i,j,k}}{\Delta x|_{i-\frac{1}{2}}},$$

$$\left(K_{xy}\frac{\Delta h}{\Delta y}s\right)|_{i-\frac{1}{2},j,k} = (K_{xy}s)|_{i-\frac{1}{2},j,k}\cdot$$
$$\left(\frac{\Delta x_{i-1}}{2\Delta x|_{i-\frac{1}{2}}}\frac{h_{i,j-1,k}-h_{i,j+1,k}}{\Delta y|_{j-\frac{1}{2}+\frac{1}{2}}} + \frac{\Delta x_i}{2\Delta x|_{i-\frac{1}{2}}}\frac{h_{i-1,j-1,k}-h_{i-1,j+1,k}}{\Delta y|_{j-\frac{1}{2}+\frac{1}{2}}}\right),$$

$$\left(K_{xx}\frac{\Delta h}{\Delta x}s\right)|_{i+\frac{1}{2},j,k} = (K_{xx}s)|_{i+\frac{1}{2},j,k}\frac{h_{i,j,k}-h_{i+1,j,k}}{\Delta x|_{i+\frac{1}{2}}},$$

$$\left(K_{xy}\frac{\Delta h}{\Delta y}s\right)|_{i+\frac{1}{2},j,k} = (K_{xy}s)|_{i+\frac{1}{2},j,k}\cdot$$
$$\left(\frac{\Delta x_i}{2\Delta x|_{i+\frac{1}{2}}}\frac{h_{i+1,j-1,k}-h_{i+1,j+1,k}}{\Delta y|_{j-\frac{1}{2}+\frac{1}{2}}} + \frac{\Delta x_{i+1}}{2\Delta x|_{i+\frac{1}{2}}}\frac{h_{i,j-1,k}-h_{i,j+1,k}}{\Delta y|_{j-\frac{1}{2}+\frac{1}{2}}}\right),$$

$$\left(K_{yx}\frac{\Delta h}{\Delta x}s\right)|_{i,j-\frac{1}{2},k} = (K_{yx}s)|_{i,j-\frac{1}{2},k}\cdot$$
$$\left(\frac{\Delta y_{j-1}}{2\Delta y|_{j-\frac{1}{2}}}\frac{h_{i-1,j,k}-h_{i+1,j,k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2}}} + \frac{\Delta y_j}{2\Delta y|_{j-\frac{1}{2}}}\frac{h_{i-1,j-1,k}-h_{i+1,j-1,k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2}}}\right),$$

$$\left(K_{yy}\frac{\Delta h}{\Delta y}s\right)|_{i,j-\frac{1}{2},k} = (K_{yy}s)|_{i,j-\frac{1}{2},k}\frac{h_{i,j-1,k}-h_{i,j,k}}{\Delta y|_{j-\frac{1}{2}}},$$

$$\left(K_{yx}\frac{\Delta h}{\Delta x}s\right)|_{i,j+\frac{1}{2},k} = (K_{yx}s)|_{i,j+\frac{1}{2},k}\cdot$$
$$\left(\frac{\Delta y_j}{2\Delta y|_{j+\frac{1}{2}}}\frac{h_{i-1,j+1,k}-h_{i+1,j+1,k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2}}} + \frac{\Delta y_{j+1}}{2\Delta y|_{j+\frac{1}{2}}}\frac{h_{i-1,j,k}-h_{i+1,j,k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2}}}\right),$$

$$\left(K_{yy}\frac{\Delta h}{\Delta y}s\right)|_{i,j+\frac{1}{2},k} = (K_{yy}s)|_{i,j+\frac{1}{2},k}\frac{h_{i,j,k}-h_{i,j+1,k}}{\Delta y|_{j+\frac{1}{2}}},$$

$$\left(K_{zz}\frac{\Delta h}{\Delta z}s\right)|_{i,j,k-\frac{1}{2}} = (K_{zz}s)|_{i,j,k-\frac{1}{2}}\frac{h_{i,j,k-1}-h_{i,j,k}}{\Delta z|_{k-\frac{1}{2}}},$$

$$\left(K_{zz}\frac{\Delta h}{\Delta z}s\right)|_{i,j,k+\frac{1}{2}} = (K_{zz}s)|_{i,j,k+\frac{1}{2}}\frac{h_{i,j,k}-h_{i,j,k+1}}{\Delta z|_{k+\frac{1}{2}}},$$

where s is the interface area between two adjacent cells. Note that for one cell, the six area sizes are possibly different in the case of a flexible grid. Conductance items $\mathbf{K}(\mathbf{x})$ as input data are defined at the interfaces between cells, $K_{xx}(\mathbf{x})$, $K_{xy}(\mathbf{x})$, $K_{yx}(\mathbf{x})$, $K_{yy}(\mathbf{x})$, and $K_{zz}(\mathbf{x})$, which can be obtained from the upscaling subroutines, the random field generators, or simply by harmonically averaging the permeability values at node centers. In the case that a scalar hydraulic conductivity field is given, the computation of the interblock conductance widely employs the weighted harmonic mean,

$$K_{\frac{1}{2}} = \frac{K_1 K_2(\Delta x_1 + \Delta x_2)}{K_1 \Delta x_2 + K_2 \Delta x_1}, \tag{2.7}$$

where Δx_1 and Δx_2 are the sizes of two adjacent blocks, and K_1 and K_2 are the block-centered conductance. Other parameters, such as $s_s(\mathbf{x})$, $\phi(\mathbf{x})$, $q(\mathbf{x})$, $\Delta x(\mathbf{x})$, $\Delta y(\mathbf{x})$, and $\Delta z(\mathbf{x})$, are totally defined at the center of grid-blocks.

Although the simulation time $t_i \in [t_0, t_e]$, $i \in [0, n_t]$, for one history event may be discretized in any way as given by the user in a file format, two common alternatives are also implemented in the code: one is the equal interval discretization and the other is the so-called time-step multiplier scheme. The first scheme just assigns an identical time increment to all time-steps,

$$\Delta t_i = \Delta t = \frac{1}{n_t}(t_e - t_0); \qquad i \in [0, n_t).$$

The latter assumes the time increment is multiplied by a constant time-step coefficient α, i.e.,

$$\Delta t_i = \alpha \Delta t_{i-1}; \qquad i \in (0, n_t),$$

and, given the starting and the ending simulation times for one history event, t_0 and t_e, the first time-step is calculated by,

$$\Delta t_0 = \frac{\alpha - 1}{\alpha^{n_t} - 1}(t_e - t_0).$$

Therefore, the simulation time is discretized as,

$$t_{i+1} = t_i + \Delta t_i; \qquad i \in [0, n_t). \tag{2.8}$$

The advantage of the second scheme is that it allows for an adequate time discretization at the early stage of simulation if $\alpha > 1$ such that the simulated transient head distribution is to the smallest degree influenced by the time discretization. Note that the simulation time steps are assumed to be identical for all history events in implementing these two schemes.

Employing an implicit time scheme which less suffers from the numerical instability caused by the error propagation during the successive simulation

times, the difference equation for an active cell (i, j, k) at the time $t_t \in (t_0, t_e]$ can be rearranged as,

$$
\begin{aligned}
&Z_m Y_c X_c \cdot h^t_{i,j,k-1} + \\
&Z_c Y_m X_m \cdot h^t_{i-1,j-1,k} + Z_c Y_m X_c \cdot h^t_{i,j-1,k} + Z_c Y_m X_p \cdot h^t_{i+1,j-1,k} + \\
&Z_c Y_c X_m \cdot h^t_{i-1,j,k} + Z_c Y_c X_c \cdot h^t_{i,j,k} + Z_c Y_c X_p \cdot h^t_{i+1,j,k} + \\
&Z_c Y_p X_m \cdot h^t_{i-1,j+1,k} + Z_c Y_p X_c \cdot h^t_{i,j+1,k} + Z_c Y_p X_p \cdot h^t_{i+1,j+1,k} + \\
&Z_p Y_c X_c \cdot h^t_{i,j,k+1} = -q_{i,j,k} - s_{s,i,j,k} \frac{h^{t-1}_{i,j,k}}{t_t - t_{t-1}} \Delta x_{i,j,k} \Delta y_{i,j,k} \Delta z_{i,j,k},
\end{aligned}
$$

where,

$$
Z_m Y_c X_c = \frac{(K_{zz} s)|_{i,j,k-\frac{1}{2}}}{\Delta z|_{k-\frac{1}{2}}},
$$

$$
\begin{aligned}
Z_c Y_m X_m &= \frac{\Delta x_{i,j,k}}{2\Delta x|_{i-\frac{1}{2}}} \frac{(K_{xy} s)|_{i-\frac{1}{2},j,k}}{\Delta y|_{i-1,j-\frac{1}{2}+\frac{1}{2},k}} + \frac{\Delta y_{i,j,k}}{2\Delta y|_{j-\frac{1}{2}}} \frac{(K_{yx} s)|_{i,j-\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2},j-1,k}}, \\
Z_c Y_m X_c &= \frac{\Delta x_{i-1,j,k}}{2\Delta x|_{i-\frac{1}{2}}} \frac{(K_{xy} s)|_{i-\frac{1}{2},j,k}}{\Delta y|_{i,j-\frac{1}{2}+\frac{1}{2},k}} - \frac{\Delta x_{i+1,j,k}}{2\Delta x|_{i+\frac{1}{2}}} \frac{(K_{xy} s)|_{i+\frac{1}{2},j,k}}{\Delta y|_{i,j-\frac{1}{2}+\frac{1}{2},k}} + \frac{(K_{yy} s)|_{i,j-\frac{1}{2},k}}{\Delta y|_{j-\frac{1}{2}}}, \\
Z_c Y_m X_p &= -\frac{\Delta x_{i,j,k}}{2\Delta x|_{i+\frac{1}{2}}} \frac{(K_{xy} s)|_{i+\frac{1}{2},j,k}}{\Delta y|_{i+1,j-\frac{1}{2}+\frac{1}{2},k}} - \frac{\Delta y_{i,j,k}}{2\Delta y|_{j-\frac{1}{2}}} \frac{(K_{yx} s)|_{i,j-\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2},j-1,k}},
\end{aligned}
$$

$$
Z_c Y_c X_m = \frac{(K_{xx} s)|_{i-\frac{1}{2},j,k}}{\Delta x|_{i-\frac{1}{2}}} + \frac{\Delta y_{i,j-1,k}}{2\Delta y|_{j-\frac{1}{2}}} \frac{(K_{yx} s)|_{i,j-\frac{1}{2},k}}{\Delta x|_{i+\frac{1}{2}+\frac{1}{2},j,k}} - \frac{\Delta y_{i,j+1,k}}{2\Delta y|_{j+\frac{1}{2}}} \frac{(K_{yx} s)|_{i,j+\frac{1}{2},k}}{\Delta y|_{i-\frac{1}{2}+\frac{1}{2},j,k}},
$$

$$
\begin{aligned}
Z_c Y_c X_c = &- \frac{(K_{xx} s)|_{i-\frac{1}{2},j,k}}{\Delta x|_{i-\frac{1}{2}}} - \frac{(K_{xx} s)|_{i+\frac{1}{2},j,k}}{\Delta x|_{i+\frac{1}{2}}} - \frac{(K_{yy} s)|_{i,j-\frac{1}{2},k}}{\Delta y|_{j-\frac{1}{2}}} - \frac{(K_{yy} s)|_{i,j+\frac{1}{2},k}}{\Delta y|_{j+\frac{1}{2}}} \\
&- \frac{(K_{zz} s)|_{i,j,k-\frac{1}{2}}}{\Delta z|_{k-\frac{1}{2}}} - \frac{(K_{zz} s)|_{i,j,k+\frac{1}{2}}}{\Delta z|_{k+\frac{1}{2}}} - s_{s,i,j,k} \frac{\Delta x_{i,j,k} \Delta y_{i,j,k} \Delta z_{i,j,k}}{t_t - t_{t-1}},
\end{aligned}
$$

$$
Z_c Y_c X_p = \frac{(K_{xx} s)|_{i+\frac{1}{2},j,k}}{\Delta x|_{i+\frac{1}{2}}} - \frac{\Delta y_{i,j-1,k}}{2\Delta y|_{j-\frac{1}{2}}} \frac{(K_{yx} s)|_{i,j-\frac{1}{2},k}}{\Delta x|_{i+\frac{1}{2}+\frac{1}{2},j,k}} + \frac{\Delta y_{i,j+1,k}}{2\Delta y|_{j+\frac{1}{2}}} \frac{(K_{yx} s)|_{i,j+\frac{1}{2},k}}{\Delta y|_{i-\frac{1}{2}+\frac{1}{2},j,k}},
$$

$$Z_c Y_p X_m = -\frac{\Delta x_{i,j,k}}{2\Delta x|_{i-\frac{1}{2}}} \frac{(K_{xy}s)|_{i-\frac{1}{2},j,k}}{\Delta y|_{i-1,j-\frac{1}{2}+\frac{1}{2},k}} - \frac{\Delta y_{i,j,k}}{2\Delta y|_{j+\frac{1}{2}}} \frac{(K_{yx}s)|_{i,j+\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2},j+1,k}},$$

$$Z_c Y_p X_c = -\frac{\Delta x_{i-1,j,k}}{2\Delta x|_{i-\frac{1}{2}}} \frac{(K_{xy}s)|_{i-\frac{1}{2},j,k}}{\Delta y|_{i,j-\frac{1}{2}+\frac{1}{2},k}} + \frac{\Delta x_{i+1,j,k}}{2\Delta x|_{i+\frac{1}{2}}} \frac{(K_{xy}s)|_{i+\frac{1}{2},j,k}}{\Delta y|_{i,j-\frac{1}{2}+\frac{1}{2},k}} + \frac{(K_{yy}s)|_{i,j+\frac{1}{2},k}}{\Delta y|_{j+\frac{1}{2}}},$$

$$Z_c Y_p X_p = \frac{\Delta x_{i,j,k}}{2\Delta x|_{i+\frac{1}{2}}} \frac{(K_{xy}s)|_{i+\frac{1}{2},j,k}}{\Delta y|_{i+1,j+\frac{1}{2}+\frac{1}{2},k}} + \frac{\Delta y_{i,j,k}}{2\Delta y|_{j+\frac{1}{2}}} \frac{(K_{yx}s)|_{i,j+\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2},j+1,k}},$$

$$Z_p Y_c X_c = \frac{(K_{zz}s)|_{i,j,k+\frac{1}{2}}}{\Delta z|_{k+\frac{1}{2}}},$$

in which the interface areas can be obtained by,

$$s|_{i-\frac{1}{2},j,k} = \frac{\Delta y_{i,j,k} + \Delta y_{i-1,j,k}}{2} \frac{\Delta z_{i,j,k} + \Delta z_{i-1,j,k}}{2},$$

$$s|_{i+\frac{1}{2},j,k} = \frac{\Delta y_{i,j,k} + \Delta y_{i+1,j,k}}{2} \frac{\Delta z_{i,j,k} + \Delta z_{i+1,j,k}}{2},$$

$$s|_{i,j-\frac{1}{2},k} = \frac{\Delta x_{i,j,k} + \Delta x_{i,j-1,k}}{2} \frac{\Delta z_{i,j,k} + \Delta z_{i,j-1,k}}{2},$$

$$s|_{i,j+\frac{1}{2},k} = \frac{\Delta x_{i,j,k} + \Delta x_{i,j+1,k}}{2} \frac{\Delta z_{i,j,k} + \Delta z_{i,j+1,k}}{2},$$

$$s|_{i,j,k-\frac{1}{2}} = \frac{\Delta x_{i,j,k} + \Delta x_{i,j,k-1}}{2} \frac{\Delta y_{i,j,k} + \Delta y_{i,j,k-1}}{2},$$

$$s|_{i,j,k+\frac{1}{2}} = \frac{\Delta x_{i,j,k} + \Delta x_{i,j,k+1}}{2} \frac{\Delta y_{i,j,k} + \Delta y_{i,j,k+1}}{2}.$$

Two types of boundary conditions are considered to build the system of linear equations: (1) non-flow boundary (Neumann's condition), i.e., no flow-connection through the external face of a boundary cell nor through the face shared with a dead cell, and (2) prescribed head (Dirichlet's condition), which is located at the center of any active grid-block (not a dead cell).

Three types of well conditions are considered: (1) prescribed head at the grid-block intersected by a well, (2) constant flow-rate wells ($q > 0$ if fluid extracted and $q < 0$ if fluid injected), and (3) observation wells ($q = 0$). Both vertical and non-vertical wells are considered in the code implementation as long as the trajectory of the well is given. The trick is that all cells that the well passes are assumed to be connected by a "high permeability path" which has a super-K value specified by the user, say, $K = 99999$. If the well has a constant flow-rate q, then q is fully assigned to the bottom-hole cell (i.e., the first well cell). If the well has a constant head h, then h is simply assigned to

all of the cells that the well penetrates. These three operations are applied to all of the well cells along the trajectory starting from the bottom-hole cell.

Given the initial head distribution $h(\mathbf{x})$ and boundary and/or (well) stress conditions, the system of difference equations can then be built up for all active cells. If there are dead cells in the reservoir, however, the corresponding difference equations are never built as above. For the purpose of easy identification, a special constant head value specified by the user, say, $h = -99999$, is simply assigned to those dead cells. Note that the interface permeability values around dead cells are also required to be set to zero, i.e., no-permeability, in making difference equations for other active cells. Once the system of linear equations is built up, a preconditioned bi-conjugated gradient method is employed to solve the equations sequentially along the given time-steps.

2.2.2 Well flow-rate

The well flow-rate can be easily computed by integrating the flux from various directions across the well (either production or injection), i.e.,

$$q_{i,j,k} = q_{i+\frac{1}{2},j,k} + q_{i-\frac{1}{2},j,k} + q_{i,j+\frac{1}{2},k} + q_{i,j-\frac{1}{2},k} + q_{i,j,k+\frac{1}{2}} + q_{i,j,k-\frac{1}{2}} - \Delta q_{i,j,k}, \quad (2.9)$$

where (i, j, k) is the well position in the mesh and the seven components are listed as follows,

$$\begin{aligned}
q_{i+\frac{1}{2},j,k} =& K_{ipxx}(h_{i,j,k} - h_{i+1,j,k}) + \\
& K_{ixyp}(h_{i+1,j-1,k} - h_{i+1,j+1,k}) + K_{ipxy}(h_{i,j-1,k} - h_{i,j+1,k}), \\
q_{i-\frac{1}{2},j,k} =& K_{imxx}(h_{i-1,j,k} - h_{i,j,k}) + \\
& K_{ixym}(h_{i-1,j-1,k} - h_{i-1,j+1,k}) + K_{imxy}(h_{i,j-1,k} - h_{i,j+1,k}), \\
q_{i,j+\frac{1}{2},k} =& K_{jpyy}(h_{i,j,k} - h_{i,j+1,k}) + \\
& K_{jyxp}(h_{i-1,j+1,k} - h_{i+1,j+1,k}) + K_{jpyx}(h_{i-1,j,k} - h_{i+1,j,k}), \\
q_{i,j-\frac{1}{2},k} =& K_{jmyy}(h_{i,j-1,k} - h_{i,j,k}) + \\
& K_{jyxm}(h_{i-1,j-1,k} - h_{i+1,j-1,k}) + K_{jmyx}(h_{i-1,j,k} - h_{i+1,j,k}), \\
q_{i,j,k+\frac{1}{2}} =& K_{kpzz}(h_{i,j,k} - h_{i,j,k+1}), \\
q_{i,j,k-\frac{1}{2}} =& K_{kmzz}(h_{i,j,k-1} - h_{i,j,k}), \\
\Delta q_{i,j,k} =& s_{s,i,j,k} \frac{h^t_{i,j,k} - h^{t-1}_{i,j,k}}{t_t - t_{t-1}} \Delta x_{i,j,k} \Delta y_{i,j,k} \Delta z_{i,j,k}.
\end{aligned}$$

The interface conductances are computed as follows,

$$K_{kmzz} = \frac{(K_{zz}s)|_{i,j,k-\frac{1}{2}}}{\Delta z|_{k-\frac{1}{2}}},$$
$$K_{jmyy} = \frac{(K_{yy}s)|_{i,j-\frac{1}{2},k}}{\Delta y|_{j-\frac{1}{2}}},$$

$$K_{imxx} = \frac{(K_{xx}s)|_{i-\frac{1}{2},j,k}}{\Delta x|_{i-\frac{1}{2}}},$$
$$K_{ipxx} = \frac{(K_{xx}s)|_{i+\frac{1}{2},j,k}}{\Delta x|_{i+\frac{1}{2}}},$$

$$K_{jpyy} = \frac{(K_{yy}s)|_{i,j+\frac{1}{2},k}}{\Delta y|_{j+\frac{1}{2}}},$$
$$K_{kpzz} = \frac{(K_{zz}s)|_{i,j,k+\frac{1}{2}}}{\Delta z|_{k+\frac{1}{2}}},$$

$$K_{ixym} = \frac{\Delta x_{i,j,k}}{2\Delta x|_{i-\frac{1}{2}}} \frac{(K_{xy}s)|_{i-\frac{1}{2},j,k}}{\Delta y|_{i-1,j-\frac{1}{2}+\frac{1}{2},k}},$$
$$K_{jyxm} = \frac{\Delta y_{i,j,k}}{2\Delta y|_{j-\frac{1}{2}}} \frac{(K_{yx}s)|_{i,j-\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2},j-1,k}},$$

$$K_{imxy} = \frac{\Delta x_{i-1,j,k}}{2\Delta x|_{i-\frac{1}{2}}} \frac{(K_{xy}s)|_{i-\frac{1}{2},j,k}}{\Delta y|_{i,j-\frac{1}{2}+\frac{1}{2},k}},$$
$$K_{jmyx} = \frac{\Delta y_{i,j-1,k}}{2\Delta y|_{j-\frac{1}{2}}} \frac{(K_{yx}s)|_{i,j-\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2}}},$$

$$K_{jpyx} = \frac{\Delta y_{i,j+1,k}}{2\Delta y|_{j+\frac{1}{2}}} \frac{(K_{yx}s)|_{i,j+\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2}}},$$
$$K_{ipxy} = \frac{\Delta x_{i+1,j,k}}{2\Delta x|_{i+\frac{1}{2}}} \frac{(K_{xy}s)|_{i+\frac{1}{2},j,k}}{\Delta y|_{i,j-\frac{1}{2}+\frac{1}{2},k}},$$

$$K_{jyxp} = \frac{\Delta y_{i,j,k}}{2\Delta y|_{j+\frac{1}{2}}} \frac{(K_{yx}s)|_{i,j+\frac{1}{2},k}}{\Delta x|_{i-\frac{1}{2}+\frac{1}{2},j+1,k}},$$
$$K_{ixyp} = \frac{\Delta x_{i,j,k}}{2\Delta x|_{i+\frac{1}{2}}} \frac{(K_{xy}s)|_{i+\frac{1}{2},j,k}}{\Delta y|_{i+1,j-\frac{1}{2}+\frac{1}{2},k}}.$$

2.2.3 Velocity fields

Once the block-centered head field $h(\mathbf{x}, t)$ is obtained, the internodal Darcian velocity field $\mathbf{q}(\mathbf{x}, t)$ can be easily computed according to the Darcy's law, i.e.,

$$\mathbf{q} = -\mathbf{K}\nabla h. \tag{2.10}$$

The corresponding pore fluid velocity field $\mathbf{v}(\mathbf{x}, t)$ is,

$$\mathbf{v} = \frac{\mathbf{q}}{\phi_i}, \tag{2.11}$$

where ϕ_i is the internodal porosity, which can be approximated by the simple linear interpolation, i.e.,

$$\phi_x = \frac{x_0 - x}{x_0}\phi_{x_1} + \frac{x}{x_0}\phi_{x_2},$$
$$\phi_y = \frac{y_0 - y}{y_0}\phi_{y_1} + \frac{y}{y_0}\phi_{y_2},$$
$$\phi_z = \frac{z_0 - z}{z_0}\phi_{z_1} + \frac{z}{z_0}\phi_{z_2},$$

in which (ϕ_x, ϕ_y, ϕ_z) is the interface porosity to be calculated, (x, y, z) is the dimension of the block interface, (x_0, y_0, z_0) is the dimension of the block center, and (ϕ_{x_1}, ϕ_{x_2}), (ϕ_{y_1}, ϕ_{y_2}), and (ϕ_{z_1}, ϕ_{z_2}) are the internodal porosity of two adjacent blocks.

If a spatially variable retardation factor $r = r(\mathbf{x}, t)$ is considered to depict the flow and transport of reactive solutes, the internodal pore fluid velocity field $\mathbf{v}(\mathbf{x}, t)$ of absorbed solutes is modified by,

$$\mathbf{v} = \frac{\mathbf{q}}{\phi_i r_i}, \tag{2.12}$$

where r_i is the internodal retardation factor, which can be approximated by the simple linear interpolation, i.e.,

$$
\begin{aligned}
r_x &= \frac{x_0 - x}{x_0} r_{x_1} + \frac{x}{x_0} r_{x_2}, \\
r_y &= \frac{y_0 - y}{y_0} r_{y_1} + \frac{y}{y_0} r_{y_2}, \\
r_z &= \frac{z_0 - z}{z_0} r_{z_1} + \frac{z}{z_0} r_{z_2},
\end{aligned}
$$

where (r_x, r_y, r_z) is the interface retardation factor to be calculated, (x, y, z) is the dimension of the block interface, (x_0, y_0, z_0) is the dimension of the block center, and (r_{x_1}, r_{x_2}), (r_{y_1}, r_{y_2}), and (r_{z_1}, r_{z_2}) are the retardation factor of two adjacent blocks.

2.3 Multi-scale-oriented Transport Simulator

The random-walk particle-tracking method (RWPT) has been widely used to simulate the conservative and reactive transport in porous media since it is free from numerical dispersion and computationally efficient. Aiming at the temporal and spatial distribution of the solute concentration, this approach simulates the behavior of solute transport by a discrete collection of particles subject to a deterministic displacement, which depends only on the local velocity field, and a random Brownian motion. If the number of particles is sufficiently large, the stochastic differential equation (SDE) of particles in a Lagrangian framework is equivalent to the advection-dispersion equation (ADE) of solutes in an Eulerian framework.

A particle in the flow domain Ω is displaced according to the stochastic differential equation, i.e., the Langevin equation (*Lichtner et al.*, 2002),

$$
\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \mathbf{a}\Delta t + \mathbf{B}\boldsymbol{\xi}\sqrt{\Delta t}, \tag{2.13}
$$

where $\mathbf{x}_t = (x_t, y_t, z_t)^T$ is the spatial position of the particle at time t and Δt is the discrete time step. The vector $\mathbf{a} = \mathbf{a}(\mathbf{x}, t)$, $[LT^{-1}]$, is responsible for the deterministic particle displacement along the flow streamlines, i.e.,

$$
\mathbf{a} = \mathbf{v} + \frac{1}{\phi}\nabla \cdot (\phi \mathbf{D}) = \mathbf{v} + \nabla \cdot \mathbf{D} + \mathbf{D} \cdot \nabla(\ln \phi), \tag{2.14}
$$

where the local dispersion tensor $\mathbf{D} = \mathbf{D}(\mathbf{x}, t)$ is defined by,

$$
\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{zx} & D_{zy} & D_{zz} \end{bmatrix}.
$$

The dispersion tensor $\mathbf{B} = \mathbf{B}(\mathbf{x}, t)$, $[LT^{-\frac{1}{2}}]$, of the random displacement part is,

$$\mathbf{B} = \begin{bmatrix} \frac{v_x}{v}\sqrt{2\alpha_L v} & -\frac{v_x v_z}{v\sqrt{v_x^2+v_y^2}}\sqrt{2\alpha_T^V v} & -\frac{v_y}{\sqrt{v_x^2+v_y^2}}\sqrt{2\frac{\alpha_T^H(v_x^2+v_y^2)+\alpha_T^V v_z^2}{v}} \\ \frac{v_y}{v}\sqrt{2\alpha_L v} & -\frac{v_y v_z}{v\sqrt{v_x^2+v_y^2}}\sqrt{2\alpha_T^V v} & \frac{v_x}{\sqrt{v_x^2+v_y^2}}\sqrt{2\frac{\alpha_T^H(v_x^2+v_y^2)+\alpha_T^V v_z^2}{v}} \\ \frac{v_z}{v}\sqrt{2\alpha_L v} & \frac{\sqrt{v_x^2+v_y^2}}{v}\sqrt{2\alpha_T^V v} & 0 \end{bmatrix}, \tag{2.15}$$

where $\mathbf{v} = (v_x, v_y, v_z)^T \in \Omega$ and α_L, α_T^H and α_T^V are the longitudinal dispersivity coefficient, the transverse dispersivity coefficient in the horizontal direction, and the transverse dispersivity coefficient in the vertical direction, respectively, [*dimensionless*]. And,

$$\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)^T \sim N(0, 1), \tag{2.16}$$

is a vector of independent Gaussian random numbers with a zero mean and a unit variance, [*dimensionless*].

The three-dimensional expressions for the one-step displacement of a particle are,

$$x_{t+\Delta t} = x_t + a_x \Delta t + b_x \sqrt{\Delta t}, \tag{2.17a}$$

$$y_{t+\Delta t} = y_t + a_y \Delta t + b_y \sqrt{\Delta t}, \tag{2.17b}$$

$$z_{t+\Delta t} = z_t + a_z \Delta t + b_z \sqrt{\Delta t}. \tag{2.17c}$$

The expressions for the deterministic displacement coefficient vector $\mathbf{a}$ of a particle are,

$$a_x = v_x + \frac{\partial D_{xx}}{\partial x} + \frac{\partial D_{xy}}{\partial y} + \frac{\partial D_{xz}}{\partial z} + D_{xx}\frac{\partial \ln\phi}{\partial x} + D_{xy}\frac{\partial \ln\phi}{\partial y} + D_{xz}\frac{\partial \ln\phi}{\partial z}, \tag{2.18a}$$

$$a_y = v_y + \frac{\partial D_{yx}}{\partial x} + \frac{\partial D_{yy}}{\partial y} + \frac{\partial D_{yz}}{\partial z} + D_{yx}\frac{\partial \ln\phi}{\partial x} + D_{yy}\frac{\partial \ln\phi}{\partial y} + D_{yz}\frac{\partial \ln\phi}{\partial z}, \tag{2.18b}$$

$$a_z = v_z + \frac{\partial D_{zx}}{\partial x} + \frac{\partial D_{zy}}{\partial y} + \frac{\partial D_{zz}}{\partial z} + D_{zx}\frac{\partial \ln\phi}{\partial x} + D_{zy}\frac{\partial \ln\phi}{\partial y} + D_{zz}\frac{\partial \ln\phi}{\partial z}. \tag{2.18c}$$

Since the particle tracking scheme requires the evaluation of velocity vector at an arbitrary position in the flow domain, i.e., $\mathbf{v} = (v_x, v_y, v_z)^T \in \Omega$, for

the computation of **a** and **B**, the velocity components within a block can be computed by the simple linear interpolation (*LaBolle et al.*, 1996), e.g.,

$$
\begin{aligned}
v_x &= \frac{x_0 - x}{x_0} v_{x_1} + \frac{x}{x_0} v_{x_2}, \\
v_y &= \frac{y_0 - y}{y_0} v_{y_1} + \frac{y}{y_0} v_{y_2}, \\
v_z &= \frac{z_0 - z}{z_0} v_{z_1} + \frac{z}{z_0} v_{z_2},
\end{aligned}
$$

where (x, y, z) is the position where the velocity components are to be calculated, (x_0, y_0, z_0) is the block size, and $(v_{x_1}, v_{y_1}, v_{z_1})$ and $(v_{x_2}, v_{y_2}, v_{z_2})$ are the interface center velocity of the given block. This scheme assumes that the velocity component varies linearly within a finite-difference cell with respect to the direction of that component. The continuous velocity field generated as such is consistent with the block-centered finite-difference formulation and conserves locally mass within each block.

The components of the local dispersion tensor $\mathbf{D} = \mathbf{D}(\mathbf{x}, t)$ are defined by (*Lichtner et al.*, 2002),

$$
\begin{aligned}
D_{xx} &= \frac{1}{v} \left[\alpha_L v_x^2 + \alpha_T^H v_y^2 + \alpha_T^V v_z^2 \frac{v_x^2}{v_x^2 + v_y^2} \right], \\
D_{yy} &= \frac{1}{v} \left[\alpha_L v_y^2 + \alpha_T^H v_x^2 + \alpha_T^V v_z^2 \frac{v_y^2}{v_x^2 + v_y^2} \right], \\
D_{zz} &= \frac{1}{v} \left[\alpha_L v_z^2 + \alpha_T^V (v_x^2 + v_y^2) \right], \\
D_{xy} = D_{yx} &= \frac{v_x v_y}{v} \left[\alpha_L - \alpha_T^H \frac{v^2}{(v_x^2 + v_y^2)} + \alpha_T^V \frac{v_z^2}{(v_x^2 + v_y^2)} \right], \\
D_{xz} = D_{zx} &= \frac{v_x v_z}{v} \left[\alpha_L - \alpha_T^V \right], \\
D_{yz} = D_{zy} &= \frac{v_y v_z}{v} \left[\alpha_L - \alpha_T^V \right].
\end{aligned}
$$

There are two sets of dispersion tensors: (i) D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{yx}, D_{xz}, D_{zx}, D_{yz}, and D_{zy} are located at the interface center of blocks, so are v_x, v_y, and v_z, and (ii) D_{xx}, D_{yy}, D_{zz}, D_{xy}, D_{yx}, D_{xz}, D_{zx}, D_{yz}, and D_{zy} are located at the block centers, so are v_x, v_y, and v_z. Both of velocity fields are approximated by the bilinear interpolation of internodal velocity values before the local dispersion tensor is calculated (*LaBolle et al.*, 1996). The first set is for the computation of the divergence of the local dispersion tensor and the second one is for the terms associated with the gradient of the porosity.

The divergence of the local dispersion tensor is calculated by the difference formulae using the second set of dispersion formulae, i.e., of the block center. The resulted divergence of the local dispersion tensor is interpolated into an arbitrary position by the linear interpolation rule. The gradient of the porosity is calculated by the difference formulae. Then it multiplies the local dispersion tensor of the interface center using the second set of dispersion formulae. The evaluation of those two set of values, i.e., the second and third terms in the vector $\mathbf{a}$, at an arbitrary position in the flow domain is propagated by the simple linear interpolation as well.

The expressions for the random displacement coefficient $\mathbf{B}\boldsymbol{\xi}$ of a particle are,

$$b_x = \xi_1 \frac{v_x}{v}\sqrt{2\alpha_L v} - \xi_2 \frac{v_x v_z}{v\sqrt{v_x^2+v_y^2}}\sqrt{2\alpha_T^V v} - \xi_3 \frac{v_y}{\sqrt{v_x^2+v_y^2}}\sqrt{2\frac{\alpha_T^H(v_x^2+v_y^2)+\alpha_T^V v_z^2}{v}}, \tag{2.19a}$$

$$b_y = \xi_1 \frac{v_y}{v}\sqrt{2\alpha_L v} - \xi_2 \frac{v_y v_z}{v\sqrt{v_x^2+v_y^2}}\sqrt{2\alpha_T^V v} + \xi_3 \frac{v_x}{\sqrt{v_x^2+v_y^2}}\sqrt{2\frac{\alpha_T^H(v_x^2+v_y^2)+\alpha_T^V v_z^2}{v}}, \tag{2.19b}$$

$$b_z = \xi_1 \frac{v_z}{v}\sqrt{2\alpha_L v} + \xi_2 \frac{\sqrt{v_x^2+v_y^2}}{v}\sqrt{2\alpha_T^V v}. \tag{2.19c}$$

The computation of these three expressions is straightforward since the velocity field has been interpolated into an arbitrary position within the flow domain.

In contrast to the constant time scheme, a constant displacement scheme is considered in the code implementation to improve the efficiency of particle transport (*Wen and Gomez-Hernandez*, 1996a). Assuming the particles move along the x direction, the time step for a constant-displacement scheme in one grid-block with the size equal to Δx is,

$$\Delta t = \frac{1}{n}\frac{\Delta x}{v_x}. \tag{2.20}$$

3

A Multi-scale Blocking McMC Method: 1. Methodology

Abstract

An adequate representation of the detailed spatial variation of parameters calls for high-resolution modeling of subsurface reservoirs or aquifers. A Monte Carlo method for such refined characterization of large-scale models typically invokes a stochastic simulation method to generate $i.i.d$ realizations that honor both hard data (e.g., conductivity) and dependent state data (e.g., piezometric head, concentration, etc.), known as conditioning and inverse-conditioning modeling, respectively. The blocking Markov chain Monte Carlo (BMcMC) method is an effective scheme to carry out such conditional and inverse-conditional simulation by sampling directly from a posterior distribution that incorporates the prior information and the posterior observations in a Bayesian framework. However, the usefulness of the BMcMC method suffers from the its efficiency in dealing with the high-resolution cases. In this study, a multi-scale blocking McMC (MsBMcMC) scheme is presented to generate high-resolution, multi-Gaussian, conditional and inverse-conditional realizations. What make this method more efficient in exploring the parameter space of high-dimensional models are that the blocking proposal kernel is an appropriate approximation to the target posterior distribution, that the fast

generation of the proposal blocking kernel is based on the LU-decomposition of the covariance matrix or the spectral decomposition of the covariance matrix with the aid of the fast Fourier transform, and that a multi-scale procedure is used to calculate the likelihood quickly. The adoption of the blocking scheme is based on three considerations: (1) it helps preserve the spatial statistics and model structure, (2) it helps improve the mixing of a Markov chain, and (3) it is more suitable for the introduction of upscaling procedures to accelerate the computation of the likelihood. The introduction of the multi-scale scheme in calculating their likelihood efficiently speeds up the construction of the Markov chain because those low probable candidates are rejected without consuming too many CPU resources. Moreover, the proposal kernel is multi-scale, which forms a coupled McMC, and the convergence velocity and the mixing speed of the Markov chain are improved simply because the information from the large scale proposal and that from the small scale proposal are shared in constructing the Markov chain such that the modes can be found quickly by the fastest scheme and the realizations are output by the most optimal scheme. The independent goestatistical realizations generated in this way are not only conditioned to the conductivity, the piezometric head, the temporal moments of solute concentration, and other measurements available, but also have the expected spatial statistics and structure.

3.1 Introduction

Collections of direct and indirect measurements provide limited but indispensable knowledge on the subsurface reservoirs or aquifers which generally calls for a stochastic method to generate conditional realizations and to characterize the inherent uncertainties. To this purpose, two steps are typically involved: model structure identification and stochastic conditional simulation. Correspondingly, model uncertainties are classified into two types: structure uncertainty that arises from uncertain hypotheses or unmodeled processes and parameter uncertainty that comes from measurement errors, inherent heterogeneity or scaling problem. There are a number of methods for assessing the impact of parameter uncertainty on performance prediction (*Dagan*, 1989; *Gelhar*, 1993). The majority of them are based on the underlying assumption of a correct model structure and hence the output uncertainty is evaluated by propagating the input uncertainty. Traditionally, however, structure uncertainty and parameter uncertainty are seldom isolated from each other which results in a problem that prediction uncertainty is generally a combined effect of both. The worth of measurements on one type of model property with respect to predicted performance is therefore hardly evaluated in a deterministic manner since prediction uncertainty may come from either model structure or

model parameters or both. The data worth evaluation based on inconsistent model structures may produce a misleading result.

A special difficulty occurs in assessing the worth of dependent state data because the problem is complicated by an ill-posed inverse problem. Classical optimization-based inverse methods for stochastic simulation generally work as follows (*Carrera and Neuman*, 1986; *RamaRao et al.*, 1995; *McLaughlin and Townley*, 1996; *Gomez-Hernandez et al.*, 1997). Assuming an initial model structure, an ensemble of equal-likely realizations constraint to hard data are obtained by geostatistically-based conditional simulation subroutines. Then each of realizations is calibrated to honor dependent state data by applying an inverse procedure. With these two stages, those generated models are considered as the independent conditional and inverse-conditional realizations. Most often, however, the parameter statistics and spatial structure have to be perturbed, either passively or actively, during the procedure of parameter calibration in order to match the dependent state data. It matters because a large number of efforts have been spent on building the model structure through model identification, e.g., collections of hard data for variogram analysis (e.g., *Deutsch and Journel*, 1998), collections of state data for model structure identification (e.g., *Sun and Yeh*, 1985; *Kitanidis*, 1996), outcrop's analogue, geological mapping (e.g., *Koltermann and Gorelick*, 1996), geophysical imaging (e.g., *Rubin et al.*, 1992), etc. But after model calibration or history matching, the identified model structure has been destroyed in an unpredicted manner in order to honor the state data. Therefore, a Monte Carlo method for uncertainty assessment entails a stochastic simulation method capable of generating independent, identically distributed (*i.i.d*) conditional realizations that share a specified *a priori* model structure, i.e., identical structure parameters.

The Markov chain Monte Carlo (McMC) method (*Oliver et al.*, 1997; *Robert and Casella*, 1999) outstands itself from other inverse approaches because it completely circumvents the numerical instability due to the fact that the inverse-conditioning is simply a sampling procedure rather than an optimization step. Moreover, the adoption of the blocking scheme is not only helpful for the mixing of the chain (*Liu*, 1996; *Roberts and Sahu*, 1997) but also useful for the preserving of spatial statistics and structure. Several challenges that the current McMC methods often face include that they fail to treat the high-resolution models due to the limitation of the LU-decomposition-based sampler and that the computation of the likelihood is extremely CPU expensive. To attack these problems, a multi-scale blocking McMC (MsBMcMC) method is presented in this sequence of papers to generate high-resolution, conditional and inverse-conditional realizations that strictly honor the specification of parameter statistics and spatial structures for physical models.

In essence, this approach just thins down, with the aid of the McMC theorem, the ensemble of candidates merely conditional to hard data by means

of judiciously selecting those models inverse-conditional to the nonlinear state data, measured by the posterior probability. It differs itself from other approaches to inverse problems in several respects: (1) the inverse problem is coined as inverse-conditioning problems in the framework of geostatistically-based conditional simulation and the inverse-conditioning problem is cast as a sampling procedure which fully circumvents the ill-posedness that the classical inverse methods often encounter, (2) the generated realizations strictly follow the prior configuration of the spatial statistics and structure for model parameters, and (3) the efficiency of the proposed method for the generation of independent high-resolution models is improved due to the incorporation of the multi-scale blocking scheme into the classical McMC method.

The purpose of this chapter is to develop and implement the MsBMcMC algorithm to generate *i.i.d* high-dimensional models that not only honor static hard data and dynamic state data but also preserve specified spatial statistics and structure. A synthetic example will be presented in Chapter 4 to demonstrate the efficiency of the proposed method.

3.2 Blocking McMC Methods

Consider a stochastic simulation at n grid nodes conditional to m hard data and k nonlinear state data where the term "nonlinear" simply means that the dependent state data are nonlinearly related to the model parameters governed by the flow and transport partial differential equations. Specifically, let $\boldsymbol{x} = (x_0, x_1, ..., x_{n-1})^T \subset R^n$ denote a realization conditional to m hard data $\boldsymbol{x}_1 = \boldsymbol{x}_{obs} = (x'_0, x'_1, ..., x'_{m-1})^T \subset R^m$ and k state data $\boldsymbol{y} = \boldsymbol{y}_{obs} = (y_0, y_1, ..., y_{k-1})^T \subset R^k$. Assuming a multi-Gaussian process, the spatial distribution of $\boldsymbol{x}$ follows, $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C_x})$, where $\boldsymbol{\mu} \subset R^n$ is the prior mean of the Gaussian process and $\boldsymbol{C_x} \subset R^{n \times n}$ describes the structure dependence of spatial points from each other. The observation errors of $\boldsymbol{x}_{obs}$ are assumed to be assimilated into the prior statistical model. Assuming a multi-normal error, the simulated observation $\boldsymbol{y}_{sim}$ for a given sample $\boldsymbol{x}$ can be expressed as, $\boldsymbol{y}_{sim}|\boldsymbol{x} \sim N(g(\boldsymbol{x}), \boldsymbol{C_y})$, where $\boldsymbol{C_y} \subset R^{k \times k}$ describes the degree of discrepancy between the transfer function $g(\boldsymbol{x})$ and the true but error-prone observation $\boldsymbol{y}$. The transfer function $g(\boldsymbol{x})$ 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. In such case, its accuracy depends on the spatial discretization of the physical model. 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 $\boldsymbol{y}$ that can be included in this statistical model. In this sense, $\boldsymbol{C_y}$ measures both the modeling errors and the measurement errors.

In summary, the objective of the stochastic simulation is to infer $\boldsymbol{x}$ from $\boldsymbol{x}_1$, $\boldsymbol{y}$, and some hyperparameters $\boldsymbol{\theta}$ that describe spatial statistics and structure of models. The most challenging part of the conditional simulation is basically an inverse problem since an inverse operator $g^{-1}(\boldsymbol{y})$ is applied to the conditioning procedure.

3.2.1 Bayesian formulation

The joint prior probability density function (pdf) of a multi-Gaussian random field $\boldsymbol{x}$ is,

$$\pi(\boldsymbol{x}|\boldsymbol{x}_1,\boldsymbol{\theta}) = (2\pi)^{-\frac{n}{2}} \|\boldsymbol{C}_{\boldsymbol{x}}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \tag{3.1}$$

This pdf represents some prior knowledge about the parameterization of a physical model $\boldsymbol{x}$ through the configuration of $\boldsymbol{\mu}$ and $\boldsymbol{C}_{\boldsymbol{x}}$ which, together with other parameters, boil down to a hyperparameter set $\boldsymbol{\theta}$. It should allow for the greatest uncertainty while obeying the constraints imposed by the prior information. The hyperparameters $\boldsymbol{\theta}$ are inferred from both the *a posteriori* measurements and the *a priori* subjective imagination.

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

$$L(\boldsymbol{x}|\boldsymbol{y}) = (2\pi)^{-\frac{k}{2}} \|\boldsymbol{C}_{\boldsymbol{y}}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(g(\boldsymbol{x})-\boldsymbol{y})^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(g(\boldsymbol{x})-\boldsymbol{y})\right\}. \tag{3.2}$$

This likelihood function is defined by the misfit between the observed data $\boldsymbol{y}$ and the predicted data $g(\boldsymbol{x})$ from a candidate parameter model $\boldsymbol{x}$ and measures the probability of observing the data $\boldsymbol{y}$ for the model $\boldsymbol{x}$. Obviously, the probability of observing the given data $\boldsymbol{y}$ becomes smaller as the misfit becomes larger.

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

$$\pi(\boldsymbol{x}|\boldsymbol{x}_1,\boldsymbol{y},\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}(g(\boldsymbol{x})-\boldsymbol{y})^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(g(\boldsymbol{x})-\boldsymbol{y})\right\}. \tag{3.3}$$

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

of presentation, $\boldsymbol{x}_1$ and $\boldsymbol{\theta}$ are dropped out such that $\pi(\boldsymbol{x}) \equiv \pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{\theta})$ and $\pi(\boldsymbol{x}|\boldsymbol{y}) \equiv \pi(\boldsymbol{x}|\boldsymbol{x}_1, \boldsymbol{y}, \boldsymbol{\theta})$. A multi-scale blocking McMC scheme is developed as below to explore the posterior distribution.

3.2.2 Markov chain Monte Carlo method

Due to the highly nonlinearity of the likelihood model, it is impossible to sample directly from this posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$. The Markov chain Monte Carlo method (*Metropolis et al.*, 1953; *Hastings*, 1970; *Geman and Geman*, 1984), however, is especially suitable for exploring the parameter space with such type of complicated posterior distribution. A typical McMC algorithm employing the Metropolis-Hastings rule to explore the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ goes as follows,

(1) Initialize the parameters $\boldsymbol{x}$;

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

- propose $\boldsymbol{x}^* \sim q(\boldsymbol{x}^*|\boldsymbol{x})$;
- accept $\boldsymbol{x}^*$ with probability $\min\{1, \alpha\}$, where $\alpha = \frac{\pi(\boldsymbol{x}^*|\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x}|\boldsymbol{y})q(\boldsymbol{x}^*|\boldsymbol{x})}$;

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

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

One of the most interesting problems in this algorithm is the configuration of the proposal transition kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$, which plays a crucial role in the computational efficiency of a Metropolis-Hastings-type McMC method.

3.2.3 Single-component McMC method

For a classical single-component McMC method, the proposal kernel $x^* \sim N(\mu, \sigma_x^2)$ (*Oliver et al.*, 1997). The acceptance rate α is computed by,

$$\alpha = \frac{\pi(\boldsymbol{x}^*|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{y})}\frac{q(x|x^*)}{q(x^*|x)} = \frac{\pi(\boldsymbol{x}^*)}{\pi(\boldsymbol{x})}\frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})}\frac{q(x|x^*)}{q(x^*|x)}. \tag{3.4}$$

Take its logarithm,

$$\ln\alpha = \ln\pi(\boldsymbol{x}^*) - \ln\pi(\boldsymbol{x}) + \ln q(x|x^*) - \ln q(x^*|x) + \ln L(\boldsymbol{x}^*|\boldsymbol{y}) - \ln L(\boldsymbol{x}|\boldsymbol{y}). \tag{3.5}$$

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

$$\ln \pi(\boldsymbol{x}^*) \propto -\frac{1}{2}(\boldsymbol{x}^* - \boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}^* - \boldsymbol{\mu}), \tag{3.6}$$

$$\ln \pi(\boldsymbol{x}) \propto -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}), \tag{3.7}$$

$$\ln q(x^*|x) \propto -\frac{1}{2}z^2; \qquad z \sim N(0,1), \tag{3.8}$$

$$\ln q(x|x^*) \propto -\frac{(x-\mu)^2}{2\sigma_x^2}, \tag{3.9}$$

$$\ln L(\boldsymbol{x}^*|\boldsymbol{y}) \propto -\frac{1}{2}(g(\boldsymbol{x}^*) - \boldsymbol{y})^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(g(\boldsymbol{x}^*) - \boldsymbol{y}), \tag{3.10}$$

$$\ln L(\boldsymbol{x}|\boldsymbol{y}) \propto -\frac{1}{2}(g(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(g(\boldsymbol{x}) - \boldsymbol{y}). \tag{3.11}$$

For a small-dimensional case when the LU-decomposition of $\boldsymbol{C}_{\boldsymbol{x}} = \boldsymbol{LU}$ is available, the above formulae can be computed straightforwardly. Formulae (6) and (7), which are often referred as the model density evaluation, can be computed from,

$$-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) = -\frac{1}{2}(\boldsymbol{Lu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{Lu}) = -\frac{1}{2}\boldsymbol{u}^T\boldsymbol{u}, \tag{3.12}$$

where $\boldsymbol{u}$ is the solution of $\boldsymbol{Lu} = \boldsymbol{x} - \boldsymbol{\mu}$. Note that $\boldsymbol{C}_{\boldsymbol{x}}$ has already been LU-decomposed and the system of linear equations about $\boldsymbol{u}$ can be solved by a simple back-substitution. The computations of Formulae (10) and (11) are straightforward since $\boldsymbol{C}_{\boldsymbol{y}}$ is generally a diagonal matrix and the forward simulator $g(\boldsymbol{x})$ is called in a black-box way. Section 3.2.6 will further present a multi-scale scheme to accelerate the evaluation of these log-likelihood models.

For a large-dimensional case, however, the LU-decomposition of $\boldsymbol{C}_{\boldsymbol{x}}$ is often unavailable which makes the single-component proposal easy to fail. There are two main challenges: the convergence problem and the evaluation of the model density.

3.2.4 Blocking McMC methods

To the convergence problem, the block scheme is well known for helping improve the mixing of a Markov chain (*Liu*, 1996; *Roberts and Sahu*, 1997). Moreover, a blocking proposal kernel that follows the correlation structure as specified *a priori* makes the candidate closer to the posterior distribution which also speeds up the convergence of the chain. Based on these two facts, a blocking scheme is employed in this study to construct the transition kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$. 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, i.e., $q(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*|\boldsymbol{x})$.

Scheme #1

Specifically, for the blocking McMC method, the proposal kernel $\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}} \sim N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{C}}_{\boldsymbol{x}})$ which has the identical spatial distribution as the prior model $\boldsymbol{x}^*|\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C}_{\boldsymbol{x}})$ except that their dimensions are different, where $\hat{\boldsymbol{x}}^* \subseteq \boldsymbol{x}^*$ denotes the proposed parameters for the updating block and $\check{\boldsymbol{x}} \subset \boldsymbol{x}$ represents the current parameters at a limited neighbor around the updating block (see Figure 3.1). The superblock that consists of the updating block and its neighbor can be defined as a template over which the proposal scheme generally works, i.e., $\ddot{\boldsymbol{x}} = (\hat{\boldsymbol{x}}^*, \check{\boldsymbol{x}})^T \subseteq \boldsymbol{x}$. The acceptance rate α reads as,

$$\alpha = \frac{\pi(\boldsymbol{x}^*)}{\pi(\boldsymbol{x})} \frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})} \frac{\pi(\hat{\boldsymbol{x}}|\check{\boldsymbol{x}})}{\pi(\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}})}. \tag{3.13}$$

Take its logarithm,

$$\ln \alpha = \ln \pi(\boldsymbol{x}^*) - \ln \pi(\boldsymbol{x}) + \ln \pi(\hat{\boldsymbol{x}}|\check{\boldsymbol{x}}) - \ln \pi(\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}}) + \ln L(\boldsymbol{x}^*|\boldsymbol{y}) - \ln L(\boldsymbol{x}|\boldsymbol{y}). \tag{3.14}$$

The computations of $\ln \pi(\boldsymbol{x}^*)$, $\ln \pi(\boldsymbol{x})$, $\ln L(\boldsymbol{x}^*|\boldsymbol{y})$ and $\ln L(\boldsymbol{x}|\boldsymbol{y})$ are in the same manner as for the single-component case. The other two items are listed as follows,

$$\ln \pi(\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}}) \propto -\frac{1}{2}(\ddot{\boldsymbol{x}}^* - \ddot{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}})^T \ddot{\boldsymbol{C}}_{\boldsymbol{x}}^{-1} (\ddot{\boldsymbol{x}}^* - \ddot{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}}), \tag{3.15}$$

$$\ln \pi(\hat{\boldsymbol{x}}|\check{\boldsymbol{x}}) \propto -\frac{1}{2}(\ddot{\boldsymbol{x}} - \ddot{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}})^T \ddot{\boldsymbol{C}}_{\boldsymbol{x}}^{-1} (\ddot{\boldsymbol{x}} - \ddot{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}}), \tag{3.16}$$

where the superblock $\ddot{\boldsymbol{x}}^* = (\hat{\boldsymbol{x}}^*, \check{\boldsymbol{x}})^T \subseteq \boldsymbol{x}^*$; $\ddot{\boldsymbol{x}} = (\hat{\boldsymbol{x}}, \check{\boldsymbol{x}})^T \subseteq \boldsymbol{x}$; $\ddot{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}}$ is the kriging estimate for the superblock from the neighbor $\check{\boldsymbol{x}}$; and $\ddot{\boldsymbol{C}}$ is the covariance matrix of the superblock.

For the blocking McMC method, the proposal kernel $\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}} \sim N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{C}}_{\boldsymbol{x}})$ 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 $\ln \pi(\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}})$, i.e.,

$$\ln \pi(\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}}) \propto -\frac{1}{2} \boldsymbol{z}_{\hat{\boldsymbol{x}}^*}^T \boldsymbol{z}_{\hat{\boldsymbol{x}}^*}, \tag{3.17}$$

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

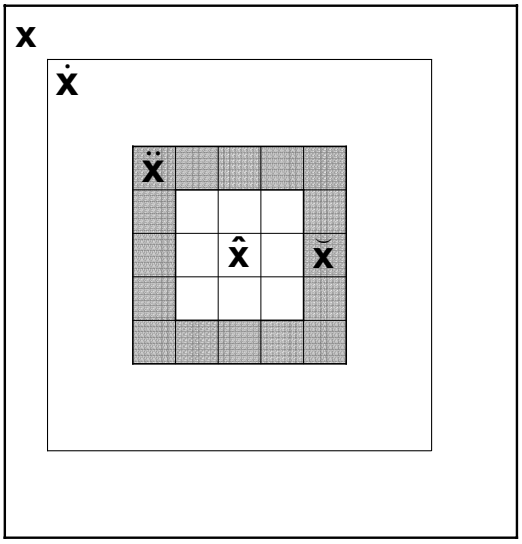


Figure 3.1: *A superblock template* $\ddot{\boldsymbol{x}}$ $(\subseteq \dot{\boldsymbol{x}} \subseteq \boldsymbol{x})$ *consists of the updating block* $\hat{\boldsymbol{x}}$ *and its neighbor* $\check{\boldsymbol{x}}$*. Note that* $\dot{\boldsymbol{x}}$*, which is a subset of* $\boldsymbol{x}$*, is used to approximate the prior density for a high-dimensional case, i.e.,* $\pi(\dot{\boldsymbol{x}}) \doteq \pi(\boldsymbol{x})$*.*

$$-\frac{1}{2}(\hat{\boldsymbol{x}}^* - \hat{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}})^T \boldsymbol{C}_{\hat{\boldsymbol{x}}^*}^{-1}(\hat{\boldsymbol{x}}^* - \hat{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}}) = -\frac{1}{2}(\boldsymbol{L}_{\hat{\boldsymbol{x}}^*\hat{\boldsymbol{x}}^*}\boldsymbol{z}_{\hat{\boldsymbol{x}}^*})^T \boldsymbol{C}_{\hat{\boldsymbol{x}}^*}^{-1}(\boldsymbol{L}_{\hat{\boldsymbol{x}}^*\hat{\boldsymbol{x}}^*}\boldsymbol{z}_{\hat{\boldsymbol{x}}^*}) = -\frac{1}{2}\boldsymbol{z}_{\hat{\boldsymbol{x}}^*}^T \boldsymbol{z}_{\hat{\boldsymbol{x}}^*}. \tag{3.18}$$

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

$$\ln \pi(\hat{\boldsymbol{x}}|\check{\boldsymbol{x}}) \propto -\frac{1}{2}\boldsymbol{u}_{\hat{\boldsymbol{x}}}^T \boldsymbol{u}_{\hat{\boldsymbol{x}}}, \tag{3.19}$$

where $\boldsymbol{u}_{\hat{\boldsymbol{x}}}$ satisfies $\boldsymbol{L}_{\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}}\boldsymbol{u}_{\hat{\boldsymbol{x}}} = \hat{\boldsymbol{x}} - \hat{\boldsymbol{\mu}}_{\check{\boldsymbol{x}}}$. Note that the computation of $\boldsymbol{L}_{\hat{\boldsymbol{x}}\hat{\boldsymbol{x}}}$ 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.

Scheme #2

For a high-dimensional case, however, it still remains a challenge because the LU-decomposition of the covariance matrix is not easy (see Formulae (6) and (7)). A numerically efficient method in computing the prior density $\pi(\boldsymbol{x})$ is

using a subset of the field $\dot{\boldsymbol{x}}$ to approximate the entire field $\boldsymbol{x}$, i.e., $\ddot{\boldsymbol{x}} \subseteq \dot{\boldsymbol{x}} \subseteq \boldsymbol{x}$ (see Figure 3.1). Therefore, the acceptance rate α is simply approximated by,

$$\alpha = \frac{\pi(\dot{\boldsymbol{x}}^*)}{\pi(\dot{\boldsymbol{x}})} \frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})} \frac{\pi(\hat{\boldsymbol{x}}|\check{\boldsymbol{x}})}{\pi(\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}})}. \tag{3.20}$$

From our experience, a subset with the size equal to one or two correlation lengths is sufficient to approximate the entire field.

Scheme #3

For a large-dimensional case with a large updating block, furthermore, considering the numerical approximation of a kriging estimate, two ways to generating the proposal kernel are treated equivalently, i.e., $\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}} \sim N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{C}}_{\boldsymbol{x}}) \doteq \boldsymbol{x}^*|\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C}_{\boldsymbol{x}})$. Therefore, the acceptance rate α can be simply set as,

$$\alpha = \frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})}, \tag{3.21}$$

which holds because $q(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*|\boldsymbol{x}) \doteq \pi(\boldsymbol{x}^*)$ and $q(\boldsymbol{x}|\boldsymbol{x}^*) = \pi(\boldsymbol{x}|\boldsymbol{x}^*) \doteq \pi(\boldsymbol{x})$. In such case, the computation of α avoids the complicated evaluation of the model density as done in Formulae (6) and (7).

3.2.5 Generation of the proposal kernel

The previous part presents a blocking proposal scheme to construct the Markov chain for exploring the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$. This part presents two numerical methods for generating geostatistical realizations of the proposal blocking kernel $\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}} \sim N(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{C}}_{\boldsymbol{x}})$ and for evaluating their density if necessary.

Although many random field generators may produce conditional realizations for the blocking proposal (*Deutsch and Journel*, 1998), a main challenge lies in the generation velocity. A fast generator is strongly recommended because a huge amount of *i.i.d* models are required for forward evaluations. The algorithm based on the LU-decomposition of the covariance matrix is preferred since it is quite efficient in generating a large number of conditional realizations (*Davis*, 1987; *Alabert*, 1987). If the LU-decomposition of the covariance matrix $\ddot{\boldsymbol{C}}_{\boldsymbol{x}} = \boldsymbol{LU}$ is available, the updated model parameters can be easily obtained by,

$$\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}} = \boldsymbol{\mu}_{\hat{\boldsymbol{x}}^*} + \boldsymbol{L}_{\hat{\boldsymbol{x}}^*\check{\boldsymbol{x}}} \boldsymbol{L}_{\check{\boldsymbol{x}}\check{\boldsymbol{x}}}^{-1} (\check{\boldsymbol{x}} - \boldsymbol{\mu}_{\check{\boldsymbol{x}}}) + \boldsymbol{L}_{\hat{\boldsymbol{x}}^*\hat{\boldsymbol{x}}^*} \boldsymbol{z}_{\hat{\boldsymbol{x}}^*}, \tag{3.22}$$

where $\boldsymbol{\mu}_{\hat{\boldsymbol{x}}^*}$ and $\boldsymbol{\mu}_{\check{\boldsymbol{x}}}$ are the prior mean for the updating block and its neighbor, respectively. If Formula (13) is used to compute the acceptance rate, the model densities of $\hat{\boldsymbol{x}}^*|\check{\boldsymbol{x}}$ and $\hat{\boldsymbol{x}}|\check{\boldsymbol{x}}$ can be efficiently evaluated by Formulae (17)

and (19), respectively. More details on the derivation of Formula (22) and its numerical implementation are given in Appendix A.

However, the LU-decomposition method tends to fail in dealing with the high-dimensional case with a large superblock due to the difficulty of the LU-decomposition of the covariance matrix. In this regard, the FFT-based spectral simulator has a computational ability more powerful than the LU-based generator. Just like the LU-decomposition method, the FFT-based spectral method is also based on the matrix decomposition technique. First the symmetric covariance matrix $\boldsymbol{C_x}$ is extended to form a symmetric circulant matrix $\boldsymbol{S}$ such that its square root can be efficiently computed by the fast Fourier transform,

$$\boldsymbol{S} = (\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}})(\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}})^T. \tag{3.23}$$

Then the stochastic realizations in the Fourier space can be generated by $\tilde{\boldsymbol{x}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z}$, where $\boldsymbol{z} = \boldsymbol{z}_1 + i\boldsymbol{z}_2$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \sim N(\boldsymbol{0}, \boldsymbol{1})$. The corresponding random fields in the spatial domain $\boldsymbol{x}$ can be obtained by applying an inverse FFT operation to $\tilde{\boldsymbol{x}}$. Appendix B gives an outline on the derivation and numerical implementation of the FFT-based spectral sampler; more details are referred to relevant literatures elsewhere (*Pardo-Iguzquiza and Chica-Olmo*, 1993; *Robin et al.*, 1993; *Dietrich and Newsam*, 1993, 1996; *Gutjahr et al.*, 1997; *Ruan and McLaughlin*, 1998).

In addition, an important step to generating conditional realizations is the configuration of model structure for candidates, i.e., the specification of $\boldsymbol{\theta}$. If the two-point geostatistics are adopted, an *ad hoc* method is to carry out experimental variogram analysis of parameters if some direct measurements are available. On the other hand, the structural parameters of candidates may benefit from fitting to the observations of state variables (e.g., *Sun and Yeh*, 1985; *Kitanidis*, 1996). The geological mapping (e.g., *Koltermann and Gorelick*, 1996), geophysical imaging (e.g., *Rubin et al.*, 1997), outcrop's analogue and other soft information are also helpful for model structure identification.

3.2.6 Multi-scale computation of the likelihood

The need of a prohibited CPU cost for running forward simulations with a complete candidate model is an obvious shortcoming in calculating the log-likelihood as expressed in Formulae (10) and (11) since numerous candidates should be tested and forward simulator $g(\boldsymbol{x})$ is generally expensive. It is especially true for an McMC scheme dealing with a high-resolution case, e.g., a refined description of aquifer with a small variance of state variable (*Wen et al.*, 1998). In such case, the model is highly discretized (i.e., a large n) and the state variable is considered to be highly confident (i.e., a small $\boldsymbol{C_y}$). The acceptance rate for a candidate may be rather low and, consequently, a large

number of candidates must be proposed for the construction of the Markov chain. A way out of this dilemma is to find an alternative to calculate the likelihood fast but accurately as possible. One possible implementation is to use a fast proxy of forward simulators. Another one is to use a coarsen version of candidates with the aid of upscaling to speedup the forward evaluation. For the latter, only those candidates potentially with higher acceptance probability are subject to complete forward simulations. For both cases, however, some candidates may be discarded unreasonably for the loss of the information in calculating the real likelihood. Superior to a single-component version of McMC, the BMcMC is especially suitable for incorporation of widely used upscaling schemes to speedup the forward computation. This is because the upscaling procedures tend to average out the effect of a single component update such that the updated model is almost the same as the old one. Hence, all the updated models tend to be accepted which makes the McMC quite inefficient. It happens more often for a high-resolution case where small prior variances are specified for state variables. In such case a very low acceptance rate, say, much less than 5%, is often observed so that a large number of candidate models have been tested but then discarded due to the small possibility of being accepted. The introduction of coarsen models helps speedup the BMcMC computation simply because a large number of computational time is saved in performing the forward simulations.

In this work, the second option is adopted for fast computation of the likelihood. If the average acceptance rate is less than a specified threshold, say 5%, then the multiscale scheme is invoked. First, a proper (economical) upscaling scheme, e.g., the geometric mean, is selected to yield less biased models. In addition, a large number of alternatives have been developed for upscaling hydraulic conductivity (*Wen and Gomez-Hernandez*, 1996b; *Renard and de Marsily*, 1997). Second, the forward simulators should be able to cope with the flow and transport problems at various scales. For this purpose, a flexible-grid full-tensor finite-difference flow simulator, which is widely compatible with the outputs from various upscaling subroutines, is developed for the fast computation of the likelihood. This flow simulator has ample abilities to accept the input of aquifer models within irregular grids and the input $\ln K$ field could be a full tensor. A constant-displacement random-walk particle-tracking approach is employed to solve the transport equations quickly and accurately. In contrast to the scale-dependent constant-time-step scheme, the scale-independent constant-displacement scheme is implemented to calculate the travel time of particles aiming at the different scale's transport problems. With this scheme, the computation of travel times of particles is only done in specified steps within one cell, by which numerous computational times are saved in solving transport problems at the coarse scale and, moreover, the results are proved quite accurate compared to the constant time scheme (*Wen*

and Gomez-Hernandez, 1996a). In summary, these two forward simulators developed in this work, i.e., the multi-scale-oriented flow and transport solvers, pay much attention to the scale problem and hence are especially suitable for the MsBMcMC sampler. More implementation details are beyond this paper. Third, if the coarse model is accepted, then the fine model is run and the acceptance rate at the fine scale is evaluated to determine whether or not the proposal is accepted. If the coarse model is rejected, then a new proposal must be generated. Note that if, after the chain is convergent, the acceptance rate is above a threshold, say 25%, the multi-scale scheme ceases to play, i.e., only the operation at the fine scale is applied.

Up to now, a workflow for implementing the MsBMcMC scheme may be summarized in Figure 3.2.

3.3 Sensitivity Analysis

3.3.1 A synthetic example

Consider a 2D transient single-phase flow test on a confined aquifer with 32×32 grid-blocks as designed in Figure 3.3 under the forced-gradient flow condition. The reference ln$K$ field is generated by the LUSIM subroutine from the GSLIB (*Deutsch and Journel*, 1998) with a prior distribution ln$K \sim N(0,1)$ and an exponential variogram type, i.e.,

$$\gamma_{\boldsymbol{x}}(r) = \sigma^2_{\boldsymbol{x}} \left\{ 1 - \exp\left[-\frac{r}{\lambda_{\boldsymbol{x}}} \right] \right\}, \tag{3.24}$$

where r is the two-point separation distance, $\sigma^2_{\boldsymbol{x}}$ is the variance, and $\lambda_{\boldsymbol{x}}$ is the correlation length. The prior correlation length is set as $\lambda_{\boldsymbol{x}} = 16$ $[cells]$, which is much longer than the well spacing ($l_w \doteq 11$ $[cells]$) such that the spatial variability of ln$K$ is well captured by the well configuration.

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 days ($t_0 = 0$ 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.

Nine wells are drilled throughout this confined aquifer (Figure 3.3): four of them are the injection wells with a constant flow-rate ($q = 20.5$ per day) and the other five are production wells with a constant pressure (the piezometric head is maintained at $h = -3.0$ for all five wells). The flow-rate data at

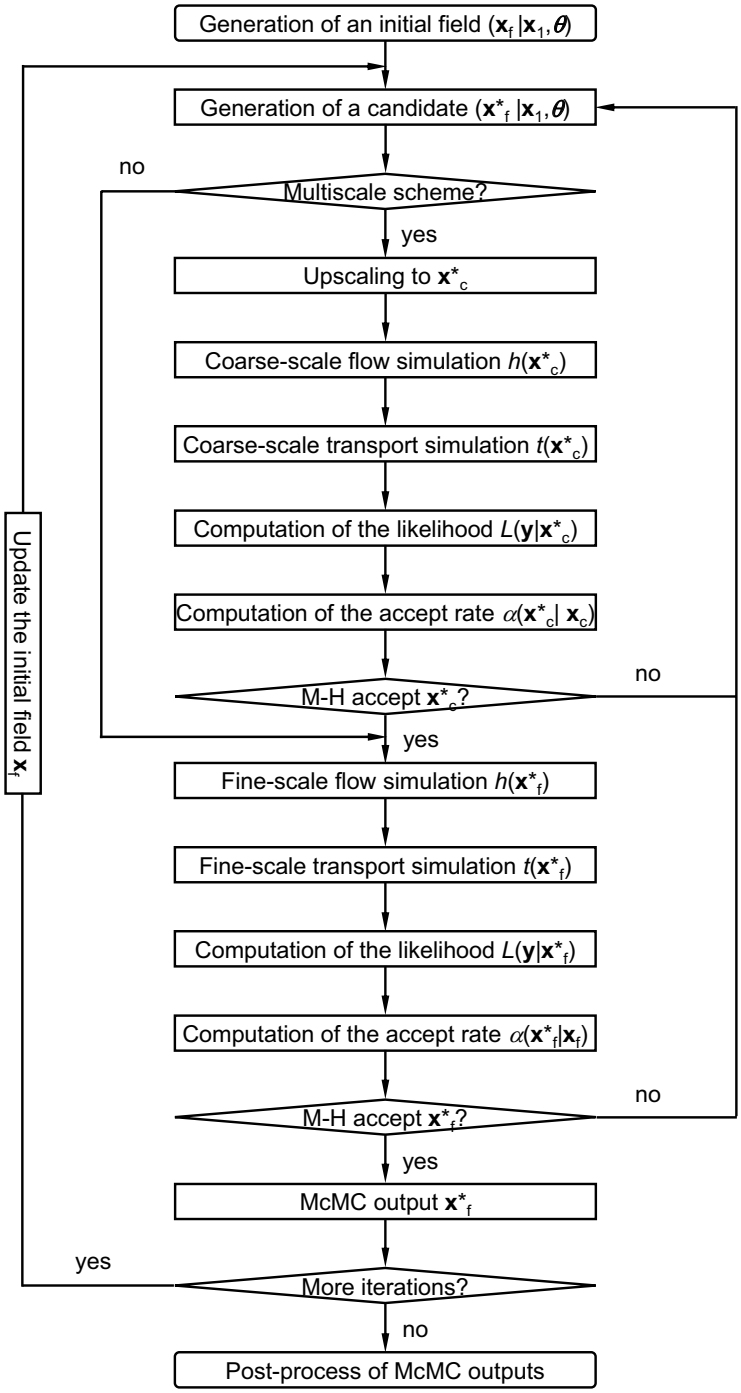


Figure 3.2: *Flowchart of the multi-scale blocking McMC scheme*

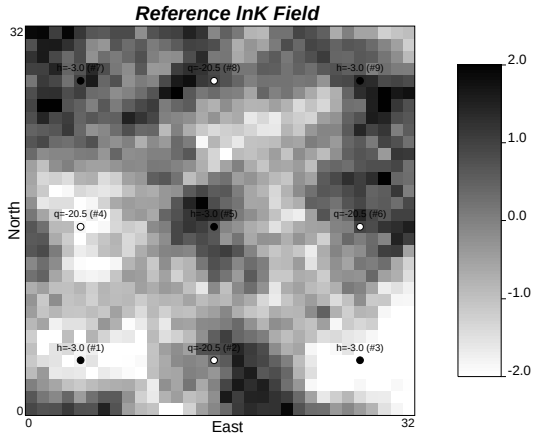


Figure 3.3: *Reference* ln$K$ *field: five constant pressure production wells* (*in bullet*) *and four constant flow-rate injection wells* (*in circle*)

the four injection wells and the bottom-hole-pressure (BHP) data at the five extraction wells are continuously collected at the first 40.1 days which consist of the first 50 time steps.

The inverse stochastic modeling problem, therefore, is to infer the permeability field (in ln$K$) according to the observed 40.1-day's flow-rate and BHP data at the well-bores. The stochastic ln$K$ fields are also required to be constraint to the given prior information, e.g., ln$K \sim N(0, 1)$ and $\lambda_{\boldsymbol{x}} = 16$. Other flow parameters are assumed to be constant and known perfectly.

3.3.2 Convergence velocity

Because the forward simulation $g(\boldsymbol{x})$ is usually very computationally demanding, a way fast convergent to the mode that the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ has is strongly recommended.

Effect of block size

It is well known that the block updating scheme helps improve the McMC convergence. Several numerical experiments are carried out to uncover the relationship between the block size and the convergence velocity. From Figure 3.4 (A), one can easily find that the block with a size equal to 8×8, which is one half of the correlation length ($\lambda_x = 16$), only needs several thousand iterations to reach the convergence, while the 1×1 (i.e., the correlated single-component case) and 2×2 block schemes almost require 10^5 iterations. Since each iteration calls for a forward evaluation, a larger updating block extremely

saves the computational resource. On the other hand, the updating block can not be large to any size. Figure 3.4 (B) shows that a block with a size over 8×8 does not improve the efficiency by too much and even becomes worse. In summary, a larger updating block with a dimension up to $\lambda_x/2$ very efficiently enhances the convergence velocity of the Markov chain.

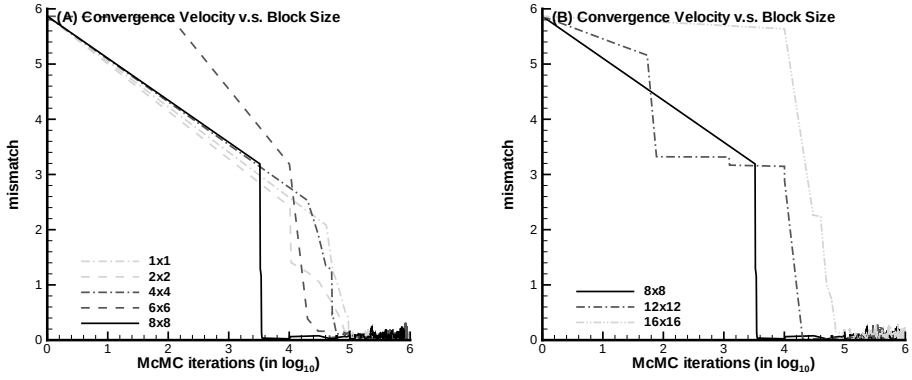


Figure 3.4: *Effect of the block size on convergence velocity of McMC: Note that the proposal of each of chain member requires a run of the forward simulator.*

Effect of McMC scheme

Figure 3.5 compares the influence of three schemes for constructing the Markov chain and computing the acceptance rate α on the convergence velocity of McMC. The case with a block size equal to 4×4 is used to evaluate such effect. First, the scheme #2 uses a subdomain only equal to one correlation length to approximate the prior density and well reproduces the result of the scheme #1. Even a slightly faster convergence velocity is observed in this case study. It shows that such approximation is reasonable even in McMC convergence efficiency. Second, the scheme #3 yields the fastest convergence result. Indeed, the scheme #3 represents the "steepest descent" path to searching and locating the region(s) of mode(s) that the posterior distribution has. As long as a model has a larger likelihood, it will be accepted into the Markov chain.

3.3.3 McMC estimation performance

After the chain converges to the target distribution, two types of problems are also deserved paying attention to: one is the estimation performance of the

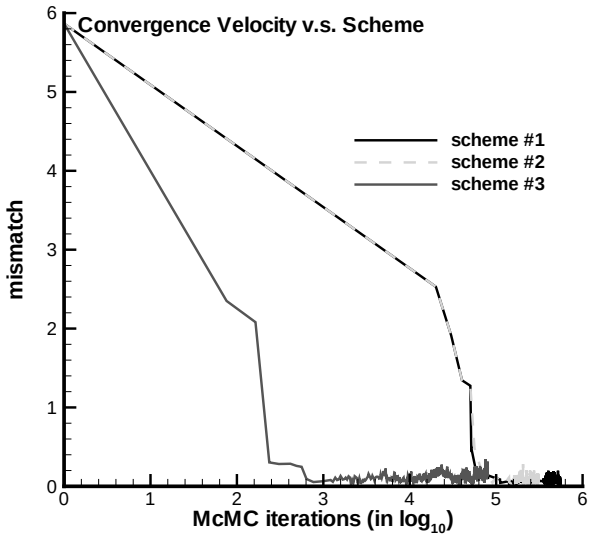


Figure 3.5: *Effect of the computation schemes for acceptance rate α on convergence velocity of McMC*

generated realizations and the other is the mixing speed of the Markov chain, i.e., its exploration efficiency.

Convergence performance measures

A natural 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, $\eta(\boldsymbol{x}) = (\eta(\boldsymbol{x}_0), \eta(\boldsymbol{x}_1), ..., \eta(\boldsymbol{x}_{n_r-1}))^T$, given a set of output realizations $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_{n_r-1})^T$ and an evaluation function $\eta(\cdot)$.

Based solely on a single replication, 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,

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

(2) Calculate the CUSUM,

$$\sigma_t = \sum_{i=0}^{t} (\eta(\boldsymbol{x}_i) - \bar{\eta}), \tag{3.25}$$

for $t = 0, 1, ..., n_r - 1$, and $\sigma_{n_r} = 0$;
(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 $i = 1, 2, ..., n_r - 1$;
(4) Calculate the hairiness indices,

$$\Sigma_t = \frac{1}{t-1} \sum_{i=1}^{t-1} \delta_i, \tag{3.26}$$

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

The key outputs are two sequential plots: $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{n_r})^T$ and $\Sigma = (\Sigma_2, \Sigma_3, ..., \Sigma_{n_r})^T$. 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.

Effect of block size

This part presents a numerical experiment to compare the efficiency of three block sizes to explore the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$. Figure 3.6 plots the mismatch between the observations and simulations, which is of main interest, after the chain is convergent. Figure 3.6 (A), (B) and (C) seemingly tell us that a smaller block produces a smaller mismatch. Note that the average mismatch of 2×2 is 0.08 which is slightly less than those of 4×4 ($\mu = 0.09$) and 8×8 ($\mu = 0.10$). On the other hand, the 4×4 block has a faster mixing speed than both the 2×2 block and the 8×8 block (see Figure 3.6 (D)).

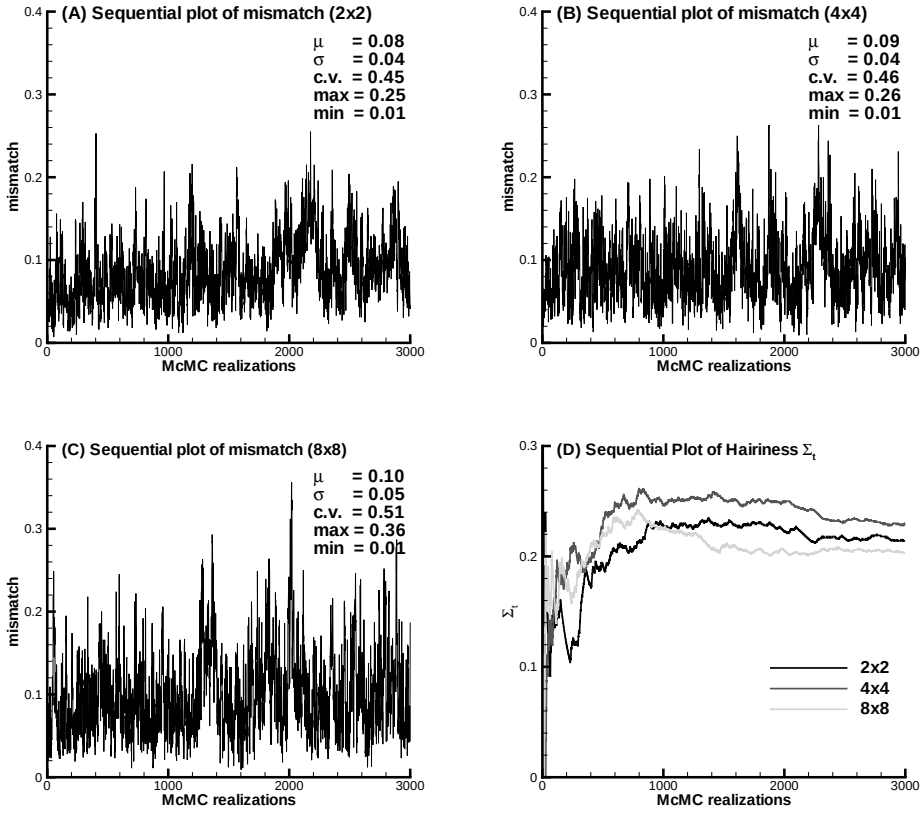


Figure 3.6: *Effect of the block size on the McMC performance*

Effect of McMC scheme

Figure 3.7 compares the influence of three blocking schemes on the McMC exploration performance and the mixing speed of the chain. One can easily find that the scheme #2 almost produces the same stable results as the scheme #1. Their basic statistics are almost the same (Figure 3.7 (A) and (B)). But the original version (i.e., scheme #1) does have a faster mixing speed than its improved version (i.e., scheme #2). However, the scheme #3 yields worse results. The simulated average mismatch ($\mu = 0.14$) is obviously larger than those of the other two schemes ($\mu = 0.09$). Moreover, its mixing speed is much less than the other two schemes (see Figure 3.7 (D)).

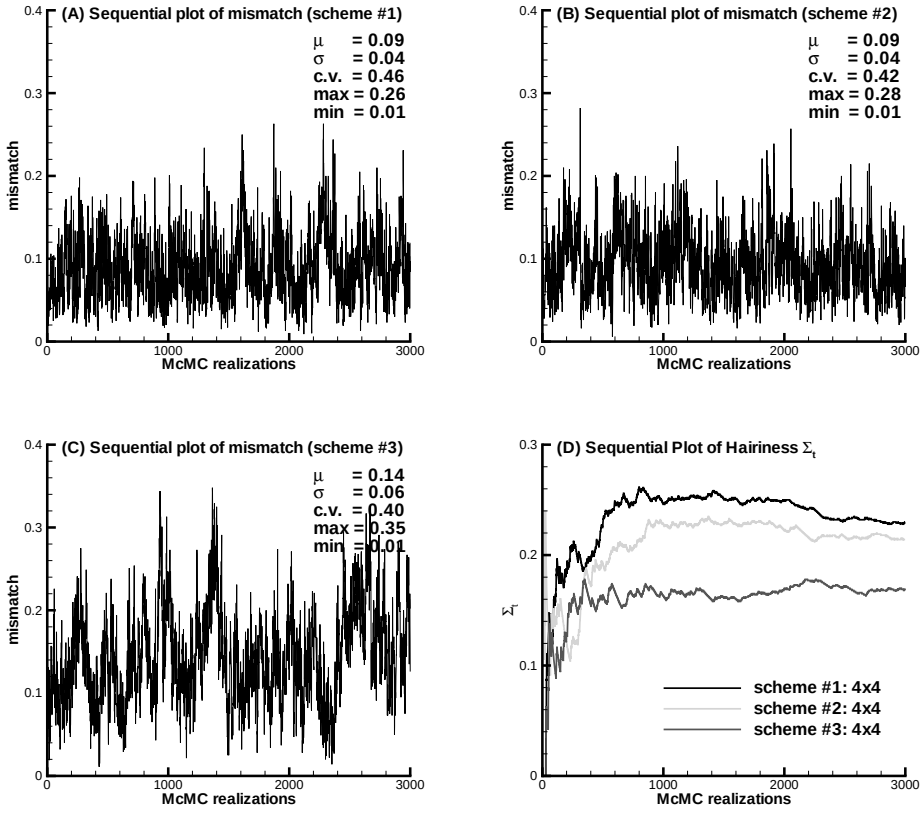


Figure 3.7: *Effect of three schemes for computing the acceptance rate on the McMC performance*

3.3.4 Sampling efficiency

It is also interesting to determine the velocity of generating *i.i.d* realizations after the chain is convergent since the forward simulation is expensive. One typical way to generating the independent realizations is to output one updated model once the updating block sweep off the whole modeling domain. Since the scanning path is random, an effective operation is to output one every n_{xyz}/b_{xyz} updates, where n_{xyz} and b_{xyz} are the dimensions of the model and the updating block, respectively. Figure 3.8 compares the relationship between the block size and the McMC iterations. The average iterations is computed on the basis of 3000 independent realizations. Each iteration calls for a forward simulation. Obviously, the 4×4 block ($b_x = \lambda_x/4$) requires the least forward evaluations while the correlated single-component proposal needs the most forward evaluations.

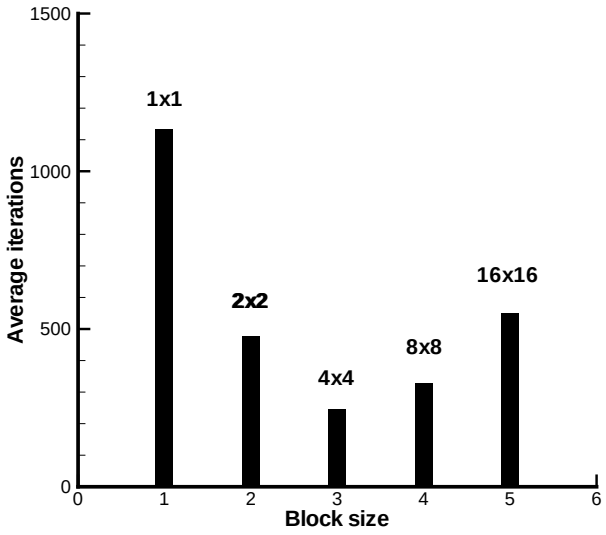


Figure 3.8: *Effect of the block size on the velocity of McMC in generating independent realizations*

In addition, from Figure 3.5 one may also find that to generating 1000 independent realizations the scheme #3 needs 8.0×10^4 iterations; the scheme #2 requires 3.0×10^5 iterations; and the scheme #1 almost needs 6.0×10^5 iterations.

3.4 Coupled Multi-scale BMcMC Methods

3.4.1 Scheme #4

Extensive numerical experiments show that a larger updating block up to $\lambda_x/2$ very efficiently improves the convergence velocity of the Markov chain. Compared to the original BMcMC scheme (scheme #1) and its improved version (scheme #2), the scheme #3 extremely speedups the convergence of the Markov chain. On the other hand, the scheme #1 (or #2) and a relatively smaller updating block ($b_x = \lambda_x/4$) give a better result in the mixing of chain and a faster velocity in generating independent realizations. Therefore, a mixed scheme may be proposed to combine the strong points. That is, use a larger updating block (e.g., $\lambda_x/2$) and the scheme #3 at the early stage in order to drive the chain to rapidly move to the region of the mode. After a "burn-in" period of trial runs and the chain is convergent, the proposal scheme switches to a smaller block (e.g., $\lambda_x/4$) and the scheme #1 or #2 to compute the acceptance rate. The advantage of such mixing multi-scale proposal scheme is that it can rapidly find the mode of the target distribution and the chain remains a fast mixing speed and a fast generation of independent realizations.

The synthetic example in Figure 3.3 is used to illustrate the efficiency of the proposed new scheme. The "burn-in" 50 iterations (almost 3000 forward simulations) are run by the scheme #3 with a 8×8 block which is a half of the correlation length. Then the blocking proposal scheme is switched to the scheme #2 with a 4×4 block which is one quarter of the correlation length. Figure 3.9 compared the combined multi-scale proposal scheme to the original single proposal schemes with various updating block sizes in generating 1000 independent realizations.

Three observations can be drawn from the picture. First, the new scheme (scheme #4) has the same convergence velocity as the scheme #3, which is much faster than the small updating block scheme (scheme #2). The scheme #3 and #4 only need 1.0×10^3 forward runs but the scheme #2 requires almost up to 1.0×10^5 forward runs. Second, the new scheme takes 1.5×10^5 forward runs to generate 1000 independent realizations, which is much faster than the scheme #2 (up to 3.2×10^5 forward runs). Third, the new scheme yields simulation results the same accurate as the scheme #2 and much better than the scheme #3. Note that the scheme #3 has a larger fluctuation than both the scheme #2 and the scheme #3 after the chains are convergent. This point may be see more clearly from Figure 3.10.

After the chains get convergent, each 3000 independent realizations are output from the scheme #2, #3, and #4 for measuring the estimate performance and the mixing speed of these three chains. The results are plotted in

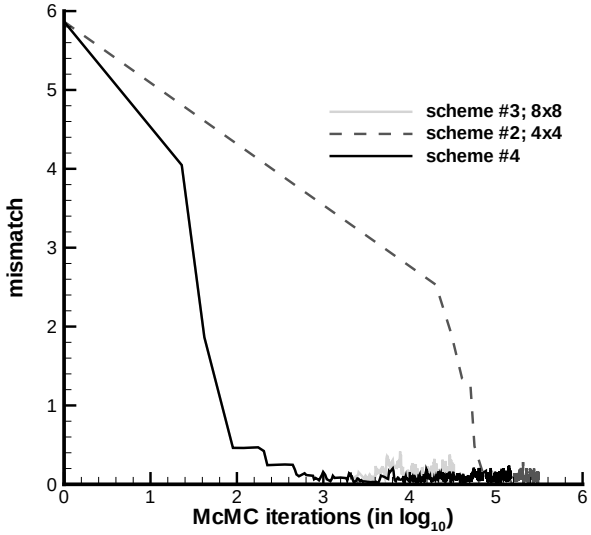


Figure 3.9: *An improvement in convergence velocity by the scheme #4*

Figure 3.10. The proposal new scheme (Figure 3.10(C)) produces the statistics of mismatch very similar to the scheme #2 (Figure 3.10(B)) but much better than the scheme #3 (Figure 3.10(A)). The index of convergence performance supports the same conclusion: the models from the scheme #4 are identical to those of the scheme #2 (see Figure 3.10(D)).

3.4.2 Scheme #5

The scheme #4 can be further extended to form a coupled BMcMC scheme: the scheme #5, that is, the information between the scheme #3 and the scheme #1 or #2 can be exchanged to improve the mixing of both.

Suppose that there are two separate chains that evolve independently: one is constructed by a large updating block with scheme #3, and the other has a small updating block with scheme #1 or #2 (see Figure 3.11). Every certain iterations, the two chains exchange information to form coupled Markov chains. Without considering the parallel implementation, the chain constructed by the scheme #3 is first run until it reaches the stable state, i.e., convergent; then the proposal scheme is switched to a smaller scale and the scheme #1 or #2 is used to output the independent realizations. After a period, the chain is switched to a larger scale proposal that uses the scheme #3 to locate a new mode and is switched back to the small scale proposal to output the realiza-

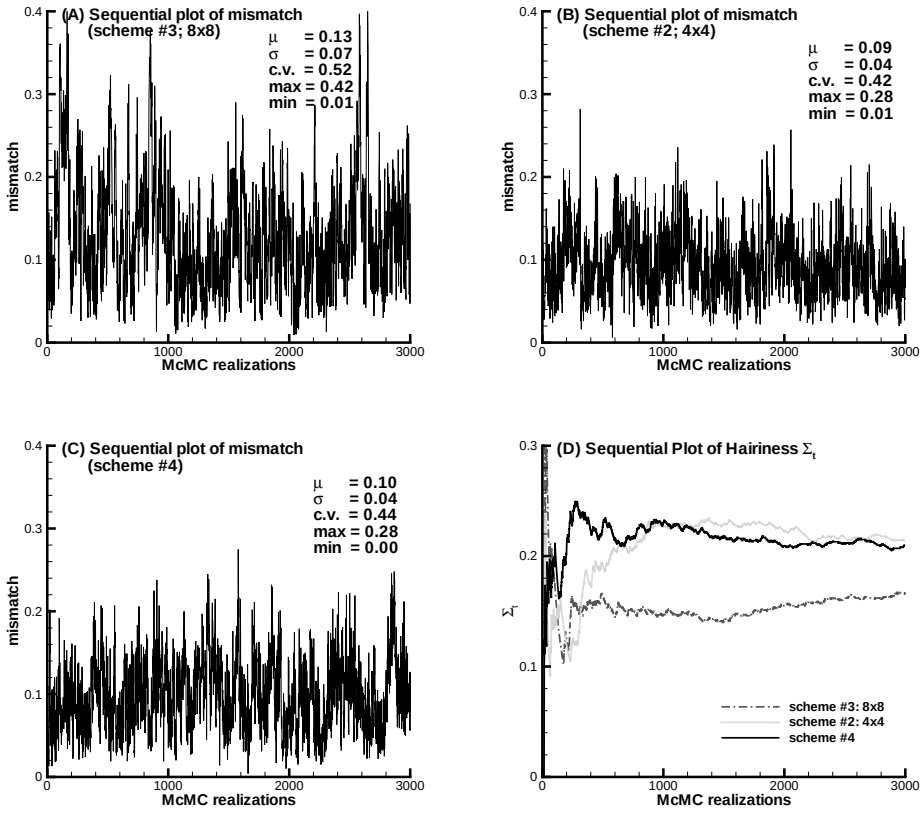


Figure 3.10: *An improvement in McMC performance by the scheme #4*

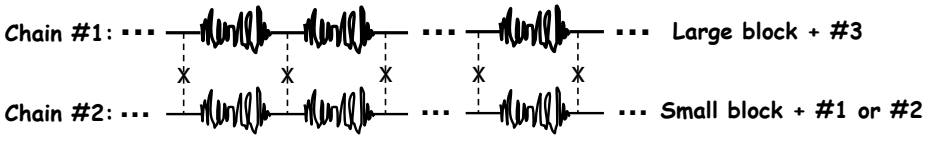


Figure 3.11: *Two coupled Markov chains*

tions. Repeating this progress that couples the large scale and the small scale proposals, a faster mixing speed is expected to attain.

Figure 3.12(A) displays the sequential plot of the mismatch produced by the scheme #5. The output realizations well match the best results as in the scheme #2 (Figure 3.12(B)). Figure 3.12(B) compares the mixing speed of the scheme #5 to that of the others. The exploration efficiency of McMC has been evidently enhanced by the coupled scheme.

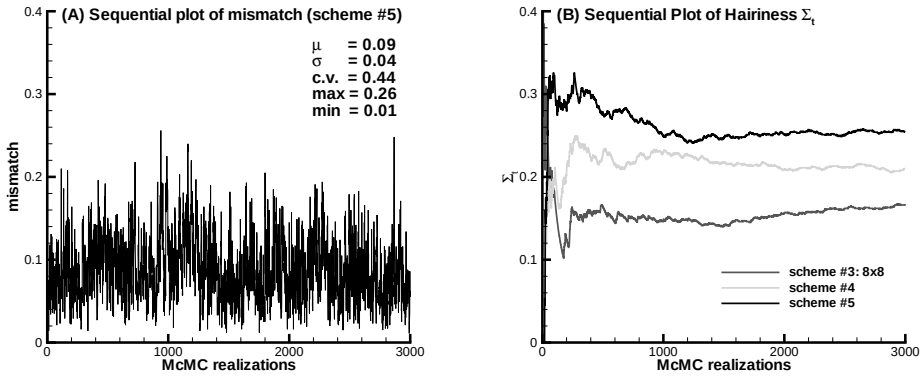


Figure 3.12: *An improvement in McMC performance by the scheme #5*

3.5 Conclusions and Discussions

A multi-scale blocking McMC method is presented to generate independent realizations that honor both the static measurements and the dependent dynamic observations. The blocking proposal kernel for the Monte Carlo integration is an appropriate approximation to the target posterior distribution due to the adoption of the blocking scheme, which efficiently improves the mixing of chain and hence saves a large number of computational times. The generation of candidate realizations is fast because the generation of the proposal kernel is based on the LU-decomposition of the covariance matrix for a small scale case and the spectral decomposition for a large scale case. The computation of the acceptance rate is economical since an upscaling procedure is integrated into the algorithm such that the coarse-scale model can be used to quickly evaluate the likelihood by the multi-scale-oriented flow and transport simulator developed in this study. The exploration ability and the mixing speed of McMC are improved owing to the coupling between the multiscale proposals.

In addition, the proposed method has several other advantages over classical inverse methods: (1) Unlike those conventional approaches, the MsBMcMC method is remarkably stable and well-behaved. It fully circumvents the difficulties of ill-posedness and identifiability as faced by the conventional optimization-based approaches. (2) It is easy to implement although a lot of cares should be taken in coding. (3) It has an ability to incorporate all linear and nonlinear data into the parameter models without invoking the failure-prone, usually highly nonlinear, optimization procedure as done as other commonly used inversion procedures. (4) It is not restricted to the assumption of a multi-Gaussian process for model parameters and can be easily extended to other stochastic process, such as, Markovian.

However, the disadvantages of the proposed method are also obvious. First, the presented MsBMcMC scheme is extremely computationally intensive. A large number of candidate realizations have to be provided to construct the Markov chain. Just like all the Monte Carlo methods, the number of realizations to be generated can not be efficiently determined in order to obtain a reliable result. Second, the efficiency of the proposed MsBMcMC method in exploring parameter space is desired to improve.

4

A Multi-scale Blocking McMC Method: 2. Uncertainty Assessment

Abstract

A multi-scale blocking McMC (MsBMcMC) method has been presented in Chapter 3 to perform conditional and inverse-conditional simulations honoring both the static data and the dynamic data. However, the purpose of the stochastic simulations to generating physical models is not only to reproduce the observations as precisely as possible but also to characterize model parameters between the observations in space and to predict model responses that have not yet been observed in space and time. Inevitably, an inherent uncertainty occurs in predicting the spatiotemporal distribution of model parameters and their responses. This study presents a synthetic example to validate the efficiency and robustness of the proposed MsBMcMC method for performing conditional and inverse-conditional simulations. Specifically, the models are configured *a priori*, either correctly or wrongly. Then the MsBMcMC approach is invoked to generate independent realizations conditioning on hard data and inverse-conditioning on state data, e.g., the head observations and the temporal moments of tracer data. Finally the estimated models and the predicted responses are compared with the real case. Results show that the hydraulic conductivity only contains local information on the spatial

distribution of model parameters while piezometric head and tracer data in conjugation with sampling network may convey regional or global information on the spatial trend of model parameters. The uncertainty propagation due to conditional and inverse-conditional mapping of aquifers is quantified in terms of the scale-dependent macrodispersion by means of solving assumed steady-state flow and ideal, conservative transport problems under a natural-gradient flow condition. It is found that inverse-conditioning to the temporal moments of BTCs, or equivalently, the spatial moments of concentration data, substantially improves the estimation on the solute spreading. The reduction of uncertainties proves not only the worth of hydraulic conductivity and piezometric head but also the significance of the temporal moments of tracer data in mapping model parameters and predicting model responses.

4.1 Introduction

In general, there are three methods for uncertainty assessment in inverse stochastic modeling: the linear (first-order second-moment) method, the nonlinear method, and the Monte Carlo method (*Carrera et al.*, 2005). Amongst them, the Monte Carlo method is probably the most widely used though computationally burdensome. The MsBMcMC is by nature a Monte Carlo method for uncertainty assessment but it also has the ability to constrain the models on both the static data and the dynamic data.

The groundwater flow inverse problem is to estimate parameters of physical model from head observations. The groundwater transport inverse problem consists of estimating parameters of physical model on the basis of concentration measurements. The concentration data for inverse stochastic modeling is attractive due to their abundance in aquifer measurements. However, it is a challenging task since the highly nonlinear relationship between the heterogeneity of hydraulic conductivity and the spatial variability of solute concentration, which are governed by the flow and transport equations.

The concentration data have been directly used to the parameter inference, e.g., characterization on the spatial correlation structure and the point estimation on local parameters, in several literatures (*Graham and McLaughlin*, 1989a, 1989b; *Sun and Yeh*, 1990a, 1990b; *Woodbury and Sudicky*, 1992; *Deng et al.*, 1993; *Anderman and Hill*, 1999; *Nowak and Cirpka*, 2006; etc.). For instance, *Ezzedine and Rubin* (1996) derived, in a geostatistical approach, the cross covariance between the tracer concentration data and the hydrogeological variables such as conductivity and head, which allows for the utilization of tracer data for estimating the spatial distribution of conductivity. *Franssen et al.* (2003) used the sequential self-calibration method to generate realizations conditional to the spatially distributed concentration data with the aid of the

adjoint-state method to calculate the sensitivity matrix. A synthetic study was presented to show the worth of concentration data.

However, the direct use of concentration data for inverse-conditional modeling generally requires a large of concentration samples extensively distributed over the entire flow domain over space and frequently sampled over time in order to obtain an exact description of spatial and temporal distribution of tracers, which is quite expensive and even impractical. An example is the Cape Cod tracer experiment where 9840 sampling points in the 656 monitoring wells were used to measure the concentration of tracers (*LeBlanc et al.*, 1991). By contrast, the travel time is a cheaper alternative for parameter inference to condition on the concentration data due to the low cost in the data acquisition. Indeed, collecting travel time of tracers only invokes couples of wells for forced-gradient flow or a series of wells distributed along a plane perpendicular to the mean flow direction for natural-gradient flow (*Rubin and Ezzedine*, 1997). *Fernandez-Garcia et al.* (2005) found that, even under a uniform, natural-gradient flow condition, only several full-penetrated wells are required to accurately estimate the first two moments of BTCs obtained from total mass fluxes passing through the control planes. Additionally, there are also some merits in the computational aspects (*Harvey and Gorelick*, 1995). For example, the travel times are scale-independent and thus avoid the disparity problem between the model resolution and the measurement scale since the travel times are typically computed in a Lagrangian framework rather than the grid-based Eulerian method when solving the forward transport problem.

Several literatures have been found for conditioning aquifer parameters on the travel time. *Vasco and Datta-Gupta* (1999) developed an asymptotic solution to the solute transport, in a single forward simulation, to calculate the sensitivities for the inversion of the tracer data. Then an iterative linearized inversion algorithm is used to infer the parameter distribution. *Wen et al.* (2002) derived sensitivity coefficients of tracer travel time with respect to the permeability by tracking streamlines between the well pairs. The sequential self-calibration method is then employed to construct geostatistical realizations conditional to concentration data. Results from a synthetic aquifer show that tracer concentration data carry important information on the spatial variation of permeability in the inter-well areas while the pressure data only provide information near the well-bore.

In contrast to the entire BTCs, a variety of statistical measures computed from the BTCs, e.g., the peak concentration arrival times, the percentiles of travel times, and the temporal moments of tracer data, can also be used for the inverse-conditional simulation and mitigate the computational effort. Several methods based on the temporal moments and statistics of the BTCs have been used for parameter inference. *Cirpka and Kitanidis* (2001) developed a sensitivity matrix of the temporal moments of tracer data with respect to

the conductivity using the adjoint-state method. On the basis of such sensitivity matrix (of the first moment), the quasi-linear geostatistical inversion or iterative cokriging method is employed to conditioning the conductivity on the tracer data. A synthetic example demonstrates a minor improvement of the integration of tracer data (in terms of the first temporal moment) into the estimate of conductivity compared to the result of head data. *Rubin and Ezzedine* (1997), *Woodbury and Rubin* (2000), and *Bellin and Rubin* (2004) proposed to use the peak concentration arrival times to infer the geostatistical models of conductivity. Actually, public officials assessing health risks associated with contaminant exposure in a drinking water supply system may be most concerned with peak concentration or the corresponding arrival time (*Lemke et al.*, 2004). Moreover, one appealing point in data acquisition is that the peak concentration arrival time is less affected by the truncated BTC records, e.g., the missing early or late arrivals due to the infrequent sampling and the insensitivity of measurements. *Wilson and Rubin* (2002) used the indicator variable of solute arrivals for the inference of parameters controlling the heterogeneous structure of conductivity and the mean flow velocity.

The meaningfulness of conditioning on the various percentiles of the BTCs is apparent in physics. The early arrivals in the BTCs follow the fastest pathways between the release source and the control plane, which are dominated by preferential flow, i.e., flow conduits. On the other hand, the late travel times reflect a more integral behavior, or even flow barriers. Therefore, different inversion results provide distinct knowledge about the flow and transport properties. High connectivity generally results in earlier breakthrough, i.e., an earlier front part of BTC. Failing to account for such case will have too conservative conclusion in risk analysis in that the real arrival time may be too much faster than that estimated (*Gomez-Hernandez and Wen*, 1998). On the other hand, low connectivity results in later breakthrough. An aquifer remediation design without considering such feature may fail because the resident contaminants will be removed more slowly than expected (*Wagner and Gorelick*, 1989). *Harvey and Gorelick* (1995) presented a method for estimating the spatial pattern of conductivity from the quartiles of solute arrive times. In a hypothetical aquifer example, they found that adding the median quartile of the BTCs to the cokriging procedure does improve the accuracy of the estimate of conductivity. But the tails of the BTCs (0.1 and 0.9 percentiles in their case) do not convey much more information about the conductivity field than the median quartile on the basis of the first-order approximation of the flow and transport equations.

The transformations of the raw measurement data of dependent state variables as the input to the inversion procedure have their own advantages. First, such preprocess decreases the inconsistency or reduces the discrepancy between the raw data, which helps to improve the stability of the inversion procedure,

but not of the methodology itself, and the identification of the subsurface reality. This is straightforward since the acquisition of raw data contain a huge number of errors. One advantage of using temporal moments as conditioning data, for example, is that the derivation of travel time moments from the raw BTC can efficiently average out the measurement errors. An alternative is to assign a pdf to define measurements instead of individual raw values. Second, it may increase the sensitivity of state data to the model parameters. Third, it may convey a more useful information for engineering design. For instance, a surfactant flooding scheme requiring delivery of surfactant at concentrations exceeding a specified threshold for a minimum time period might be more concerned with second temporal moment, i.e., dispersion. The first purpose of this study, therefore, is to constrain the stochastic models on the moments of BTCs by the MsBMcMC.

The construction of the 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. Chapter 3 has already presented a detailed description on the MsBMcMC approach to this problem. Of equal importance, on the other hand, is to carry out 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 (*Vasco et al.*, 1997; *Deutch and Journel*, 1998), so as to measure how much the property parameters can depart from the conditional realizations. Note that the purpose of quantitative uncertainty analysis is not to reduce uncertainties which can only be achieved by collecting additional effective information. However, not all conditioning algorithms are sensitive to, and thus are capable of detecting such uncertainty reduction introduced by, additional effective information. In this regard, quantification of uncertainty can adversely check the efficiency of a conditioning algorithm.

The second purpose of this work, therefore, is to investigate the worth of various types of data and to assess the uncertainty reduction caused by the conditional and inverse-conditional simulations in a synthetic example under a natural-gradient flow condition. First, the worth of local conductivities, piezometric head and travel time data is evaluated by the errors of the generated realizations deviated from the real model in terms of the spatial distribution of the hydraulic conductivity following the line of *Franssen et al.* (2003). Then, the predicted uncertainties of head distribution due to the conditioning to various types of data are assessed. Finally, the uncertainty propagation of conditional simulations is quantified in terms of the longitudinal macrodispersion to further validate the worth of various measurements. It is worth pointing out that since the model structure of realizations generated by the MsBMcMC are identical, the worth of conditioning data from diverse sources are evaluated exclusively without structure uncertainty involved.

4.2 A Synthetic Example

The first purpose of this exercise is to examine the efficiency of the proposed MsBMcMC algorithm to generating independent conditional realizations. It should be mentioned that although the initial and boundary conditions are subject to the uncertainty as well, we assume they are known or estimated reasonably in order to ease the computational burden. In addition, although the head and concentration data may provide information for identifying the spatial distribution of other parameters, such as porosity and retardation factor, this study only considers the inverse-conditional problem of hydraulic conductivity since the spatial variability of conductivity predominantly controls the flow of fluid and transport of solutes. In other words, porosity, retardation factor and other parameters are treated as the constants or known.

4.2.1 Reference model and conditioning data set

A synthetic 2D highly-correlated multi-Gaussian confined aquifer under a uniform, natural-gradient flow condition, as shown in Figure 4.1 (A), serves as the reference field to illustrate the effectiveness of the proposed method for inverse-conditional simulation.

The multi-Gaussian field for reference is chosen because of its simplicity and traceability in performing input and output analysis. Its spatial structure can be described completely by its first two moments. This reference $\ln K$ field is generated by the program SGSIM from GSLIB (*Deutsch and Journel*, 1998). The grid size of the entire computational domain is 100×100, i.e., $n_x = n_y = 100$. The size of each square cell is one by one unit-free, i.e., $\Delta x = \Delta y = 1$. The flow domain thus is $l = l_x = l_y = 100$. The mean value is set to zero and its variance is set to one, that is, $x \sim N(0, 1)$. An exponential correlation structure without nugget effect is specified. The $\ln K$ field is isotropic and the correlation length of both x and y direction is 50, i.e., $\lambda_x = \lambda_y = 50$, which is half of the domain size and hence highly correlated.

The high correlation of the aquifer model is preferred in this study for two reasons: (1) it provides more straightforward visual comparison between the simulations and the reference due to the block property of images and (2) only a few observations are required to capture the critical features of aquifer for easing the computational burden although the proposed method has potential to deal with a large number of data. In this paper, the high correlation means a high ratio of the correlation length to the size of the computational domain while the grid resolution refers to the cell number per correlation length. For the $\ln K$ field in this study this ratio is $l : \lambda_{\boldsymbol{x}} = 100 : 50$. The grid resolution is 50 by 50 per correlation length, i.e., $R_\lambda = 50 \times 50$, which is enough to characterize the spatial variability in details. However, the high correlation

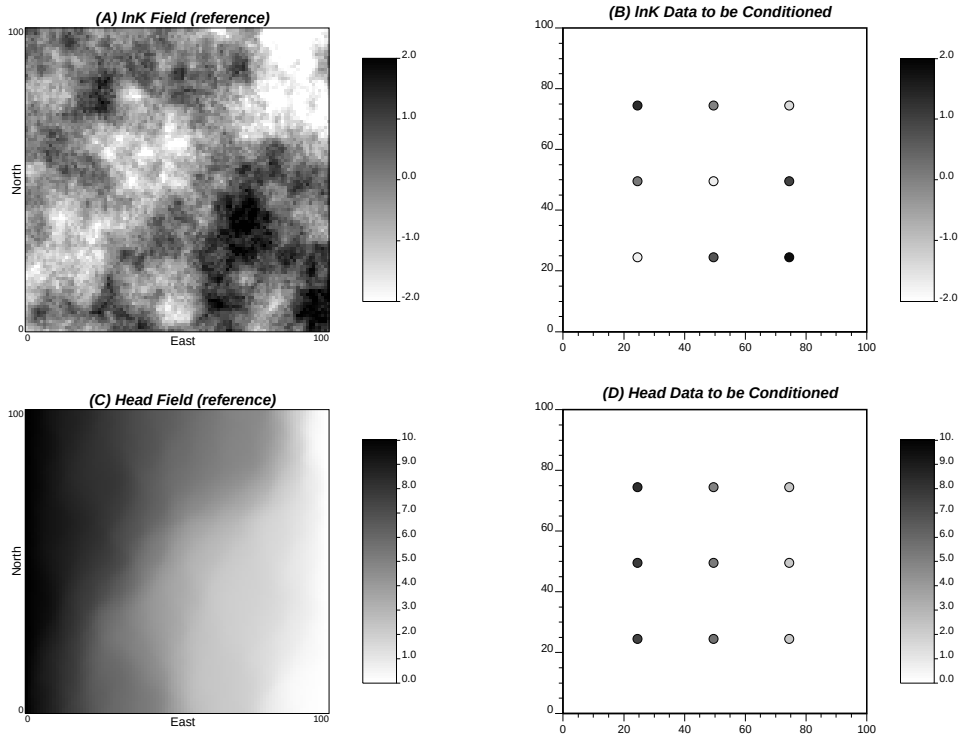


Figure 4.1: *The conditioning data set and the reference* $\ln K$ *and head fields* $(\mu = 0, \lambda_{\boldsymbol{x}} = 50, \sigma_x^2 = 1.0)$

tightens the dependence of parameters and decreases the degree of freedom of aquifer models, which makes more limited the admissible parameter space and hence increases difficulty for the MsBMcMC algorithm to search for acceptable models.

Another factor to be considered is the ratio of the correlation length to the sampling spacing because the sampling density may be an important factor for model identification. Generally speaking, compared to the sampling spacing, a smaller prior correlation length yields an inverse estimate closer to the prior mean except in the immediate vicinity of measurements (*McLaughlin and Townley*, 1996). In principle, parameter variability with a correlation length smaller than the sampling spacing cannot be determined. Conversely, a larger prior correlation length yields a smooth estimate which varies gradually over the region of interest. Besides, it was recognized that the uncertainties of models and their responses remain significant even with a large number of hard conditioning data (*Harter and Yeh*, 1996; *van Leeuwen et al.*, 2000; *Scheibe and Chien*, 2003). *Eggleston et al.* (1996) found that for the Cape Code aquifer the estimation error is relatively insensitive to the number of hard data above a threshold of three measurements per integral volume. For the $\ln K$ field in this study, this ratio is $\lambda_{\boldsymbol{x}} : \Delta s = 50 : 25$ which keeps two sampling points per correlation length so that the main characteristics of the real aquifer are captured by these measurements.

A single-phase flow experiment is set up as follows. The upper and lower boundaries are set as impermeable, i.e., non-flow boundaries. The left and right sides are set as constant heads equal to ten and zero, respectively, i.e., $h_{-x} = 10$ and $h_{+x} = 0$. The confined steady-state flow problem is solved by a multiscale-oriented block-center finite-difference simulator developed in this study. The reference head field is shown in Figure 4.1 (C). Nine points uniformly distributed in the flow domain are selected as the conditioning positions. The sampling interval thus is $\Delta s = \Delta s_x = \Delta s_y = 25$. Nine $\ln K$ values are shown in Figure 4.1 (B) as hard data to be conditioned. Nine head values at the same locations are shown in Figure 4.1 (D) as state data to be inverse-conditioned.

Therefore, the problem of interest is to infer the spatial distribution of $\ln K$ from nine head observations under the same flow and transport conditions by the MsBMcMC method. Moreover, the spatial structure specified for models should be preserved for all inverse-conditional realizations.

4.2.2 An inverse-conditional estimate

This part presents a preliminary result by inverse-conditioning to nine head observations just to show the effectiveness of the proposed MsBMcMC scheme. A critical pattern of the reference model, as shown in Figure 4.1 (A), is that

the low $\ln K$ values, mostly locating at the left-lower corner, spread along the north-west direction while the high $\ln K$ values locate mainly at the right-lower corner and some of them at the left-right corner. The main feature of the head distribution, as shown in Figure 4.1 (C), is that the flow breakthrough of the lower part is much faster than that of the upper part. Nine head observations uniformly scattered over the field, as shown in Figure 4.1 (D), basically captures the main flow pattern. Therefore, the identified $\ln K$ field by inverse-conditioning on the head data is expected to be capable of reproducing the critical patterns of the reference models.

The main input parameters to the MsBMcMC program for inverse-conditioning are configured as follows. Except for the nine head data, the $\ln K$ is assumed as the unknown. Therefore, the output realizations are only inverse-conditioned to head observations. The expected value is set to zero and the variance is set to one, $x \sim N(0, 1)$. The model structure of $\ln K$ is assumed to be estimated unbiasedly, i.e., an isotropic field is assumed and the covariance is simulated by an exponential function with a correlation length equal to 50, i.e., $\lambda_{\boldsymbol{x}} = 50$, whence all parameters keep consistent with the reference model. The parameter configuration for flow simulations is also assumed to be known perfectly, that is, a steady-state single-phase flow is simulated with boundary conditions consistent with the reference model.

Three thousand independent realizations are generated by the MsBMcMC program. Figure 4.2 well displays the Gaussian distribution of $\ln K$ values, i.e., $\ln K \sim N(0, 1)$. Figure 4.3 shows the variograms in two orthogonal directions. Obviously the generated realizations have an identical variogram as specified *a priori*. Note that the inverse conditional simulations perfectly match the prior model at a small separate distance. Over a large separate distance, the generated realizations tend to slightly adjust their variograms to match the real case (reference).

The ensemble mean $\ln K$ field over these 3000 realizations may represent the most probable estimate by the MsBMcMC method. As shown in Figure 4.4 (A), the main distribution pattern of the reference $\ln K$ field (see Figure 4.1 (A)) has been well identified in visual though the amplitude of the mean estimate is weaker than that of the reference field. The most probable estimate is obviously over-smoothed and underestimated since local details tend to be smoothed out during the procedure of inverse-conditioning with only several sparse data, which is just as done as kriging.

The norm of the posterior covariance matrix represents the actual error of parameters in uncertainty quantification. Figure 4.4 (B) plots the variance of the most probable estimate over an ensemble of three thousand realizations from the MsBMcMC output. It just shows the reliability distribution of the $\ln K$ estimate since a smaller variance demonstrates a narrower region that $\ln K$ is allowed to fluctuate over. In turn, it also displays the degree of reliability of

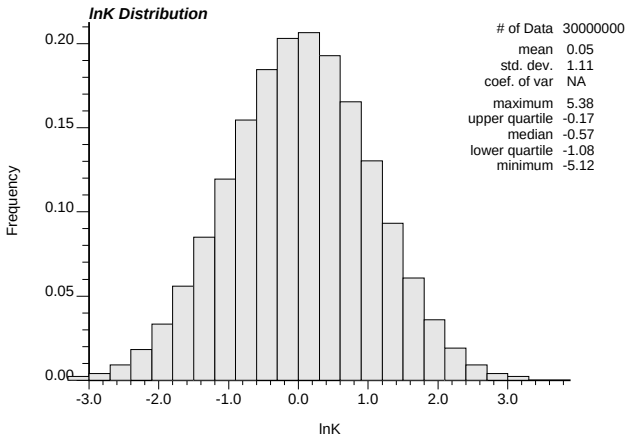


Figure 4.2: $\ln K$ *distribution of the inverse-conditional simulation*

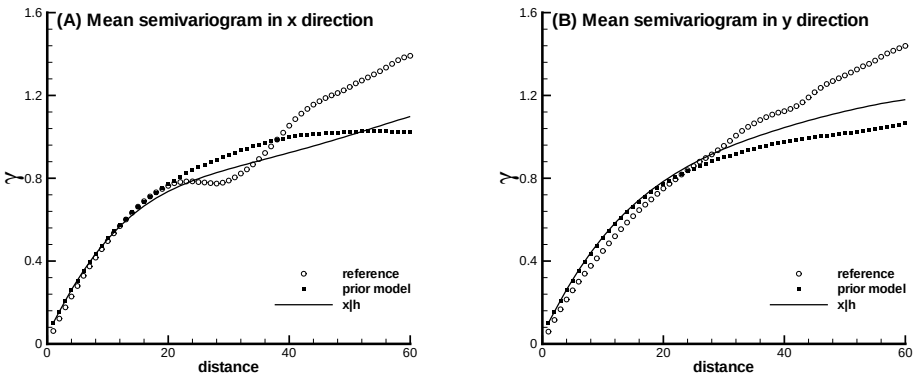


Figure 4.3: *Theoretical semivariograms of the inverse-conditional simulation*

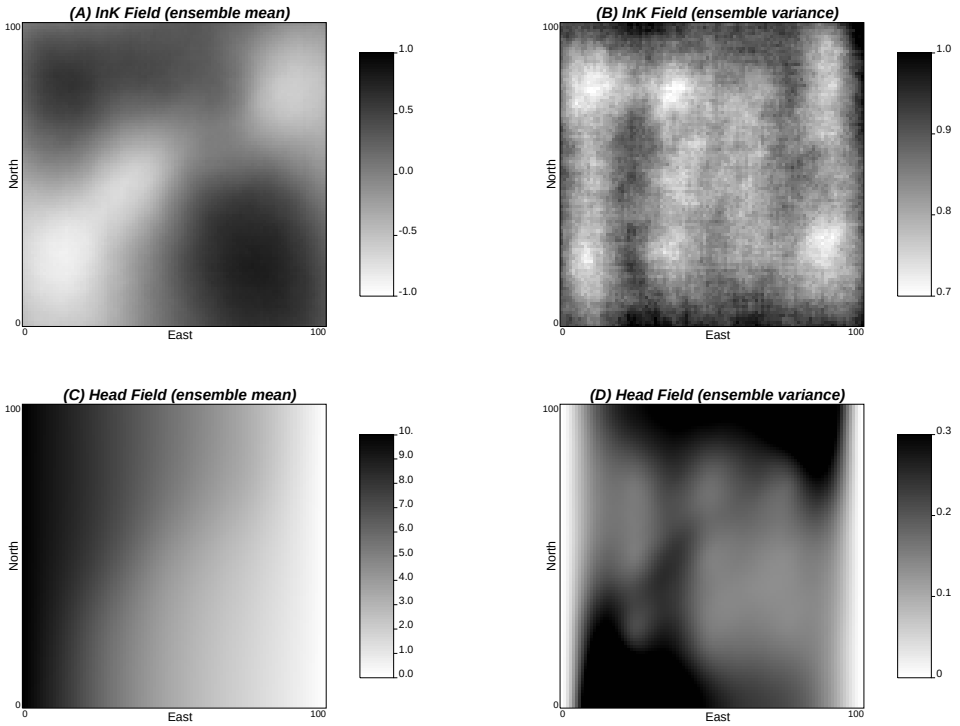


Figure 4.4: *An MsBMcMC estimate by inverse-conditioning to head observations*

the most probable estimate of $\ln K$. The posterior variance obviously decreases compared to the prior one due to inverse-conditioning. Such decrease is not uniform in space, however, because of the non-uniform sensitivity of head observations to model parameters.

One could easily find from Figure 4.4 (B) that the inter-well region has a more reliable estimate than the well-bore region (observation locations for state variables) and that the worst estimates have been obtained close to the boundary regions. This is because in this case study both the well-bore region and boundaries have specified heads that inverse-conditional models have to honor. The sensitivity of heads to parameters is very low near the locations with constant heads since there is no change of heads responding to the fluctuation of parameters. Therefore, the uncertainties at those locations are not significantly reduced by inverse-conditioning to their own head observations. But the uncertainties can be reduced as benefiting from inverse-conditioning to head observations at other positions.

The same flow experiments as done to the reference model are conducted for the ensemble of 3000 realizations. The mean head field from the batch flow experiments is plotted in Figure 4.4 (C), which is quite similar to the reference head field, e.g., the lower part has earlier breakthrough than the upper part yet the estimated distribution is more regular and smooth. The variance of estimated head fields is plotted in Figure 4.4 (D), which displays the fluctuation sizes of heads. Note that the values on the left and the right boundaries, are not the real cases due to the boundary effect. The reasoning is that because the head field is highly correlated, just as a natural result of the highly correlated $\ln K$ field, and these two boundaries have specified values, i.e., ten for the left and zero for the right, the head values on the regions close to these boundaries are more determined from the boundary specification rather than the stochastic fluctuation identified by inverse-conditioning. The constant head configuration prevents the simulated head from changing significantly for different parameter values. In other words, the sensitivity of heads to parameters is low near the boundaries with constant heads. The regions of high head variance are those locations that new observations should be placed on in the new network design for reducing the uncertainty.

In summary, head observations, albeit sparse in space, do contain important flow information for identifying the large-scale trend of $\ln K$, which can be effectively extracted by the proposed MsBMcMC inverse-conditioning algorithm.

4.3 Uncertainty Assessment

Two types of uncertainties are considered in this section, model uncertainty and prediction uncertainty. The model uncertainty is important at the spatiotemporal scale not only because the generated models form the basis for future performance prediction at the existing wells but also because they serve as risk analysis of candidate wells. The prediction uncertainty directly measures the prediction ability of models at the spatiotemporal scale. In addition, the purpose of the assessment on uncertainty propagation is straightforward, that is, to determine which sources of uncertainty contribute mostly to the uncertainties of the predicted responses and hence to further design experiments to reduce the most critical components. Assuming that the uncertainty from numerical simulations is negligible and considering that the uncertainty of data acquisition is uncontrollable, the model uncertainty (e.g., from different model structures, such as, Gaussian vs. Markovian, and from different structure parameters, such as, mean, variance, correlation length, sill, and anisotropy for the Gaussian case) is essentially responsible for the resulted prediction uncertainty.

4.3.1 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 ($I(\boldsymbol{x})_1$) and the standard deviation of the ensemble average error ($I(\boldsymbol{x})_2$), which are defined as the L_1-norm and L_2-norm between the simulated models and the mean models, i.e.,

$$I(\boldsymbol{x})_1 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r} - \bar{x}_i|, \tag{4.1a}$$

$$I(\boldsymbol{x})_2^2 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r} - \bar{x}_i)^2, \tag{4.1b}$$

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

In case of the synthetic example like 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 (*Deng at al.*, 1993; *Frassen*

et al., 2003). In such case, the L_1-norm and L_2-norm between the simulated models and the real models are defined as,

$$I(\boldsymbol{x})_3 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r}^{sim} - x_i^{ref}|, \tag{4.2a}$$

$$I(\boldsymbol{x})_4^2 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r}^{sim} - x_i^{ref})^2, \tag{4.2b}$$

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

4.3.2 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, $I(\boldsymbol{y})_1$, $I(\boldsymbol{y})_2$, $I(\boldsymbol{y})_3$, and $I(\boldsymbol{y})_4$, are defined as follows,

$$I(\boldsymbol{y})_1 = \|\boldsymbol{y}_{sim} - \bar{\boldsymbol{y}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r} - \bar{y}_i|, \tag{4.3a}$$

$$I(\boldsymbol{y})_2^2 = \|\boldsymbol{y}_{sim} - \bar{\boldsymbol{y}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r} - \bar{y}_i)^2, \tag{4.3b}$$

$$I(\boldsymbol{y})_3 = \|\boldsymbol{y}_{sim} - \boldsymbol{y}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r}^{sim} - y_i^{ref}|, \tag{4.3c}$$

$$I(\boldsymbol{y})_4^2 = \|\boldsymbol{y}_{sim} - \boldsymbol{y}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r}^{sim} - y_i^{ref})^2, \tag{4.3d}$$

respectively.

In essence, $I(\boldsymbol{x})_1$, $I(\boldsymbol{x})_2$, $I(\boldsymbol{y})_1$, and $I(\boldsymbol{y})_2$ measure the degree of precision that the McMC simulations could render, that is, how narrow the confidence interval of McMC simulations is. $I(\boldsymbol{x})_3$, $I(\boldsymbol{x})_4$, $I(\boldsymbol{y})_3$, and $I(\boldsymbol{y})_4$ measure the degree of accuracy that the McMC simulations may attain, that is, how they are close to the true model and it response. From the standpoint of estimate and uncertainty, $I(\boldsymbol{x})_1$, $I(\boldsymbol{x})_3$, $I(\boldsymbol{y})_1$, and $I(\boldsymbol{y})_3$ measure the reliability of the estimated models and their responses while $I(\boldsymbol{x})_2$, $I(\boldsymbol{x})_4$, $I(\boldsymbol{y})_2$, and $I(\boldsymbol{y})_4$ measure the uncertainty of the estimates and their responses.

4.3.3 Macrodispersion

In addition, the macrodispersion coefficient can be viewed as an important parameter to represent the spatial variability of hydraulic conductivity. Indeed, experimental and theoretical results have suggested that macrodispersion of solutes is essentially produced by the spatial variation of the fluid velocity resulting from the heterogeneity of hydraulic conductivity. Because the physical models inverse-conditional to the concentration data and the head data are expected to identify the main flow and transport pattern, the objective of this study is to validate the capability of the conditioning algorithm employing the McMC scheme to capture such features by comparing the macrodispersion of the conditional realizations with that of the unconditional realizations.

The scale- or time-dependent macrodispersion is defined as the change rate of the second-order moment of a solute plume. Extensive studies on the effects of hydraulic conductivity on macrodispersion of solutes have shown that, under steady-state flow conditions with a uniform mean hydraulic gradient in the statistically stationary media of finite correlation length of hydraulic conductivity, macrodispersion increases with time from the point at which the solute body first enters the flow domain, until after the solute cloud has traveled a few tens of correlation length of the hydraulic conductivity and then reaches a constant asymptotic value (*Dagan*, 1984; *Khaleel*, 1994).

One possible way to estimating the macrodispersion of solutes is to carry out a series of Monte Carlo numerical simulations to calculate the temporal moments at various displacement distances using the statistics of the BTCs and transverse drifts. The longitudinal and transverse macrodispersion can be then computed from the particle arrival times and arrival transverse positions to those distances, respectively. In this study, only the longitudinal macrodispersion is considered. Specifically, given an ensemble of aquifer realizations, say, n_r, one can get a set of BTCs at different control planes, say, n_c, perpendicular to the mean fluid velocity, i.e., $\{\boldsymbol{t}(x_0), \boldsymbol{t}(x_1), ..., \boldsymbol{t}(x_{n_c-1})\}_r, r \in [0, n_r)$, by solving the steady flow and transport problems. The longitudinal macrodispersion for a given control plane can be calculated from the temporal moments of an ensemble of BTCs by (*Fernandez-Garcia et al.*, 2005),

$$A_L(x) = \frac{x}{2} \frac{\langle \sigma_t^2(x) \rangle + \langle \sigma_T^2(x) \rangle}{\langle T(x) \rangle^2}, \tag{4.4}$$

where x denotes the travel distance (i.e., the distance between the control plane and the tracer source in the x direction) and $\langle\rangle$ denotes the ensemble average operator. In the formula above, $\langle \sigma_t^2(x) \rangle$ is the expected variance of travel time, $\langle \sigma_T^2(x) \rangle$ is the expected variance of mean travel time, and $\langle T(x) \rangle^2$ is the expected average travel time. A detailed procedure for computing these three values is given as follows.

(1) Expected average travel time $\langle T(x)\rangle$

$T(x)$ is the first-order moment of travel time for one single realization, i.e., the average arrival time of a given realization, which can be approximated by,

$$T_r(x) = \frac{1}{n_{p,r}} \sum_{i=0}^{n_{p,r}-1} t_{r,i}(x); \qquad r \in [0, n_r),$$

where $n_{p,r}$ is the total number of tracer particles that arrive at the control plane and $t_{r,i}(x)$ is the travel time of the ith particle for the rth realization. $\langle T(x)\rangle$ is the ensemble mean of the average arrival time over all the realizations, which can be approximated by,

$$\langle T(x)\rangle = \frac{1}{n_r} \sum_{r=0}^{n_r-1} T_r(x).$$

(2) Expected variance of travel time $\langle \sigma_t^2(x)\rangle$

$\sigma_t^2(x)$ is the second-order moment of travel time for one single realization, i.e.,

$$\sigma_{t,r}^2(x) = \frac{1}{n_{p,r}} \sum_{i=0}^{n_{p,r}-1} t_{r,i}^2(x) - T_r^2(x); \qquad r \in [0, n_r),$$

and, then, the ensemble mean $\langle \sigma_t^2(x)\rangle$ is,

$$\langle \sigma_t^2(x)\rangle = \frac{1}{n_r} \sum_{r=0}^{n_r-1} \sigma_{t,r}^2(x).$$

(3) Expected variance of average travel time $\langle \sigma_T^2(x)\rangle$

$\sigma_T^2(x)$ is the variance of the first temporal moment, i.e., the squared deviation of average arrival time of a single realization from the ensemble average arrival time,

$$\sigma_{T,r}^2(x) = \left(T_r(x) - \langle T(x)\rangle\right)^2; \qquad r \in [0, n_r),$$

and, then, the ensemble mean $\langle \sigma_T^2(x)\rangle$ is,

$$\langle \sigma_T^2(x)\rangle = \frac{1}{n_r} \sum_{r=0}^{n_r-1} \sigma_{T,r}^2(x).$$

An alternative way to calculating the apparent longitudinal macrodispersion is to use the coefficient of variation of the BTCs (*Kreft and Zuber*, 1978; *Desbarats and Srivastava*, 1991; *Wen and Gomez-Hernandez*, 1998),

$$A_L(x) = \frac{x}{2}\frac{\sigma_t^2(x)}{m_t^2(x)}, \tag{4.5}$$

where $m_t(x)$ and $\sigma_t^2(x)$ are the mean and variance of travel times at the displacement distance x, respectively. To overcome the sensitivity of $\sigma_t^2(x)$ to the presence of outlier travel times, the distribution of log travel time was suggested to calculate the temporal moments,

$$m_t(x) = \exp\left\{m_{\ln t}(x) + \frac{1}{2}\sigma_{\ln t}^2(x)\right\},$$
$$\sigma_t^2(x) = m_t^2(x)\left(\exp\left(\sigma_{\ln t}^2(x)\right) - 1\right),$$

following *Khaleel* (1994) and *Wen and Gomez-Hernandez* (1998). $m_{\ln t}(x)$ and $\sigma_{\ln t}^2(x)$ are the mean and variance of log travel times at the displacement distance x, respectively.

4.4 A Synthetic Experiment

The stochastic conditional simulation to generating models is not only to reproduce the observations as precisely as possible but, even more importantly sometimes, to characterize parameters where we do not know such as at those points between the observations and to make inference to the future performance of wells, either old or virtual. Correspondingly, an inherent uncertainty occurs in predicting the spatiotemporal distribution of parameters and their responses, which are called model uncertainty and prediction uncertainty, respectively. A classical method for inverse stochastic modeling includes two stages, model identification and parameter estimation. Although the McMC method may merge these two stages into one procedure by automatic model selection, this study intentionally separates them for comparing and validating the efficiency and robustness of the proposed MsBMcMC in uncertainty assessment. Specifically, the models are configured *a priori*, either correctly or wrongly. Then the McMC approach is invoked to generate independent realizations conditioning on the hard data and inverse conditioning on the state data, e.g., the head observations and temporal moments of tracer data. Finally the model uncertainty and the prediction uncertainty are quantified in comparison with the assumed real case.

4.4.1 Experimental configurations

Starting from the $\ln K$ field, once the head field is obtained by solving the flow problem, the Darcian velocity field at cell's interface can be established by applying the Darcy's law. The porosity field is assumed to be homogenous with a constant, $\phi = 0.3$. The conservative transport problem is then solved by the constant-displacement random-walk particle-tracking algorithm in the Lagrange framework as implemented in Chapter 2. The effect of pore-scale dispersion is neglected, i.e., $\alpha_L = \alpha_T = 0$. In such case, the solute particles are convected along the streamlines of the steady velocity field. Hence the solute plume is confined transversally by the two no-flow boundaries. Molecular diffusion is also neglected in this example problem. Two thousand particles, randomly uniformly distributed on the left boundary, are tracked until they arrive at the control plane located at the right boundary. The travel time is recorded at the control plane and forms a pdf to describe the transport property of conservative mass. The statistics of the BTC computed from the reference travel time distribution are shown in Table 4.1 as state data to be inverse-conditioned. The tracer test designed as such is based on two considerations: (1) it allows to fully capture the globally spatial variability of $\ln K$ and (2) the observation well system is rejected since the natural-gradient flow pattern may cause a rather low particle capture rate at the well-bores.

Table 4.1: *Statistics of reference travel time*

Statistics	Travel time
05 percentile of the BTC	912.39520
25 percentile of the BTC	1013.97559
50 percentile of the BTC	1081.44373
75 percentile of the BTC	1232.37927
95 percentile of the BTC	1823.62537
1st moment	1192.83350
2nd moment	98379.28125

Therefore, the aim of conditional and inverse-conditional problems is to infer the spatial distribution of $\ln K$ from nine $\ln K$ measurements, nine head observations and statistics of travel time under the identical flow and transport conditions by the MsBMcMC method.

Totally six scenarios of stochastic simulations and numerical experiments are carried out as listed in Table 4.2. To ease the computational burden, the relative errors of head and travel time data are assumed to be rather large. The relative resolution of head observations is set as $\sigma_h^2 = 0.2$ which means

that the head field may fluctuate around the given head observations with standard deviation equal to 0.45 times observations. In other words, the head observations h_i $(i \leq k)$ are corrupted with noises $0.45 \times h_i$. The relative resolution of travel time is set as $\sigma_t^2 = 1.0$ which means that travel times may fluctuate around the given residual time observations with standard deviation equal to 1.0 times observations. Owing to the multi-Gaussian property of the $\ln K$ field, the moments of BTC may simply replace the entire BTC as the conditioning data. This set of parameter configurations for MsBMcMC make possible all of the simulations and computations to run with a cheap PC.

Table 4.2: *Parameter configuration for stochastic simulations*

Scenario	Model	McMC configuration
1	$\boldsymbol{x}\|-$	$\mu_{\boldsymbol{x}} = \mathbf{0}$, $\sigma_x^2 = 1$, $\lambda_{\boldsymbol{x}} = 50$
2	$\boldsymbol{x}\|\boldsymbol{x}_1$	$\mu_{\boldsymbol{x}} = \mathbf{0}$, $\sigma_x^2 = 1$, $\lambda_{\boldsymbol{x}} = 50$, $\sigma_{x_1}^2 = 0.0$
3	$\boldsymbol{x}\|\boldsymbol{h}$	$\mu_{\boldsymbol{x}} = \mathbf{0}$, $\sigma_x^2 = 1$, $\lambda_{\boldsymbol{x}} = 50$, $\sigma_h^2 = 0.2$
4	$\boldsymbol{x}\|\boldsymbol{t}$	$\mu_{\boldsymbol{x}} = \mathbf{0}$, $\sigma_x^2 = 1$, $\lambda_{\boldsymbol{x}} = 50$, $\sigma_t^2 = 1.0$
5	$\boldsymbol{x}\|\boldsymbol{h}, \boldsymbol{t}$	$\mu_{\boldsymbol{x}} = \mathbf{0}$, $\sigma_x^2 = 1$, $\lambda_{\boldsymbol{x}} = 50$, $\sigma_h^2 = 0.2$, $\sigma_t^2 = 1.0$
6	$\boldsymbol{x}\|\boldsymbol{x}_1, \boldsymbol{h}, \boldsymbol{t}$	$\mu_{\boldsymbol{x}} = \mathbf{0}$, $\sigma_x^2 = 1$, $\lambda_{\boldsymbol{x}} = 50$, $\sigma_{x_1}^2 = 0.0$, $\sigma_h^2 = 0.2$, $\sigma_t^2 = 1.0$

For each scenario, the MsBMcMC is invoked to generate 100 *i.i.d* $\ln K$ realizations by assuming that the other parameters are free from uncertain and known perfectly. A batch of flow and transport simulations that are configured in the same manner as the reference case are run to build the spatiotemporal distribution of head fields and travel times. The model uncertainty metrics, $I(\boldsymbol{x})_1$, $I(\boldsymbol{x})_2$, $I(\boldsymbol{x})_3$, and $I(\boldsymbol{x})_4$, the response uncertainty metrics, $I(\boldsymbol{h})_1$, $I(\boldsymbol{h})_2$, $I(\boldsymbol{h})_3$, and $I(\boldsymbol{h})_4$, and the scale-dependent macrodispersion coefficient $A_L(x)$ are computed by the Equation 1 through 5, respectively. The importance of hydraulic conductivity, piezometric head, and travel time (moments) on reduction of model uncertainty and prediction uncertainty is analyzed as follows.

4.4.2 Model uncertainty

Figure 4.5 shows the spatial distributions of the mean $\ln K$ due to conditioning data from various sources. In visual, inverse-conditioning to head data does give rise to a correct large-scale trend about the $\ln K$ distribution, e.g., Scenario 3 and 5, although the amplitudes are much less than the reference case which has been plotted in Figure 4.1(A). As predicted, the conditioning data of $\ln K$ only convey some local information even though the field is highly correlated (see Scenario 2). Inverse-conditioning on travel time moments should help

identify the regionally spatial pattern of physical models, but Scenario 4 does not give an ideal image as expected. One possible reason is due to the network design of tracer experiment for recording the travel time. Since the tracer test is designed to cover the entire flow region, fully from the left to the right boundary, the travel time record only reflects an average transport property of the physical model. No further information about spatial configuration of model parameters could be extracted from one single BTC record. However, it adversely proves the importance of sampling density and sampling network design in model identification and uncertainty reduction.

Figure 4.6 shows the uncertainty of the estimate given in Figure 4.5. Conditioning on $\ln K$ measurements only makes the uncertainty of the estimate reduced at the local regions around the measurement points (Scenario 2). The uncertainty reduction due to inverse-conditioning to head and/or travel time observations is fully regional (Scenario 3-5). Jointly conditioning to conductivity and inverse-conditioning to head and/or travel times does reduce both the local and global uncertainties (Scenario 6).

It is worth noting that Figure 4.5 and Figure 4.6 seemingly tell us that both the head data and the tracer data may convey useful information on the spatial trend of $\ln K$ distribution fully at the regional scale. However, *Wen et al.* (2002) found that the tracer breakthrough data carry out important information on the spatial variation of $\ln K$ in the inter-well areas while the pressure data provide information at near well-bore areas. This discrepancy is probably because (1) *Wen et al.*'s observations are based on the radial, forced-gradient flow condition while the example presented in this study simulates the uniform, natural-gradient flow condition, (2) the sampling network design for recording the BTCs is in part responsible for such disagreement, and (3) the $\ln K$ field is highly correlated such that the model responses at the observation points is also "kriged" or propagated to the neighborhood. Indeed, the pressure (head) transmission and tracer transport are inherently diffusive processes and hence are largely governed by the average conditions rather than by the local heterogeneities. For a natural-gradient case, the sensitivity of h and t to $\ln K$ is completely areally effective. For a forced-gradient case, however, the effectively propagated region is apparently limited to a certain vicinity near the wellbores because the local conditions prevail and the sensitivity of h and t to $\ln K$ is relatively undetectable.

Besides, from the Scenario 4 and 5 in Figure 4.6, one can find that the uncertainty reduction in the upstream is larger than that in the downstream after inverse-conditioning to temporal moments, which agrees to the finding presented by *Franssen et al.* (2003) whose observation is based on inverse-conditioning to spatial concentration data using the sequential self-calibration method. The reason why it happens is still not very clear to the authors but

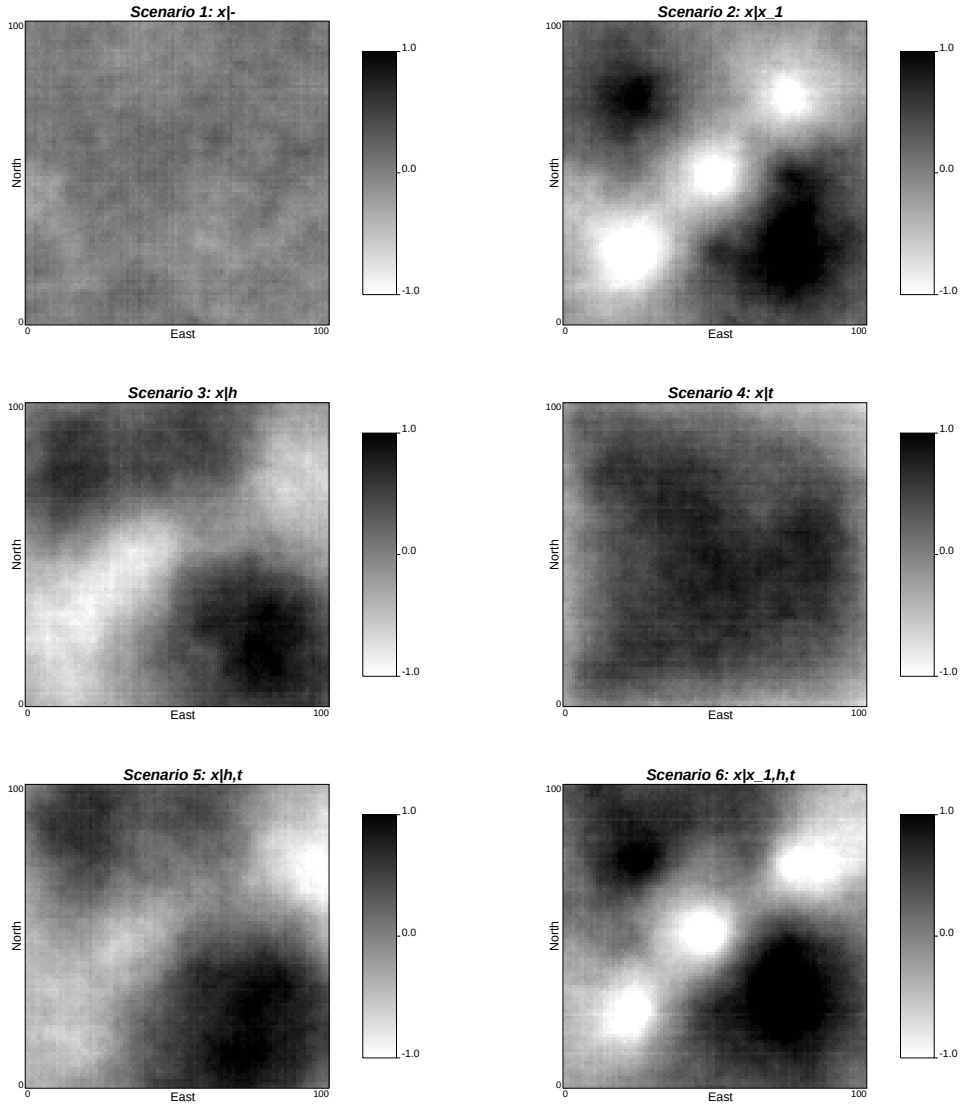


Figure 4.5: *The mean* $\ln K$ *fields due to conditioning to various source data*

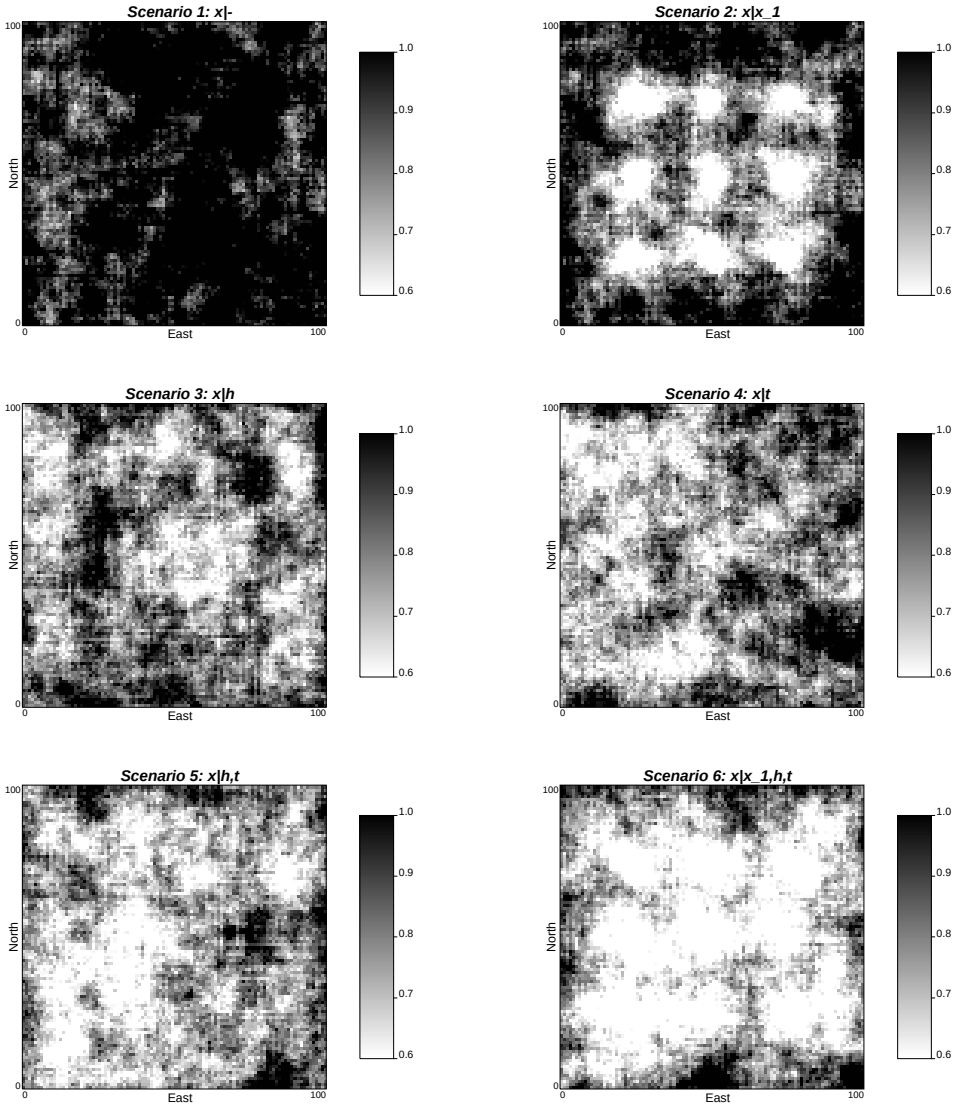


Figure 4.6: *The variances of* $\ln K$ *fields due to conditioning to various source data*

Franssen et al. (2003) contributed this phenomenon to more sensitivity of concentration to flow velocity in the upstream zone.

Table 4.3 summarizes four metrics for measuring the credibility and uncertainty of conditional and inverse-conditional simulations. Although it is unfair to compare the absolute worth of distinct source data since the measurement errors are most often involved, it is important to know their essential significance on model identification and uncertainty reduction. Basically, the collection of direct measurements of $\ln K$ is extremely important for the configuration of the prior model parameters which most effectively reduce model uncertainty while the collection of state data helps identify the spatial trend of physical models. From Table 4.3, it is hard to tell which data source is more important than the other although hard data (Scenario 2), head data (Scenario 3), and travel time moments (Scenario 4) do enhance the reliability of estimates ($I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_3$) and decrease the uncertainty of models ($I(\boldsymbol{x})_2$ and $I(\boldsymbol{x})_4$) in comparison with the unconditional case (Scenario 1). Besides, the sampling density may also play an important role in uncertainty reduction. In this example problem, head data (Scenario 3) seems to reduce the model uncertainty slightly larger than hard data (Scenario 2) due to a rather dense sampling spacing, i.e., two samples per correlation length. Inverse-conditioning on travel time moments (Scenario 4) gives a rather unreliable estimate, e.g., $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2$ are inconsistent with $I(\boldsymbol{x})_3$ and $I(\boldsymbol{x})_4$. But just because the data from diverse sources contain distinct information on physical models, joint integration of multi-source data can efficiently improve the estimates to the greatest extent, e.g., Scenario 5 and 6.

Table 4.3: *Mean absolute error and mean variance of* $\ln K$

Scenario	Model	$I(\boldsymbol{x})_1$	$I(\boldsymbol{x})_2^2$	$I(\boldsymbol{x})_3$	$I(\boldsymbol{x})_4^2$
1	$\boldsymbol{x}\|-$	0.8170	1.0210	1.1539	1.4484
2	$\boldsymbol{x}\|\boldsymbol{x}_1$	0.7262	0.9135	0.9836	1.2444
3	$\boldsymbol{x}\|\boldsymbol{h}$	0.7122	0.8913	0.9749	1.2244
4	$\boldsymbol{x}\|\boldsymbol{t}$	0.7032	0.8812	1.1364	1.4241
5	$\boldsymbol{x}\|\boldsymbol{h},\boldsymbol{t}$	0.6800	0.8537	0.9450	1.1871
6	$\boldsymbol{x}\|\boldsymbol{x}_1,\boldsymbol{h},\boldsymbol{t}$	0.6446	0.8127	0.8905	1.1297

In addition, correct prior configurations for spatial statistics and model structure (especially the variance of parameter variables and the spatial correlation length) also play important roles in reducing model uncertainty. From the experience of the authors, the correct configurations for the variance of parameter variables and the spatial correlation length obviously render the better reproduction of the reference models compared to the wrong specifications.

Figure 4.7 and Figure 4.8 plot the histograms and variograms of $\ln K$ distributions before and after conditioning on diverse source data in this case study. The $\ln K$ fields generated by MsBMcMC obviously follow the prior specification, e.g., $\ln K \sim N(0,1)$, which significantly differ itself from the traditional inverse methods. The theoretic variograms in two orthogonal directions also have identical distribution as the prior model especially at smaller separate distance. A slight deviation from the prior specification may be due to the fact that the structure parameters of the reference model are not perfectly identified by the prior model. Therefore the generated realizations slightly adjusts their variogram parameters to match the reference model, which is more obvious at the larger separate distance. But, in general, the reference model, the generated realizations and the prior model share identical structure parameters.

4.4.3 Response uncertainty

Once the physical models of $\ln K$ are generated by the MsBMcMC, the flow and transport experiments may apply to these realizations to predict the model responses at the spatiotemporal scale, e.g., the piezometric head distribution. Figure 4.9 compares the predicted steady-state head distribution. Conditioning on hard data (Scenario 2) seems to reproduce the main points of the reference head distribution, but it does not reflect the real worth of hard data in model response prediction. The truth is that nine samples uniformly distributed over the field produce a kriging field that can sufficiently catch the main pattern of the real field since the $\ln K$ field is highly correlated ($\lambda_{\boldsymbol{x}} = 50$). Inverse-conditioning on travel time moments (Scenario 4) obviously fails to reproduce the spatial pattern of head. It does not mean that the travel times do not carry out much useful information for model response prediction. The sampling density and network configuration is mostly responsible for such failure since the travel time under the uniform flow condition is only an integrated response of $\ln K$ variability in space. One single BTC does not have enough ability to recover the spatial configuration of model parameters. As expected, joint integration of diverse source data does improve the prediction of model responses to the greatest extent, e.g., Scenario 5 and 6.

Figure 4.10 plots prediction uncertainties of model responses. Conditioning merely on hard data (Scenario 2) and inverse-conditioning only to travel time moments produce responses that remain rather large uncertainties. Inverse-conditioning to head observations, either solely or jointly, reduces the uncertainty much more than other types of data do, which seems to say that the head observations have the largest worth in reducing the uncertainty of head prediction.

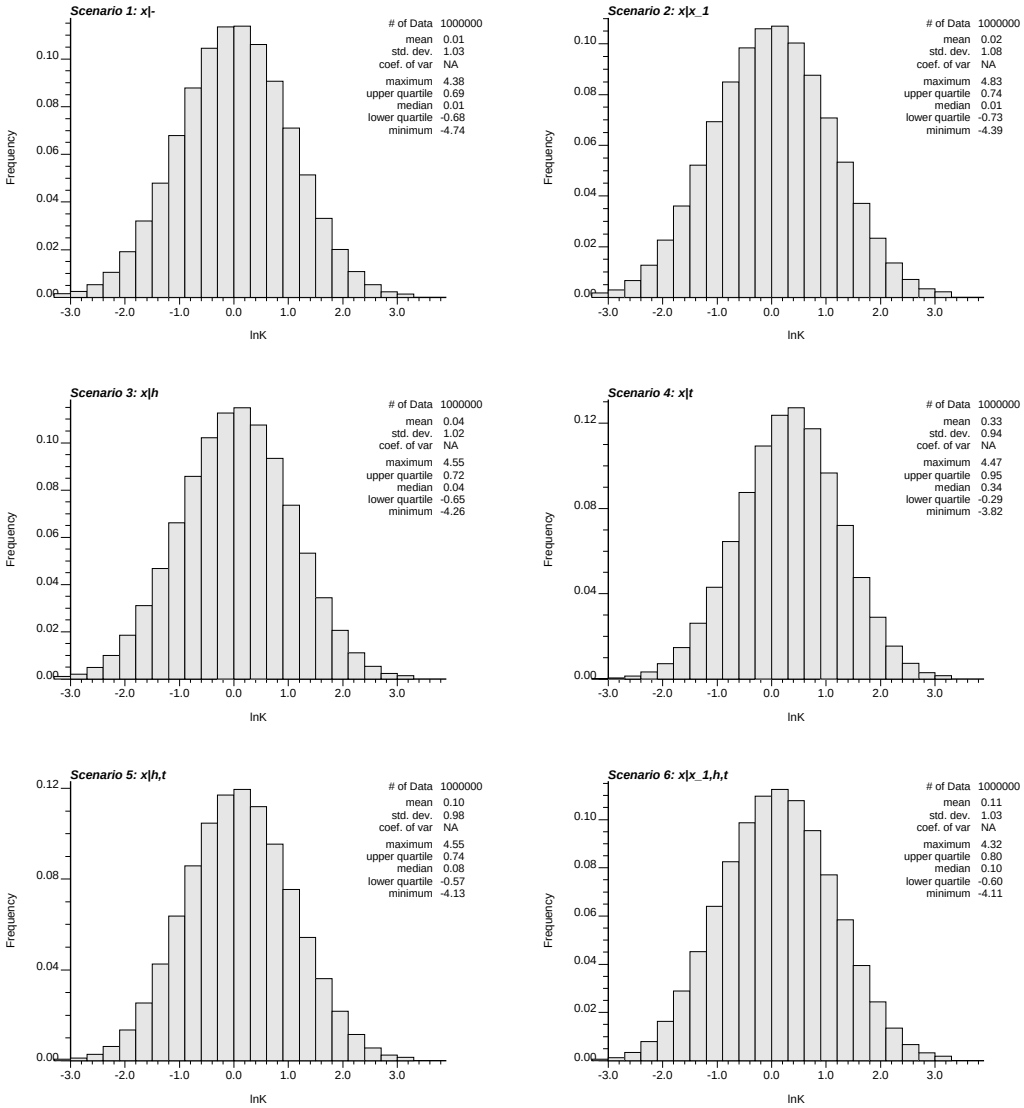


Figure 4.7: *The histograms of* $\ln K$ *distribution after conditioning to data from various sources*

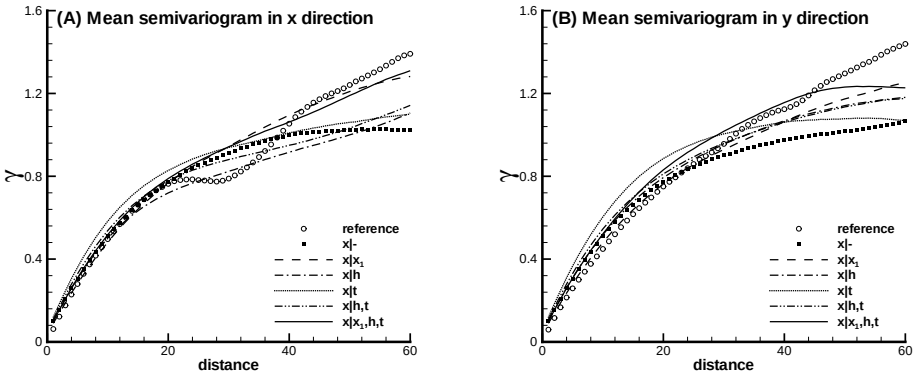


Figure 4.8: *The variograms of* ln$K$ *distribution after conditioning to data from various sources*

Table 4.4 summarizes four metrics for measuring the reliability and uncertainty of predicted head distributions. It seemingly shows that inverse-conditioning on the head observations gives better reliability and less uncertainty than conditioning on ln$K$ and inverse-conditioning on travel time. If both the head data and the others (ln$K$ and travel time) are honored, the uncertainty reduces to the largest extent.

Table 4.4: *Mean absolute error and mean variance of predicted head*

Scenario	Model	$I(\boldsymbol{h})_1$	$I(\boldsymbol{h})_2^2$	$I(\boldsymbol{h})_3$	$I(\boldsymbol{h})_4^2$
1	$\boldsymbol{x}\|-$	0.7499	0.9889	1.0021	1.3594
2	$\boldsymbol{x}\|\boldsymbol{x}_1$	0.5344	0.7292	0.6514	0.8837
3	$\boldsymbol{x}\|\boldsymbol{h}$	0.3017	0.3961	0.4110	0.5504
4	$\boldsymbol{x}\|\boldsymbol{t}$	0.5071	0.6532	0.8550	1.1164
5	$\boldsymbol{x}\|\boldsymbol{h},\boldsymbol{t}$	0.3196	0.4130	0.4378	0.5749
6	$\boldsymbol{x}\|\boldsymbol{x}_1,\boldsymbol{h},\boldsymbol{t}$	0.2703	0.3620	0.3429	0.4712

In addition, from Table 4.3 and Table 4.4, one can easily find that the McMC simulations always underestimate the real uncertainties and give more optimistic predictions than they are. Note that $I(\boldsymbol{x})_1 < I(\boldsymbol{x})_3$, $I(\boldsymbol{x})_2 < I(\boldsymbol{x})_4$, $I(\boldsymbol{h})_1 < I(\boldsymbol{h})_3$, and $I(\boldsymbol{h})_2 < I(\boldsymbol{h})_4$.

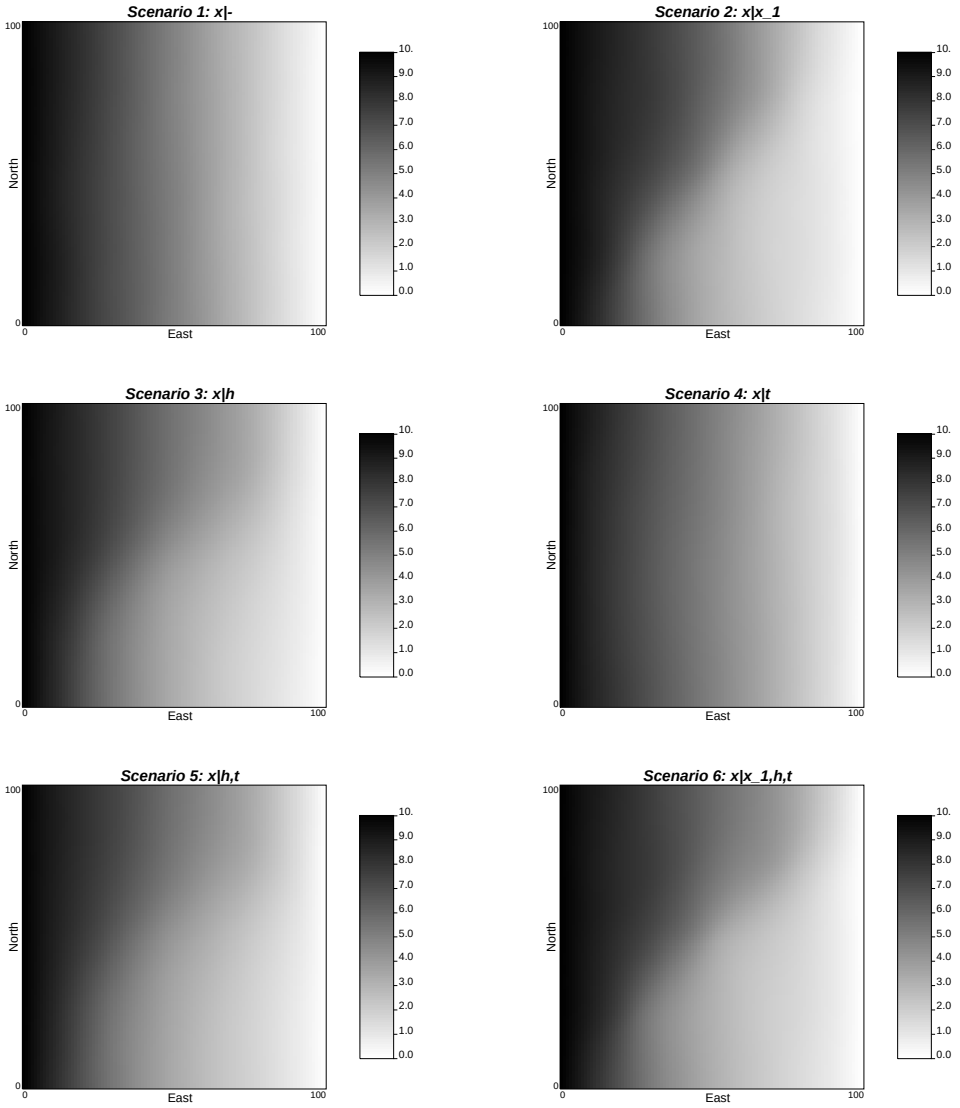


Figure 4.9: *The mean head fields due to conditioning to data from various sources*

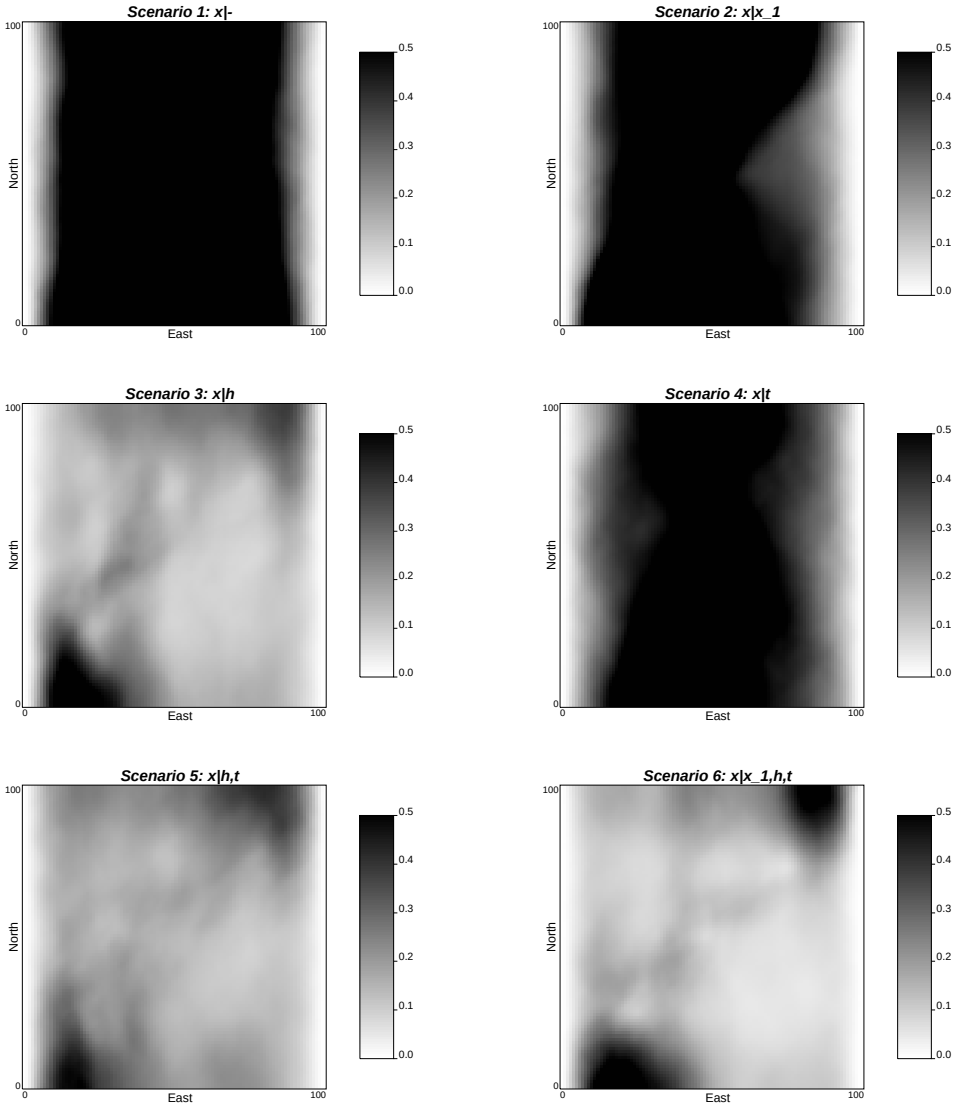


Figure 4.10: *The variances of head fields due to conditioning to data from various sources*

4.4.4 Uncertainty propagation

Rubin (1991) had examined the uncertainty propagation due to conditioning on the ln$K$ and h measurements with respect to the prediction of tracer plume migration. This study tries to include the temporal moments of BTCs into the conditioning and inverse-conditioning procedure and to investigate its impact on the prediction of solute plume spreading. Two underlying assumptions for the *Rubin*'s or other traditional methods should be pointed out: one is the linearization of the flow and transport equations and the other is that structure uncertainty is often involved. Our method is free from these two assumptions and hence some more general results are expected to be reached.

By setting various control planes in the mean flow direction, the unconditional and conditional macrodispersion coefficients may be computed by Equation 4 or 5 from the temporal moments of the BTCs. Figure 4.11 compares the scale-dependent macrodispersion coefficients with the reference curve. Due to the limitation of experimental design, only the early time, small-scale macrodispersions have been computed and plotted, i.e., $x/\lambda_{\boldsymbol{x}} \leq 2$.

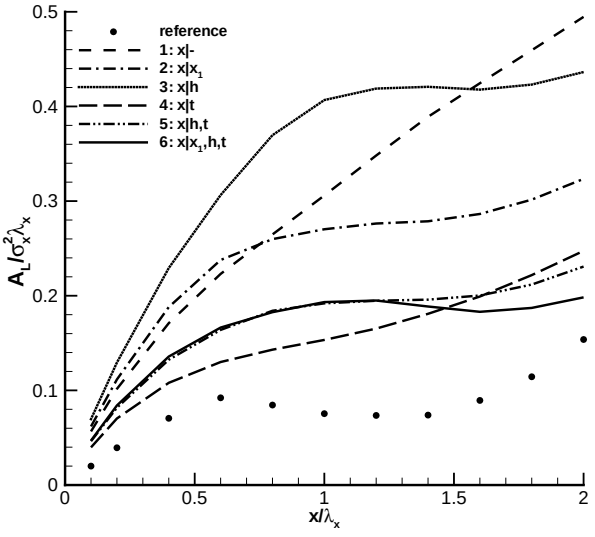


Figure 4.11: *A comparison of simulated macrodispersions due to conditioning to data from various sources*

Several observations may be drawn from the Figure 4.11. First, inverse-conditioning to travel time moments substantially improves the estimate of macrodispersion. Note the decrease in $A_L(x)$ by comparing Scenario 4-6 to

Scenario 1-3. This finding based on inverse-conditioning to temporal moments of BTCs is totally consistent with the result based on inverse-conditioning to spatial distribution of tracer concentration which was provided by *Woodbury and Sudicky* (1992). On the basis of the bromide and chloride tracer tests performed at the Borden aquifer in Ontario, Canada, they found that inverse-conditioning to spatial moments of concentration data considerably enhances the estimation on the rates of plume spreading in the longitudinal direction compared to the unconditional case (e.g., Figure 6 over 4 in their article). Therefore one can conclude from this experiment that even though inverse-conditioning to temporal moments of BTC may not reduce the uncertainties of $\ln K$ and head by much, it may serve to identify other components such as the spreading of solute plume, which is even more of interest for the water resource management.

Second, the $\ln K$ measurements (Scenario 2) seemingly contain more information for determining macrodispersion than the head observations (Scenario 3). *Rubin* (1991) also found that introduction of head data does not cause any considerable improvement for estimation on the solute plume migration. The travel time data (Scenario 4) may convey more information on macrodispersion than both the $\ln K$ measurements and the head observations.

Third, jointly conditional and inverse-conditional simulations tend to render the most abundant information for macrodispersion determination after a certain distance from the source, i.e., $x > 1.5\lambda_{\boldsymbol{x}}$ (Scenario 5 and 6). At the early stage when the control plane is close to the source, however, the impact of conditioning on determining the macrodispersion coefficient is complex. A general conclusion is hardly to arrive at since the results are highly dependent on the local conditions of $\ln K$ distribution (*Rubin and Dagan*, 1992). Moreover, unlike the results from the linearization of flow and transport equations, our numerical results are fully the nonlinear responses to the $\ln K$ fields which makes the distribution of macrodispersion coefficient less regular than the linear method, e.g., *Rubin*'s (1991).

By comparing Table 4.3, Table 4.4, and Figure 4.11, one can find that conditioning data of state variables may have not too much influence on the uncertainty reduction of model parameters, but the uncertainties of model responses may be reduced remarkably. For example, inverse-conditioning on the moments of trave time t (Scenario 4) does not reduce the uncertainties of $\ln K$ and h distributions by too much (e.g., the relative uncertainty to the unconditional case is less than 14% and 32%, respectively), but the prediction on the rate of solute plume spreading is significantly improved (Scenario 4 in Figure 4.11). Similarly, collection of h only slightly enhances the estimation of model, e.g., the model uncertainty decreases only 14% compared to the unconditional case. It also has not too much influence on the estimation of $A_L(x)$, e.g., very slightly better than the unconditional case by comparing

Scenario 3 to 1 in Figure 4.11. But the prediction of heads has been highly improved, almost 60%. That is because the method presented in this study strictly preserves the model statistics and structure, the model uncertainty may not be reduced to very high degree even though a large number of hard data are available. But inverse-conditional simulations do remarkably decrease the uncertainties of model responses. On the other hand, the uncertainty reduction of predicted responses is aim-dependent, i.e, the collection of state data improves the prediction of response of the same type to the greatest extent which is consistent with the intuition.

In addition, the evolution of the plume front and tail is of interest for the risky analysis on a radioactive waste repository. The ensemble mean of the 5 and 95 percentile of BTCs, denoted by $t_{5\%}$ and $t_{95\%}$, represent the behavior of contaminant plumes at the early time (plume front) and the late time (plume tail), respectively. $t_{5\%}$ reflects the earliest times when a location of interest is polluted by the contaminant plume while $t_{95\%}$ estimates the operational time to remove most of contaminants from the aquifer. Figure 4.12 plots the evolution history of contaminant plume's front and tail along the mean plume trajectories before and after conditioning to diverse data.

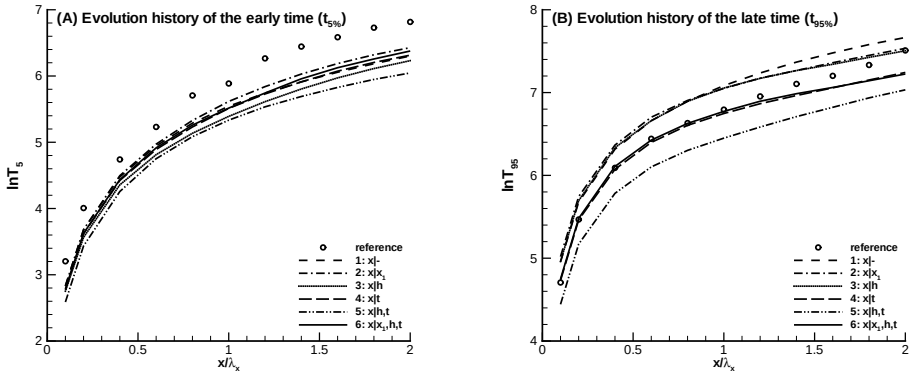


Figure 4.12: *A comparison on evolution history of contaminant plume* (A) *at the early time* $t_{5\%}$ *and* (B) *at the late time* $t_{95\%}$ *due to conditioning to data from various sources*

Two observations from Figure 4.12 (A) are that conditioning and non-conditioning make no evident difference to the estimation on the plume's front ($t_{5\%}$) and that the stochastic simulations obviously underestimate the estimation on the early arrival time. From Figure 4.12 (B), one may find that inverse-conditioning to the temporal moments of BTCs underestimates the

plume's tail ($t_{95\%}$), non-conditioning to the temporal moments overestimates the late time, and jointly conditioning to hard data, head data and temporal moments yields the best estimate. But in general, hard data, head data and temporal moments of BTCs seemingly do not contain too much information on the estimation of the early and late time. This result may come from two facts: (1) conditioning to the given data, especially the temporal moments of BTCs, is only an integrated, average reflection of aquifer properties while $t_{5\%}$ represents the fastest flow track and $t_{95\%}$ represents the slowest path, both of which are the extreme responses of aquifer properties, and (2) the isotropic multi-Gaussianity of the random fields is partly responsible for it since, unlike a non-Gaussian, multi-modal field, the isotropic multi-Gaussian model tends to give rise to an extremity-free response.

4.5 Conclusions and Discussions

The MsBMcMC method is an useful tool for stochastic conditional and inverse-conditional simulation and quantitative uncertainty assessment. A synthetic example under a uniform, natural-gradient flow condition is presented to show its usefulness in uncertainty assessment and data worth evaluation. One of the most appealing points is that this method is capable of generating *i.i.d* conditional and inverse-conditional realizations that have an identical model structure as specified *a priori*, which makes parameter uncertainty separated from the model structure uncertainty such that uncertainty assessment on parameter variability may be carried out exclusively without other factors involved. As a result, it can give a more accurate measure on the data worth from diverse sources in quantifying uncertainty reduction due to conditioning and inverse-conditioning. In this study, three types of distinct data, hydraulic conductivity $\ln K$, piezometric head h, and travel time t (or concentration c), have been compared to examine their worth on the uncertainty reduction. The results are summarized as follows.

First, as for the effect of $\ln K$, h and t upon the model uncertainty of $\ln K$, the $\ln K$ measurements play a major role in reducing such uncertainty compared to h and t, especially for the prior configuration on model parameters. Jointly inverse-conditioning on h and t does improve the model estimation on $\ln K$ compared to conditioning on $\ln K$ solely. However, the information on model extracted from various source data may be totally different. For example, the measurements on h and t are informative on the large-scale trend of $\ln K$. The measurements on $\ln K$ do not carry too much information on the spatial trend of $\ln K$ but can essentially reduce the local fluctuation.

Second, as for the effect of $\ln K$, h and t upon the prediction uncertainty of h, the measurement on h plays a major role in reducing such uncertainty

compared to that of $\ln K$ and t. Although the prediction of h is quite insensitive to local $\ln K$ values and temporal moments of t, jointly conditioning on $\ln K$ and t does help improve the model prediction of h compared to inverse-conditioning on h solely.

Third, as for the effect of $\ln K$, h and t upon the uncertainty of the predicted macrodispersion coefficient, inverse-conditioning to the temporal moments of t substantially reduces such uncertainty. Both the measurements on $\ln K$ and h may play positive roles in reducing such uncertainty, but it is hard to distinguish the majority from the other as for the example problem in this study. Hard data, head data and the first two temporal moments of BTCs seemingly do not contain too much information for the estimation on the early and late time in an isotropic multi-Gaussian case.

In addition, the sampling density and network design for the collection of conditioning data may have a certain influence on the uncertainty reduction. The improvement is more considerable if the measurements are carried out at sampling spacing smaller than the correlation length of $\ln K$, which was also observed by *Dagan* (1985) and *Rubin and Dagan* (1992). Relatively larger sampling spacing on state data yields inverse estimates closer to the prior mean while relatively larger correlation length of $\ln K$ yields inverse estimates smoother (e.g., *McLaughlin and Townley*, 1996).

One of the main findings from this study is that the measurements on aquifers mostly improve the estimation on model parameters and/or responses of the same type. In other words, the main error source of aquifer parameters and/or responses is the scarcity of measurements from parameter and/or response themselves. For example, the data acquisition on $\ln K$ mostly improves the estimation on spatial distribution of $\ln K$. The observations of h extremely improve the estimate of h itself. Along the same line, the measurements on c are more beneficial to the depiction on c distribution. On the other hand, the measurements on dependent state variables do provide invaluable, complementary information for estimating model parameters, that is, the coupled/joint inversion is helpful in pattern recognition of aquifers. Actually, head observations and travel time moments are informative only about the large-scale variability of parameters while direct measurements of parameters are informative only about their small-scale fluctuation. The variances of model parameters may not necessarily be reduced significantly by inverse-conditioning to head observations and travel time moments but the identified large-scale trend may be the most important information for regional groundwater management.

This finding has some important applications, e.g., in the network design for improving the reliability of groundwater modeling. Since the goal of an optimal network design scheme is to minimize sampling costs of aquifer parameters while estimating aquifer responses to a specified precision (*James and Gorelick*, 1994), the worth of installing a monitoring well (sampling cost) is

linked to how much the information (on uncertainty reduction) that it provides can be used to reduce the remediation cost (of aquifer responses). Apart from the economical factor in the network design (*James and Gorelick*, 1994; *Criminisi et al.*, 1997), most of algorithms construct the objective function either based on the error reduction of state variables or dependent on the decrease of prediction error (e.g., *McKinney and Loucks*, 1992), but seldom both. Our results from the evaluation on uncertainty propagation show that the conditional and inverse-conditional simulation may not always meet all the ends, that is, the uncertainty of some responses may be reduced more significantly than the others and some of them even not at all. More reliable selection of the "best" aquifer parameter measurement locations should comprehensively account for various uncertainties if no individual aim is specified ahead.

Since the relative sensitivity of one type of state variable to the spatial variability of $\ln K$, the estimation performance may not improve dramatically when more measurements on the same type of property are included. A smaller number of measurements on other state variables such as the solute concentration may provide more valuable information, which is known as the coupled inverse problem. Even more, the estimation on $\ln K$ may benefit more from the economical measurements on other model parameters which is known as the joint inverse problem. The geophysical measurements, e.g., the seismic, the ground-penetrating radar, and the electrical resistivity, may also provide indirect, supplementary determination of $\ln K$. Any emphasis on decreasing only a certain type of uncertainty and ignoring the other may not result in an improvement of prediction accuracy (*Gaganis and Smith*, 2001). Actually, as observed by *Sun and Yeh* (1992), additional head observations decrease the model prediction errors, but only with a minor improvement in parameter estimation. *Dagan* (1985) also observed that the effect of conditioning to head on $\boldsymbol{x}$ is much weaker than that of conditioning to $\boldsymbol{x}$ itself and moreover, head measurements, no matter how dense, reduce σ_x^2 to no less than 30% for his study in absence of neighboring $\ln K$ measurements.

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 uncertainty reduction of a response from different sources, e.g., the model structure, the parameter estimation, and the inherent variation of aquifers. Ranking of uncertainty sources is extremely useful for experimental network design and uncertainty reduction since a new optimal experiment design aiming at reducing the most critical uncertainties depends on the identification of uncertainty sources giving rise to the worst predictions. Inevitably, the relative significance of different sources is problem specific and it is not expected that a general conclusion can be drawn from one single case study.

5

Does Model Structure Preservation not Matter? A Comparison on Two Methods for Inverse Stochastic Modeling

Abstract

Dynamic data provide invaluable information for identifying the spatial pattern of reservoir or aquifer parameters (e.g., hydraulic conductivity) and for reducing the uncertainties of models and their responses. To extract such spatial information from the measurements, a Monte Carlo method typically calls for generating a large number of realizations that are inverse-conditional to dynamic data and honor the known prior information. In contrast to some classical nonlinear optimizers, the blocking Markov chain Monte Carlo (BMcMC) method is capable of generating *i.i.d* realizations that are not only constraint dynamic observations but also preserve expected spatial statistics and model structure as specified *a priori*. By comparing the ensemble Kalman filtering (EnKF) and the BMcMC, this study presents two synthetic examples to demonstrate the importance of honoring the prior information for inverse stochastic modeling. Numerical simulations show that, even though the

EnKF method may efficiently provide a better reproduction of observed dynamic data than the BMcMC method, the preservation of spatial statistics and model structure makes the BMcMC simulations competitive for some cases in predicting accurately and reliably the future performance of reservoirs particularly at new well locations. Therefore, including the prior information in the inverse stochastic simulation is of significance for accurate assessment of model uncertainties and response uncertainties if the prior parameter information effectively reflects the underground reality.

5.1 Introduction

Collections of time-dependent dynamic data provide indispensable knowledge on the subsurface reservoirs or aquifers which generally entails a stochastic method capable of generating independent, identically distributed (*i.i.d*) conditional realizations that honor the known dynamic measurements. A large number of analytical and numerical inverse methods have been developed to address this problem, which typically consist of two stages, model structure identification and stochastic conditional simulation (e.g., *Kitanidis and Vomvoris*, 1983; *Dagan*, 1985; *Sun*, 1985; *Yeh*, 1986; *Carrera and Neuman*, 1986; *RamaRao et al.*, 1995; *McLaughlin and Townley*, 1996; *Gomez-Hernandez et al.*, 1997; *Evensen*, 2003; *Carrera et al.*, 2005). The model structure identification is to build the main spatial features on model properties through, e.g., variogram analysis by collecting hard data (e.g., *Deutsch and Journel*, 1998), model identification by measuring dependent state variables (e.g., *Sun and Yeh*, 1985; *Kitanidis*, 1996), outcrop's analogue, geological mapping (e.g., *Koltermann and Gorelick*, 1996), geophysical imaging (e.g., *Rubin et al.*, 1992), etc. Stochastic conditional simulation is then simply to assign parameter values to the models once their structure has been correctly identified.

Amongst the typical inverse methods, the ensemble Kalman filtering (EnKF) proposed by *Evensen* (1994, 2003) is an extremely computationally efficient, purely statistically based Monte Carlo method for nonlinear dynamic data assimilation, which gains popularity recently in the petroleum engineering for history matching problems (e.g., *Nævdal et al.*, 2003; *Gu and Oliver*, 2004; *Liu and Oliver*, 2005a, 2005b; *Wen and Chen*, 2005; *Zafari and Reynolds*, 2005; *Gao et al.*, 2005; *Skjervheim et al.*, 2005; *Dong et al.*, 2006). It is essentially a sequential Bayesian inverse method where an ensemble of reservoir models are used to construct the error statistics for both model parameters and model responses. The ensemble of models evolves in a Markov chain fashion over the parameter space with the ensemble mean as the best estimate and the ensemble variance as the error spreading. The prediction of the estimate

and uncertainty is performed by integrating the ensemble of reservoir models. The most important contributions of this method perhaps are (1) that the data integration is designed in a Markov process in time which allows for online, continuous model updating for assimilating a large number of dynamic data and (2) that an ensemble of models are updated extremely efficiently in computation simply because only the statistics of inputs and outputs for an ensemble of forward simulations are needed.

Aiming at preserving the parameter statistics and spatial structure for models, Chapter 3 proposes a blocking Markov chain Monte Carlo (BMcMC) method to perform such stochastic conditional simulation. In essence, this approach just thins down, with the aid of the McMC theorem, the ensemble of proposed candidates by means of judiciously selecting those models that honor the dynamic data under the scope of errors, measured by the posterior probability. The adoption of the blocking scheme is based on three considerations: (1) it helps preserve the prior spatial statistics and structure, (2) it helps improve the mixing of Markov chain and enhance the computational efficiency compared to the traditional single-component McMC methods, and (3) it is more suitable for the introduction of upscaling procedures to accelerate the computation of the likelihood which is especially useful in handling the high-dimensional cases.

This study presents two synthetic examples to demonstrate the importance of preserving spatial structure for models in performing inverse stochastic simulations by comparing two distinct methods, the EnKF and the BMcMC. The article is organized as follows. Section 2 gives some details on the implementation of the two methods used in this study. In Section 3 and Section 4, two scenarios of stochastic simulations by these two methods are carried out to generate realizations that are conditional on the observed pressure profiles and well flow-rate histories under forced-gradient transient flow conditions. Then, these generated realizations subject to identical flow and transport experiments are used to evaluate the model uncertainty and response uncertainty. The significance of preserving model's spatial structure for history matching are analyzed according to its effect on the reservoir performance prediction and the importance for uncertainty assessment. Finally, several main conclusions are summarized in Section 5.

5.2 Methodology

Consider a stochastic modeling at n grid nodes conditional to k nonlinear dynamic data. Specifically, let $\boldsymbol{x} = (x_0, x_1, ..., x_{n-1})^T \subset R^n$ denote a realization conditional to k dynamic data $\boldsymbol{y} = \boldsymbol{y}_{obs} = (y_0, y_1, ..., y_{k-1})^T \subset R^k$. Assuming a multi-Gaussian process, the spatial distribution of $\boldsymbol{x}$ follows, $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C_x})$,

where $\boldsymbol{\mu} \subset R^n$ is the prior mean of the Gaussian process and $\boldsymbol{C_x} \subset R^{n \times n}$ describes the structure dependence of spatial points. Assuming a multi-normal error, the simulated observation $\boldsymbol{y}_{sim} \subset R^k$ for a given sample $\boldsymbol{x}$ can be expressed as, $\boldsymbol{y}_{sim}|\boldsymbol{x} \sim N(g(\boldsymbol{x}), \boldsymbol{C_y})$, where $\boldsymbol{C_y} \subset R^{k \times k}$ describes the degree of discrepancy between the transfer function $g(\boldsymbol{x})$ and the true but error-prone observation $\boldsymbol{y}$. The transfer function $g(\boldsymbol{x})$ is error-prone since most often an analytical expression is not available. One generally resorts to some complex computer models to simulate the physical process. In such case, its accuracy depends on the spatial discretization of the physical model. 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 $\boldsymbol{y}$ that can be included in this statistical model. In this sense, $\boldsymbol{C_y}$ measures both the modeling errors and the measurement errors.

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

5.2.1 Blocking McMC method

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

$$\pi(\boldsymbol{x}|\boldsymbol{\theta}) = (2\pi)^{-\frac{n}{2}} \|\boldsymbol{C_x}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C_x}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}. \tag{5.1}$$

This pdf represents some prior knowledge about the parameterization of a physical model $\boldsymbol{x}$ through the configuration of $\boldsymbol{\mu}$ and $\boldsymbol{C_x}$ which, together with other parameters, boil down to a hyperparameter set $\boldsymbol{\theta}$. It should allow for the greatest uncertainty while obeying the constraints imposed by the prior information. The hyperparameters $\boldsymbol{\theta}$ may be inferred from both the *a posteriori* measurements and the *a priori* subjective imagination.

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

$$L(\boldsymbol{x}|\boldsymbol{y}) = (2\pi)^{-\frac{k}{2}} \|\boldsymbol{C_y}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(g(\boldsymbol{x})-\boldsymbol{y})^T \boldsymbol{C_y}^{-1}(g(\boldsymbol{x})-\boldsymbol{y})\right\}. \tag{5.2}$$

This likelihood function, which is defined by the misfit between the observed data $\boldsymbol{y}$ and the predicted data $g(\boldsymbol{x})$ from a candidate parameter model $\boldsymbol{x}$,

measures the probability of observing the data $\boldsymbol{y}$ for the model $\boldsymbol{x}$. Obviously, the probability of observing the given data $\boldsymbol{y}$ becomes smaller as the misfit becomes larger.

A normalizing factor that makes the integral of the posterior pdf equal to unity is called the marginal likelihood, i.e., $\int L(\boldsymbol{x}|\boldsymbol{y})\pi(\boldsymbol{x}|\boldsymbol{\theta})d\boldsymbol{x}$. Since the marginal likelihood is not a function of $\boldsymbol{x}$, it is typically ignored in the parameter estimation problem. Using the Bayes' rule and dropping the normalizing constant, the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$ and $\boldsymbol{\theta}$ may be written as $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \propto L(\boldsymbol{x}|\boldsymbol{y}) \times \pi(\boldsymbol{x}|\boldsymbol{\theta})$, i.e.,

$$\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\theta}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}(g(\boldsymbol{x})-\boldsymbol{y})^T\boldsymbol{C}_{\boldsymbol{y}}^{-1}(g(\boldsymbol{x})-\boldsymbol{y})\right\}. \tag{5.3}$$

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

Due to the highly nonlinearity of the likelihood model, it is impossible to sample directly from this posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$. The Markov chain Monte Carlo method (*Metropolis et al.*, 1953; *Hastings*, 1970; *Geman and Geman*, 1984), however, is especially suitable for exploring the parameter space with such type of complicated posterior distribution. A typical McMC algorithm employing the Metropolis-Hastings rule to explore the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ goes as follows,

(1) Initialize the parameters $\boldsymbol{x}$;

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

- propose $\boldsymbol{x}^* \sim q(\boldsymbol{x}^*|\boldsymbol{x})$;
- accept $\boldsymbol{x}^*$ with probability $\min\{1,\alpha\}$, where $\alpha = \frac{\pi(\boldsymbol{x}^*|\boldsymbol{y})q(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x}|\boldsymbol{y})q(\boldsymbol{x}^*|\boldsymbol{x})}$;

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

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

One of the most interesting problems in this algorithm is the configuration of the proposal transition kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$, which plays a crucial role in the computational efficiency of a Metropolis-Hastings-type McMC method. The classical McMC method constructs the Markov chain by a single-component proposal $x^* \sim N(\mu, \sigma_x^2)$ (*Oliver et al.*, 1997). It tends to fail in handling

the high-dimensional case since the numerical evaluation of the posterior pdf $\pi(\boldsymbol{x}^*|\boldsymbol{y})$ and $\pi(\boldsymbol{x}|\boldsymbol{y})$ is a tough task: the inverse of the covariance matrix $\boldsymbol{C_x}$ is a big challenge on one hand, the single-component updating makes the forward simulations $g(\boldsymbol{x}^*)$ and $g(\boldsymbol{x})$ look alike on the other hand.

In these two regards, the blocking proposal scheme $\hat{\boldsymbol{x}}^*|\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C_x})$, $\hat{\boldsymbol{x}}^* \subset \boldsymbol{x}^*$, which has the identical spatial distribution as the prior model $\boldsymbol{x}^* \sim N(\boldsymbol{\mu}, \boldsymbol{C_x})$, is obviously advantageous over the single-component scheme. 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, i.e., $q(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*|\boldsymbol{x})$. The advantages of the blocking proposal kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$ over the single-component proposal kernel $q(x^*|x)$ can be summarized as follows. First, the block updating scheme is well known for improving the mixing of a Markov chain which makes the McMC simulations more efficient (*Liu*, 1996; *Roberts and Sahu*, 1997). Second, it is more suitable for the introduction of the upscaling subroutines to accelerate the computation of the likelihood as in Formula (2). Moreover, $L(\boldsymbol{x}^*|\boldsymbol{y})$ will obviously differ from $L(\boldsymbol{x}|\boldsymbol{y})$, which is advantageous over the single-component scheme. Third, maybe the most interesting, it helps preserve the spatial statistics and model structure as specified *a priori* since $q(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*|\boldsymbol{x})$ holds always during the model updating. Fourth, if $q(\boldsymbol{x}^*|\boldsymbol{x}) = \pi(\boldsymbol{x}^*|\boldsymbol{x}) \doteq \pi(\boldsymbol{x}^*)$ and $q(\boldsymbol{x}|\boldsymbol{x}^*) = \pi(\boldsymbol{x}|\boldsymbol{x}^*) \doteq \pi(\boldsymbol{x})$, the acceptance rate α can be simply computed by,

$$\alpha = \frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})}. \tag{5.4}$$

In such case, the complicated evaluation on the model density as in Formula (1) is totally avoided for the high-dimensional case.

In addition, an important implementation detail is the fast generation of the proposal transition kernel 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 be maintained for 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. However, for a high-dimensional case, the FFT-based spectral simulator is preferred since it has a computational ability more powerful than the LU-based generator (*Pardo-Iguzquiza and Chica-Olmo*, 1993; *Robin et al.*, 1993; *Dietrich and Newsam*, 1993, 1996; *Gutjahr et al.*, 1997; *Ruan and McLaughlin*, 1998). To save the computational time, furthermore, the evaluation of the likelihood can be efficiently accelerated by running the forward simulations at a coarse scale which invokes an economical but accurate upscaling method and a multi-scale-oriented black-box-like forward simulator that is widely compatible with the

output from upscaling subroutines. More details on the development of the method have been presented in Chapter 3.

5.2.2 Ensemble Kalman filtering

Without considering hard data, the objective of the EnKF for dynamic data assimilation is to draw samples $\boldsymbol{x} \sim \pi(\boldsymbol{x}|\boldsymbol{y})$. Define a joint vector $\boldsymbol{\psi}_t$, $t \in [0, n_t]$, whose components are given by,

$$\boldsymbol{\psi}_{t,r} = \begin{pmatrix} \boldsymbol{x}_{t,r} \\ \boldsymbol{y}_{t,r} \end{pmatrix}; \qquad r \in [0, n_r), \tag{5.5}$$

where $\boldsymbol{x}_t = (\boldsymbol{x}_{t,r} \subset R^{n_{xyz}}; r \in [0, n_r), n_{xyz} = n)^T$ is the vector of static model parameters to be estimated; $\boldsymbol{y}_t = (\boldsymbol{y}_{t,r} \subset R^{n_{d,t}}; r \in [0, n_r))^T$ is the vector of dynamic model responses at the given time t which either have been observed in the actual field ($\boldsymbol{y}_t^{obs}$) or have been predicted on the basis of the given models ($\boldsymbol{y}_t^{sim}$); the subscript $r \in [0, n_r)$ denotes the index of realizations; n_r is the total number of realizations; and the subscript $t \in [0, n_t] = [t_0, t_e]$ indicates the index of times (n_t is the total time steps, and t_0 and t_e are the initial and last times for dynamic history observations, respectively.). The initial condition $\boldsymbol{\psi}_0$ is given as follows,

$$\boldsymbol{\psi}_{0,r} = \begin{pmatrix} \boldsymbol{x}_{0,r} \\ \boldsymbol{y}_0 \end{pmatrix}; \qquad r \in [0, n_r),$$

where $\boldsymbol{x}_0 = (\boldsymbol{x}_{0,r}; r \in [0, n_r))^T$ is an ensemble of initial realizations of static model parameters which are generated by stochastic simulation subroutines; $\boldsymbol{y}_0$ is the given initial conditions which may be the same for all realizations. Then, the EnKF for joint updating of the models and their responses can be viewed as a Markov chain in time,

$$\boldsymbol{\psi} = (\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_t, ...)^T. \tag{5.6}$$

Following this Markov process, therefore, the dynamic history observation data $\boldsymbol{y}_{obs} = \left(\boldsymbol{y}_0^{obs}, \boldsymbol{y}_1^{obs}, ..., \boldsymbol{y}_t^{obs}, ...\right)^T$ can be sequentially assimilated into the models $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_t, ...)^T$ which are of the most interest in reservoir engineering. The performance prediction of reservoir is always based on the latest models.

There are two steps involving the implementation of the EnKF method: (1) a forecast based on current model parameters and (2) the updating of model parameters by assimilating nonlinear data. For example, for $t = 1$, aiming at the first set of dynamic observations $\boldsymbol{y}_1^{obs}$, the predicted dynamic data after running the forward simulator are $\boldsymbol{y}_{1,r}^{sim} = g(\boldsymbol{x}_{0,r})$, $r \in [0, n_r)$, and the joint vector thus becomes,

$$\boldsymbol{\psi}_{1,r}^{sim} = \begin{pmatrix} \boldsymbol{x}_{0,r} \\ \boldsymbol{y}_{1,r}^{sim} \end{pmatrix}; \qquad r \in [0, n_r).$$

However, a problem arises since there is a mismatch between the observed and the computed responses,

$$\Delta \boldsymbol{y}_{1,r} = \boldsymbol{y}_1^{obs} - \boldsymbol{y}_{1,r}^{sim} = \boldsymbol{y}_1^{obs} - g(\boldsymbol{x}_{0,r}) \neq 0,$$

which holds almost for all realizations $r \in [0, n_r)$. The current models $\boldsymbol{x}_0$ have to be updated to $\boldsymbol{x}_1$ in order to make the model responses $\boldsymbol{y}_1^{sim} = g(\boldsymbol{x}_1)$ agree with the given observations $\boldsymbol{y}_1^{obs}$.

In the EnKF, the models are updated by minimizing the error variance of models and their responses. One of the crucial assumptions for the EnKF is that the pdf of the joint vector $\boldsymbol{\psi}_t$ can be approximated by a multi-Gaussian distribution whose basis is that the dynamic data $\boldsymbol{y}_{t,r}$ are linearly related to the model parameters $\boldsymbol{x}_{t,r}$. The details on development of the updating formula are given by *Evensen* (1994, 2003). Starting from the time $t = 1$ when the nonlinear dynamic observations are available, the joint vector is updated by,

$$\begin{aligned} \boldsymbol{\psi}_{t,r} &= \boldsymbol{\psi}_{t,r}^{sim} + \boldsymbol{G}_t \left(\boldsymbol{y}_t^{obs} + \boldsymbol{\epsilon}_t - \boldsymbol{H}_t \boldsymbol{\psi}_{t,r}^{sim} \right) \\ &= \boldsymbol{\psi}_{t,r}^{sim} + \boldsymbol{G}_t \left(\boldsymbol{y}_t^{obs} + \boldsymbol{\epsilon}_t - \boldsymbol{y}_{t,r}^{sim} \right), \end{aligned} \tag{5.7}$$

where $\boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, \boldsymbol{\sigma}_{\boldsymbol{y}}^2) \subset R^{n_{d,t}}$ is the random observation error vector and the Kalman gain matrix $\boldsymbol{G}_t \subset R^{(n_{xyz}+n_{d,t}) \times n_{d,t}}$ is computed by,

$$\boldsymbol{G}_t = \boldsymbol{C}_{\boldsymbol{\psi},t}^{sim} \boldsymbol{H}_t^T \left(\boldsymbol{H}_t \boldsymbol{C}_{\boldsymbol{\psi},t}^{sim} \boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t} \right)^{-1}, \tag{5.8}$$

where $\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim} \subset R^{(n_{xyz}+n_{d,t}) \times (n_{xyz}+n_{d,t})}$ is the covariance matrix of the joint vector $\boldsymbol{\psi}_{t,r}$ at the time $t = 1$,

$$\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim} = E\left[\left(\boldsymbol{\psi}_t^{sim} - \boldsymbol{\psi}_t^{true} \right) \left(\boldsymbol{\psi}_t^{sim} - \boldsymbol{\psi}_t^{true} \right)^T \right],$$

which can be estimated by,

$$\begin{aligned} \hat{\boldsymbol{C}}_{\boldsymbol{\psi},t}^{sim} &= E\left[\left(\boldsymbol{\psi}_t^{sim} - \bar{\boldsymbol{\psi}}_t^{sim} \right) \left(\boldsymbol{\psi}_t^{sim} - \bar{\boldsymbol{\psi}}_t^{sim} \right)^T \right] \\ &= \frac{1}{n_r - 1} \sum_{r=0}^{n_r - 1} \left(\boldsymbol{\psi}_{t,r}^{sim} - \bar{\boldsymbol{\psi}}_t^{sim} \right) \left(\boldsymbol{\psi}_{t,r}^{sim} - \bar{\boldsymbol{\psi}}_t^{sim} \right)^T, \end{aligned} \tag{5.9}$$

where $\boldsymbol{\psi}_t^{sim} \subset R^{(n_{xyz}+n_{d,t}) \times n_r}$ is a vector which consists of an ensemble of simulated joint vectors at the time $t = 1$ and $\bar{\boldsymbol{\psi}}_t^{sim} \subset R^{n_{xyz}+n_{d,t}}$ is a vector which is

the ensemble mean of the simulated joint vectors $\boldsymbol{\psi}_t^{sim}$; $\boldsymbol{H}_t \subset R^{n_{d,t}\times(n_{xyz}+n_{d,t})}$ is a matrix operator that maps the joint vector $\boldsymbol{\psi}_t$ to the dynamic observation vector $\boldsymbol{y}_t$, which has a form of,

$$\boldsymbol{H}_t = [\mathbf{0}, \mathbf{1}],$$

where $\mathbf{0} \subset R^{n_{d,t}\times n_{xyz}}$ and $\mathbf{1} \subset R^{n_{d,t}\times n_{d,t}}$; $\boldsymbol{C}_{\boldsymbol{y},t} \subset R^{n_{d,t}\times n_{d,t}}$ is the covariance matrix of the dynamic observation $\boldsymbol{y}_t^{obs}$ which is a diagonal matrix if the observation errors of state variables are assumed to be independent from each other. It should be pointed out that if different measurements are assimilated simultaneously, the observed data need to be nondimensional or scaled to have the similar variabilities (*Evensen and van Leeuwen*, 2000; *Evensen*, 2003). By doing so, the eigenvalues of the matrix $\boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t}$ corresponding to each of the measurement types have the same magnitude. Note that not all of the elements of $\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}$ are necessary to be calculated and stored since only $\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T$ is needed for the computation of the Kalman gain. Actually, one can easily find that in the formula of the Kalman gain for the time t,

$$\begin{aligned}\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T &= \boldsymbol{C}_{\boldsymbol{\psi}\boldsymbol{y},t}^{sim},\\ \boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T &= \boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y},t}^{sim},\end{aligned}$$

where $\boldsymbol{y}_{t,r}^{sim}$ is the simulated dynamic response for the realization r; $\boldsymbol{C}_{\boldsymbol{\psi}\boldsymbol{y},t}^{sim}$ denotes the simulated covariance between the joint vector $\boldsymbol{\psi}$ and the response $\boldsymbol{y}$; and $\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y},t}^{sim}$ denotes the simulated covariance between the responses $\boldsymbol{y}$. The preceding procedure can be directly applied to the numerical implementation if $\boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t}$ is positive-defined. However, the potential singularity of the matrix $\boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t}$ entails the use of the pseudo-inverse of a matrix, e.g., the singular value decomposition (SVD) method, which involves the computation of the eigenvalues and eigenvectors from the matrix, i.e.,

$$\begin{aligned}\boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t} &= \boldsymbol{V}_t\boldsymbol{\Lambda}_t\boldsymbol{V}_t^T,\\ \left(\boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t}\right)^{-1} &= \boldsymbol{V}_t\boldsymbol{\Lambda}_t^{-1}\boldsymbol{V}_t^T.\end{aligned}$$

This two-step procedure for model updating is repeatedly advanced for $t = 2, 3, ..., n_t$ until all the observations are assimilated into the physical models. One can easily find that the main computational challenge focuses on the updating of the models and their responses by,

$$\boldsymbol{\psi}_{t,r} = \boldsymbol{\psi}_{t,r}^{sim} + \boldsymbol{G}_t\left(\boldsymbol{y}_t^{obs} + \boldsymbol{\epsilon}_t - \boldsymbol{y}_{t,r}^{sim}\right), \tag{5.10}$$

$$\boldsymbol{G}_t = \boldsymbol{C}_{\boldsymbol{\psi}\boldsymbol{y},t}^{sim}\left(\boldsymbol{C}_{\boldsymbol{y}\boldsymbol{y},t}^{sim} + \boldsymbol{C}_{\boldsymbol{y},t}\right)^{-1}, \tag{5.11}$$

for $r \in [0, n_r)$ and $t \in (0, n_t]$. The computational burden is therefore rather small compared to other methods, which makes this method very competitive for model calibration and history matching problems.

5.3 Case Study 1

5.3.1 Experimental configuration

Consider a 2D transient single-phase flow test on a confined aquifer with 32×32 grid-blocks as designed in Figure 5.1 under the forced-gradient flow condition. The reference $\ln K$ field is generated by the LUSIM subroutine from the GSLIB (*Deutsch and Journel*, 1998) with a prior distribution $\ln K \sim N(0, 1)$ and an exponential variogram type, i.e.,

$$\gamma_{\boldsymbol{x}}(r) = \sigma^2_{\boldsymbol{x}} \left\{ 1 - \exp\left[-\frac{r}{\lambda_{\boldsymbol{x}}} \right] \right\}, \tag{5.12}$$

where r is the two-point separation distance, $\sigma^2_{\boldsymbol{x}}$ is the variance, and $\lambda_{\boldsymbol{x}}$ is the correlation length. The prior correlation length is set as $\lambda_{\boldsymbol{x}} = 16$ $[cells]$, which is much longer than the well spacing ($l_w \doteq 11$ $[cells]$) such that the spatial variability of $\ln K$ is well captured by the well configuration. The experimental statistical parameters, however, show a certain deviation from the prior configuration, e.g., $\ln K \sim N(-0.33, 1.07^2)$ (see Figure 5.3 and Figure 5.4 for the histogram and variogram of the reference $\ln K$ field).

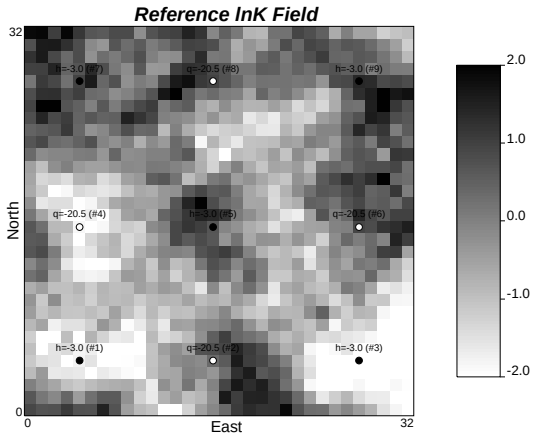


Figure 5.1: *Case study 1: Reference* $\ln K$ *field and the well configuration: five constant pressure production wells* (*in bullet*) *and four constant flow-rate injection wells* (*in circle*)

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 days ($t_0 = 0$ 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.

Nine wells are drilled throughout this confined aquifer (Figure 5.1): four of them are the injection wells with a constant flow-rate ($q = 20.5$ per day) and the other five are production wells with a constant pressure (the piezometric head is maintained at $h = -3.0$ for all five wells). The flow-rate data at the four injection wells and the bottom-hole-pressure (BHP) data at the five extraction wells are continuously collected at the first 40.1 days which consist of the first 50 time steps.

The inverse stochastic modeling problem, therefore, is to infer the permeability field (in $\ln K$) according to the observed 40.1-day's flow-rate and BHP data at the well-bores. The stochastic $\ln K$ fields are also required to be constraint to the given prior information, e.g., $\ln K \sim N(0, 1)$ and $\lambda_{\boldsymbol{x}} = 16$, even though they have rather large departure from the real case. Other flow parameters are assumed to be constant and known perfectly.

5.3.2 Inverse-conditional modeling by EnKF

First, the LUSIM subroutine from the GSLIB (*Deutsch and Journel*, 1998) is invoked to generate one hundred seed fields that obey all of the specified prior information. Then, the history matching problem is performed by the EnKF method to constrain on the 40.1-day's dynamic observations which include the flow-rate and the BHP data from all nine wells. Figure 5.2 compares the the flow-rate and BHP curves at the two typical wells before and after history matching. The history matching remarkably reduces the uncertainties of predicted flow-rate and BHP. Moreover, the real case (the circle line) is contained in one hundred stochastic estimates (the black lines). The accuracy and precision of predicted BHP (up to 500 days) are highly improved after matching the first 40.1 days' data.

But what will happen to the updated models after history matching? One of the facts is that the models generated as such are far from perfect as expected, e.g., the parameter statistics and model structures may severely deviate from the prior specification for models, even though the prediction on model responses at the known well locations seems to attain to an ideal result

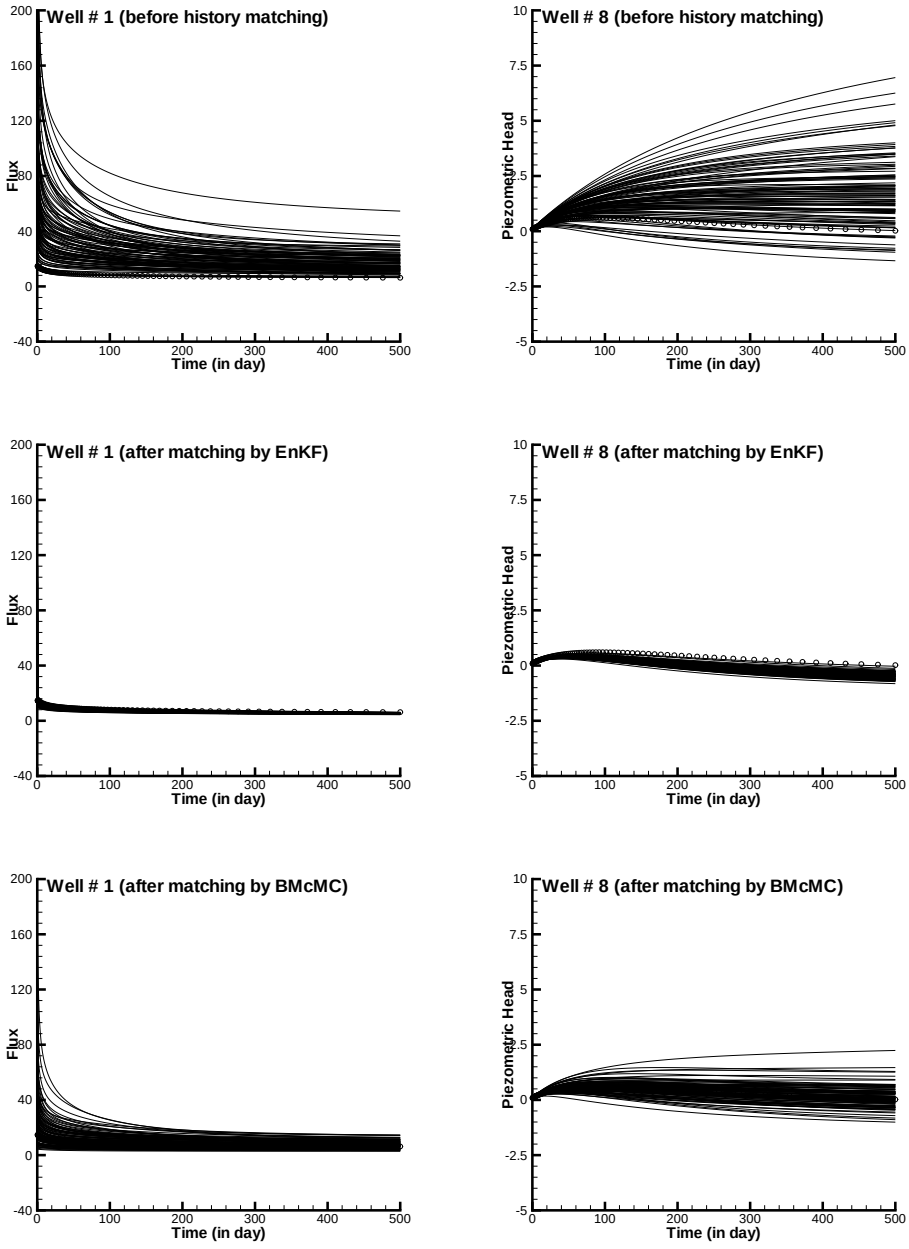


Figure 5.2: *Well performance reproduction by EnKF and BMcMC: The circle line denotes the reference data.*

at the time scale and the updated models do look more similar to the reference field than the initial seed fields.

Figure 5.3 displays the evolution of parameter histograms during data assimilation. One can easily find that the spreading of $\ln K$ at the early stage is obviously much wider than expected. At the day 0.2 and the day 2.4, for example, some unreasonably large values and some extremely small values frequently occur in the physical models which obviously do no more follow the prior specification, $\ln K \sim N(0,1)$. Note that the permeability is shown at the natural logarithm scale, i.e., $\ln K$, which implies a small fluctuation in $\ln K$ will cause a huge change in K. This phenomenon is not unique and is also observed by other researches known as "overshooting" and "undershooting" problems (e.g., *Gu and Oliver*, 2006). The problem might come from a fact that the prior formation is never used to constrain the models during the history matching.

Figure 5.4 shows the evolution of theoretic variograms of models during the data assimilation. The mean semivariogram of the initial 100 seed fields (the dashed line) is very close to the prior specification (the square line). As dynamic data are assimilated into the models, the mean semivariogram in an unpredicted matter gradually departs from the prior model. The x direction (A) displays an obvious random fluctuation but the y direction (B) seemingly tells us that the more newly updated models (e.g., the solid line) are closer to the reference field (the circle line). The EnKF method automatically compensates for the unprecise specification for the prior model by adjusting model parameters. Indeed, the reference field (Figure 5.1) along the y direction demonstrates a high-value distribution on the upper part and low values on the lower part. The dynamic observations from the nine wells are capable of reflecting effectively such a main feature which makes the variogram reproduction look quite well by the EnKF method.

The two observations from this experiment (Figure 5.3-5.4) indicate that the EnKF method may be well applied to the case that a huge number of dynamic data are required to assimilate for individual history events, otherwise the generated models after history matching may severely deviate from the real case. One explanation for it might be that the EnKF method is only accurate in the sense of statistics which entails a statistically huge amount of conditioning data to assimilate before producing a reasonable result. This problem matters since it is hard to determine how much data are required to assimilate before yielding reasonable results. Even worse, this method might fail for the reservoir evaluation at the early stage when the dynamic data are far from abundance.

In addition, the generated realizations may be far from *i.i.d* although they are considered to be *a posteriori i.i.d* with respect to the 100 initial seed fields. These realizations are extremely similar and the variance between the models

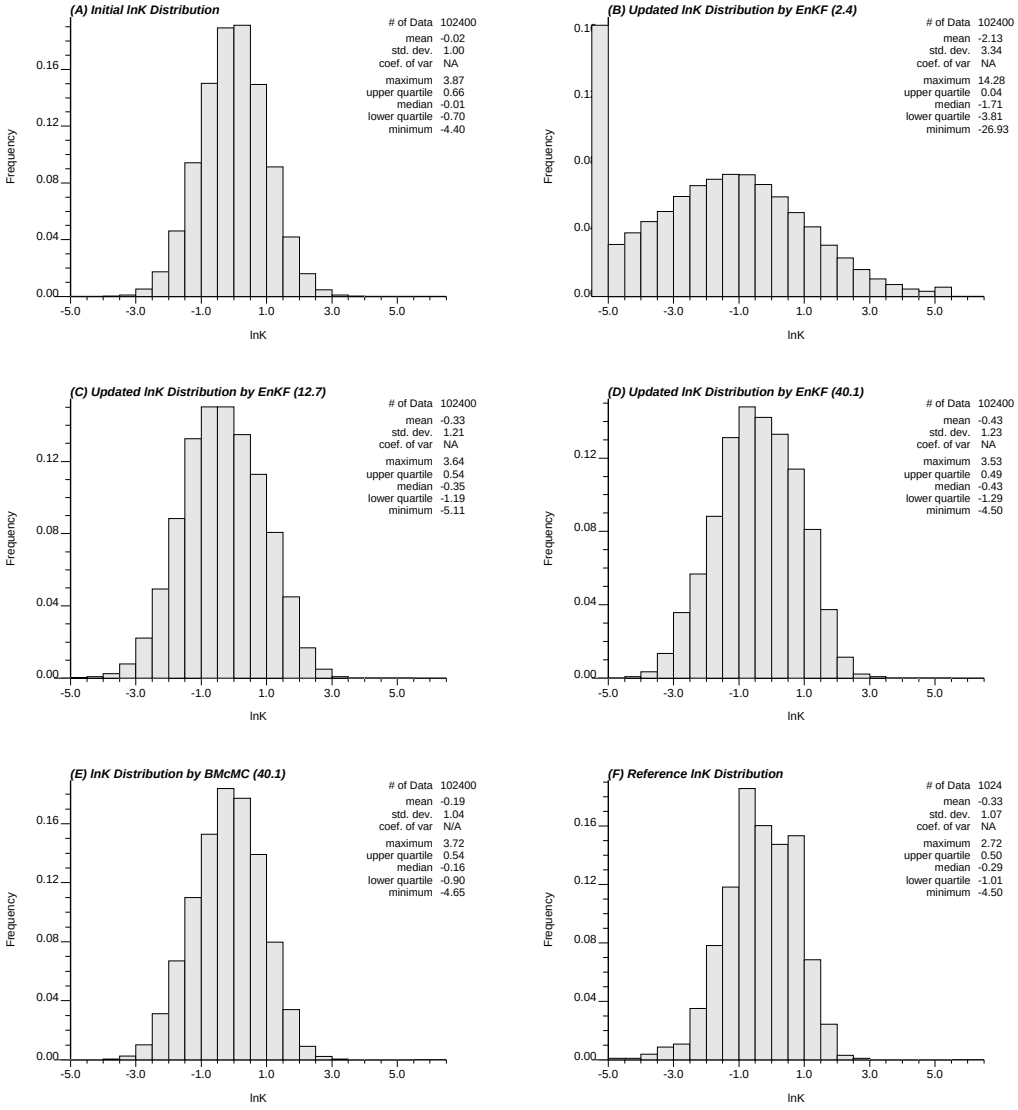


Figure 5.3: *Evolution of histograms of* ln$K$ *during history matching up to different times by the EnKF. The reference, the BMcMC simulation, and the initial seed fields are also listed for comparison.*

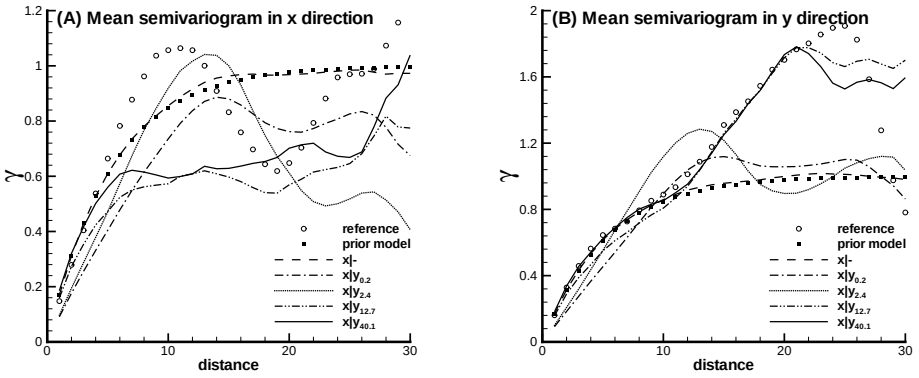


Figure 5.4: *Evolution of mean semivariograms of* ln$K$ *over 100 realizations during history matching up to different times by the EnKF.*

is much smaller than expected. It may come from the fact that the EnKF is essentially a variance minimization method, i.e., the variance between the realizations decreases as more dynamic data are assimilated into the models.

5.3.3 Inverse-conditional modeling by BMcMC

Since the BMcMC is not designed in a Markov process to assimilate the dynamic data, the history matching problem is performed only one time to constrain on the 40.1-day's dynamic observations which include the flow-rate and BHP data from all nine wells. Figure 5.2 also plots the flow-rate and BHP curves at two typical wells after history matching by BMcMC. The history matching obviously reduces, though not much as EnKF in magnitude, the uncertainties of predicted flow-rate and BHP. Moreover, the real case (the circle line) is well contained in the 100 stochastic estimates (the solid lines). The accuracy and precision of predicted BHP (up to 500 days) are improved after matching the first 40.1-day's data. Figure 5.3 and Figure 5.5 display the histogram and mean variogram of the 100 realizations, respectively. The BMcMC method also demonstrates a certain but limited self-adjust in spatial statistics and model structure in order to fit the real case if the prior information is incorrectly specified.

5.3.4 Uncertainty assessment

Although the realizations from the EnKF seem to yield a better reproduction on well histories than those of the BMcMC, the model structure and spatial

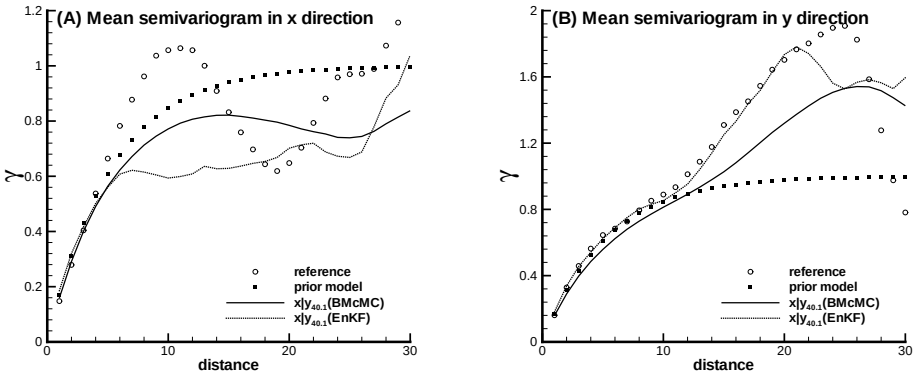


Figure 5.5: *A comparison on the mean semivariograms of* $\ln K$ *by the EnKF and BMcMC. The prior model and the reference field are also plotted for comparison.*

statistics from the EnKF turn out to be worse on the other hand. This part will quantify this uncertainty of models and further point out its side-effect on the future performance prediction at both time and space scale through several new scenarios of reservoir operations.

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 to estimate the simulation errors. Two parameters are computed as the metrics of performance measure to this end, the ensemble average error ($I(\boldsymbol{x})_1$) and the standard deviation of the ensemble average error ($I(\boldsymbol{x})_2$), which are defined as the L_1-norm and L_2-norm between the simulated models and the mean models, i.e.,

$$I(\boldsymbol{x})_1 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r} - \bar{x}_i|, \tag{5.13a}$$

$$I(\boldsymbol{x})_2^2 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r} - \bar{x}_i)^2, \tag{5.13b}$$

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

In case of the synthetic example like this study, however, the model uncertainty can be measured by the simulated errors to validate the efficiency of the inverse methods since the real model is available. In such case, the L_1-norm and L_2-norm between the simulated models and the real models are defined as,

$$I(\boldsymbol{x})_3 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r}^{sim} - x_i^{ref}|, \tag{5.14a}$$

$$I(\boldsymbol{x})_4^2 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r}^{sim} - x_i^{ref})^2, \tag{5.14b}$$

respectively. Note that $\boldsymbol{x}_{ref}$ is the vector of reference attribute values.

In essence, $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2$ measure the degree of precision that the stochastic simulations could render, that is, how narrow the confidence interval of stochastic simulations is. The smaller $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2$ mean the more precise simulation results. $I(\boldsymbol{x})_3$ and $I(\boldsymbol{x})_4$ measure the degree of accuracy that the stochastic simulations may attain, that is, how they are close to the true model. Obviously, the smaller $I(\boldsymbol{x})_3$ and $I(\boldsymbol{x})_4$ are, the closer to the real model the generated realizations are. From the standpoint of estimate and uncertainty, on the other hand, $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_3$ measure the reliability of the estimated models while $I(\boldsymbol{x})_2$ and $I(\boldsymbol{x})_4$ measure the uncertainty of the estimates.

The evolution of the four metrics on model uncertainty from the EnKF method has been compared to the model uncertainties estimated by the BMcMC method in Figure 5.6 from which two observations can be drawn. First, the EnKF simulations at the early stage of data assimilation seriously deviate from the real case. After 30 time steps (i.e., at day 12.7), the estimates from the stochastic simulations begin to settle down to a stable result. Note that $I(\boldsymbol{x})_3$ and $I(\boldsymbol{x})_4^2$ arrive at the best estimated at 12.7 day after which more data assimilation seemingly does not help improve the model uncertainty even though the stochastic simulations from EnKF still show a certain improvement. $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2^2$ are still dropping down after 12.7 days while $I(\boldsymbol{x})_3$ and $I(\boldsymbol{x})_4^2$ almost remain unchanged. Second, BMcMC yields almost the same real stable estimate as EnKF but the EnKF simulations seriously underestimate the model uncertainty compared to the BMcMC simulations. Note that at 40.1 day $I(\boldsymbol{x})_3(\text{EnKF}) = I(\boldsymbol{x})_3(\text{BMcMC})$ and $I(\boldsymbol{x})_4^2(\text{EnKF}) = I(\boldsymbol{x})_4^2(\text{BMcMC})$ but $I(\boldsymbol{x})_1(\text{EnKF}) \ll I(\boldsymbol{x})_1(\text{BMcMC}) < I(\boldsymbol{x})_3$

and $I(\boldsymbol{x})_2^2(\text{EnKF}) \ll I(\boldsymbol{x})_2^2(\text{BMcMC}) < I(\boldsymbol{x})_4^2$. One of the main reasons is that the realizations from the EnKF simulations are extremely similar to each other (i.e., not *i.i.d*) which makes the model uncertainty greatly underestimated.

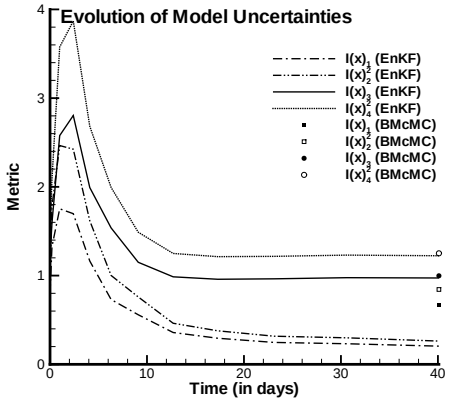


Figure 5.6: *Evolution of model uncertainties predicted by the EnKF. The estimates by BMcMC are also plotted for comparison.*

Uncertainty propagation: Transport performance prediction

Assume that a conservative tracer test operation under a steady-state, natural-gradient, uniform flow is applied to the aquifer due to the alternation of experimental conditions. The prescribed heads are imposed on the left and right sides of the aquifer and impermeable boundaries on the other two faces. A five-point block-centered finite-difference method is employed to solve the steady flow problem. A constant-displacement random-walk particle-tracking algorithm, which proves more efficient than the constant-time-step scheme in numerical computation (*Wen and Gomez-Hernandez*, 1996), is used to solve the conservative solute transport problems. For each realization, continuous 4 000 particles released from the left boundary are tracked until they exit the computational domain. The release source of particles covers the whole left boundary and the particles almost distribute over the entire domain of the field when they move along with the fluid flow such that the particles can largely sample the entire $\ln K(\mathbf{x})$ field and their overall spatial variability has been captured. The breakthrough curves observed at the right side of the aquifer are recorded and the ensemble mean and its 95% confidence interval are plotted in Figure 5.7. One can easily find that the prediction from the BMcMC simulations are better than that of the EnKF simulations. Although

the EnKF simulations provide a narrower confidence interval and the reference BTC is effectively contained in the prediction region, the BMcMC simulations yield a more accurate mean BTC reproduction.

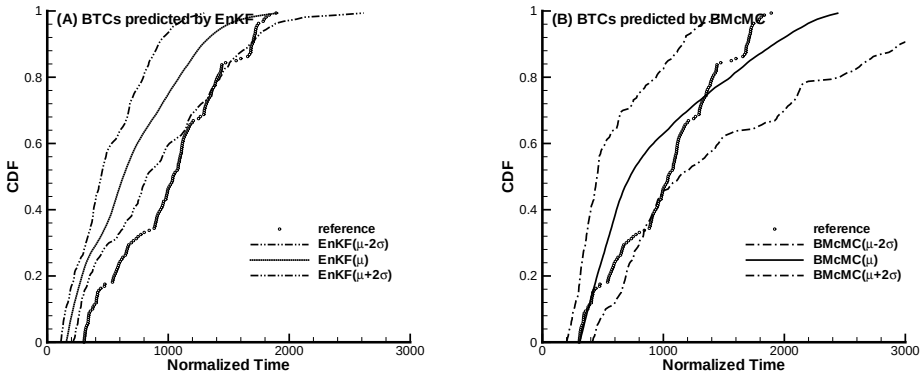


Figure 5.7: *A comparison on the breakthrough curves predicted by the EnKF and BMcMC*

5.4 Case Study 2

Case study 1 presents an example that the prior model has a rather large deviation from the real model. Results show that the EnKF provides a better well performance reproduction but yields a worse transport property prediction. This section will present another example with a more sparse well system but almost perfectly identified prior model for inverse stochastic modeling.

5.4.1 Experimental configuration

Consider a 2D transient single-phase flow test over a confined aquifer with 32×32 grid-blocks under the forced-gradient flow condition as designed in Figure 5.8. The four boundaries are set to be non-flow. The initial head field is assumed to be zero everywhere in the aquifer. The simulation time of total 500 days ($t_0 = 0$ and $t_e = 500$) is discretized into 100 steps, i.e., $n_t = 100$, with α equal to 1.05. Five wells are drilled throughout this confined aquifer: one located at the center is the injection well with a constant flow-rate ($q = -50$) and the other four are production wells with a constant pressure (the piezometric head is maintained at $h = -3.0$ for all four wells). The flow-rate data at the injection well and the BHP data at the four extraction wells are continuously collected at 500 days.

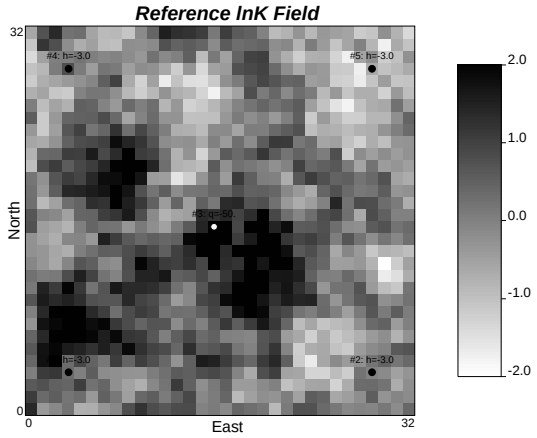


Figure 5.8: *Case study 2: Reference* ln$K$ *field and the well configuration: four constant pressure production wells* (*in bullet*) *and one constant flow-rate injection wells* (*in circle*)

The prior model specification for inverse stochastic simulation is almost perfectly consistent with the real model, e.g., $\ln K \sim N(0, 0.9^2)$ and $\lambda_{\boldsymbol{x}} = 10$ $[cells]$. The correlation length is much shorter than the well spacing ($l_w \doteq 18$ $[cells]$) such that the five wells provide limited information for identifying spatial pattern of model which leaves rather large free space for parameter allocation. On the other hand, the 10-cell correlation length also ensures that the spatial variability is captured by the random models.

The inverse stochastic modeling problem, therefore, is to infer the permeability field (in ln$K$) according to the observed 500-day's flow-rate and BHP data at the five well-bores. The stochastic ln$K$ fields are also required to constrain to the given prior information that is identical to the reference field. Other flow parameters are assumed to be known without uncertainty. Two scenarios of inverse stochastic simulations are carried out by the EnKF and the BMcMC to infer the ln$K$ distribution from the wells' dynamic observations.

5.4.2 Inverse-conditional modeling by EnKF

First, the LUSIM subroutine from GSLIB (*Deutsch and Journel*, 1998) is invoked to generate one hundred seed fields that honor all of the specified prior information. Then, the history matching problem is performed by the EnKF method to sequentially constrain on the 500-day's dynamic observations which include the flow-rate and the BHP data from all five wells. Figure 5.9 compares the flow-rate histories and BHP curves at two typical wells before and after history matching. The models after history matching to the 500-day's dynamic

data obviously reduce the uncertainties of the BHP and flow-rate histories at the wellbores.

Figure 5.10 displays the evolution of parameter histograms during data assimilation. One can easily find that the spreading of $\ln K$ by the EnKF is obviously much wider than expected ($\sigma_{\boldsymbol{x}}^2 = 0.9^2$) as more data are assimilated, e.g., the models at 186.1 days ($\sigma_{\boldsymbol{x}}^2 = 1.36^2$) and at 500 days ($\sigma_{\boldsymbol{x}}^2 = 1.44^2$). The models after 500 days have larger standard deviation than those of 186.1, 40.1, 12.7, 2.4-day's and the initial models. These phenomena are also observed by other researches known as "filter divergence" (e.g., *Nævdal et al.*, 2003) and "overshooting" and "undershooting" problems (e.g., *Gu and Oliver*, 2006).

Figure 5.11 shows the evolution of mean variograms of models during the data assimilation by the EnKF. The mean semivariogram of the initial 100 seed fields (the dashed line) is very close to the prior specification (the square line). As dynamic data are assimilated into the models, however, the mean semivariogram in an unpredicted matter departs from the prior model. Note that the mean variograms in both the x direction (A) and the y direction (B) display an obvious random fluctuation.

5.4.3 Inverse-conditional modeling by BMcMC

The last row in Figure 5.9 plots the flow-rate and BHP curves at the two typical wells after constraining to the 500-day's dynamic observations by the BMcMC. The history matching obviously reduces, though not much as the EnKF in magnitude, the uncertainties of the BHP and flow-rate. Moreover, the real case (the circle line) is well contained in the 100 stochastic estimates (the solid lines). The accuracy and precision of the BHP and flow-rate history reproduction are improved after matching the 500-day's data compared to the unconstraint case (the first column).

More importantly, the generated models follow the prior specification of spatial statistics and model structure. Figure 5.10 and 5.11 also display the histogram and mean semivariogram of the 100 BMcMC simulations, respectively. The BMcMC simulation results perfectly match the prior specification and the reference model, which is obviously closer to the real case and the prior than the EnKF simulations.

5.4.4 Uncertainty assessment

Model uncertainty

The evolution of the four metrics on model uncertainty from the EnKF method has been compared to the model uncertainties estimated by the BMcMC method in Figure 5.12 from which several observations can be drawn. First, as more dynamic data are assimilated into the models, the precision of the

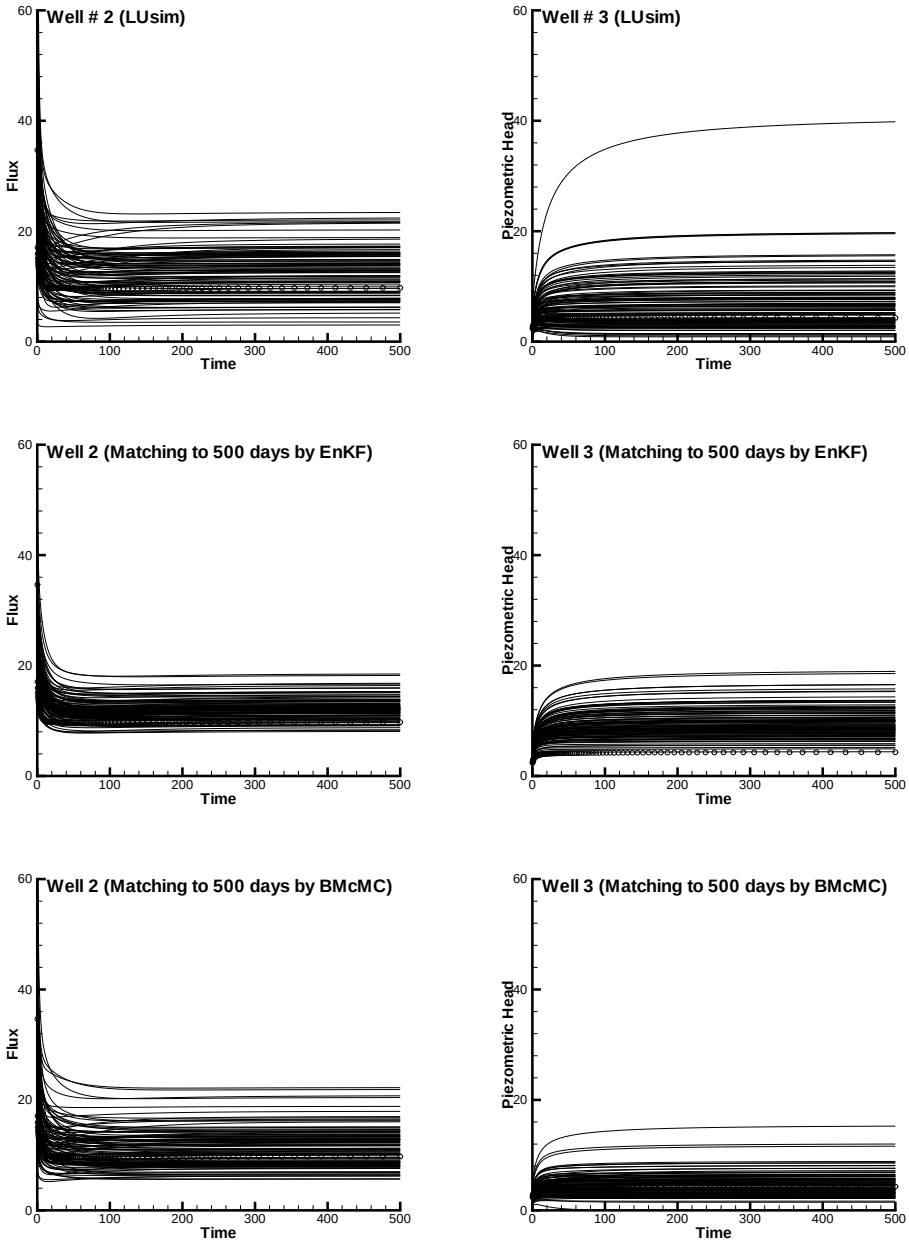


Figure 5.9: *Well performance reproduction by EnKF and BMcMC: The circle line denotes the reference data.*

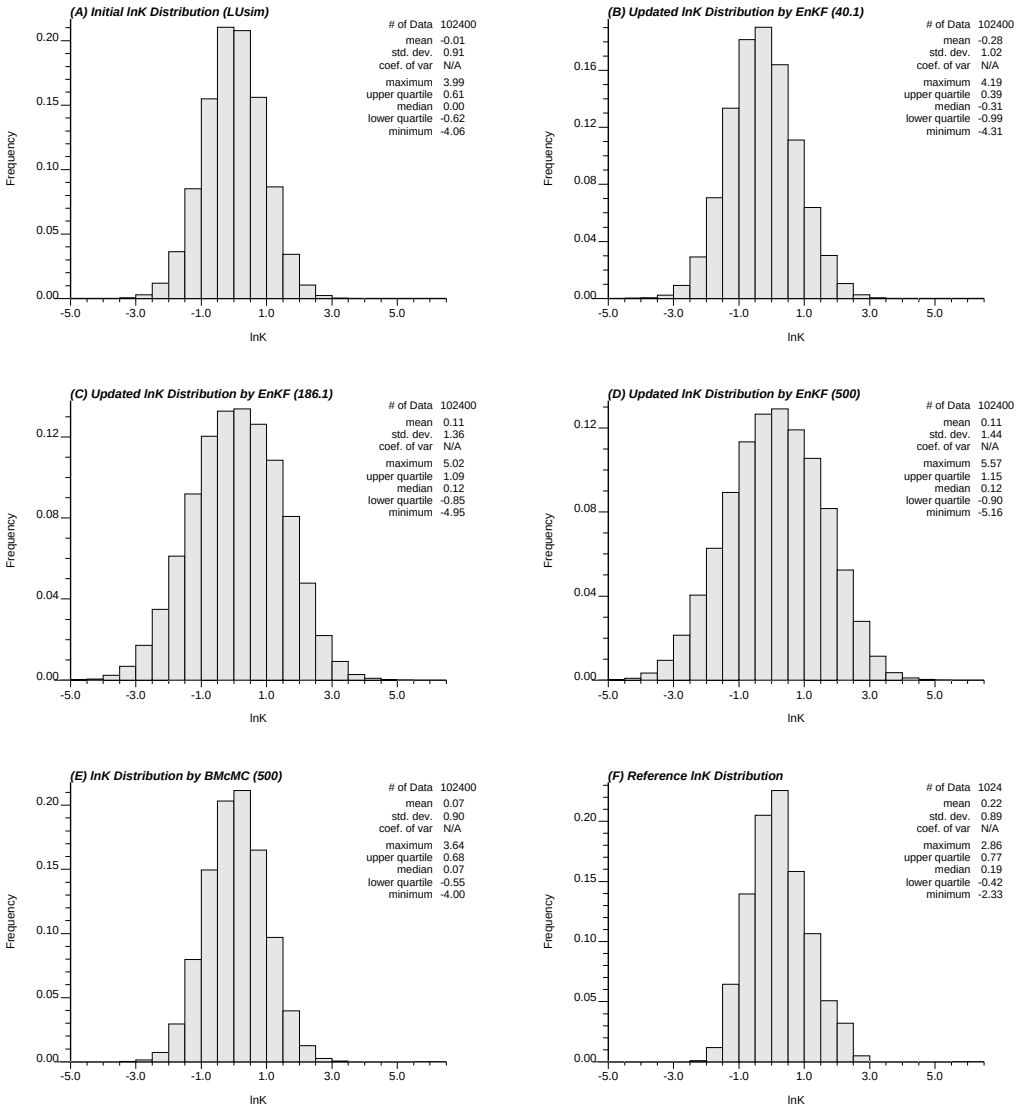


Figure 5.10: *Evolution of histograms of* ln$K$ *during history matching up to different times by the EnKF. The reference, the BMcMC simulation, and the initial seed fields are also listed for comparison.*

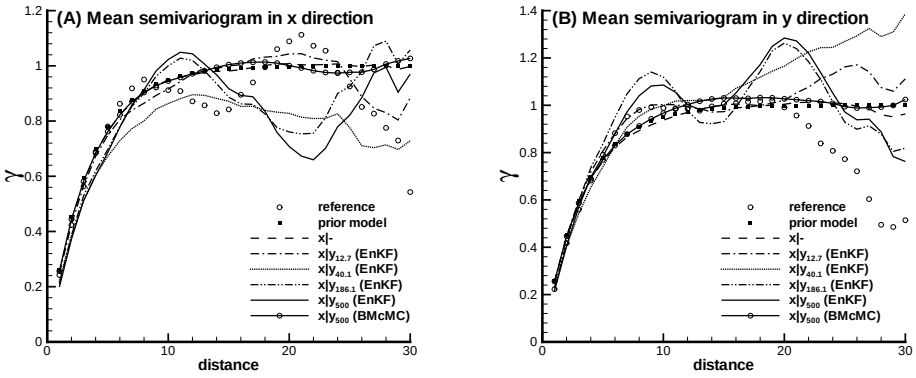


Figure 5.11: *Evolution of mean semivariograms of* $\ln K$ *over 100 realizations during history matching up to different times by the EnKF. The prior model, the reference field, the initial seed fields, and the BMcMC simulations are also plotted for comparison.*

EnKF simulations increases but their accuracy does not increase even worsens in this case study. Note that although both $I(\boldsymbol{x})_3$ and $I(\boldsymbol{x})_4^2$ decrease, which means the precision of simulations increases, $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2^2$ increase at the same time which means the estimates more severely deviate from the real case. Second, the EnKF simulations seriously underestimate the model uncertainty compared to the BMcMC simulations. Note that $I(\boldsymbol{x})_1$(EnKF) $\ll I(\boldsymbol{x})_3$(EnKF) and $I(\boldsymbol{x})_2^2$(EnKF) $\ll I(\boldsymbol{x})_4^2$(EnKF). Furthermore, at the 500 day, $I(\boldsymbol{x})_1$(EnKF) $< I(\boldsymbol{x})_1$(BMcMC) and $I(\boldsymbol{x})_2^2$(EnKF) $< I(\boldsymbol{x})_2^2$(BMcMC) but $I(\boldsymbol{x})_3$(EnKF) $> I(\boldsymbol{x})_3$(BMcMC) and $I(\boldsymbol{x})_4^2$(EnKF) $> I(\boldsymbol{x})_4^2$(BMcMC). In summary, compared to the EnKF, the BMcMC provides a set of more reliable models in terms of model uncertainty.

Uncertainty propagation: Flow performance prediction

Assume that several new wells are added in the reservoir operation system as displayed in Figure 5.13. Three wells in the middle row are designed as the injection wells (in circle) with constant flow rates equal to 30, 80, and 25. Totally six wells in the upper and lower two rows are specified as the pumping wells (in bullet) with constant piezometric heads equal to 4 or 5. Note that well #2, #4, #6, and #8 are newly drilled. Old well conditions change but the aquifer boundaries remain impermeable. The new flow performances of one old well and one new well are plotted in Figure 5.14. For comparison, the unconditional simulations by LUSIM are listed in the first row; and conditional

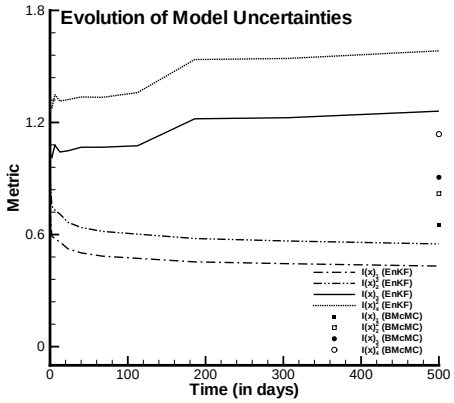


Figure 5.12: *Evolution of model uncertainties predicted by the EnKF. The estimates by BMcMC are also plotted for comparison.*

simulations by the EnKF and BMcMC are listed in the second and third row, respectively. In general, the EnKF yields a better prediction on the flow performance of old wells (e.g., well #3); but at the new well locations, the BMcMC provides a more accurate prediction (e.g., well #6). This result shows that the BMcMC has a more reliable ability in predicting new wells' performance at the spatial scale.

Uncertainty propagation: Transport performance prediction

An identical conservative tracer test operation under a steady-state, natural-gradient, uniform flow as in the case study 1 is applied to the aquifer. The breakthrough curves observed at the right side of the aquifer are recorded and the ensemble mean and its 95% confidence interval are plotted in Figure 5.15. One can easily find that the prediction from the BMcMC simulations are better than that of the EnKF simulations. Although the EnKF simulations provide a narrower confidence interval and the reference BTC is effectively contained in the prediction region, the BMcMC simulations yield a more accurate mean BTC reproduction.

5.5 Discussions and Conclusions

Two stochastic inverse methods, the blocking Markov chain Monte Carlo (BMcMC) and the ensemble Kalman filtering (EnKF), are implemented and compared to perform the model calibration and history matching. The synthetic

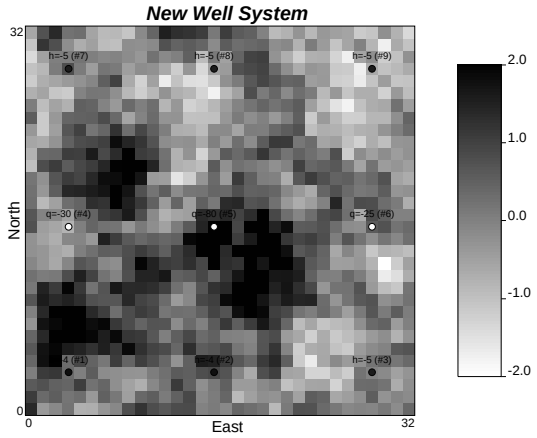


Figure 5.13: *New well system configuration for Case study 2: Six constant pressure production wells* (*in bullet*) *and three constant flow-rate injection wells* (*in circle*)

examples show that both methods provide an improved flow and transport performance prediction than the unconditional case. Detailed examination on generated models, however, shows that the BMcMC can effectively preserve the model structure and spatial statistics as specified *a priori* but the EnKF seemingly fails to do so. If the spatial correlation is short and the well spacing is large, the models after history matching by the EnKF might produce severely biased model structure and spatial statistics even though they can more effectively reproduce the known dynamic observations at the wellbores. The side effect is that it might yield biased prediction on future performance of reservoir and aquifer particularly at the spatial scale. The synthetic example in the case study 2 shows that, as more new wells are added in, the flow performance prediction of the EnKF is much worse than that of the BMcMC although the BMcMC might not provide a better result than the EnKF in predicting the old well performance. Such side effect is further magnified by a conservative tracer test where the BMcMC obviously yields more reliable estimate of the breakthrough curve than the EnKF. In summary, the comparison shows that preserving model structure and spatial statistics for inverse stochastic simulations is of importance for further predicting future reservoir performance in some cases.

The BMcMC method developed in Chapter 3 is extended to an application on a synthetic example under a forced-gradient transient flow condition. This method is proved extremely computationally intensive not only because the McMC itself is CPU demanding but also because the plausible solution set is

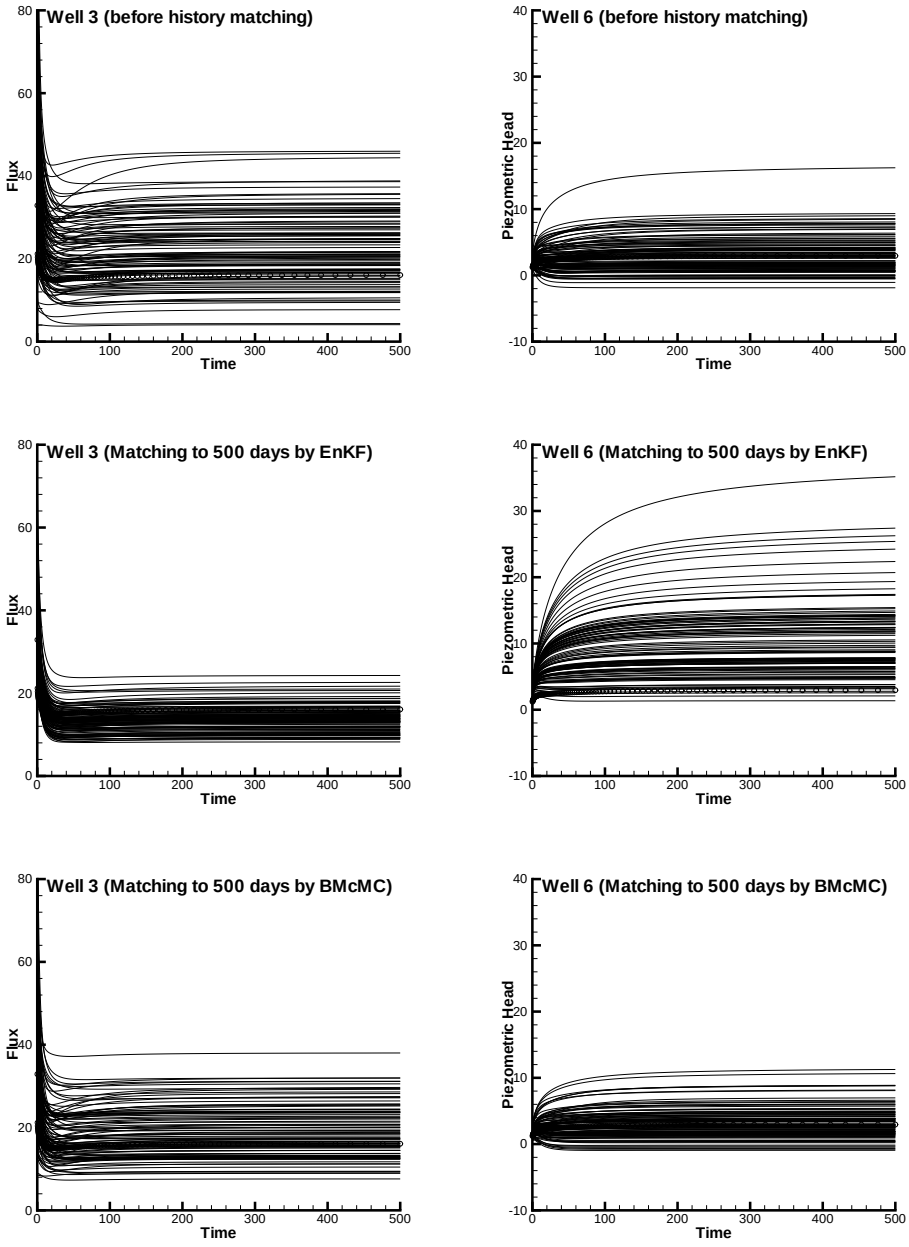


Figure 5.14: *Well performance prediction by EnKF and BMcMC: Note that Well #3 is an old well while well #6 is a new well.*

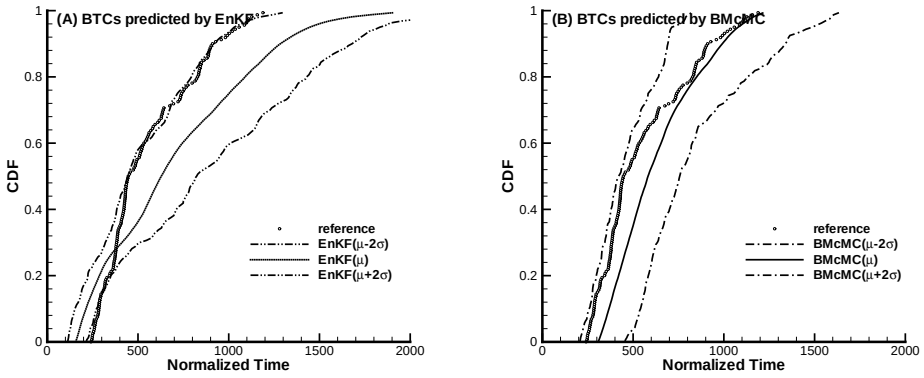


Figure 5.15: *A comparison on the breakthrough curves predicted by the EnKF and BMcMC*

very limited and should be constraint to two-fold factors, the dynamic data and model structure specification. Especially, the model structure preservation essentially reduces the size of the plausible solution set, which often makes unsuccessful the conventional algorithms that resort to the optimization-based model updating. On the other hand, depending on the inconsistence between the prior model and the reality, BMcMC is also capable of adjusting the statistical parameters of models to match to the reality.

Although highly computationally efficient for the history matching problem, the EnKF is insufficient to characterize the uncertainties after model calibration especially in accuracy. First, the updated models have unpredicted spatial structures and parameter statistics, which does not meet the end for refined reservoir characterization since the structure parameters are also a set of data to be honored. The structure instability may make the predicted uncertainties overestimated and deviated from the real case in accuracy. Second, the generated realizations are far from *i.i.d* which results in the predicted uncertainties severely underestimated. For these two reasons, the predicted uncertainty by the EnKF may be biased. The key point of this problem is that the prior information is never explicitly included in the model updating during the model calibration.

6

Conclusions and Suggestions

6.1 Summary

A Markov chain Monte Carlo method for stochastic mappings of physical models is presented in this study to honor both static and dynamic observations on the subsurface reservoir or aquifer properties and to quantitatively assess uncertainties. Several main, novel contributions from this work are summarized as follows.

Any forward solver to the flow and transport problems may be simply integrated into the proposed algorithm as a black-box which makes it very attractive since the computation of the sensitivity coefficient is no more necessary which saves a lot of time in coding and computing. In dealing with a high-dimensional case, however, the limitation of the present computer power available entails a coarser model for forward simulations which calls for an upscaling process that can generate a coarse-scale model from the fine scale. In order to be widely compatible with various upscaling subroutines, a multi-scale-oriented flow and transport simulator is developed to fast but accurately perform the forward simulations at a lower computational cost. Chapter 2 presents a complete description on the implementation details of the flexible-grid full-tensor finite-difference flow simulator and the constant-displacement random-walk particle-tracking transport simulator.

Although some classic McMC methods have been applied into stochastic simulations for honoring both the linear and nonlinear data, their efficiency deserves more improvement in order to ease the intensively computational bur-

den of inverse problems. A blocking McMC scheme is presented in this work to improve the computational efficiency and to preserve the specified spatial statistics and model structure at the same time. This scheme is implemented in an example problem under the framework of multi-Gaussian process. The proposal kernel for the McMC integration are generated very fast on the basis of the LU-decomposition of the covariance matrix owing to its Gaussian property. In addition, an FFT-based spectral conditional simulator, which is known as one of the fastest random field generator, is employed to deal with the high-dimensional case. For the fast computation of the likelihood, an upscaling scheme is used to generate a coarser version of the fine-scale model as the input to the multi-scale-oriented flow and transport simulator presented in Chapter 2 for the fast computation of the likelihood. Chapter 3 gives a detailed description on the development of the proposed MsBMcMC method.

However, the objective of inverse stochastic simulations is not only to generate *i.i.d* realizations that honor all of the given information but also to quantitatively predict reservoir performances and assess their uncertainties. By examining the reduction of uncertainties due to conditioning, the worth of data from various sources can be evaluated and ranked for further sampling design, aquifer remediation operation and reservoir management. Three types of data sources, i.e., static hard data ($\ln K$), piezometric head (h) or pressure data (p), spatial concentration data (c) or temporal moments (t) of tracer data, are systematically evaluated in Chapter 4 with the aid of a synthetic example under a uniform, natural-gradient flow condition. One of the main findings from these experiments is that the uncertainty reduction mostly benefits from measurements on the property of the same type. In other words, the main error source of model or prediction uncertainty comes from the scarcity of measurements on parameters or responses themselves. The measurements on other properties play a secondary but complementary role in reducing the uncertainty of interest by means of jointly conditional simulations. Second, conditioning on static hard data only makes the uncertainties of estimates reduced at the local regions around the measurement points. The uncertainty reduction due to inverse-conditioning to head observations or travel time moments is fully regional. Third, inverse-conditioning to travel time moments significantly improves the estimation on the spreading of solute plume, i.e., the apparent macrodispersion coefficient. Fourth, uncertainty propagation due to conditioning on various types of data is examined and its meaningfulness on the risk assessment of radioactive waste repository is also pointed out.

Besides its stability, well-posedness and flexibility, one of the most striking features, which is also one of the main objectives of this thesis, is that the proposed algorithm can generate *i.i.d* realizations that strictly preserve the specified spatial statistics and model structure while most of inverse methods fail to do so. It matters since the spatial statistics and model structure may

be one of the main error sources for performance prediction and uncertainty assessment at the spatiotemporal scale. Stochastic models should honor all of the given information extracted from geological settings or other sources. Traditional inverse methods tends to destroy these parameters in order to match the observed state data during the procedure of model calibration and history matching. Chapter 5 ascertains such importance by comparing two distinct inverse problem: the BMcMC and the EnKF. Numerical simulations show that, even though the EnKF method may efficiently provide a better reproduction of observed dynamic data than the BMcMC method, the preservation of spatial statistics and model structure makes the BMcMC simulations competitive for some cases in predicting accurately and reliably the future performance of reservoirs particularly at new well locations. Therefore, including the prior information in the inverse stochastic simulation is of significance for accurate assessment of model uncertainties and response uncertainties if the prior parameter information effectively reflects the underground reality.

6.2 Recommendations for Further Research

This study mainly focuses on evaluating the feasibility, the correctness and the usefulness of the proposed McMC scheme under the framework of the multi-Gaussian process. Further extensions to other stochastic processes, more applications to real case studies and detailed comparisons to other inverse methods are expected to carry out in the near future. Specifically, several challenges and interesting topics that deserve more investigations are listed as follows.

- **Improvement of the computational efficiency of McMC**. Although this study made some trials and advances on the McMC in dealing with the inverse problem, the computational efficiency is far from perfect especially for the high-dimensional and high-resolution case which has been proved extremely demanding in computation. Actually, the McMC method is quite an active topic in statistics community up to now. Further advances in easing the computational burden are expected to be integrated into the proposed methods in this work.

- **Extend to real-world field applications**. In this study, all experiments are performed on the basis of several synthetic examples. Accordingly, some general conclusions are hardly drawn only from those toy case studies. The efficiency and effectiveness of the methods are needed to be checked in real applications. Especially the significance of preserving the statistics and spatial structure for models is expected to be investigated in more applications on some real cases.

- **Extend to 3D multi-phase, transient flow problem**. This study only presents the results from the steady-state single-phase flow experiments. Application on transient problems is expected. The extension to two-phase and multi-phase flow problems is also of interest to both petroleum engineering and groundwater communities.

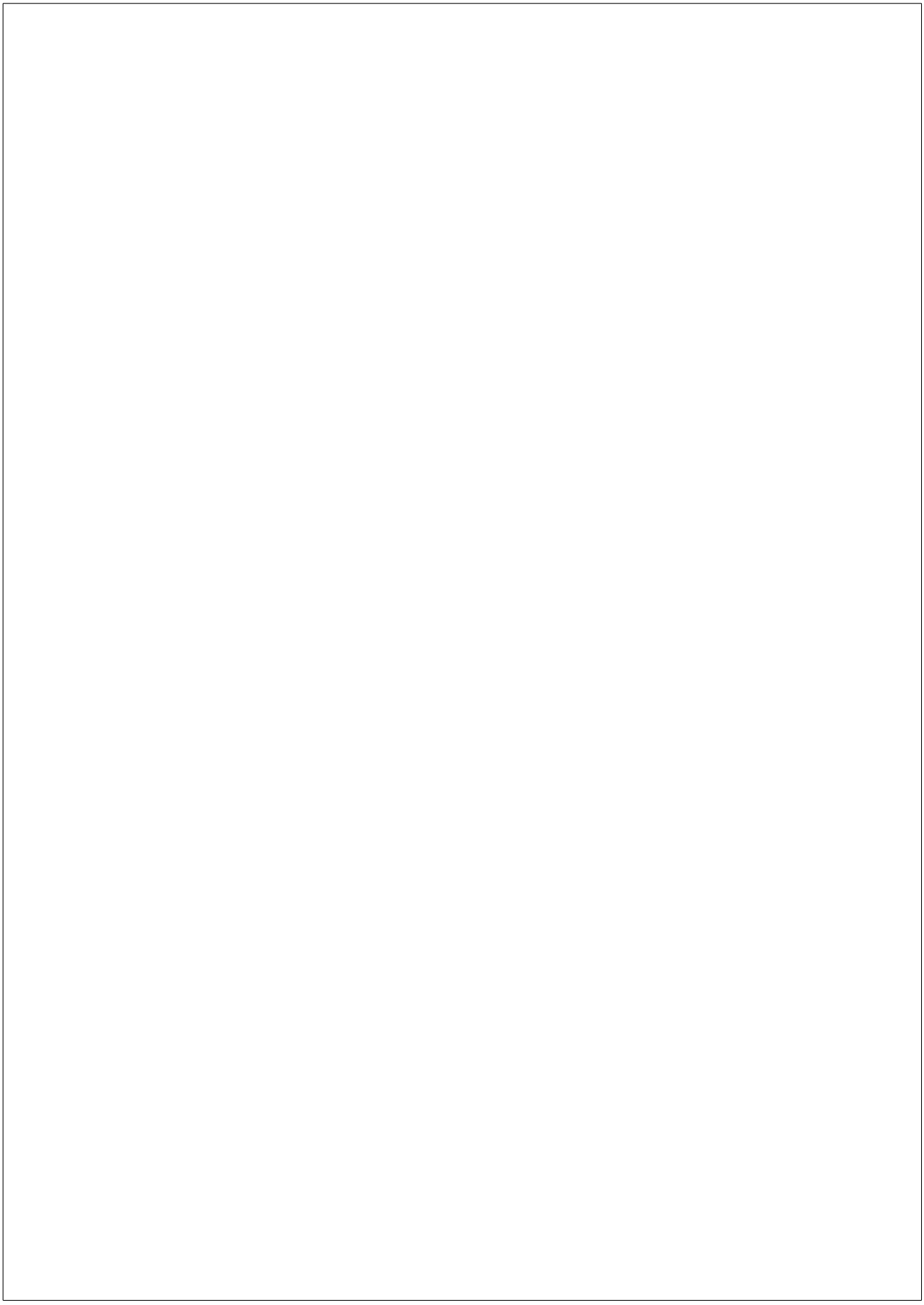
- **Extend to non-Gaussian, multi-modal cases**. One of the important requirements in practice is to constrain the non-Gaussian, multi-modal permeability model on the geological mapping which is based on outcrop's analogue, sedimentary study, geophysical imaging, etc. In other words, use the geological mapping as a training image to generate *i.i.d* non-Gaussian, multi-modal physical models for flow simulations. Even though the proposal kernel is no longer multi-Gaussian, the BMcMC scheme still applies. To this end, however, a new method for generating non-Gaussian, multi-modal candidate fields is needed. In addition, the extension to fracture and fault zones is also a challenging but active topic.

- **Extend to integrate information from other sources**, e.g., discharge or recharge, grain size analysis, temperature (*Woodbury et al.*, 1987), groundwater age (*Varni and Carrera*, 1998), indirect geophysical measurements including remote sensing, seismic and ground-penetrating radar data (*Rubin et al.*, 1992; *Hyndman et al.*, 1994; *Copty and Rubin*, 1995; *Hyndman and Gorelick*, 1996), etc. For example, seismic data may provide densely sampled but low-resolution spatial information which complements the inability of hard data and state data that are most often collected at sparsely distributed well sites. However, a striking problem is that, unlike well logging data, seismic data are highly uncertainly related to hydraulic properties which generally entails a statistical method to extract the spatial information of physical models from the seismic properties. For this reason, indirect geophysical measurements are often called soft data. Integration of soft data into the stochastic simulation is expected to remarkably reduce the uncertainties.

- **Extend to simulate other properties**, e.g., porosity ϕ, distribution coefficient K_d or retardation factor R (*Huang et al.*, 2004), local dispersivity coefficients α_L and α_T (*Nowak and Cirpka*, 2006), mass transfer rate β, etc. For the high-dimensional cases, however, some basic rules should be built up for upscaling these flow and transport parameters, see *Gomez-Hernandez et al.* (2005) for R, *Held and Celia* (2001) and *Christ et al.* (2006) for β, etc. The joint inversion on the combination of permeability ($\ln k$) and other physical parameters is also useful in practice.

- **Integrate into the cost-effective, on-line sampling network design**. The sampling design in the synthetic example presented in this study is purely based on the empirical consideration. It was found that the installation of sampling network (including the sampling density and the network orientation) for collecting conditioning data might have remarkable influence on uncertainty reduction. For example, for the collection of state data (e.g., h, c and t), a sampling spacing less than the correlation scale of physical model along the mean flow direction and mean travel path can greatly reduce prediction uncertainty (*Dagan*, 1985). For the collection of static hard data (e.g., $\ln K$ and ϕ), model uncertainty becomes less important as the number of conditioning data increases since substantial uncertainties remain even at a very high sampling density (*Harter and Yeh*, 1996). But if the economic of the dynamic constraint is considered, the sampling design on uncertainty reduction could be different than that based on the pure scientific value. The meaningfulness of the economical constraint is twofold: one is the cost of data acquisition and the other is the economical value of the uncertainty reduction that the data provide. A cost-effective network design should provide a balance between these two costs. To such problem, a complex, sometimes real-time, optimization procedure is involved to seek the best solution (see *Knopman and Voss*, 1987, 1988a, 1988b, 1989; *Freeze et al.*, 1992; *James and Freeze*, 1993; *James and Gorelick*, 1994; *Wagner*, 1995; *Freeze and Gorelick*, 1999).

- **Integrate into the closed-loop aquifer remediation strategy or reservoir production scheme**. An optimal decision-making process for a closed-loop remediation strategy in contaminant reclamation or a closed-loop reservoir development scheme, which includes well-site selection, pumping-injection rate, etc., should take into account the uncertainties of the models and the predicted responses (*Wagner and Gorelick*, 1987, 1989). The information value of conditioning data (including, $\ln K$, h, c, t, etc.) also depends on to what extent they affect the optimal decision (*Wagner et al.*, 1992). Seamless integration of stochastic conditional and inverse-conditional simulation into such decision-making system is expected to be capable of improve its reliability (*Loaiciga and Mariño* 1987; *Bakr et al.*, 2003).



A

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,

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

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

The objective is to draw randomly a large number of equi-probable realizations from the distribution $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C}_{\boldsymbol{x}})$. 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.

1. Unconditional Sampler

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

(1) Cholesky decompose $\boldsymbol{C_x} = \boldsymbol{L}\boldsymbol{L}^T$, where $\boldsymbol{L} \subset R^{n \times n}$;

(2) Randomly draw $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{1}) \subset R^n$;

(3) Calculate $\boldsymbol{v} = \boldsymbol{L}\boldsymbol{z} \subset R^n$;

(4) Generate an unconditional sample $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{v}$.

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., $\boldsymbol{L}\boldsymbol{z}$, as in the step 3.

2. Conditional Sampler

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

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

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

$$\boldsymbol{x}_2 \sim N(\boldsymbol{\mu}^*, \boldsymbol{C}^*),$$

where,

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_2 + \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1),$$

and,

$$\boldsymbol{C}^* = \boldsymbol{C}_{22} - \boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}.$$

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,

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{C}_{11} & \boldsymbol{C}_{12} \\ \boldsymbol{C}_{21} & \boldsymbol{C}_{22} \end{bmatrix} = \boldsymbol{L}\boldsymbol{U}$$
$$= \begin{bmatrix} \boldsymbol{L}_{11} & 0 \\ \boldsymbol{L}_{21} & \boldsymbol{L}_{22} \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ 0 & \boldsymbol{U}_{22} \end{bmatrix} = \begin{bmatrix} \boldsymbol{L}_{11}\boldsymbol{U}_{11} & \boldsymbol{L}_{11}\boldsymbol{U}_{12} \\ \boldsymbol{L}_{21}\boldsymbol{U}_{11} & \boldsymbol{L}_{21}\boldsymbol{U}_{12} + \boldsymbol{L}_{22}\boldsymbol{U}_{22} \end{bmatrix}.$$

Therefore, $\boldsymbol{L}_{11}$, $\boldsymbol{L}_{21}$, and $\boldsymbol{L}_{22}$ can be obtained by $\boldsymbol{C}_{11} = \boldsymbol{L}_{11}\boldsymbol{U}_{11}$, $\boldsymbol{C}_{21} = \boldsymbol{L}_{21}\boldsymbol{U}_{11}$, and $\boldsymbol{C}_{22} = \boldsymbol{L}_{21}\boldsymbol{U}_{12} + L_{22}\boldsymbol{U}_{22}$, 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 $\boldsymbol{x}$ is obtained by the multiplication of $\boldsymbol{L}$ with a column vector $\boldsymbol{z} \subset R^n$,

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

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

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

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

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

(2) Calculate $\boldsymbol{L}_{21} \subset R^{(n-m)\times m}$ from $\boldsymbol{C}_{21} = \boldsymbol{L}_{21}\boldsymbol{L}_{11}^T$;

(3) Calculate $\boldsymbol{L}_{22} \subset R^{(n-m)\times(n-m)}$ from $\boldsymbol{C}_{22} = \boldsymbol{L}_{21}\boldsymbol{U}_{12} + \boldsymbol{L}_{22}\boldsymbol{L}_{22}^T$;

(4) Calculate the simple kriging field $\boldsymbol{v}_1 = \boldsymbol{L}_{21}\boldsymbol{L}_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1) \subset R^{n-m}$;

(5) Randomly draw $\boldsymbol{z}_2 \sim N(\boldsymbol{0}, \boldsymbol{1}) \subset R^{n-m}$;

(6) Calculate $\boldsymbol{v}_2 = \boldsymbol{L}_{22}\boldsymbol{z}_2 \subset R^{n-m}$;

(7) Generate a conditional sample $\boldsymbol{x}_2 = \boldsymbol{\mu}_2 + \boldsymbol{v}_1 + \boldsymbol{v}_2$.

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., $\boldsymbol{L}_{22}\boldsymbol{z}_2$, as in the step 3.

B

An FFT-based Sampler

Despite the merit of quite cheaply repeatedly generating realizations, the LU-based sampler has rather limited capability of generating high-dimensional random fields due to the expensive, even unavailable, computational resources for the LU-decomposition of a huge covariance matrix. In this regard, the FFT-based spectral simulation algorithm can handle with quite large scale random field. Moreover, it is well known that the FFT-based sampler is one of the fastest generators, which is quite suitable for the BMcMC computation.

Again, our objective is to draw randomly a large number of equi-probable realizations from the distribution $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C_x})$ whose joint prior density is,

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

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

1. Basic theory of FFT

A physical process can be described either in the time domain by its attribute value h as a function of time t, i.e., $h(t)$, or in the frequency domain by its amplitude H as a function of frequency f (or $\omega = 2\pi f$), i.e., $H(f)$ (or

$H(\omega)$). One can go back and forth between these two representations of the same function by means of Fourier transform equation pairs,

$$H(f) = \int_{-\infty}^{\infty} h(t) e^{2\pi i f t} dt,$$
$$h(t) = \int_{-\infty}^{\infty} H(f) e^{-2\pi i f t} df,$$

where,

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}.$$

The Fourier transform of a function can be estimated from a finite number of sampled points, which is known as discrete Fourier transform equations,

$$H_k = \sum_{j=0}^{n-1} h_j e^{-2\pi i \frac{j}{n} k},$$
$$h_j = \frac{1}{n} \sum_{k=0}^{n-1} H_k e^{-2\pi i \frac{k}{n} j}.$$

One appealing point is that these two Fourier transform can be quite efficiently computed using the FFT algorithm.

Since it often happens that a physical process can be performed more efficiently in the frequency domain than attacking it in the time domain, one can take advantage of the fact that this physical process can be transformed to and from using the Fourier transform, typically by FFT, and that the computation is efficiently performed in the frequency domain. For example, the convolution of two digital signals can be computed much more efficiently by transforming them to the frequency domain and performing an element-wise multiplication there instead of a series of scalar products in the time domain.

Similarly, a stochastic process $\boldsymbol{x}$, which is assumed to be second-order stationary in the space domain, can be constructed in terms of its spectral representation $\boldsymbol{\omega}$ in the "Fourier" domain by,

$$\boldsymbol{x} = \int_{-\infty}^{\infty} e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\boldsymbol{\omega}.$$

The Fourier increment of the attribute value, $d\boldsymbol{\omega}$, must satisfy the following orthogonal conditions,

$$
\begin{aligned}
E[d\boldsymbol{\omega}] &= \mathbf{0}, \\
E[d\boldsymbol{\omega} \cdot d\boldsymbol{\omega}^*] &= \mathbf{0}; \qquad \boldsymbol{\omega} \neq \boldsymbol{\omega}^*, \\
E[d\boldsymbol{\omega} \cdot d\boldsymbol{\omega}^*] &= |S(\boldsymbol{\omega}) \cdot d\boldsymbol{\omega}|; \qquad \boldsymbol{\omega} = \boldsymbol{\omega}^*,
\end{aligned}
$$

where $E[]$ indicates the mathematical expectation operator, * denotes the complex conjugate operator, the differential vector $d\boldsymbol{\omega}$ represents the n-dimensional differential wave number volume elements, $d\boldsymbol{\omega} = (d\omega_0, d\omega_1, ..., d\omega_{n-1})^T \subset R^n$, and $S(\boldsymbol{\omega})$ is the density spectral function which is linked to the covariance function $C(\boldsymbol{h})$ through the Fourier transform,

$$
\begin{aligned}
C(\boldsymbol{h}) &= \int_{-\infty}^{\infty} e^{i\boldsymbol{\omega}\cdot\boldsymbol{h}} \cdot S(\boldsymbol{\omega}) \cdot d\boldsymbol{\omega}, \\
S(\boldsymbol{\omega}) &= \int_{-\infty}^{\infty} e^{-i\boldsymbol{\omega}\cdot\boldsymbol{h}} \cdot C(\boldsymbol{h}) \cdot d\boldsymbol{h}.
\end{aligned}
$$

If one can generate a stochastic process $\boldsymbol{\omega}$ in the "Fourier" domain, which has the statistics satisfying orthogonal conditions, then one can generate its counterpart $\boldsymbol{x}$ in the space domain, e.g., by the numerical integration through the discrete Fourier transform. From this fact we can do the simulation in the "Fourier" domain but retain the structure feature of the specified covariance in the space domain. Why and how should we do simulations in the "Fourier" domain? The trick lies in that we only need to multiply the amplitude, which is derived from the specified covariance function, by a randomly drawn phase, which has a uniform distribution between 0 and 1 (or between 0 and 2π in the angular frequency), to form a stochastic process in the "Fourier" domain, which can be easily and quickly mapped to the space domain through the fast Fourier transform.

2. Unconditional Sampler

The classical spectral representation theorem shows that any sequence $\boldsymbol{x}$ can be expressed as a finite series of Fourier coefficients, α and β:

$$
x_k = F^{-1}(\boldsymbol{a}) = \sum_{j=0}^{n-1} a_j e^{2\pi i \frac{k}{n} j} = \sum_{j=0}^{n-1} \left[\alpha_j \cos\left(2\pi \frac{k}{n} j\right) + i\beta_j \sin\left(2\pi \frac{k}{n} j\right) \right],
$$

where $k \in [0, n)$, $a_j = \alpha_j - i\beta_j = |a_j| e^{-i\phi_j}$ is the jth complex Fourier coefficient, $|a_j| = \sqrt{\alpha_j^2 + \beta_j^2}$ is the amplitude, and $\phi_j = \arctan(-\beta_j/\alpha_j)$ is the

phase of the jth Fourier coefficient. The amplitude $|a_j|$ is related to the discrete spectral density s_j by $|a_j|^2 = s_j$, $j \in [0, n)$. The complex Fourier coefficient is given by,

$$a_j = |a_j|e^{-i\phi_j} = |a_j|\cos\phi_j - i|a_j|\sin\phi_j,$$

where $j \in [0, n)$, and the phase ϕ_j is drawn randomly from the uniform distribution between 0 and 2π.

The idea of generating an unconditional random field $\boldsymbol{x}$ given a covariance $C(\boldsymbol{h})$ is as follows. Inverse Fourier transform of $\boldsymbol{a}$ provides a discrete finite realization of $\boldsymbol{x}$ with a specified covariance spectrum $\boldsymbol{s}$. Perturbing the phase $\boldsymbol{\phi}$ will produce a series of such realizations. This procedure can be done very fast with prime factor FFT.

Assume that a stationary Gaussian process is imposed over an equi-spaced regular grid, a symmetric covariance matrix $\boldsymbol{C_x}$ is constructed by the two-point geostatistics as follows,

$$\boldsymbol{C_x} = \begin{bmatrix} C_0 & C_1 & C_2 & \cdots & C_{n-1} \\ C_1 & C_0 & C_1 & \cdots & C_{n-2} \\ C_2 & C_1 & C_0 & \cdots & C_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n-1} & C_{n-2} & C_{n-3} & \cdots & C_0 \end{bmatrix}.$$

A stochastic realization $\boldsymbol{x}$ can be generated by $\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{Lz}$, $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{1})$, using the LU-decomposition of the matrix $\boldsymbol{C_x} = \boldsymbol{LL}^T$, i.e., the square root decomposition of the matrix. However, a high-dimensional case makes the matrix decomposition easy to fail due to the limitation of computer resources and the ubiquitousness of numerical round-off errors. Fortunately, the symmetric matrix $\boldsymbol{C_x}$ can be extended to form a larger positive definite matrix $\boldsymbol{S} \subset R^{2n\times 2n}$ with a symmetric circulant structure,

$$\boldsymbol{S} = \begin{bmatrix} C_0 & C_1 & C_2 & \cdots & C_{n-1} & C_n & C_{n-1} & C_{n-2} & \cdots & C_1 \\ C_1 & C_0 & C_1 & \cdots & C_{n-2} & C_{n-1} & C_n & C_{n-1} & \cdots & C_2 \\ C_2 & C_1 & C_0 & \cdots & C_{n-3} & C_{n-2} & C_{n-1} & C_n & \cdots & C_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{n-1} & C_{n-2} & C_{n-3} & \cdots & C_0 & C_1 & C_2 & C_3 & \cdots & C_n \\ C_n & C_{n-1} & C_{n-2} & \cdots & C_1 & C_0 & C_1 & C_2 & \cdots & C_{n-1} \\ C_{n-1} & C_n & C_{n-1} & \cdots & C_2 & C_1 & C_0 & C_1 & \cdots & C_{n-2} \\ C_{n-2} & C_{n-1} & C_n & \cdots & C_3 & C_2 & C_1 & C_0 & \cdots & C_{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C_1 & C_2 & C_3 & \cdots & C_n & C_{n-1} & C_{n-2} & C_{n-3} & \cdots & C_0 \end{bmatrix},$$

such that the square root of this matrix can be efficiently computed by the fast Fourier transform, i.e.,

$$\boldsymbol{S} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T = (\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}})(\boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}})^T,$$

where $\boldsymbol{V}$ is the eigenvector matrix; $\boldsymbol{\Lambda}$ is a diagonal matrix consisting of the nonnegative eigenvalues; and the superscript T means the conjugate transpose. Accordingly, a stochastic realization in the Fourier space $\tilde{\boldsymbol{x}}$ can be generated by $\tilde{\boldsymbol{x}} = \boldsymbol{V}\boldsymbol{\Lambda}^{\frac{1}{2}}\boldsymbol{z}$, where $\boldsymbol{z} = \boldsymbol{z}_1 + i\boldsymbol{z}_2$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \sim N(\boldsymbol{0}, \boldsymbol{1})$. Its counterpart in the spatial domain $\boldsymbol{x}$ can be obtained simply by applying an inverse FFT operation to $\tilde{\boldsymbol{x}}$. Note that only the first column $\boldsymbol{s} = (C_0, C_1, C_2, ..., C_{n-1}, C_n, C_{n-1}, ..., C_1)^T$ is needed for the numerical implementation.

Sample Algorithm 1. An FFT-based unconditional simulation algorithm typically includes,

(1) Construct the covariance matrix $\boldsymbol{C_x}$;

(2) Construct the vector $\boldsymbol{s} = (C_0, C_1, C_2, ..., C_{n-1}, C_n, C_{n-1}, ..., C_1)$;

(3) Compute the FFT $\tilde{\boldsymbol{s}}$ of the vector $\boldsymbol{s}$;

(4) Compute the vector $\sqrt{\tilde{\boldsymbol{s}}}$;

(5) Generate a random vector $\boldsymbol{z} = \boldsymbol{z}_1 + i\boldsymbol{z}_2$, where $\boldsymbol{z}_1, \boldsymbol{z}_2 \sim N(\boldsymbol{0}, \boldsymbol{1})$;

(6) Construct the Fourier coefficients $\boldsymbol{a} = \sqrt{\tilde{\boldsymbol{s}}}\boldsymbol{z}$;

(7) Generate a realization through the inverse FFT of $\boldsymbol{a}$;

(8) Go to step (5) for the next realization.

Since the implementation of the spectral simulation as above will yield some deviations of mean and variance from the real case, some corrections will have to be applied to generate the desired physical models.

Sample Algorithm 2. The procedure of the unconditional simulation consists of,

(1) Sample the discrete covariance, $\boldsymbol{C}$. There are some additional details for correcting the mean and variance deviations:

- Shift the covariance: First calculate the covariance values by $C(h) = C(h_0) - \gamma(h)$. Then calculate their mean $\mu(C)$. Shift all covariance values by the current ones, $C^*(h) = C(h) - \mu(C)$, so that the shifted covariance has a zero mean.

- Rescale the covariance by $C(h_0)/(C(h_0) - \mu(C))$, so that the generated realizations have the correct variance.

(2) Calculate the discrete spectral density, $\boldsymbol{s}$. This procedure can be done by FFT transforming of the sampled covariance sequence, i.e.,

$$s_j = \frac{1}{n}\sum_{k=0}^{n-1} C_k e^{-2\pi i \frac{j}{n}k}; \qquad j \in [0, n).$$

(3) Retrieve the amplitude $|\boldsymbol{a}|$ by $|a_j| = \sqrt{s_j}$, $j \in [0, n)$.

(4) Randomly draw the phase $\boldsymbol{\phi}$ by,

$$\phi_j \sim U(0, 2\pi); \qquad j \in [0, n).$$

(5) Build up the Fourier coefficients $\boldsymbol{a}$, and thus $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, by,

$$a_j = |a_j|e^{-i\phi_j} = |a_j|\cos\phi_j - i|a_j|\sin\phi_j; \qquad j \in [0, n).$$

(6) Generate the random field $\boldsymbol{x}$ by inverse FFT transforming $\boldsymbol{a}$, e.g.,

$$\begin{aligned} x_k = F^{-1}(\boldsymbol{a}) &= \sum_{j=0}^{n-1} a_j e^{2\pi i \frac{k}{n}j} \\ &= \sum_{j=0}^{n-1}\left[\alpha_j \cos\left(2\pi\frac{k}{n}j\right) + i\beta_j \sin\left(2\pi\frac{k}{n}j\right)\right]; \qquad k \in [0, n). \end{aligned}$$

(7) Go to step (4) for the next realization.

3. Conditional Sampler

The realizations output from the FFT-based unconditional simulator are independent, identically distributed (*i.i.d*) which need to be constraint to the hard data, *a posteriori* observations, that is, to minimize the objective function,

$$J = \sum_{i=0}^{m-1}(x_i^{sim} - x_i^{obs})^2,$$

where $x_i = x_i^{sim}$ is the unconditional simulation value at the observation location and x_i^{obs} is the corresponding observed value. It can be viewed as to tune the unconditional realizations to honor the observation.

Kriging tunes the realizations simply by setting the objective function to zero. In this sense, Kriging is a minimum (square) error estimator. The critical step in the kriging-based conditional simulation is to perform a smooth interpolation of the difference between conditioning data and the unconditional simulation values. A traditional way to conditioning on local data is simply to add an independently simulated residual by kriging into the unconditional simulation (*Journel and Huijbregts*, 1978). This process calls for solving one kriging system per location. In the code implementation, the basic conditional simulation can be expressed by the kriging mean and error components conditional to linear data plus a stochastic component through any unconditional simulation. Mathematically it can be written as follows,

$$\boldsymbol{x} = \boldsymbol{x}_u + (\boldsymbol{x}_K^* - \boldsymbol{x}_{u,K}^*),$$

where $\boldsymbol{x}_u$ is the stochastic component through the FFT-based sampler and $\boldsymbol{x}_K^* - \boldsymbol{x}_{u,K}^*$ is the kriging error.

Sample Algorithm 3. Steps to obtaining a conditional simulation are as follows,

(1) Generate an unconditional realization $\boldsymbol{x}_u$, e.g., by an FFT-based method;

(2) Obtain the kriging estimate $\boldsymbol{x}_K^*$ using the data $\boldsymbol{x}_{obs}$;

(3) Obtain the kriging estimate $\boldsymbol{x}_{u,K}^*$ using the unconditional simulated data $\boldsymbol{x}_u$ at the corresponding conditioning locations;

(4) Add the kriging error $\boldsymbol{x}_u - \boldsymbol{x}_{u,K}^*$ to the kriging estimate $\boldsymbol{x}_K^*$.

Bibliography

Alabert, F., 1987. The practice of fast conditional simulations through the LU decomposition of the covariance matrix, *Mathematical Geology*, 19(5), 369-386.

Anderman, E.R., and M.C. Hill, 1999. A new multistage groundwater transport inverse method: Presentation, evaluation, and implications, *Water Resources Research*, 35(4), 1053-1063.

Bakr, M.I., C.B.M. te Stroet, and A. Meijerink, 2003. Stochastic groundwater quality management: Role of spatial variability and conditioning, *Water Resources Research*, 39(4), 1078.

Beck, M.B., 1987. Water quality modeling: A review of the analysis of uncertainty, *Water Resources Research*, 23(8), 1393-1442.

Bellin, A., and Y. Rubin, 2004. On the use of peak concentration arrival time for the inference of hydrogeological parameters, *Water Resources Research*, 40(7), W07401.

Brooks, S.P., 1998. Quantitative convergence assessment for Markov chain Monte Carlo via cusums, *Statistics and Computing*, 8(3), 267-274.

Carrera, J., A. Alcolea, A. Medina, J. Hidalgo, and L.J. Slooten, 2005. Inverse problem in hydrogeology, *Hydrogeology Journal*, 13, 206-222.

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.

Cirpka, O.A., and P.K. Kitanidis, 2001. Sensitivity of temporal moments calculated by the adjoint-state method and joint inversing of head and tracer data, *Advances in Water Resources*, 24(1), 89-103.

Christ, J.A., C.A. Ramsburg, K.D. Pennell, and L.M. Abriola, 2006. Estimating mass discharge from dense nonaqueous phase liquid source zones using upscaled mass transfer coefficients: An evaluation using multiphase numerical simulations, *Water Resources Research*, 42(11), W11420.

Copty, N., and Y. Rubin, 1995. A stochastic approach to the characterization of lithofacies from surface seismic and well data, *Water Resources Research*, 31(7), 1673-1686.

Criminisi, A., T. Tucciarelli, and G.P. Karatzas, 1997. A methodology to determine optimal transmissivity measurement locations in groundwater quality management models with scarce field information, *Water Resources Research*, 33(6), 1265-1274.

Dagan, G., 1984. Solute transport in heterogeneous porous formations, *Journal of Fluid Mechanics*, 145, 151-177.

Dagan, G., 1985. Stochastic modeling of groundwater flow by unconditional and conditional probabilities: the inverse problem, *Water Resources Research*, 21(1), 65-72.

Dagan, G., 1989. *Flow and Transport in Porous Formations*, Springer-Verlag, Berlin, pp465.

Datta-Gupta, A., L.W. Lake, and G.A. Pope, 1995. Characterizing heterogeneous permeable media with spatial statistics and tracer data using sequential simulated annealing, *Mathematical Geology*, 27(6), 763-787.

Davis, M.W., 1987. Production of conditional simulations via the LU triangular decomposition of the covariance matrix, *Mathematical Geology*, 19(2), 91-98.

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.

Desbarats, A.J., and R.M. Srivastava, 1991, Geostatistical characterization of groundwater flow parameters in a simulated aquifer, *Water Resources Research*, 27(5), 687-698.

Deutsch, C.V., and A.G. Journel, 1998. *GSLIB: Geostatistical software library and user's guide*, second edition, Oxford University Press, pp369.

Eggleston, J.R., S.A. Rojstaczer, and J.J. Peirce, 1996. Identification of hydraulic conductivity structure in sand and gravel aquifers: Cape Cod data set, *Water Resources Research*, 32(5), 1209-1222.

Evensen, G., 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *Journal of Geophysical Research*, 99(C5), 10143-10162.

Evensen, G., 2003. The ensemble Kalman fiter: theoretical formulation and practical implementation, *Ocean Dynamics*, 53, 343-367.

Evensen, G., and P.J. van Leeuwen, 2000. An ensemble Kalman smoother for nonlinear dynamics, *Monthly Weather Review*, 128, 1852-1867.

Ezzedine, S., and Y. Rubin, 1996. A geostatistical approach to the conditional estimation of spatially distributed solute concentration and notes on the use of tracer data in the inverse problem, *Water Resources Research*, 32(4), 853-861.

Fernandez-Garcia, D., T.H. Illangasekare, and H. Rajaram, 2005. Differences in the scale-dependence of dispersivity estimated from temporal and spatial moments in chemically and physically heterogeneous porous media, *Advances in Water Resources*, 28(7), 745-759.

Franssen, H.J.H., J.J. Gomez-Hernandez, and A. Sahuquillo, 2003. Coupled inverse modeling of groundwater flow and mass transport and the worth of concentration data, *Journal of Hydrology*, 281, 281-295.

Freeze, R.A., and S.M. Gorelick, 1999. Convergence of stochastic optimization and decision analysis in the engineering design of aquifer remediation, *Ground Water*, 37(6), 934-954.

Freeze, R.A., B. James, J. Massmann, T. Sperling, and L. Smith, 1992. Hydrogeological decision analysis: 4. The concept of data worth and its use in the development of site investigation strategies, *Ground Water*, 30(4), 574-588.

Gaganis, P., and L. Smith, 2001. A Bayesian approach to the quantification of the effect of model error on the predictions of groundwater models, *Water Resources Research*, 37(9), 2309-2322.

Gao, G., and A.C. Reynolds, 2004. An improved implementation of the LBFGS algorithm for automatic history matching, *SPE Annual Technical Conference and Exhibition*, SPE 90058.

Gao, G., M. Zafari, and A.C. Reynolds, 2005. Quantifying the uncertainty for the PUNQ-S3 problem in a Bayesian setting with the RML and EnKF, *SPE Reservoir Simulation and Symposium*, SPE 93324.

Gelhar, L.W., 1993. *Stochastic Subsurface Hydrology*, Prentice Hall, New Jersery, pp390.

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.

Gomez-Hernandez, J.J. and X.-H. Wen, 1997. To be or not to be multi-Gaussian? A reflection on stochastic hydrology, *Advances in Water Resources*, 21(1), 47-61.

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.

Gomez-Hernandez, J.J., Jianlin Fu, and D. Fernandez-Garcia, 2005. Upscaling retardation factors in 2-D porous media, *Calibration and Reliability in Groundwater Modelling: From Uncertainty to Decision Making*, IAHS Publ. 304, 130-136.

Graham, W., and D. McLaughlin, 1989a. Stochastic analysis of nonstationary subsurface solute transport: 1. Unconditional Moments, *Water Resources Research*, 25(2), 215-232.

Graham, W., and D. McLaughlin, 1989b. Stochastic analysis of nonstationary subsurface solute transport: 2. Conditional Moments, *Water Resources Research*, 25(11), 2331-2355.

Gutjahr, A., B. Bullard, and S. Hatch, 1997. General joint conditional simulations using a Fast Fourier Transform Method, *Mathematical Geology*, 29(3), 361-389.

Harter, T., and T.-C.J. Yeh, 1996. Conditional stochastic analysis of solute transport in heterogeneous, variably saturated soils, *Water Resources Research*, 32(6), 1597-1609.

Harvey, C.F., and S.M. Gorelick, 1995. Mapping hydraulic conductivity: Sequential conditioning with measurements of solute arrival time, hydraulic head, and local conductivity, *Water Resources Research*, 31(7), 1615-1626.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their application, *Biometrika*, 57(1), 97-109.

Held, R.J., and M.A. Celia, 2001. Pore-scale modeling and upscaling of nonaqueous phase liquid mass transfer, *Water Resources Research*, 37(3), 539-549.

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.

Hu, L.Y., 2000. Gradual deformation and iterative calibration of Gaussian-related stochastic models, *Mathematical Geology*, 32(1), 87-108.

Huang, H., B.X. Hu, X.-H. Wen, C. Shirley, 2004. Stochastic inverse mapping of hydraulic conductivity and sorption partitioning coefficient fields conditioning on nonreactive and reactive tracer test data, *Water Resources Research*, 40(1), W01506.

Hyndman, D.W., and S.M. Gorelick, 1996. Estimating lithologic and transport properties in three dimensions using seismic and tracer data: The Kesternson aquifer, *Water Resources Research*, 32(9), 2659-2670.

Hyndman, D.W., J.M. Harris, and S.M. Gorelick, 1994. Coupled seismic and tracer test inversion for aquifer property characterization, *Water Resources Research*, 30(7), 1965-1977.

James, B.R., and R.A. Freeze, 1993. The worth of data in predicting aquitard continuity in hydrogeological design, *Water Resources Research*, 29(7), 2049-2065.

James, B.R., and S.M. Gorelick, 1994. When enough is enough: The worth of monitoring data in aquifer remediation design, *Water Resources Research*, 30(12), 3499-3513.

Journel, A.G., and C. Huijbregts, 1978. *Mining Geostatsitics*, Academic Press, New York, pp600.

Khaleel, R., 1994. Scale and directional dependence of macrodispersivities in colonnade networks, *Water Resources Research*, 30(12), 3337-3355.

Kitanidis, P.K., 1986. Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resources Research*, 22(4), 499-507.

Kitanidis, P.K., 1996. On the geostatistical approach to the inverse problem, *Advances in Water Resources*, 19(6), 333-342.

Kitanidis, P.K., and E.G. Vomvoris, 1983. A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resources Research*, 19(3), 677-690.

Knopman, D. S., and C. L. Voss, 1987. Behavior of sensitivities in the one-dimensional advection-dispersion equation: Implications for parameter estimation and sampling design, *Water Resources Research*, 23(2), 253-272.

Knopman, D. S., and C. L. Voss, 1988a. Further comments on sensitivities, parameter estimation and sampling design in one-dimensional analysis of solute transport in porous media, *Water Resources Research*, 24(2), 225-238.

Knopman, D. S., and C. L. Voss, 1988b. Discrimination among one-dimensional models of solute transport in porous media: Implications for sampling design, *Water Resources Research*, 24(11), 1859-1876.

Knopman, D. S., and C. L. Voss, 1989. Multiobjective sampling design for parameter estimation and model discrimination in groundwater solute transport, *Water Resources Research*, 25(10), 2245-2258.

Kreft, A., and A. Zuber, 1978. On the physical meaning of dispersion equation and its solutions for different initial and boundary conditions, *Chemical Engineer Sciences*, 33, 1471-1480.

LaBolle, E.M., G.E. Fogg, A.F.B. Tompson, 1996. Random-walk simulation of transport in heterogeneous porous media: Local mass-conservation problem and implementation methods, *Water Resources Research*, 32(3), 585-593.

LeBlanc, D.R., S.P. Garabedian, K.M., Hess, L.W. Gelhar, R.D. Quadri, K.G. Stollenwerk, and W.W. Wood, 1991. Large-scale natural gradient tracer test in sand and gravel, Cape Cod, Massachusetts: 1. Experimental design and observed tracer movement, *Water Resources Research*, 27(5), 895-910.

Lemke, L.D., W.A. Barrack II, L.M. Abriola, and P. Goovaerts, 2004. Matching solute breakthrough with deterministic and stochastic aquifer models, *Ground Water*, 42(6), 920-934.

Li, R., A.C. Reynolds, and D.S. Oliver, 2003. History matching of three-phase flow production data, *SPE Journal*, 8, 328.

Lichtner, P.C., S. Kelkar, and B. Robinson, 2002. New form of dispersion tensor for axisymmetric porous media with implementation in particle tracking, *Water Resources Research*, 38(8), 1146.

Lin, Z.Y., 1992. On the increments of partial sums of a -mixing sequence, *Theoretical Probability and its Applications*, 36, 316-326.

Liu, J.S., 1996. Metropolized independent sampling with comparisons to rejection sampling and importance sampling, *Statistics and Computing*, 6(2), 113-119.

Loaiciga, H.A., and M.A. Mariño, 1987. Parameter estimation in groundwater: classical, Bayesian, and deterministic assumptions and their impact on management policies, *Water Resources Research*, 23(6), 1027-1035.

McKinney, D.C., and D.P. Loucks, 1992. Network design for predicting groundwater contamination, *Water Resources Research*, 28(1), 133-147.

McLaughlin, D., and L.R. Townley, 1996. A reassessment of the groundwater inverse problem, *Water Resources Research*, 32(5), 1131-1161.

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.

Nowak, W., and O.A. Cirpka, 2006. Geostatistical inference of hydraulic conductivity and dispersivities from hydraulic heads and tracer data, *Water Resources Research*, 42(8), W08416.

Oliver, D.S., N. He, and A.C. Reynolds, 1996. Conditioning permeability fields to pressure data, *European Conference for the Mathematics of Oil Recovery, V*.

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.

Omre, H., and H. Tjelmeland, 1996. Petroleum geostatistics. Technical Reports, s-8. Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.

Pardo-Iguzquiza, E., and M. Chica-Olmo, 1993. The Fourier integral method: An efficient spectral method for simulation of random fields, *Mathematical Geology*, 25(4), 177-217.

Pritchett, J.W., and S.K. Garg, 1980. Determination of effective well block radii for numerical reservoir simulations, *Water Resources Research*, 16(4), 665-674.

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.

Robert, C.P., and G. Casella, 1999. *Monte Carlo Statistical Methods*, Springer-Verlag, pp. 507.

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.

Robin, M.J.L., A.L. Gutjahr, E.A. Sudicky, and J.L. Wilson, 1993. Cross-correlated random field generation with the direct Fourier transform method, *Water Resources Research*, 29(7), 2385-2397.

Ruan, F., and D. McLaughlin, 1998. An efficient multivariate random field generator using the fast Fourier transform, *Advances in Water Resources*, 21(5), 385-399.

Rubin, Y., 1991. Prediction of tracer plume migration in disordered porous media by the method of conditional probabilities, *Water Resources Research*, 27(6), 1291-1308.

Rubin, Y., and G. Dagan, 1992. Conditional estimation of solute travel time in heterogeneous formations: Impact of transmissivity measurements, *Water Resources Research*, 28(4), 1033-1040.

Rubin, Y., and S. Ezzedine, 1997. The travel times of solutes at the Cape Cod tracer experiment: Data analysis, modeling, and structural parameters inference, *Water Resources Research*, 33(7), 1537-1547.

Rubin, Y., G. Mavko, and J. Harris, 1992. Mapping permeability in heterogeneous aquifers using hydrologic and seismic data, *Water Resources Research*, 28(7), 1809-1816.

Spear, R.C., 1970. The application of Kolmogorov-Renyi statistics to problems of parameter uncertainty in system design, *International Journal of Control*, 11, 771-778.

Scheibe, T.D., and Y.-J. Chien, 2003. An evaluation of conditioning data for solute transport prediction, *Ground Water*, 41(2), 128-141.

Sun, N.-Z., and W.W.-G. Yeh, 1985. Identification of parameter structure in groundwater inverse problem, *Water Resources Research*, 21(6), 869-883.

Sun, N.-Z., and W.W.-G. Yeh, 1990a. Coupled inverse problems in groundwater modeling: 1. Sensitivity analysis and parameter identification, *Water Resources Research*, 26(10), 2507-2525.

Sun, N.-Z., and W.W.-G. Yeh, 1990b. Coupled inverse problems in groundwater modeling: 2. Identifiability and experimental design, *Water Resources Research*, 26(10), 2527-2540.

Vasco, D.W., and A. Datta-Gupta, 1999. Asymptotic solutions for solute transport: A formalism for tracer tomography, *Water Resources Research*, 35(1), 1-16.

Vasco, D.W., A. Datta-Gupta, J.C.S. Long, 1997. Resolution and uncertainty in hydrologic characterization, *Water Resources Research*, 33(3), 379-397.

Van Leeuwen, M., A.P. Butler, C.B.M. te Stroet, and J.A. Tompkins, 2000. Stochastic determination of well capture zones conditioned on regular grids of transmissivity measurements, *Water Resources Research*, 36(4), 949-957.

Wagner, B.J., and S.M. Gorelick, 1987. Optimal groundwater quality management under parameter uncertainty, *Water Resources Research*, 23(7), 1162-1174.

Wagner, B.J., and S.M. Gorelick, 1989. Reliable aquifer remediation in the presence of spatial variable hydraulic conductivity: from data to design, *Water Resources Research*, 25(10), 2211-2225.

Wagner, B.J., U. Shamir, and H.R. Nemati, 1992. Groundwater quality management under uncertainty: stochastic programming approaches and the value of information, *Water Resources Research*, 28(5), 1233-1246.

Wagner, B.J., 1995. Sampling design methods for groundwater modeling under uncertainty, *Water Resources Research*, 31(10), 2581-2591.

Wen, X.-H., and J.J. Gomez-Hernandez, 1996a. The constant displacement scheme for tracking particles in heterogeneous aquifers, *Ground Water*, 34(1), 135-142.

Wen, X.-H., and J.J. Gomez-Hernandez, 1996b. Upscaling hydraulic conductivities in heterogeneous media: an overview, *Journal of Hydrology*, 183(1-2), ix-xxxii.

Wen, X.-H., and J.J. Gomez-Hernandez, 1998. Numerical modeling of macrodispersion in heterogeneous media: a comparison of multi-Gaussian and non-multi-Gaussian models, *Journal of Contaminant Hydrology*, 30, 129-156.

Wen, X.-H., C.V. Deutsch, and A.S. Cullick, 2002. Construction of geostatistical aquifer models integrating dynamic flow and tracer data using inverse technique, *Journal of Hydrology*, 255, 151-168.

Wilson, A., and Y. Rubin, 2002. Characterization of aquifer heterogeneity using indicator variables for solute concentrations, *Water Resources Research*, 38(12), 1283.

Woodbury, A.D., and Y. Rubin, 2000. A full-Bayesian approach to parameter inference from tracer travel time moments and investigation of scale effects at the Cape Cod experimental site, *Water Resources Research*, 36(1), 159-171.

Woodbury, A.D., and E.A. Sudicky, 1992. Inversion of the Borden Tracer Experimental Data: Investigation of stochastic moment models, *Water Resources Research*, 28(9), 2387-2398.

Woodbury, A.D., J.L. Smith, and W.S. Dunbar, 1987. Simultaneous inversion of temperature and hydraulic data: 1. Theory and application using hydraulic head data, *Water Resources Research*, 23(8), 1586-1606.

Yeh, W.W.-G., 1986. Review of parameter identification procedures in groundwater hydrology: The inverse problem, *Water Resources Research*, 22(1), 95-108.

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.

Zafari, M., and A.C. Reynolds, 2005. Assessing the uncertainty in reservoir description and performance predictions with the ensemble Kalman filter, *SPE Annual Technical Conference and Exhibition*, SPE 95750.

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.