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

Doctoral Thesis submitted by
Jianlin Fu

Advisor:
J. Jaime Gómez-Hernández

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

PhD Thesis submitted by
Jianlin Fu

Advisor:
J. Jaime Gómez-Hernández

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

February 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

A classical two-stage method to stochastic inverse problems in groundwater and petroleum engineering starts from the generation of a series of independent seed fields and then calibrates those fields to inverse-condition on nonlinearly dependent state data from different sources, which is known as model calibration or history matching. However, an inherent deficiency exists in this type of method: the spatial structure and statistics are not preserved during the procedure of model calibration and history matching. While 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. Aiming at this problem, a stochastic approach is presented in this study to generate independent, identically distributed ($i.i.d$) realizations which are not only conditional on static linear data and inverse-conditional on dynamic nonlinear data but also have the specified spatial structure and statistics.

Unlike those traditional two-stage methods, a conditional and inverse-conditional simulation approach may directly generate $i.i.d$ 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 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 main characteristics of the blocking McMC (BMcMC) scheme is that it fully preserves the prior spatial structure and statistics as users specified. In addition, it improves the mixing of the Markov chain and hence enhances the computational efficiency of the McMC. The LU-decomposition-based sampler is used

to rapidly generate a huge number of candidates. A synthetic example is presented to demonstrate the performance of the proposed BMcMC.

Second, in order to make the BMcMC method capable of dealing with the high-dimensional cases, a multi-scale 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 computational efforts are spent on the black-box-liked 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. The generation of the proposal kernel is also fast even for the high-dimensional case owing to the employment of an FFT-based spectral sampler. The performance of the proposed multi-scale blocking McMC (MsBMcMC) scheme is widely evaluated on the basis of a synthetic example.

Finally, 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 constraint 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.

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 for three years 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. *Teresa Martinez*, *Javier Moreno* and other sectaries 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, on 2006. The EnKF algorithm and the case study in the introduction was carried out 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 mem-

orable period to share an 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 find so hard. My former tutor *Dr. Zhuoheng Chen* also gave me a lot of supports. The development of the FFT-based conditional sampler in this thesis benefited from the discussions with *Dr. Chen*.

Finally, I would like to thank *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

CONTENTS   xiii

List of Figures

List of Tables

LIST OF TABLES

<div align="right">

1

</div>

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 $p(\boldsymbol{x})$ 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, in order to further reduce the uncertainty 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 mathematic, 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 on the measurement scale typically with a higher precision, such as permeability measured at the core scale or derived from well-logging data, and soft data at a larger scale with a lower precision, such as seismic data. There also is another type of soft data available 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 mathematical tools currently available. The widely used non-conditioning and linear conditioning algorithms include: the LU-decomposition algorithm (*Davis*, 1987; *Alabert*, 1987), the sequential Gaussian simulation, the sequential indicator simulation, the p-field simulation, the simulated annealing algorithm (*Deutsch and Journel*, 1998), the FFT-based spectral method (*Pardo-Iguzquiza and Chica-Olmo*, 1993; *Robin et al.*, 1993; *Gutjhar et al.*, 1997; *Ruan and McLaughlin*, 1998), etc.

The other type is the nonlinear data 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 crite-

rion. 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 widely used inverse-conditioning algorithms 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.

But what is wrong with these traditional model calibration and history matching methods to stochastic inverse-conditional simulation for the purpose of generating *i.i.d* realizations that should strictly honor all information available? One of the main deficiencies is that the realizations generated as such do not have the specified spatial statistics and structure for parameters. The key point is that all these methods to generating conditional and inverse-conditional realizations overemphasize on minimizing the objective function which are defined by the mismatch between the simulated and the observed dynamic state data. The problem arises during the optimization of physical models where the model parameters (and thus the parameter statistics and model structure) have to be mathematically adjusted (and thus the prior parameters are destroyed) to honor the nonlinear observations. The point is that the prior information is not obeyed during the model update. Such weakness of the classical two-stage methods is inherent and may be unavoidable since the imposition of two or more images, for example, tends to modify the structure properties very easily. Even worse, some of them, e.g., the newly developed and popular ensemble Kalman filtering (EnKF) method, pay too much attention to the mathematics or statistics of model inputs and outputs where the physical laws and prior geological information are intentionally ignored in order to accelerate the computation. One of the shortcomings is that some unreasonable overshooting (unrealistically large values) and undershooting (unreasonably small values) are easily introduced to the mathematically or statistically optimized models (e.g., *Li et al.*, 2003; *Gao and Reynolds*, 2004; *Gao et al.*, 2005).

A simple synthetic experiment might be carried out to demonstrate the problems that the classical history matching methods encounter. Consider a 2D transient flow test on a confined aquifer with 32×32 grid-blocks as designed in Figure 1.1 under the forced-gradient flow condition. The four boundaries are set to be non-flow. The initial head field is assumed to be zero everywhere in the aquifer. Nine wells are drilled throughout this aquifer: five of them are the injection wells with a constant flow-rate (20.5 per day) and the other four are production wells with a constant pressure (the piezometric head is maintained at -3.0 for all four wells). The flow-rate data (q) of the five injection wells and the bottom-hole-pressure data (BHP) of the four extraction wells are continuously collected at the first 40.1 days. The inverse stochastic simulation problem, therefore, is to infer the permeability field ($\ln K$) according to the observed well flow-rate and BHP data. The stochastic $\ln K$ fields are also required to be constraint to some given prior information, e.g., $\ln K \sim N(0, 1)$. Other flow parameters are assumed to be constant and known perfectly.
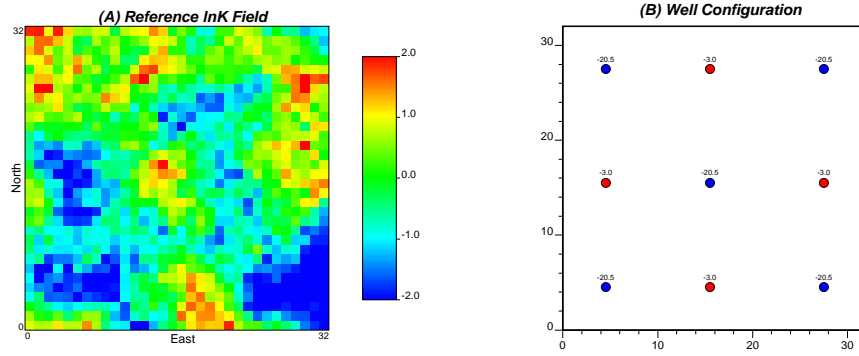
Figure 1.1: *A 2D synthetic example (A) and well configuration (B): constant pressure production wells (red) and constant flow-rate injection wells (blue)*

First, the LUSIM subroutine (*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 extremely efficiently performed by the EnKF method (Appendix C gives some details on one of its implementations for this study) to constrain on the 40.1-day's dynamic observations which include the flow-rate and BHP data of all nine wells. Figure 1.2 compares the BHP curves at the well # 6 before and after history matching. The history matching obviously reduces the uncertainty of predicted BHP. Moreover, the real case (the red line) is well contained in the one hundred stochastic estimates (the black lines). The accuracy and precision of predicted BHP (up to 500 days) are highly improved only after matching the first 40.1 days' data.
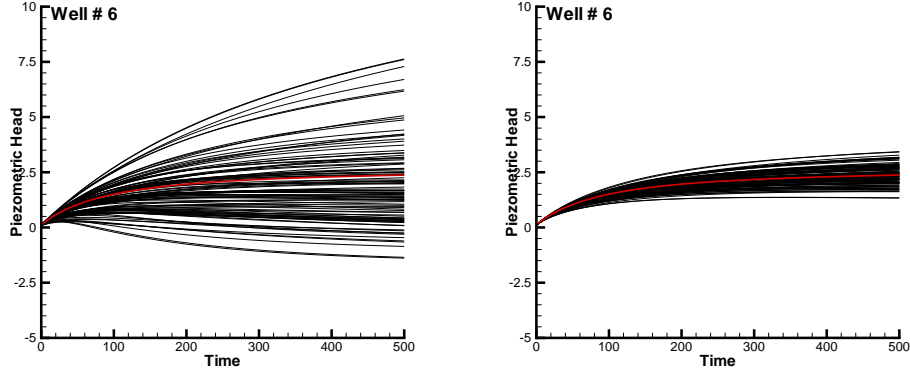
Figure 1.2: *Well # 6 bottom-hole-pressure (BHP) history: before history matching (left) and after history matching up to 40.1 days (right). The red line denotes the reference BHP curve and the black lines represent those of 100 stochastic realizations.*

A similar uncertainty reduction is also observed at the well # 1 whose flow-rate history is simulated and plotted in the Figure 1.3. One can easily find that the EnKF extremely efficiently and effectively completes the pure history matching problem. The reference well flow-rate history is perfectly reproduced after history matching even though the prediction from the initial seed fields severely deviate from the real case. Note that the simulated flow-rates from the initial 100 realizations (the black lines) are much higher than the real case (the red line).

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 severely deviate from the prior specification for models, even though the prediction of model responses at the known well locations seems to attain to an ideal result at the time scale and the updated models do look more similar to the reference field than the initial seed fields. Figure 1.4 shows the evolution of one realization (# 1) during the dynamic data assimilation by the EnKF. At the early stage, e.g, at the day 0.2 and the day 2.4, some unreasonably big parameters and some extremely small values frequently occur in the physical models which obviously do no more follow the prior specification, $\ln K \sim N(0, 1)$.

This phenomenon may be seen more clearly in Figure 1.5 which displays the evolution of parameter histograms during the data assimilation. One can easily find that the spreading of $\ln K$ at the early stage is obviously much wider than expected. Note that the permeability is shown at the natural logarithm
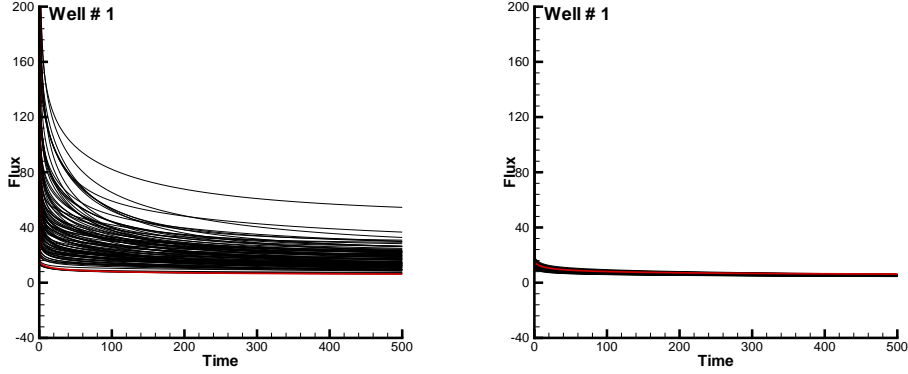
Figure 1.3: *Well # 1 flow-rate history: before history matching (left) and after history matching up to 40.1 days (right). The red line denotes the reference curve and the black lines represent those of 100 stochastic realizations.*

scale, i.e., $\ln K$, which implies a small fluctuation of $\ln K$ will cause a huge change of K. The observations from this experiment indicate that the EnKF may be only 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 this might be that the EnKF is only accurate in terms of statistics which calls for a statistically huge number of conditioning data to assimilate before producing a reasonable result.

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 initial 100 seed fields. Figure 1.6 randomly shows six realizations out of 100 updated models after history matching up to the 40.1 days. These realizations are extremely similar and the variance between the models is much smaller than expected. The influence is that it my severely underestimate the uncertainty of future performance prediction. Moreover, the spatial statistics and structure of models might disobey the given prior information since the prior is never used to constrain the models during the history matching.

Actually, the spatial statistics and structure of models, which form the basis of the representation of the subsurface heterogeneity and should strictly obey the geologic background, may be one of the more important error sources for reservoir characterization and performance prediction. This study tries to draw attention to the spatial structure problem and points out the effects on
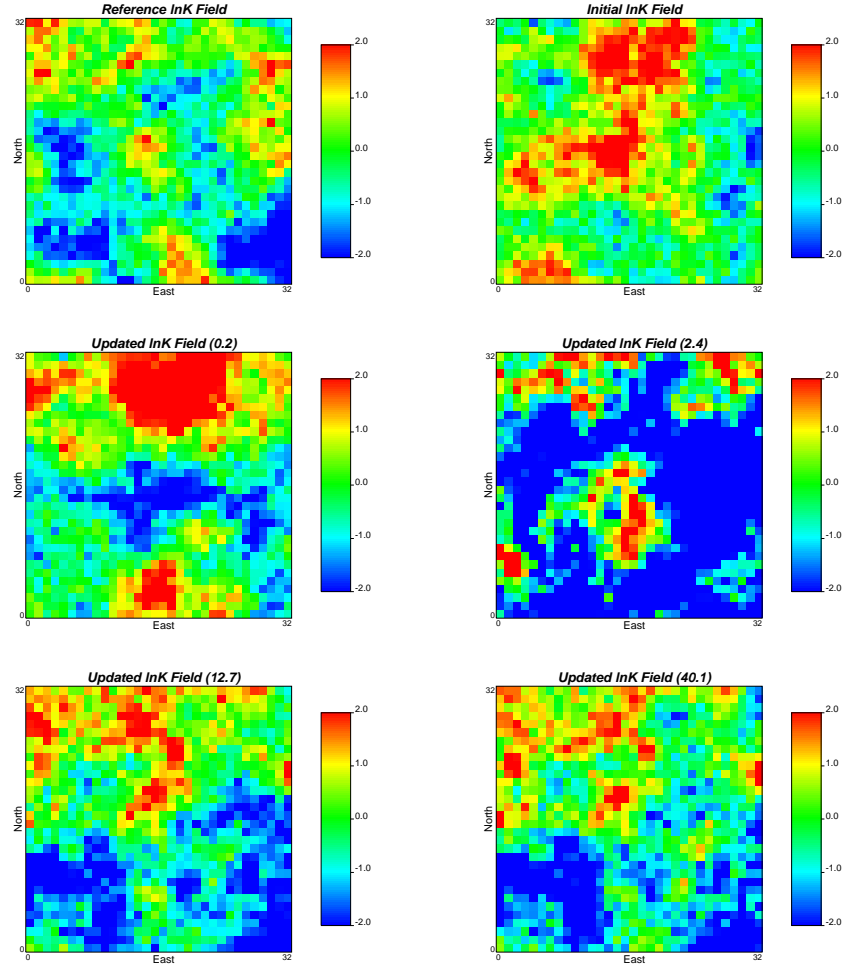
Figure 1.4: *Evolution of a* $\ln K$ *realization (#1) during history matching up to different times (0.2, 2.4, 12.7, and 40.1 days). The reference model and the initial seed field are listed in the first row.*
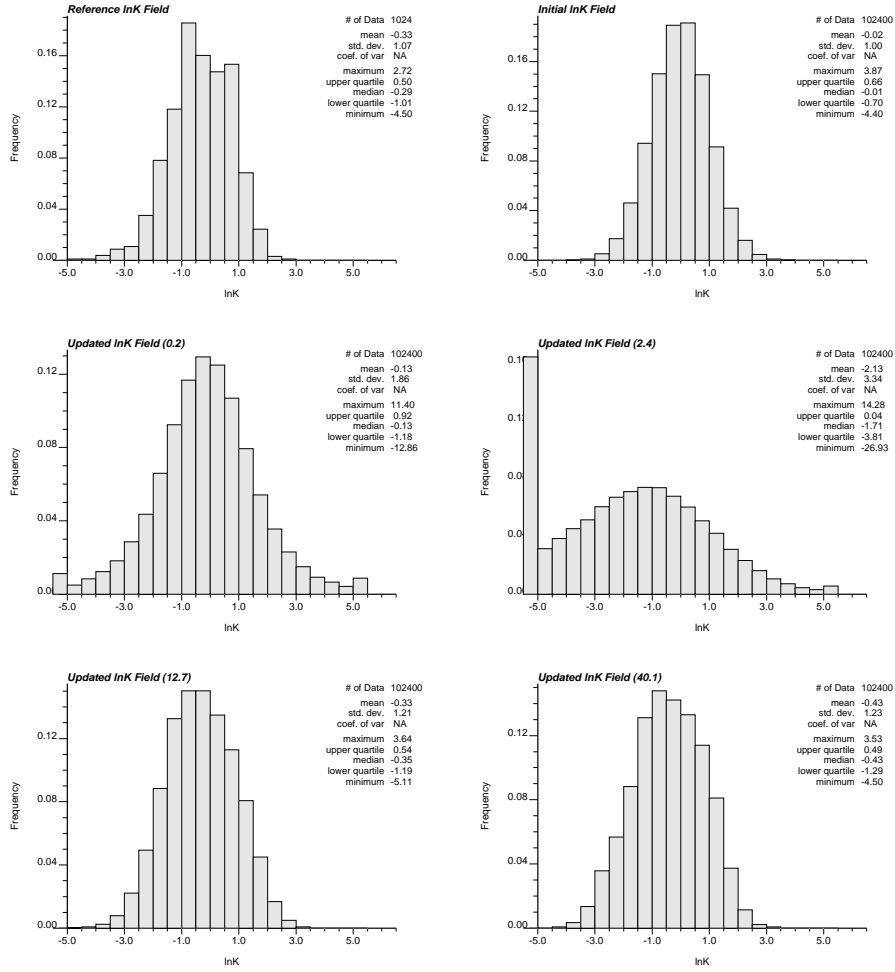
Figure 1.5: *Evolution of histograms of* $\ln K$ *during history matching up to different times (0.2, 2.4, 12.7, and 40.1 days). The reference model and the initial seed fields are listed on the first row.*
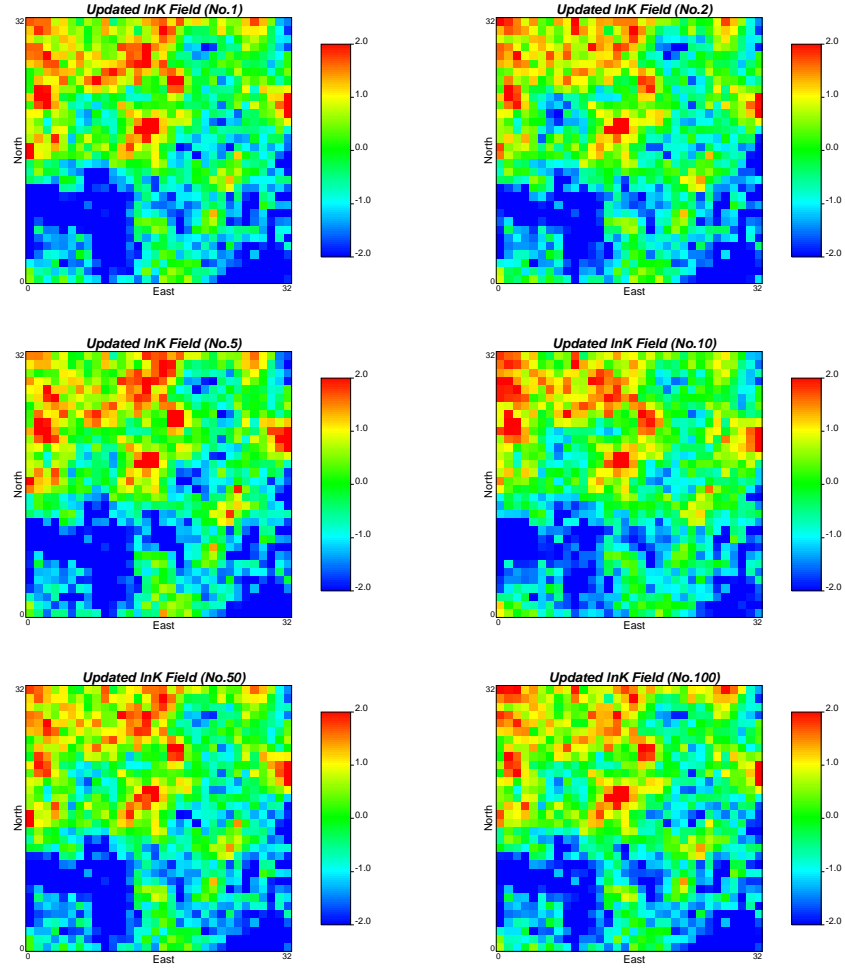
Figure 1.6: *Updated models after history matching up to 40.1 days (Only six out of one hundred realizations are listed.)*

the performance prediction and uncertainty assessment due to the ignorance of the model structure by means of several synthetic examples.

The objective of this thesis, therefore, 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. Instead of the two-stage method, a better choice is to generate realizations in one step which are both conditioned on the linear data and inverse-conditioned on the nonlinear data. With 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 approaches 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 with 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 is strictly 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. A synthetic example is presented to demonstrate the efficiency of the proposed BMcMC. The performance of the BMcMC is also widely evaluated on the basis of the toy example. The meaningfulness of emphasizing on spatial structure is also examined by evaluating the uncertainty reduction due to incorporating distinct data from different sources.

Aiming at enabling the BMcMC method to deal with the high-resolution cases, a multi-scale BMcMC (MsBMcMC) method is presented in Chapter 4 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 5 presents a complete assessment on uncertainty reduction owing to conditioning and inverse-conditioning on various types of data from different 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.

Finally, in Chapter 6, several important contributions and conclusions are summarized or repeated 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.,

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

$h = h(\mathbf{x}, t)$ 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, $[T]$; $\nabla \cdot = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})^T$ 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,

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

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, $[T^{-1}]$; $\phi(\mathbf{x})$ 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, $[T^{-1}]$.

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

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

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,

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

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

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

$$s_{s,i,j,k}\frac{\Delta h_t}{\Delta t_t}\Delta x_{i,j,k}\Delta y_{i,j,k}\Delta z_{i,j,k}, \tag{2.6}$$

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,

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

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

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

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,

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

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,

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

where,

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

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

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

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

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

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

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

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

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

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

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

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,

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

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.,

$$r_x = \frac{x_0 - x}{x_0} r_{x_1} + \frac{x}{x_0} r_{x_2},$$

$$r_y = \frac{y_0 - y}{y_0} r_{y_1} + \frac{y}{y_0} r_{y_2},$$

$$r_z = \frac{z_0 - z}{z_0} r_{z_1} + \frac{z}{z_0} r_{z_2},$$

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,

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

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

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

$$(2.15)$$

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, the transverse dispersivity in the horizontal direction, and the transverse dispersivity 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,

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

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

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

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

CHAPTER 2. FORWARD SIMULATORS   25

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.,

$$v_x = \frac{x_0 - x}{x_0} v_{x_1} + \frac{x}{x_0} v_{x_2},$$

$$v_y = \frac{y_0 - y}{y_0} v_{y_1} + \frac{y}{y_0} v_{y_2},$$

$$v_z = \frac{z_0 - z}{z_0} v_{z_1} + \frac{z}{z_0} v_{z_2},$$

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),

$$D_{xx} = \frac{1}{v} \left[\alpha_L v_x^2 + \alpha_T^H v_y^2 + \alpha_T^V v_z^2 \frac{v_x^2}{v_x^2 + v_y^2} \right],$$

$$D_{yy} = \frac{1}{v} \left[\alpha_L v_y^2 + \alpha_T^H v_x^2 + \alpha_T^V v_z^2 \frac{v_y^2}{v_x^2 + v_y^2} \right],$$

$$D_{zz} = \frac{1}{v} \left[\alpha_L v_z^2 + \alpha_T^V (v_x^2 + v_y^2) \right],$$

$$D_{xy} = D_{yx} = \frac{v_x v_y}{v} \left[\alpha_L - \alpha_T^H \frac{v^2}{(v_x^2 + v_y^2)} + \alpha_T^V \frac{v_z^2}{(v_x^2 + v_y^2)} \right],$$

$$D_{xz} = D_{zx} = \frac{v_x v_z}{v} \left[\alpha_L - \alpha_T^V \right],$$

$$D_{yz} = D_{zy} = \frac{v_y v_z}{v} \left[\alpha_L - \alpha_T^V \right].$$

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,

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

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

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

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

3

A Blocking McMC Method

Abstract

The collections of the observations of piezometric head and the measurements of local hydraulic conductivity (or transmissivity) provide invaluable information to identify the spatial pattern of aquifer parameters, either flow paths or flow barriers, and to reduce the uncertainty of aquifer models. To extract such spatial information from the measurements and accurately quantify the uncertainty, a Monte Carlo method typically calls for generating a large number of *i.i.d* realizations of aquifer parameters conditioning to hard data (conductivity) and inverse-conditioning to state data (piezometric head). However, inverse-conditional simulation of aquifer parameters is very computationally intensive, since it involves a nonlinear optimization problem for the generation of each inverse-conditional realization. In contrast to some classical nonlinear optimizers and searching algorithms, a blocking Markov chain Monte Carlo scheme (BMcMC) is presented in this study to generate multi-Gaussian conditional and inverse-conditional realizations by sampling directly from a posterior distribution that incorporates *a priori* information and *a posteriori* observations in a Bayesian framework. What makes this method quite efficient in exploring the model parameter space is that the proposal kernel is an appropriate approximation to the target posterior distribution and the generation of candidate realizations is very fast on the basis of the LU-decomposition of the covariance matrix. The realizations generated in this way are not only conditioned to hard data and inverse-conditioned to state data but also have expected spatial structures. A synthetic example which simulates the case of

natural gradient flow is presented to demonstrate the efficiency of the proposed method. The performance of the proposed BMcMC scheme is widely evaluated on the basis of the synthetic example. The model uncertainty and the uncertainty of predicted responses due to conditional and inverse-conditional mapping of aquifer models are then quantified by means of solving an assumed steady-state flow and ideal, conservative transport problem.

3.1   Introduction

The parameters of interest to both groundwater and petroleum engineering communities include hydraulic conductivity or permeability, storage coefficient or porosity, dispersivity, retardation factor, mass transfer rate, etc. Amongst them, the spatial variation of hydraulic conductivity or permeability predominantly controls flow and transport of solutes. It is also recognized that this key variable can vary significantly over a short distance showing strong heterogeneity and log-normal distribution in space. Furthermore, the experimental distribution of the parameters is highly non-symmetric which makes most scholars use log-conductivity as the unknown in order to treat the problem in a Gaussian framework. The Gaussian distribution is employed because the stochastic process can be fully described by the mean and the variance though the method developed in this study is absolutely not limited only to the Gaussian process. Therefore, how to identify the logarithm of this physical parameter, $x = \ln K$, has become one of the main focuses to both hydrologists and petroleum engineers.

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

However, this set of equally probable realizations generated should be constrained to all information available in order to further reduce the uncertainties of underground reservoirs and aquifers. This information may include some prior concepts, e.g., those from the experts' subjective imagination, and some posterior observations, e.g., those from the measurements in the field. It is possible that those observations are linearly related to the parameters of interest, e.g., those space-dependent static hard and soft data, or even nonlinearly related, e.g., those time-dependent dynamic observations known as state data. It is also possible that the measurements are available at different scales with different precisions. In this case, the simulation is cast as a conditioning problem, that is, to make inference to the physical model from the given observations. Mathematically, it can be expressed as,

$$\boldsymbol{y} = g(\boldsymbol{x}) + \boldsymbol{\epsilon}, \tag{3.1}$$

where $\boldsymbol{x} \subset R^n$ is the parameterized physical model of interest; $\boldsymbol{y} = \boldsymbol{y}_{obs} \subset R^k$ is the error-prone observation; $g(\boldsymbol{x}) \subset R^k$ is the transfer operator that maps the parameter of interest $\boldsymbol{x}$ to the observation attribute $\boldsymbol{y}$; $\boldsymbol{\epsilon} \subset R^k$ is the system error generally correlated $\boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma}$ is the covariance matrix. The objective of the conditional simulation is to predict the physical model $\boldsymbol{x}$ given the observation $\boldsymbol{y}$ by assuming some prior error $\boldsymbol{\epsilon}$. In this sense, the conditional simulation can be coined as an inverse problem, whereas $g(\boldsymbol{x})$ is the forward operator governing the transfer between the observation attribute and the parameter of interest. Here, the procedure of conditioning to the nonlinear data is termed as "inverse-conditioning problem" just because of the existence of such inverse problem.

The construction of the physical models honoring the prior information, the linear data and the nonlinear data is only one aspect of the geostatistical conditional and inverse-conditional simulation. Of equal importance is to perform uncertainty analysis, e.g., to quantify the reliability of those models, to identify key uncertainty resources, and to assess the resolutions and confidence regions of the conditional realizations so as to measure how much the property parameters can depart from the conditional realizations. Note that the purpose of the quantitative uncertainty analysis is not to reduce uncertainty which can only be achieved by collecting additional effective information. Additional effective information essentially reduces the uncertainty of model. However, not all conditioning algorithms can detect such uncertainty reduction introduced by additional effective information. Therefore, quantification of uncertainty can adversely prove the efficiency of a conditioning algorithm. Assuming that the uncertainty from numerical simulations is neglected, the uncertainty can originate from the following three sources: (1) input uncertainty, e.g., from the data acquisition, (2) conceptual 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), and (3) prediction uncertainty resulting from model input.

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

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

3.2   Blocking McMC Method

Consider a stochastic simulation at n grid nodes conditional to m hard data and k nonlinear state data. 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 $\boldsymbol{\mu}$ is the prior mean of the Gaussian process and $\boldsymbol{C_x}$ 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 $\boldsymbol{C_y}$ 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 inverse-conditional simulation is to infer $\boldsymbol{x}$ from $\boldsymbol{y}$ by assuming some spatial statistical structures, where $\boldsymbol{y}$ is nonlinearly related to $\boldsymbol{x}$ through a forward operator $g(\boldsymbol{x})$ and $\boldsymbol{x}$ may be partly observed. The most challenging part of the conditional simulation is basically an inverse problem since an inverse operator $g^{-1}(\boldsymbol{x})$ is applied to the conditioning procedure.

3.2.1   Bayesian formulation

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

$$\pi(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} \|\boldsymbol{C_x}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{C_x}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}, \qquad (3.2)$$

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., $m < n$. The prior pdf represents some prior knowledge about the parameterization of a physical model $\boldsymbol{x}$ through the configuration of $\boldsymbol{\mu}$ and $\boldsymbol{C_x}$. It should allow for the greatest uncertainty while obeying the constraints imposed by the prior information.

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

$$L(\boldsymbol{x}|\boldsymbol{y}) = (2\pi)^{-\frac{k}{2}} \|\boldsymbol{C_y}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(g(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{C_y}^{-1}(g(\boldsymbol{x}) - \boldsymbol{y})\right\}, \qquad (3.3)$$

where $\boldsymbol{y} = \boldsymbol{y}_{obs} \subset R^k$ represents the values of the observations; $g(\boldsymbol{x})$ is the transfer function of $\boldsymbol{x}$, generally highly nonlinear, by which $\boldsymbol{x}$ relates to $\boldsymbol{y}$ in the same attribute representation; and $\boldsymbol{C_y} \subset R^{k \times k}$ is the covariance matrix of the vector $\boldsymbol{y}$. Note that if the observation errors of $\boldsymbol{y}$ are fully independent from each other, then $\boldsymbol{C_y}$ is a diagonal matrix. The 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., $\int L(\boldsymbol{x}|\boldsymbol{y})\pi(\boldsymbol{x})d\boldsymbol{x}$. Since the marginal likelihood is not a function of $\boldsymbol{x}$, it is typically ignored in the parameter estimation problem. Using the Bayes' theorem and dropping the normalizing constant, the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$, $\pi(\boldsymbol{x}|\boldsymbol{y})$, may be written as $\pi(\boldsymbol{x}|\boldsymbol{y}) \propto L(\boldsymbol{x}|\boldsymbol{y}) \times \pi(\boldsymbol{x})$, i.e.,

$$\pi(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{C}_{\boldsymbol{x}}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}(g(\boldsymbol{x})-\boldsymbol{y})^T \boldsymbol{C}_{\boldsymbol{y}}^{-1}(g(\boldsymbol{x})-\boldsymbol{y})\right\}.$$
(3.4)

The 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 transferred to draw independent, identically distributed (*i.i.d*) realizations from this posterior distribution.

3.2.2   Blocking McMC scheme

Due to the highly nonlinearity of the log-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), however, is especially suitable for exploring the parameter space of such type of complicated posterior distribution. A typical McMC algorithm employing the Metropolis-Hastings rule to explore the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ 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. If the

proposal kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$ is chosen to equal the prior model $\pi(\boldsymbol{x})$, then the acceptance probability is,

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

If the proposal kernel does not depend on the current values of $\boldsymbol{x}$, the chain constructed is called an independent Markov chain (*Liu*, 1996). Obviously, the acceptance probability may be quite low for a high-dimensional problem and the chain may mix rather slowly, since,

$$\alpha = \frac{\pi(\boldsymbol{x}^*)\pi(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})\pi(\boldsymbol{x}^*|\boldsymbol{x})}\frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})} = \frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})}, \tag{3.6}$$

where the candidate model $\boldsymbol{x}^*$ is accepted simply according to the ratio of the likelihood, which is rather random.

On the other hand, if the proposal kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$ is chosen to be the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$, then the acceptance probability of $\boldsymbol{x}^*$ equals one,

$$\alpha = \frac{\pi(\boldsymbol{x}^*|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{y})}\frac{q(\boldsymbol{x}|\boldsymbol{x}^*)}{q(\boldsymbol{x}^*|\boldsymbol{x})} = 1, \tag{3.7}$$

if the Metropolis-Hastings rule is used to calculate the acceptance rate. In such case, the candidate model $\boldsymbol{x}^*$ is fully accepted, which is also called the Gibbs sampler (*Geman and Geman*, 1984). An assumption for this case is that the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ can be sampled directly from which most often is not the real case for the multi-Gaussian field. Actually, this is just the McMC explanation for the classical ML method, argmax$\{\pi(\boldsymbol{y}|\boldsymbol{x})\}$, and the MAP method, argmax$\{\pi(\boldsymbol{x}|\boldsymbol{y})\}$, which is regularized by the prior $\pi(\boldsymbol{x})$.

In the ML method, a seed field $\boldsymbol{x}$ is generated according to the prior distribution $\pi(\boldsymbol{x})$, then $\boldsymbol{x}$ is calibrated to maximize the likelihood $\pi(\boldsymbol{y}|\boldsymbol{x})$ to produce a new field $\boldsymbol{x}^*$, which can be considered as a sample from the $\pi(\boldsymbol{x}|\boldsymbol{y})$ if the prior structure remains unchanged during the calibration. In the MAP method, the prior structure is explicitly forced to follow since the target model has been already regularized by the prior model in the Bayes' rule. In both methods, therefore, the generated realizations can be viewed as a direct sample from the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ using some optimization techniques. More realizations can be obtained by repeating the same procedure. If those realizations are used to construct the Markov chain, then all of them are accepted to be the output of the McMC samples. In this sense, the McMC sampler is a superset of the ML method and the MAP method. However, the challenge of the ML method and the MAP method lies in the extreme

expensiveness of the optimization routines and the forward operator, generally high-dimensional and nonlinear. That makes it impossible to use the output of these two methods as the proposal kernel of the McMC computation.

Therefore, the trade-off between the independent Markov chain and the MAP method is to use an approximate posterior distribution as the proposal kernel to construct a Markov chain. The acceptance probability fully depends on how closely the proposal kernel approximates the posterior distribution. This idea efficiently improves the mixing of the chain, but also decreases the expensiveness of the high-dimensional, nonlinear optimization routines. This is just the origin of the algorithms developed in this study. The proposed method adopts the blocking update scheme as that in the simple implementation of the independent Markov chain, but the proposals are more efficient since they approximate the posterior distribution more closely. Indeed, it is well known that the blocking scheme helps improve the mixing of Markov chain (*Liu*, 1996; *Roberts and Sahu*, 1997).

The methods developed are more efficient than some algorithms resorting to the simple Gibbs sampler for the high-dimensional spatiotemporal problem, e.g., those inverse-conditioning problems haunting both the hydrogeology and the petroleum engineering community. Generally a Gibbs sampler updates the attribute field one cell by one cell using the full conditional distribution. For a high-dimensional nonlinear spatiotemporal problem, the likelihood perturbation introduced by the difference of one single cell's attribute value is ignorable in the computer implementation. Most often it may be less than the rounding error of the computation. It makes the simple Gibbs sampler quite ineffective. A blocking scheme perturbs the attribute field with more elements and thus results in much more change of the likelihood.

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 repeated generation of candidates because the decomposition operator is applied only once. Appendix A gives an outline on the implementation of the LU-decomposition-based sampler.

3.2.3   Performance assessment of the BMcMC

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, $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_{n_r-1})^T$.

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,

$$\sigma_t = \sum_{i=0}^{t} (\eta(\boldsymbol{x}_i) - \bar{\eta}), \tag{3.8}$$

for $t = 0, 1, \ldots, n_r - 1$, 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 $i = 1, 2, \ldots, n_r - 1$;

(4) Calculate the hairiness indices,

$$\Sigma_t = \frac{1}{t-1} \sum_{i=1}^{t-1} \delta_i, \tag{3.9}$$

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

The key outputs are two sequential plots: $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{n_r})^T$ and $\Sigma = (\Sigma_0, \Sigma_1, ..., \Sigma_{n_r})^T$. The first one, 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 second one, Σ_t, presents a quantitative measure of smoothness, hairiness index, to evaluate the convergence performance of a chain. An ideal convergence sequence will be centered at around 0.5.

3.3   A Synthetic Example

3.3.1   Reference models

A synthetic two-dimensional aquifer is presented to demonstrate the efficiency and effectiveness of the BMcMC method. Figure 3.1(A) shows the spatial distribution of the reference 16×16 $\ln K$ field generated by LUSIM (*Deutsch and Journel*, 1998), having a multi-Gaussian characteristic. The mean and variance of the $\ln K$ are set to zero and one, respectively. The spatial correlation structure is specified as an exponential function with a correlation length equal to 16, which means that the field is highly correlated. An obvious pattern of the $\ln K$ distribution is that the lower values, mainly locating at the left-lower corner, spread along the north-west direction while the higher values locate mainly at the right-lower corner and some of them at the left-right corner.
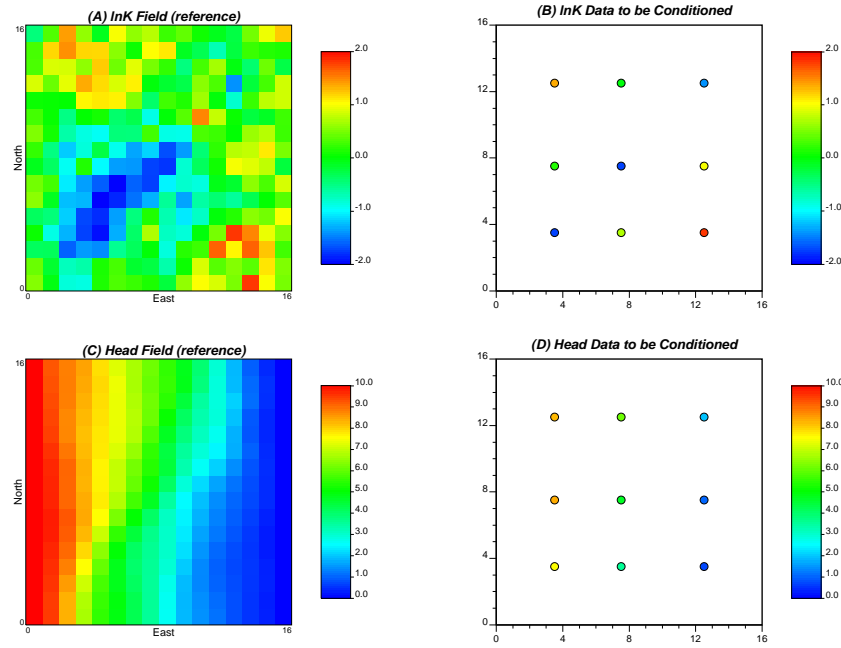
Figure 3.1: *Conditioning data set and reference* $\ln K$ *and head fields* ($\mu = 0, \lambda_{\boldsymbol{x}} = 16, \sigma_x^2 = 1.0$)

An assumed steady flow experiment is performed on this aquifer. The left and right boundary conditions are set as a prescribed head with fixed values equal to ten and zero, respectively. Both the upper and the lower bounds are set as non-flow. The resulting head field is shown in the Figure 3.1(C). The key pattern of the head distribution is that the flow breakthrough of the lower

part is much faster than that of the upper part, which is a natural result of the $\ln K$ distribution. Nine head observations as the conditioning data set of state variables are uniformly sampled from the head field as shown in Figure 3.1(D). Nine $\ln K$ observations as hard data to be conditioned are also sampled from the $\ln K$ field at the same locations as head data, which are shown in Figure 3.1(B).
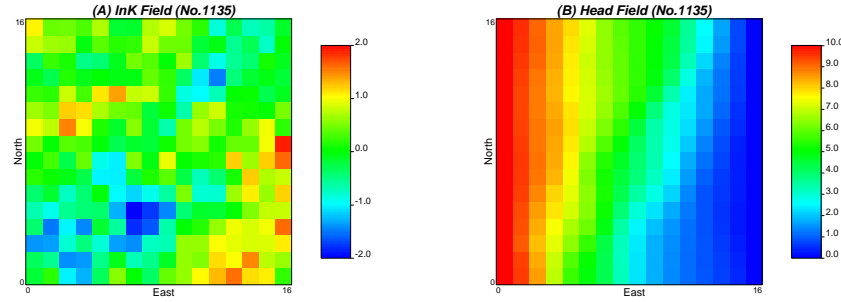
Figure 3.2: *A typical BMcMC realization and predicted head field*

3.3.2   An inverse-conditioning experiment

The objective of the inverse-conditional simulation is to generate *i.i.d* realizations inverse-conditional to the nine head samples using the proposed BMcMC method. Assume that the flow boundary conditions and the mean and covariance of $\ln K$ are correctly observed, i.e., $\mu = 0$, $\sigma_x^2 = 1.0$, and $\lambda_{\boldsymbol{x}} = 16$, although they may subject to uncertainty as well. The relative measurement error of state variables is set as 0.1, i.e., $\sigma_y^2 = 0.1$. Figure 3.2 shows a typical BMcMC realization of $\ln K$ and the corresponding head field. Both fields reveal some of main patterns of reference models.

Figure 3.3 shows the most probable estimate which is obtained by averaging over 3000 BMcMC realizations. The estimated $\ln K$ distribution shown in (A) well matches the distribution of the reference $\ln K$ field: the lower values, mainly locating at the left-lower corner, spread along the north-west direction while the higher values locate mainly at the right-lower corner, although the reproduction of the left-upper corner is not that so clear. The precision of such estimate is plotted in (B). One can find that the most uncertain parts are close to the non-flow boundaries. The mean head field shown in (C) reasonably reflects the main characteristics of the reference head field (Figure 3.1(C)). The distribution of head uncertainty as plotted in (D) also demonstrates that the regions near the non-flow boundaries are subject to the most uncertainties.

The realizations generated strictly follow the prior configuration for model parameters. For example, the experimental mean and variance of the $\ln K$
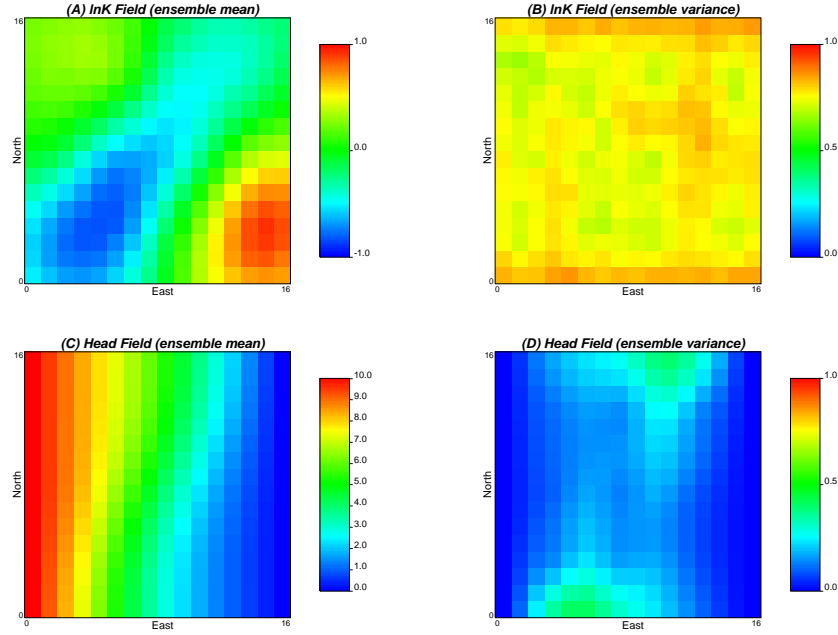
Figure 3.3: *A BMcMC estimate by inverse-conditioning to head observations* ($\lambda_{\boldsymbol{x}} = 16, \sigma_x^2 = 1.0, \sigma_y^2 = 0.10$)
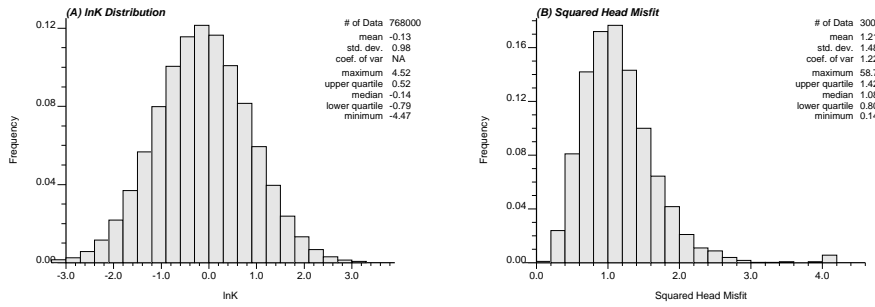
Figure 3.4: $\ln K$ *distribution and squared head misfit* (*3000 realizations*)

are -0.13 and 0.98, respectively (Figure 3.4(A)) which approximate to the corresponding prior specifications, i.e., $\mu = 0$ and $\sigma_x^2 = 1$, and those of the reference field, i.e., $\mu = -0.05$ and $\sigma_x^2 = 0.84$. Figure 3.4 shows that the $\ln K$ distribution obeys on the Gaussian assumption and the squared head mismatch follows the log-Gaussian assumption as implied in Equation 3.3.

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

3.3.3   Factors that affect the performance of the BMcMC

Three important factors are considered in this section to examine their influences on the performance of the BMcMC method: the correlation length ($\lambda_{\boldsymbol{x}}$), conditioning on the hard data ($\boldsymbol{x}|\boldsymbol{x}_{obs}$), and the relative measurement error of state variables (σ_y^2). Several performance measures considered in this study include the reproduction of reference models, the acceptance rate, and the mixing of Markov chain.
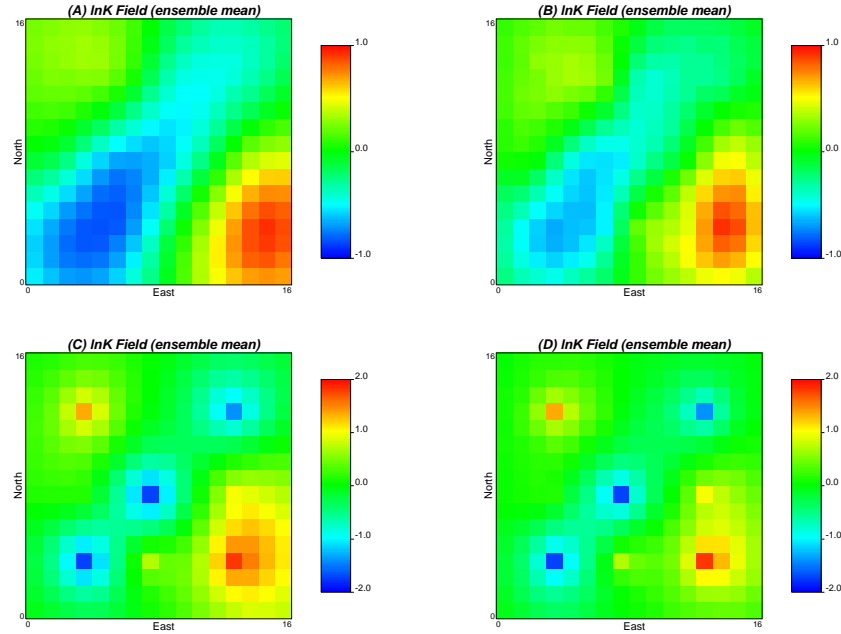
Figure 3.5: *Factors affecting the reproduction of reference models ($\sigma_x^2 = 1.0$, $\sigma_y^2 = 0.10$, (A)$\boldsymbol{x}|\boldsymbol{y}, \lambda_{\boldsymbol{x}} = 16$, (B)$\boldsymbol{x}|\boldsymbol{y}, \lambda_{\boldsymbol{x}} = 4$, (C)$\boldsymbol{x}|\boldsymbol{x}_{obs}, \boldsymbol{y}, \lambda_{\boldsymbol{x}} = 16$, (D)$\boldsymbol{x}|\boldsymbol{x}_{obs}, \boldsymbol{y}, \lambda_{\boldsymbol{x}} = 4$)*

Four BMcMC simulations with four types of configurations are run on the basis of the same conditioning data set as above. The most probable estimates are plotted in Figure 3.5. Obviously, the magnitude of ln$K$ is underestimated by a smaller prior correlation length for both the unconditional and the conditional cases as plotted in (B) and (D), respectively. The identified mean ln$K$ field for the conditional case does contain the contributions from both the kriging and inverse-conditional estimate, though not a simple linear superposition of them.

Figure 3.6 compares the effects of various BMcMC configurations on the performance of Markov chain. The effect of the prior correlation length is rather minor: as for the unconditional case in this study, the correct correlation length ($\lambda_{\boldsymbol{x}} = 16$) yields a slightly better squared head mismatch (as shown in (A) and (B)) but a slightly worse mixing of chain (a larger excursion size (E) but better hairiness indices (F)); as for the conditional case, the shorter correlation length produces better results (from (C) through (F)). Conditioning to hard data has considerate influence on the performance of Markov chain. The error of squared head misfit obviously decreases, for example, by comparing (C) to (A), and the chain mixes more rapidly, i.e., a smaller excursion size and a larger hairiness indices.

In addition, the acceptance rate is quite sensitive to the prior correlation length and whether or not conditioning on hard data. Conditioning to hard data significantly enhances the acceptance rate as shown in Table 3.1. The effect of correlation length is quite complicated: for unconditional case, a correct correlation length helps increase the acceptance rate; but for the conditional case, the correct correlation length decreases the acceptance rate.

The quality of measurements for state variables essentially affects the BMcMC simulation both in accuracy and in performance aspects. In this case study, the measurement errors are intentionally assumed to be small so that such influence can be examined in more details. First, high-quality measurements in essence decrease the acceptance rate (Table 3.1). This is just because fewer candidates may meet a smaller squared head mismatch. Second, the chain may mix extremely slowly as plotted in Figure 3.7. The excursion size is quite large (C) and hairiness indices are rather small (D). Third, many candidates with a very small likelihood may be also sampled as $i.i.d$ output such that the posterior mean estimate may be not accurate as expected as shown in (A). It should be emphasized that this is not the weak point of BMcMC. On the contrary, it just demonstrates the width of potential exploration scope that the Markov chain visits, which is also a measure of efficiency of an McMC scheme.

Towards this problem, one way out of it is to prune the output of the BMcMC according to their posterior probabilities, that is, to discard those realizations with extremely small posterior probabilities. A possible imple-
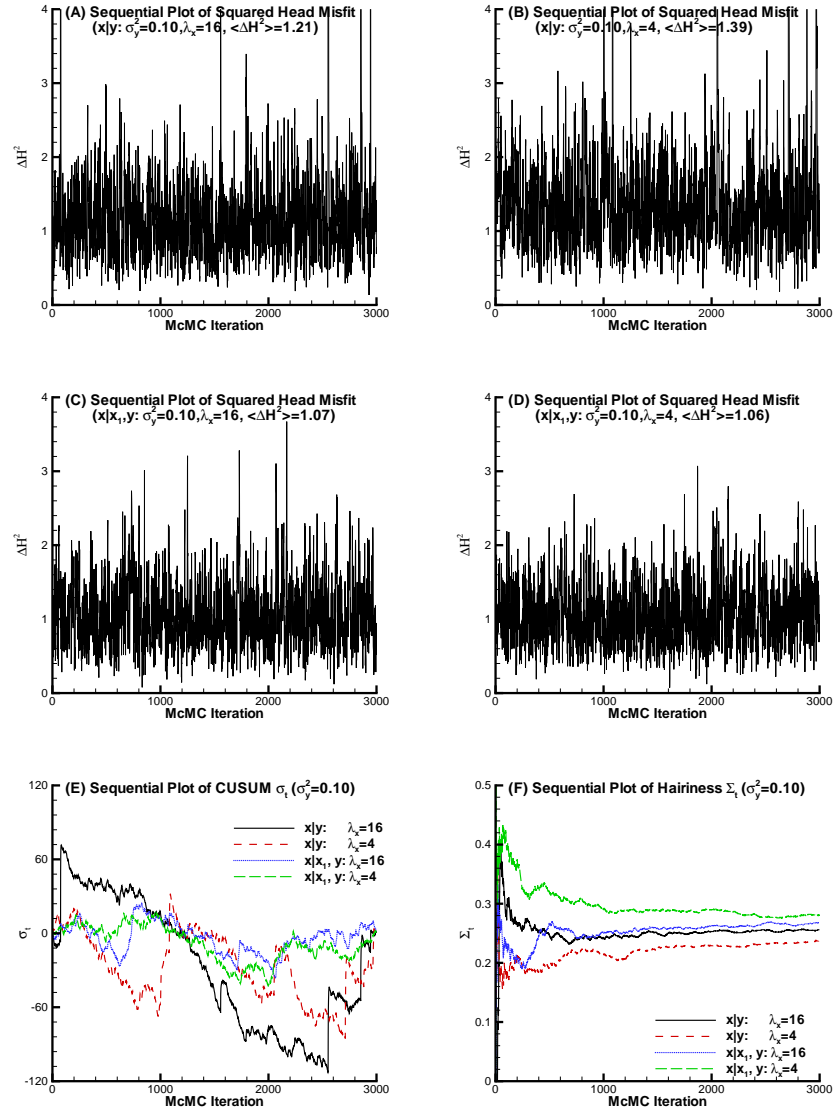
Figure 3.6: *Factors affecting the performance of Markov chain*

Table 3.1: *BMcMC configurations and performance indices*

Scenario	BMcMC configuration				Before pruning		After pruning	
	$\boldsymbol{x}$	$\lambda_{\boldsymbol{x}}$	σ_y^2	$\alpha(\%)$	$\langle\Delta y^2\rangle$	Σ_t	$\langle\Delta y^2\rangle$	Σ_t
1	$\boldsymbol{x}\|\boldsymbol{y}$	16	0.10	0.1273	1.21	0.256	0.59	0.417
2	$\boldsymbol{x}\|\boldsymbol{y}$	16	0.05	0.0175	1.24	0.047	0.37	0.405
3	$\boldsymbol{x}\|\boldsymbol{y}$	16	0.02	0.0030	2.75	0.101	0.67	0.144
4	$\boldsymbol{x}\|\boldsymbol{y}$	4	0.10	0.0903	1.39	0.238	0.81	0.391
5	$\boldsymbol{x}\|\boldsymbol{y}$	4	0.05	0.0106	1.34	0.072	0.53	0.367
6	$\boldsymbol{x}\|\boldsymbol{y}$	4	0.02	0.0028	2.38	0.108	1.05	0.135
7	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	16	0.10	0.8381	1.07	0.268	0.32	0.481
8	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	16	0.05	0.1128	0.63	0.247	0.33	0.417
9	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	16	0.02	0.0082	0.64	0.049	0.19	0.391
10	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	4	0.10	1.4622	1.06	0.281	0.30	0.492
11	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	4	0.05	0.1985	0.65	0.264	0.30	0.448
12	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	4	0.02	0.0109	0.52	0.059	0.18	0.407

mentation to this alternative is to set a threshold for the posterior probability. Only those realizations with probabilities above this threshold are retained. A difficulty that a user often faces, however, is that it is quite hard to determine the threshold of the posterior probability. Another way is to prune outputs from a very long McMC and retain those realizations as required in quantity by ranking their posterior probabilities. Figure 3.7 compares the effect of such pruning procedure on the performance of McMC. After pruning, the squared head mismatch decreases obviously (from 1.24 to 0.63 as in (A) and (B), respectively) and the chain mixes much faster (a smaller excursion size (C) and larger hairiness indices (D)). Table 3.1 also summarizes the changes of squared head mismatch and hairiness indices owing to pruning.

3.4   Uncertainty Assessment

On the basis of the synthetic example as presented above, the uncertainty reduction owing to conditioning and inverse-conditioning is assessed quantitatively in terms of several measures. Two types of uncertainties are considered in this section: model uncertainty and response uncertainty. The model uncertainty is important at the spatiotemporal scale: not only because the models generated form the basis for future performance prediction at the existing wells but also because they serve as risk analysis of candidate wells. The response

CHAPTER 3.   A BLOCKING MCMC METHOD   43
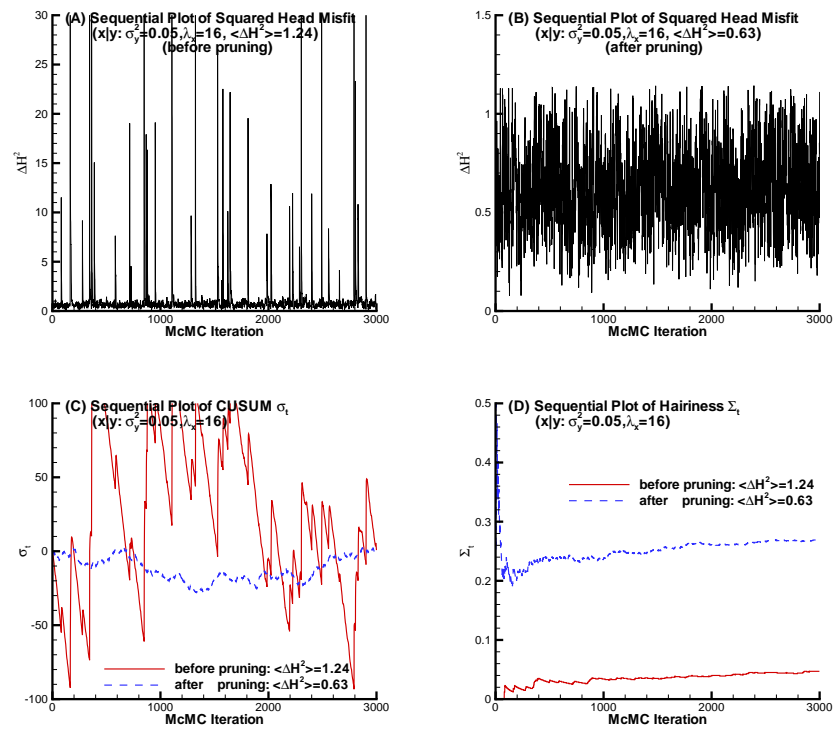
Figure 3.7: *Effect of pruning on the performance of BMcMC*

uncertainty directly measures the prediction ability of models at the time scale. Several uncertainty sources are also ranked according to their influences on the uncertainty reduction.

3.4.1   Model uncertainty

Although the reference model is well defined and observable in this study, we generally do not know what it is in advance 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 ($I(\boldsymbol{x})_3$) and the standard deviation of the ensemble average error ($I(\boldsymbol{x})_4$), which are defined as the L_1-norm and L_2-norm between the simulated models and the mean models, i.e.,

$$I(\boldsymbol{x})_1 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r} - \bar{x}_i|, \qquad (3.10a)$$

$$I(\boldsymbol{x})_2^2 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r} - \bar{x}_i)^2, \qquad (3.10b)$$

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,

$$I(\boldsymbol{x})_3 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r}^{sim} - x_i^{ref}|, \qquad (3.11a)$$

$$I(\boldsymbol{x})_4^2 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r}^{sim} - x_i^{ref})^2, \qquad (3.11b)$$

respectively. Note that $\boldsymbol{x}_{ref}$ is the vector of reference attribute values. Obviously, the smaller $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2$ are, the closer to the real model the generated realizations are.

3.4.2 Uncertainty of model responses

A method to examining the effect of conditioning to head data on the uncertainty reduction of the spatial distribution of hydraulic conductivity is to examine the decrease of the L_1-norm and L_2-norm of the predicted model responses (*Hoeksema and Kitanidis*, 1984; *Kitanidis*, 1986). The four metrics for the model responses, $I(\boldsymbol{y})_1$, $I(\boldsymbol{y})_2$, $I(\boldsymbol{y})_3$, and $I(\boldsymbol{y})_4$, are defined as follows,

$$I(\boldsymbol{y})_1 = \|\boldsymbol{y}_{sim} - \bar{\boldsymbol{y}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r} - \bar{y}_i|, \tag{3.12a}$$

$$I(\boldsymbol{y})_2^2 = \|\boldsymbol{y}_{sim} - \bar{\boldsymbol{y}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r} - \bar{y}_i)^2, \tag{3.12b}$$

$$I(\boldsymbol{y})_3 = \|\boldsymbol{y}_{sim} - \boldsymbol{y}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r}^{sim} - y_i^{ref}|, \tag{3.12c}$$

$$I(\boldsymbol{y})_4^2 = \|\boldsymbol{y}_{sim} - \boldsymbol{y}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r}^{sim} - y_i^{ref})^2, \tag{3.12d}$$

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.

3.4.3 How many realizations are needed?

The number of realizations required to approximate the ensemble statistics to a sufficient degree of accuracy may be determined either by using the *Kolmogorov-Smirnov and Renyi* statistics (*Spear*, 1970) or, more arbitrarily, by experience (*Beck*, 1987). Most often, the latter decides the number of realizations by monitoring the degree of convergence of the output statistics. One of key factors to controlling the number of realizations is the degree of heterogeneity of aquifer, e.g., the variance and the correlation length. Generally, a larger variance and a higher correlation ask for many more realizations.

Actually, a secondary usage of the uncertainty measures as outlined as above is that they serve to determine the number of realizations that are required to yield a stable and reliable result in the uncertainty evaluation. In
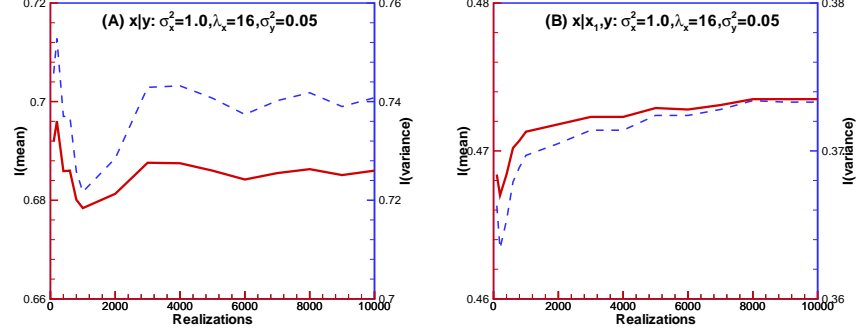
Figure 3.8: *Relationship between the uncertainty measures and the number of realizations*

other words, how long the Markov chain should be run in order to yield a reliable result? This question is important since the Markov chain can not be run forever. The chain should be terminated at a point that more realizations do not offer better results even though, in theory, more realizations give better results. This question may be answered by monitoring the uncertainty measures computed from the output of BMcMC simulation. Figure 3.8 plots the experimental relationships between the uncertainty measures ($I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2$) and the number of realizations. For both unconditional (A) and conditional (B) cases, the error measures do not fluctuate too much if the number of realizations is over 100. But if a high reliability is required, 3000 and 1000 realizations are enough to guarantee a stable estimate for unconditional and conditional cases, respectively. For a smaller variance and shorter correlation length, fewer realizations are required.

3.4.4   Results

Totally fourteen scenarios of numerical experiments are carried out to systematically evaluate the uncertainty reduction owing to conditioning on $\ln K$ and inverse-conditioning on head data. Twelve of them are listed in Table 3.1 and the other consists of two base simulations conditional and unconditional to $\ln K$.

Table 3.2 summarizes the model uncertainty reduction in terms of four model metrics, $I(\boldsymbol{x})_1$, $I(\boldsymbol{x})_2$, $I(\boldsymbol{x})_3$, and $I(\boldsymbol{x})_4$. First, an evident observation is that the real uncertainties indicated by $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2$ are uniformly larger than the estimated uncertainties represented by $I(\boldsymbol{x})_3$ and $I(\boldsymbol{x})_4$. It means that the model uncertainty is always underestimated by the BMcMC

simulation. But the latter two metrics can be used as a substitute of the former two for uncertainty assessment since they completely reflect the common rules that the former two reveal. Second, as the measurement errors of state variables decrease, the model accuracy does not increase too much. One of main reasons is that a high-resolution indirect measurement is assumed, i.e., σ_y^2 is sufficiently small, such that the model estimation suffers from the limitation of exploration ability that the BMcMC scheme has. The post-process resorting to pruning the output of McMC simulations may help improve the accuracy for model estimation. Another reason is that, after a certain point of accuracy, the overemphasis of measurement errors does not help improve the model estimation.

Table 3.2: *Mean absolute error and mean variance of* $\ln K$

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

Third, the bottom line in Table 3.2 ranks the importance of BMcMC configurations on the model uncertainty reduction. Compared to the unconditional case, conditional and/or inverse-conditional simulations do reduce the model uncertainty. The case that has a correct specification for model structure, e.g., $\lambda_{\boldsymbol{x}} = 16$, conditioning to $\ln K$ and inverse-conditioning to head data reduces the model uncertainty to the largest extent. Conditioning to $\ln K$ more efficiently reduces the uncertainty than inverse-conditioning to head data which means that the direct measurement of $\ln K$ is more helpful for the model parameter estimation than the indirect measurement of dependent state variables. However, correct configuration for spatial model, e.g., the correlation

length $\lambda_{\boldsymbol{x}}$, may play the most important role in the model uncertainty re-
duction. Due to the wrong specification for correlation length, i.e., $\lambda_{\boldsymbol{x}} = 4$,
the conditional and inverse-conditional result only ranks the fourth, while the
correct one ranks the first. In summary, the correct configuration for model
structure plays a crucial role in model estimate. The direct measurement of
parameters plays a more important role than the indirect measurement of
dependent state data for the model parameter estimation.

Table 3.3 summarizes the uncertainty reduction of model response in terms
of four response metrics, $I(\boldsymbol{y})_1$, $I(\boldsymbol{y})_2$, $I(\boldsymbol{y})_3$, and $I(\boldsymbol{y})_4$. First, just as the ob-
servation as for the model uncertainty, $I(\boldsymbol{y})_1$ and $I(\boldsymbol{y})_2$ are uniformly larger
than $I(\boldsymbol{y})_3$ and $I(\boldsymbol{y})_4$, which also means that the response uncertainty is always
underestimated by the BMcMC simulation. Second, as the measurement er-
rors of state variables decrease, the model accuracy does not increase too much.
The pruning procedure helps improve the accuracy of estimation. Third, the
bottom line in Table 3.3 ranks the importance of BMcMC configurations on
uncertainty reduction of model response. Again, conditional and/or inverse-
conditional simulations reduce the response uncertainty. The case that has
a correct specification for model structure, conditioning to $\ln K$ and inverse-
conditioning to head data reduces the response uncertainty to the largest ex-
tent. However, different from the model uncertainty, inverse-conditioning to
head data more efficiently reduces the uncertainty than conditioning to $\ln K$
which means that the indirect measurement of dependent state variables is
more helpful for the response uncertainty reduction than the direct measure-
ment of $\ln K$. Moreover, correct configuration for spatial model plays a less
important role in reducing the response uncertainty. In summary, the re-
sponse uncertainty is more sensitive to the measurements of model response
than those of model parameters.

It is worth mentioning that, even though the configurations for model
parameter are not correct, the realizations generated well reproduce the cru-
cial patterns of reference models (e.g., Figure 3.5 (B) and (D)) and both the
model uncertainty and the response uncertainty are reduced owing to inverse-
conditioning and/or conditioning (Table 3.2 and Table 3.3), which proves the
robustness of the proposed BMcMC method.

3.5   Summary

A blocking McMC scheme is presented to generate *i.i.d* realizations which
are conditional to hard data and inverse-conditional to state data. The pro-
posal kernel for the construction of Markov chain is an appropriate (block-
ing) approximation to the target posterior distribution in the Bayesian frame-
work. The generation of candidate realizations is very fast on the basis of

Table 3.3: *Mean absolute error and mean variance of predicted head*

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

the LU-decomposition of the covariance matrix. A numerical experiment on a synthetic aquifer is carried out to demonstrate the efficiency of the proposed method in performing inverse-conditional simulation. The performance of this method is also widely evaluated based on the synthetic aquifer. Numerical experiments show that the relative measurement errors of state data (σ_y^2), the degree of heterogeneity of aquifer, e.g., correlation length ($\lambda_{\boldsymbol{x}}$), and whether or not conditioning on the hard data ($\boldsymbol{x}|\boldsymbol{x}_{obs}$) may play important roles in the results of inverse-conditional simulations and also affect the performance of BMcMC itself.

The model uncertainty and the response uncertainty are also assessed. Both types of uncertainties are reduced due to conditioning on hard data and/or inverse-conditioning on state data. However, their influences on the uncertainty reduction may have different weights. As for the effect of the piezometric head (state data) and the hydraulic conductivity (hard data) upon the model uncertainty of $\ln K$, the measurement of $\ln K$ plays a major role in reducing such uncertainty compared to the head data. Conditioning to head data, however, does improve the model estimation of $\ln K$ compared to conditioning to $\ln K$ solely. This conclusion is completely consistent with the result of *Dagan* (1985). The measurement of head is informative on the large-scale configuration of $\ln K$. As for the effect of the piezometric head and the hydraulic conductivity upon the prediction uncertainty of head distribution, the

measurement of head plays a major role in reducing such uncertainty compared to that of $\ln K$. Although the prediction of head is quite insensitive to local $\ln K$ values, the joint conditioning to $\ln K$ does improve the prediction of head compared to inverse-conditioning to head data solely. The local measurement of $\ln K$ does not carry too much information about the spatial trend of $\ln K$.

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

4

A Multi-scale Blocking McMC Method: 1. Methodology

Abstract

An adequate representation of the detailed spatial variation of subsurface parameters calls for high-resolution modeling of aquifers. A Monte Carlo method for such refined characterization of large-scale models typically invokes a stochastic simulation 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 has been proved to be an effective and efficient 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 previous BMcMC method suffers from the limited capability of the LU-decomposition of the covariance matrix in dealing with the high-resolution cases. In this study, a new multi-scale blocking McMC (MsBMcMC) scheme is presented to generate high-resolution, multi-Gaussian, conditional and inverse-conditional realizations. What make this method quite efficient in exploring the parameter space of high-dimensional models are that the proposal (blocking) kernel

is an appropriate approximation to the target posterior distribution, that the fast generation of candidate realizations is based on the spectral decomposition of the covariance matrix with aid of 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 prior spatial statistics and structure, (2) it helps improve the mixing of 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 efficiently accelerates the construction of Markov chain because those low probable candidates are rejected without consuming too many CPU resources in calculating their likelihood. 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.

4.1   Introduction

Collections of direct and indirect measurements provide indispensable but limited knowledge on the subsurface reservoirs or aquifers which mostly calls for a stochastic method to characterize their uncertainties. Model uncertainties can be basically classified into two types: one is the structure uncertainty which may arise from uncertain hypotheses or unmodeled processes, the other is parameter uncertainty which may come 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 the output uncertainty is evaluated by propagating the input uncertainty. Traditionally, however, the structure uncertainty and the parameter uncertainty are seldom isolated from each other which results in that the prediction uncertainty is generally a combined effect of both. The worth of measurements on one type of parameter property with respect to predicted performance is therefore hardly evaluated in a deterministic manner since the 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. Therefore a Monte Carlo method for uncertainty assessment entails a stochastic method capable of generating independent, identically distributed (*i.i.d*) realizations that share a specified *a priori* model structure, i.e., identical structure parameters.

A special difficulty occurs in assessing the worth of dependent state data since a complex inverse problem is involved. Classical optimization-based inverse methods for stochastic simulation generally work as follows (*Carrera and Neuman*, 1986; *RamaRao et al.*, 1995; *Gomez-Hernandez et al.*, 1997). Assume 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 models generated 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. A problem arises since the perturbed structure of each realization may differ itself from another and all of them are considered as the identified correct structure.

Although *Gomez-Hernandez et al.* (1997) made some advances in retaining the spatial structures through kriging, the superimposition of one or more perturbated images over the initial seed field tends to modify the given parameter statistics and structure properties, e.g., mean, variance, sill, angle orientation, anisotropic rate, correlation length, etc. Such weakness of the two-stage methods is inherent and hence is unavoidable. In addition, the inverse problem is often ill-posed and a unique stable solution can not always be assured. That means, an inverse-conditional realization is not guaranteed to be obtained starting from an initial seed field. Actually, for some seed fields, it does not always hold to attain a convergent result in several iterations because of the numerical instability. Such instability is worsened by the severe nonlinearity for the coupled flow and transport problems.

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) but also useful for the preserving of spatial statistics and structure (see Chapter 3). 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 which strictly obey the specification of parameter statistics and spatial structures for physical models.

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

of judiciously selecting those models inverse-conditional to the nonlinear 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 numerical 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 proposed method is effective and efficient for generation of independent high-resolution models due to the incorporation of the multi-scale blocking scheme into the classical McMC method.

The purpose of this article is to develop and implement the MsBMcMC algorithm to generate $i.i.d$ high-resolution realizations which not only honor static hard data and dynamic state data but also preserve specified spatial statistics and structure. A synthetic example under a uniform, natural-gradient flow condition is presented to demonstrate the efficiency of the proposed method. The performance and influential factors of the MsBMcMC scheme are also evaluated widely in lieu of the synthetic example. The quantitative assessment on model uncertainty, uncertainty of predicted responses and uncertainty propagation due to conditional and inverse-conditional mapping of aquifer models by the proposed method will be presented the second part of this series. Since the model structure of generated realizations are identical, the worth of conditioning data from diverse sources may be evaluated exclusively without structure uncertainty involved.

4.2   Methodology

Consider a stochastic simulation at n grid nodes conditional to m hard data and k nonlinear state data. 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 $\boldsymbol{\mu}$ is the prior mean of the Gaussian process and $\boldsymbol{C_x}$ 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 $\boldsymbol{C_y}$ 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, e.g., in Chapter 2.

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, where $\boldsymbol{y}$ is nonlinearly related to $\boldsymbol{x}$ through a forward operator $g(\boldsymbol{x})$ and $\boldsymbol{x}$ may be partly observed. The most challenging part of the conditional simulation is basically an inverse problem since an inverse operator $g^{-1}(\boldsymbol{x})$ is applied to the conditioning procedure.

4.2.1　Bayesian formulation

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

$$\pi(\boldsymbol{x}) = (2\pi)^{-\frac{n}{2}} \|\boldsymbol{C_x}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{C_x}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\right\}, \qquad (4.1)$$

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., $m < n$. The prior PDF represents some prior knowledge about the parameterization of a physical model $\boldsymbol{x}$ through the configuration of $\boldsymbol{\mu}$ and $\boldsymbol{C_x}$. It should allow for the greatest uncertainty while obeying the constraints imposed by the prior information.

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

$$L(\boldsymbol{x}|\boldsymbol{y}) = (2\pi)^{-\frac{k}{2}} \|\boldsymbol{C_y}\|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(g(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{C_y}^{-1}(g(\boldsymbol{x}) - \boldsymbol{y})\right\}, \qquad (4.2)$$

where $\boldsymbol{y} = \boldsymbol{y}_{obs} \subset R^k$ represents the values of the observations; $g(\boldsymbol{x})$ is the transfer function of $\boldsymbol{x}$, generally highly nonlinear, by which $\boldsymbol{x}$ relates to $\boldsymbol{y}$ in the same attribute representation; and $\boldsymbol{C_y} \subset R^{k \times k}$ is the covariance matrix of the vector $\boldsymbol{y}$. Note that if the observation errors of $\boldsymbol{y}$ are fully independent from each other, then $\boldsymbol{C_y}$ is a diagonal matrix. The 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., $\int L(\boldsymbol{x}|\boldsymbol{y})\pi(\boldsymbol{x})d\boldsymbol{x}$. Since the marginal likelihood is not a function of $\boldsymbol{x}$, it is typically ignored in the parameter estimation problem. Using the Bayes' theorem and dropping the normalizing constant, the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$, $\pi(\boldsymbol{x}|\boldsymbol{y})$, may be written as $\pi(\boldsymbol{x}|\boldsymbol{y}) \propto L(\boldsymbol{x}|\boldsymbol{y}) \times \pi(\boldsymbol{x})$, i.e.,

$$\pi(\boldsymbol{x}|\boldsymbol{y}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{C_x}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}) - \frac{1}{2}(g(\boldsymbol{x})-\boldsymbol{y})^T\boldsymbol{C_y}^{-1}(g(\boldsymbol{x})-\boldsymbol{y})\right\}.$$
$$(4.3)$$

The posterior pdf measures how well a parameter model $\boldsymbol{x}$ agrees with the prior information and the observed data $\boldsymbol{y}$.

It should be pointed out that we deliberately leave alone the prior statistics or the hyperparamters $\theta = (\theta_{\boldsymbol{x}}, \theta_{\boldsymbol{y}})^T$ for configuring model parameters $\boldsymbol{x}$ *a priori*. Different from the geostatistical approach of *Kitanidis* (1986) which aims at $\pi(\boldsymbol{x}, \theta|\boldsymbol{y})$, our objective is $\pi(\boldsymbol{x}, \boldsymbol{y}|\theta)$, where θ may be allowed to fluctuate over a limited scope, $\theta \in [\theta_{min}, \theta_{max}]$. Basically, *Kitanidis*'s geostatistical approach derives the hyperparameters from the posterior observations $\boldsymbol{y}$ and $\boldsymbol{x}_{obs}$ while our method treats it as prior configurations. More discussions on the hyperparameters will be available later in the context of a specific application for a case study. If model selection is considered, the final distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$ is obtained by integrating out θ, $\pi(\boldsymbol{x}|\boldsymbol{y}) = \int \pi(\boldsymbol{x}|\boldsymbol{y}, \theta)d\theta$, which is out of the scope of this work.

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 $\pi(\boldsymbol{x}|\boldsymbol{y})$. A multi-scale blocking McMC scheme is presented below to explore the posterior distribution.

4.2.2 Blocking McMC scheme

Due to the highly nonlinearity of the log-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), however, is especially suitable for exploring the parameter space of such type of complicated posterior distribution. A typical McMC algorithm employing the Metropolis-Hastings rule to explore the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ 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. If the proposal kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$ is chosen to equal the prior model $\pi(\boldsymbol{x})$, then the acceptance probability is,

$$\alpha = \frac{\pi(\boldsymbol{x}^*|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{y})}\frac{q(\boldsymbol{x}|\boldsymbol{x}^*)}{q(\boldsymbol{x}^*|\boldsymbol{x})} = \frac{\pi(\boldsymbol{x}^*)L(\boldsymbol{x}^*|\boldsymbol{y})}{\pi(\boldsymbol{x})L(\boldsymbol{x}|\boldsymbol{y})}\frac{\pi(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x}^*|\boldsymbol{x})} = \frac{\pi(\boldsymbol{x}^*)\pi(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})\pi(\boldsymbol{x}^*|\boldsymbol{x})}\frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})}.$$
$$(4.4)$$

If the proposal kernel does not depend on the current values of $\boldsymbol{x}$, the chain constructed is called an independent Markov chain (*Liu*, 1996). Obviously, the acceptance probability may be quite low for a high-dimensional problem and the chain may mix rather slowly, since,

$$\alpha = \frac{\pi(\boldsymbol{x}^*)\pi(\boldsymbol{x}|\boldsymbol{x}^*)}{\pi(\boldsymbol{x})\pi(\boldsymbol{x}^*|\boldsymbol{x})}\frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})} = \frac{L(\boldsymbol{x}^*|\boldsymbol{y})}{L(\boldsymbol{x}|\boldsymbol{y})},$$
$$(4.5)$$

where the candidate model $\boldsymbol{x}^*$ is accepted simply according to the ratio of the likelihood, which is rather random.

On the other hand, if the proposal kernel $q(\boldsymbol{x}^*|\boldsymbol{x})$ is chosen to be the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$, then the acceptance probability of $\boldsymbol{x}^*$ equals one,

$$\alpha = \frac{\pi(\boldsymbol{x}^*|\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{y})}\frac{q(\boldsymbol{x}|\boldsymbol{x}^*)}{q(\boldsymbol{x}^*|\boldsymbol{x})} = 1,$$
$$(4.6)$$

if the Metropolis-Hastings rule is used to calculate the acceptance rate. In such case, the candidate model $\boldsymbol{x}^*$ is fully accepted, which is also called the Gibbs sampler (*Geman and Geman*, 1984). An assumption for this case is that the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ can be sampled directly from which most often is not the real case for the multi-Gaussian field. Actually, this is just the McMC explanation for the classical ML method, $\text{argmax}\{\pi(\boldsymbol{y}|\boldsymbol{x})\}$, and the MAP method, $\text{argmax}\{\pi(\boldsymbol{x}|\boldsymbol{y})\}$, which is regularized by the prior $\pi(\boldsymbol{x})$.

In the ML method, a seed field $\boldsymbol{x}$ is generated according to the prior distribution $\pi(\boldsymbol{x})$, then $\boldsymbol{x}$ is calibrated to maximize the likelihood $\pi(\boldsymbol{y}|\boldsymbol{x})$ to produce a new field $\boldsymbol{x}^*$, which can be considered as a sample from the $\pi(\boldsymbol{x}|\boldsymbol{y})$

if the prior structure remains unchanged during the calibration. In the MAP method, the prior structure is explicitly forced to follow since the target model has been already regularized by the prior model in the Bayes' rule. In both methods, therefore, the generated realizations can be viewed as a direct sample from the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ using some optimization techniques. More realizations can be obtained by repeating the same procedure. If those realizations are used to construct the Markov chain, then all of them are accepted to be the output of the McMC samples. In this sense, the McMC sampler is a superset of the ML method and the MAP method. However, the challenge of the ML method and the MAP method lies in the extreme expensiveness of the optimization routines and the forward operator, generally high-dimensional and nonlinear. That makes it impossible to use the output of these two methods as the proposal kernel of the McMC computation.

Therefore, the trade-off between the independent Markov chain and the MAP method is to use an approximate posterior distribution as the proposal kernel to construct a Markov chain. The acceptance probability fully depends on how closely the proposal kernel approximates the posterior distribution. This idea efficiently improves the mixing of the chain, but also decreases the expensiveness of the high-dimensional, nonlinear optimization routines. This is just the origin of the algorithms developed in this study. The proposed method adopts the blocking update scheme as that in the simple implementation of the independent Markov chain, but the proposals are more efficient since they approximate the posterior distribution more closely. Indeed, it is well known that the blocking scheme helps improve the mixing of Markov chain (*Liu*, 1996; *Roberts and Sahu*, 1997).

The methods developed are more efficient than some algorithms resorting to the simple Gibbs sampler for the high-dimensional spatiotemporal problem, e.g., those inverse-conditioning problems haunting both the hydrogeology and the petroleum engineering community. Generally a Gibbs sampler updates the attribute field one cell by one cell using the full conditional distribution. For a high-dimensional nonlinear spatiotemporal problem, the likelihood perturbation introduced by the difference of one single cell's attribute value is ignorable in the computer implementation. Most often it may be less than the rounding error of the computation. It makes the simple Gibbs sampler quite ineffective. A blocking scheme perturbs the attribute field with more elements and thus results in much more change of the likelihood. It helps the introduction of upscaling procedures to accelerate the computation of the likelihood, which will be explained with more details below.

4.2.3 Generation of candidates

The starting point for generation of geostatistical realizations is to sample the prior distribution $\pi(\boldsymbol{x})$ which is an appropriate approximation, at least in terms of model structure, to the posterior distribution $\pi(\boldsymbol{x}|\boldsymbol{y})$ by which the likelihood part is discarded. The meaningfulness of this approximation is twofold: one is to retain the specified model structure and the other is for the fast generation of candidates. First, the posterior distribution which combines the prior and the likelihood does not necessarily mean that, owing to the introduction of the likelihood, the posterior model structure should be different than the prior one. It just means that the posterior density is proportional to the product of the prior density and the likelihood. In other words, the posterior density weighs a prior model configuration by its likelihood. The second point is that, for the sake of the repeatable generation of candidates, the BMcMC prefers a fixed model structure since the decomposition operation for the covariance matrix is applied once for all and all of the candidates generated share the same spatial structure. In this sense, a single-component McMC is more suitable for the case of inverse problem with model selection.

Although many random field generators may generate candidate realizations conditional to hard and soft data (*Deutsch and Journel*, 1998), a main challenge lies in the generation velocity. A fast generator is strongly recommended since a huge number of *i.i.d* models are required for forward evaluations. The algorithm based on the LU-decomposition of the covariance matrix was preferred in Chapter 3 since it is quite efficient and effective in generating a large number of conditional realizations. However, it fails to deal with a large dimension of models due to the difficulty of the LU-decomposition of the huge covariance matrix. In this regard, the FFT-based spectral simulator has a computational ability more powerful than the LU-based generator. It is well known that the FFT-based sampler is one of the fastest generators for unconditional simulation, which is quite suitable for the McMC computation. More implementation details on the FFT-based sampler are referred to relevant literatures elsewhere (*Pardo-Iguzquiza and Chica-Olmo*, 1993; *Robin et al.*, 1993; *Gutjahr et al.*, 1997; *Ruan and McLaughlin*, 1998). Appendix B also gives some more details on the implementation of the FFT-based sampler.

An important step to generating conditional realizations is the configuration of model structure for candidates. If the two-point geostatistics are adopted, an *ad hoc* method is to carry out experimental variogram analysis of parameters when some direct measurements are available. On the other hand, the structural parameters of candidates may benefit from fitting to the observations of state variables (*Sun and Yeh*, 1985; *Kitanidis*, 1996). For the point estimation on model parameters, however, we do not encourage to pay too much attention on inferring structural parameters from the observations

of state variables. First, this enhancement is minor when those observations are sparse in space, which is most often the case, and they are subject to large uncertainties by themselves. Second, the uncertainty of model structure, in essence, comes from the parameter model itself and is independent on state variables although the observations of state variables may adversely help improve the estimate of model structure. Third, one may require conditional realizations with specified structures which are not allowed to be modified. In summary, our aim is to generate realizations that strictly follow the specified prior configuration for model structure. It should be emphasized that the proposed BMcMC scheme is not merely limited to generate realizations with given structural parameters but also has a capability of fitting to the given observations by configuring models automatically, named automatic model selection, which is a topic of our future research.

4.2.4   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 likelihood since numerous candidates should be tested and forward simulators are 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. In such case, the model is highly discretized and the state variable is considered to be highly confident. The acceptance rate for a candidate may be extremely low and, consequently, a dramatically long Markov chain, which may be beyond the computer power available currently, has to be run in order to yield a reliable output. 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 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 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 0.1% for unconditional case, is often observed as shown as before. In such case, a large number

of candidate models are 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. First, an appropriate (economical) upscaling scheme is selected to yield less biased models. A large number of approaches have been developed for upscaling hydraulic conductivity (*Wen and Gomez-Hernandez*, 1996b). 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 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 to the scale problem and hence are especially suitable for the MsBMcMC sampler. More implementation details are beyond this paper (see Chapter 2).

Another technical detail that is worth emphasizing on is the scale discrepancy for the point value assignment of hydrological parameters between the original fine scale and the upscaled coarse scale. The problem arises because the observation and the computation are performed in two distinct support scales. The upscaled parameters should be transferred to the scale that the observations have when evaluating the likelihood, or vice versa. For a fully penetrated well under transient flow conditions, the well-block head may be calculated from the following form of the Thiem equations,

$$h_w = h_e - \frac{Q}{2\pi T} \ln \left(\frac{r_e}{r_w} \right), \qquad (4.7)$$

where h_w is the well-block head; h_e is the grid-block head calculated by the finite-difference model; Q is the total pumping/injection rate from the well; r_e is the effective grid-block radius for h_e; r_w is the well-block radius; and T is the grid-block transmissivity. With this relation, the heads at the fine

and coarse scales can be transformed into the head at the well-block scale. In the formula above, the effective grid-block radius r_e should be estimated before the well-block head is calculated. For a square grid, $\Delta x = \Delta y$, it can be calculated by,

$$r_e = \frac{\Delta x}{\exp(\frac{\pi}{2})} = 0.208 \Delta x. \tag{4.8}$$

For a non-square grid, however, there is no uniform formula applicable. *Pritchett and Garg* (1980) presented more discussions on this point.

4.2.5   Post-process of the MsBMcMC outputs

Suppose that an ensemble of *i.i.d* realizations conditional to observations has been generated, $\boldsymbol{x} \sim \pi(\boldsymbol{x}|\boldsymbol{y})$, the first two moments of a prediction variable ξ, which are often of the most interest, can be obtained by the Monte Carlo integration,

$$\mu_\xi = E[\xi(\boldsymbol{x})] = \int \xi(\boldsymbol{x}) d\boldsymbol{x}, \tag{4.9a}$$

$$\sigma_\xi^2 = E[(\xi(\boldsymbol{x}) - \mu_\xi)^2] = \int (\xi(\boldsymbol{x}) - \mu_\xi)^2 d\boldsymbol{x} = \int \xi^2(\boldsymbol{x}) p(\boldsymbol{x}) d\boldsymbol{x} - \mu_\xi^2, \tag{4.9b}$$

where $p(\boldsymbol{x}) = \pi(\boldsymbol{x}|\boldsymbol{y})$ is the pdf of the random field $\boldsymbol{x}$. The conventional Monte Carlo method generally evaluates the moments by,

$$\hat{\mu}_\xi = \frac{1}{n_r} \sum_{i=0}^{n_r-1} \xi(\boldsymbol{x}_i), \tag{4.10a}$$

$$\hat{\sigma}_\xi^2 = \frac{1}{n_r} \sum_{i=0}^{n_r-1} \xi^2(\boldsymbol{x}_i) - \hat{\mu}_\xi^2, \tag{4.10b}$$

if the realizations are assumed to be equal-likely and hence may be assigned to an equal weight of $\frac{1}{n_r}$.

An ideal Markov chain attaining such stable estimate has been observed with an acceptance rate around 15-25%. However, it is hard to obtain such ideal acceptance rate for a BMcMC in exploring the parameter space for high precision cases where a low acceptance rate is usually observed in practice. Due to such limitation, some extremely low probable models may also be sampled as the output of *i.i.d* realizations. It should be emphasized that this is not a weak point of the BMcMC scheme. On the contrary, it demonstrates the large exploration scope that the BMcMC scheme has, that is, a large number

of potential regions are visited by the Markov chain. However, those samples do have extremely low probabilities and should be assigned the weights less than $\frac{1}{n_r}$. A way out of this problem is to prune the output of the BMcMC according to their posterior probabilities, that is, to discard those candidates with extremely small posterior probabilities. A possible implementation is to set a threshold for the posterior probability. Only those realizations with probabilities above this threshold are retained. A difficulty that the user faces, however, is that it is quite hard to determine the threshold of the posterior probability. Further investigations are expected to automatically quantify the threshold to attain a stable estimate. An alternative way is to truncate outputs from a very long BMcMC and retain those realizations as required in quantity by ranking their posterior probabilities.

Up to now, a complete workflow for implementing the MsBMcMC scheme may be summarized in Figure 4.1.

Once an ensemble of *i.i.d* realizations is output from the program, the most probable estimate and its covariance can be computed from,

$$E[\boldsymbol{x}] = \bar{\boldsymbol{x}} = \int \boldsymbol{x} p(\boldsymbol{x}) d\boldsymbol{x}, \tag{4.11a}$$

$$Cov[\boldsymbol{x}, \boldsymbol{x}'] = E[(\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x}' - \bar{\boldsymbol{x}})^T] = \int \int (\boldsymbol{x} - \bar{\boldsymbol{x}})(\boldsymbol{x}' - \bar{\boldsymbol{x}})^T p(\boldsymbol{x}, \boldsymbol{x}') d\boldsymbol{x} d\boldsymbol{x}', \tag{4.11b}$$

where $p(\boldsymbol{x}, \boldsymbol{x}')$ is the joint pdf of $\boldsymbol{x}$ and $\boldsymbol{x}'$. The uncertainty of one model parameter, i.e., its variance, can be easily determined by the norm of the covariance matrix, $Var[\boldsymbol{x}] = \|Cov[\boldsymbol{x}, \boldsymbol{x}]\|$.

4.2.6 Performance assessment of MsBMcMC

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, $\boldsymbol{x} = (\boldsymbol{x}_0, \boldsymbol{x}_1, ..., \boldsymbol{x}_{n_r-1})^T$.

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,

Generation of an initial field $\mathbf{x}_f$

Generation of a candidate $\mathbf{x}^*_f$

Multiscale scheme?

no

yes

Upscaling to $\mathbf{x}^*_c$

Flow simulation $h(\mathbf{x}^*_c)$

Transport simulation $t(\mathbf{x}^*_c)$

Computation of the likelihood $L(\mathbf{y}|\mathbf{x}^*_c)$

Computation of the accept rate $\alpha(\mathbf{x}^*_c|\mathbf{x}_c)$

M-H accept $\mathbf{x}^*_c$?

no

yes

Flow simulation $h(\mathbf{x}^*_f)$

Transport simulation $t(\mathbf{x}^*_f)$

Computation of the likelihood $L(\mathbf{y}|\mathbf{x}^*_f)$

Computation of the accept rate $\alpha(\mathbf{x}^*_f|\mathbf{x}_f)$

M-H accept $\mathbf{x}^*_f$?

no

yes

McMC output $\mathbf{x}^*_f$

More realizations?

yes

no

Post-process of McMC outputs

Update the initial field $\mathbf{x}_f$

Figure 4.1: *Flowchart of the multi-scale blocking McMC scheme*

(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,

$$\sigma_t = \sum_{i=0}^{t} (\eta(\boldsymbol{x}_i) - \bar{\eta}), \tag{4.12}$$

for $t = 0, 1, \ldots, n_r - 1$, 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 $i = 1, 2, \ldots, n_r - 1$;
(4) Calculate the hairiness indices,

$$\Sigma_t = \frac{1}{t-1} \sum_{i=1}^{t-1} \delta_i, \tag{4.13}$$

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

The key outputs are two sequential plots: $\sigma = (\sigma_0, \sigma_1, ..., \sigma_{n_r})^T$ and $\Sigma = (\Sigma_0, \Sigma_1, ..., \Sigma_{n_r})^T$. The first one, 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 second one, Σ_t, presents a quantitative measure of smoothness, hairiness index, to evaluate the convergence performance of a chain. An ideal convergence sequence will be centered at around 0.5.

4.3   A Synthetic Example

The purpose of this exercise is to examine the efficiency and performance 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 these conditions 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 conditional and 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, the porosity, retardation factor and other parameters are treated as the constants or known.

4.3.1   Reference models and conditioning data set

A synthetic 2D highly-correlated multi-Gaussian confined aquifer under a uniform, natural-gradient flow condition, as shown in Figure 4.2 (A), serves as the reference field to illustrate the efficiency of the proposed method for conditional and inverse-conditional simulation.
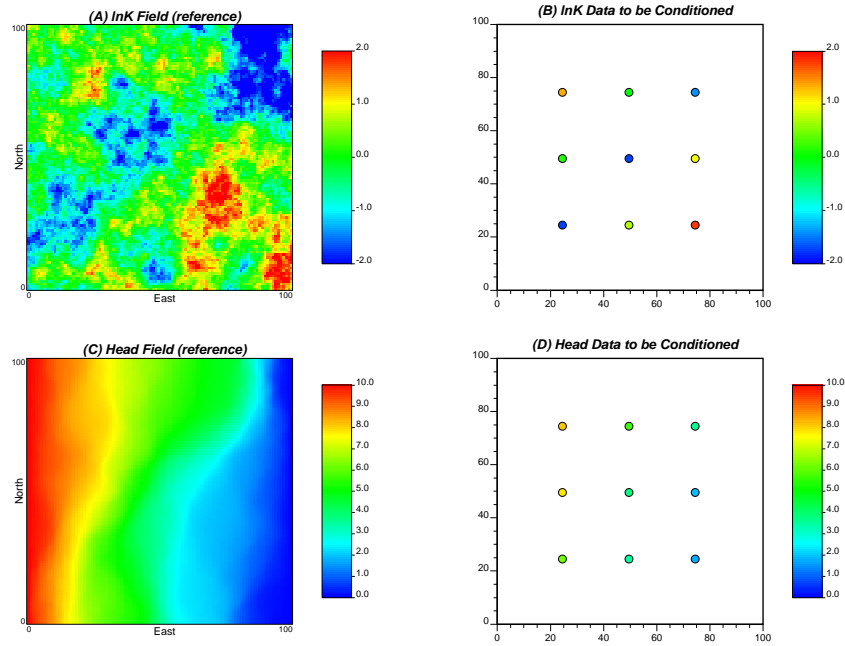
Figure 4.2: *The conditioning data set and the reference* $\ln K$ *and head fields* $(\mu = 0, \lambda_{\boldsymbol{x}} = 50, \sigma_x^2 = 1.0)$

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 $l : \lambda_x = 100 : 50$. The grid resolution is 50 by 50 per correlation length, i.e., $R_\lambda = 50 \times 50$, which is enough to characterize the spatial variation in details. However, the highly correlation 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. One side-effect of this is that the acceptance rate is dramatically lowered for the MsBMcMC scheme which will be discussed with more details in Section 4.

Another factor to be considered is the ratio of the correlation length to the sampling spacing since the sampling density may be an important factor for model identification. Generally speaking, as stated by *McLaughlin and Townley* (1996), compared to the sampling spacing, a smaller prior correlation length yields an inverse estimate that is closer to the prior mean except in the immediate vicinity of measurements. 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 $\lambda_x : \Delta s = 50 : 25$ which keeps two sampling points per correlation length which is deemed to be capable of capture the main characteristics of the real aquifer.

The single-phase flow problem 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 ten and zero, respectively, i.e., $h_{-x} = 10$

and $h_{+x} = 0$. The confined steady-state flow problem is solved by a blocking-center finite-difference simulator developed in this study. The reference head field is shown in Figure 4.2 (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 4.2 (B) as hard data to be conditioned. Nine head values at the same locations are shown in Figure 4.2 (D) as state data to be inverse-conditioned.

Therefore, the problem is to infer the spatial distribution of $\ln K$ from nine $\ln K$ measurements and nine head observations under the same flow and transport conditions by the MsBMcMC method.

4.3.2 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 4.2 (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 4.2 (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 head field, as shown in Figure 4.2 (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 developed 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. Note that the experimental mean and variance may deviate slightly from the input values due to some technical modifications in implementing the FFT-based random sampler. 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. The absolute resolution for parameter calibration is set as $\sigma_y^2 = 0.17$ which means that the head field may fluctuate around the given head observations with standard deviation equal to 0.41. In other words, the head observations h_i ($i \leq k$) are corrupted by noises 0.41, i.e., $h_i \pm 0.41$ ($i \leq k$).

Three thousand independent realizations, whose $\ln K$ values closely follow the given Gaussian distribution, i.e., $\ln K \sim N(0,1)$, are generated to calculate the ensemble mean $\ln K$ field, which represents the most probable estimate by the MsBMcMC method. As shown in Figure 4.3 (A), the main distribution pattern of the reference $\ln K$ field (see Figure 4.2 (A)) has been well identified in visual though the amplitude of the mean estimate is weaker than that of the reference field. Note that the scale of estimate is less than that of the reference as shown in the legends. 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.
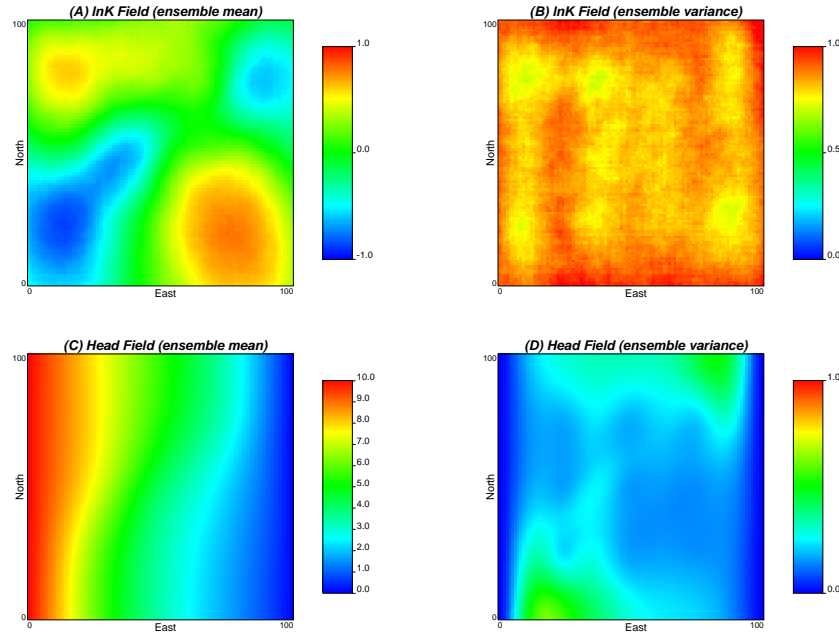
Figure 4.3: *An MsBMcMC estimate by inverse-conditioning to head observations* $(\lambda_{\boldsymbol{x}} = 50, \sigma_x^2 = 1.0, \sigma_y^2 = 0.17)$

The norm of the posterior covariance matrix represents the actual error of parameters in uncertainty quantification. Since the MsBMcMC method does not resort to the sensitivity matrix for the inverse-conditioning simulation, the uncertainty assessment based on sensitivity is no longer applicable. Without considering the correlation, however, the ensemble variance (Equation 4.9b) is in reality an alternative representation of uncertainties. Figure 4.3 (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 simply displays the degree of reliability of 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 4.3 (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. As a result, the estimated $\ln K$ near the center has less uncertainty than that close to other eight points. It is worthwhile pointing out that the conclusion for reliability analysis may only apply to this example under the steady-state natural-gradient flow condition. More numerical experiments subject to different flow configurations are required to reach a more common conclusion for reliability evaluation. Moreover, the conclusion drawn from the proposed approach may not be applicable to other inverse methods.

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 4.3 (C), which is quite similar to the reference head field, i.e., the lower part has earlier breakthrough than the upper part though the estimated distribution is more regular and smooth. The variance of estimated head fields is plotted in Figure 4.3 (D), which displays the fluctuation sizes of heads. Note that the values on two boundaries, i.e., the left and the right side, 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 efficiently extracted, though underestimated in magnitude, by the proposed MsBMcMC inverse-conditioning algorithm. However, the effect of such underestimation in uncertainty assessment may be compensated for by generating a large number of independent realizations with aid of the Monte Carlo simulation which is just what the proposed MsBMcMC method aims at.

4.4   Performance Assessment of MsBMcMC

Two key measures for evaluating the performance of the MsBMcMC method are the acceptance rate and the accuracy of the identified $\ln K$ field. The former aims at the evaluation of the efficiency of the MsBMcMC method in exploring the parameter space. The acceptance rate is one of measures that are directly related to such efficiency. It also measures the CPU efficiency of the MsBMcMC method for conditioning and inverse-conditioning procedure. Since the inverse problem is notoriously computationally intensive a high acceptance rate is strongly desired for generation of conditional realizations in a limited CPU time. The accuracy of the identified $\ln K$ field measures the goodness-of-fit of conditional and inverse-conditional simulations by the proposed MsBMcMC method. Unfortunately, such accuracy is hardly quantified systematically. An *ad hoc* way is to compare, in visual, the simulated results with the reference fields. On the basis of these measures, several factors that affect the performance of Markov chains have been examined and analyzed in total twelve scenarios as shown in Table 4.1.

4.4.1   Effects of spatial correlation

As for classical inverse methods, it is recognized that incorrectly specified models may have an adverse impact on inverse results because the parameter estimates are forced to compensate for unacknowledged errors (*McLaughlin and Townley*, 1996). The experiments of this and the next subsection are designed to examine such influence from incorrect prior configurations for model structure, i.e., model uncertainties, on the inverse-conditioning simulation by the proposed MsBMcMC scheme.

Suppose $\boldsymbol{x}$ is characterized by a constant mean (stationary random field) and isotropic, exponential covariance, i.e.,

$$E[\boldsymbol{x}] = \mu, \tag{4.14a}$$

$$Cov[\boldsymbol{x}] = \sigma_x^2 \exp\left(-\frac{d_{\boldsymbol{x}}}{\lambda_{\boldsymbol{x}}}\right), \tag{4.14b}$$

Table 4.1: *MsBMcMC configurations and performance indices*

Scenario	$\boldsymbol{x}$	$\lambda_{\boldsymbol{x}}$	σ_x^2	σ_y^2	$\alpha(\%)$	$\langle\Delta y^2\rangle$	Σ_t
1	$\boldsymbol{x}\|\boldsymbol{y}$	50	1.0	1.00	9.50	7.98	0.308
2	$\boldsymbol{x}\|\boldsymbol{y}$	50	1.0	0.25	0.33	3.26	0.243
3	$\boldsymbol{x}\|\boldsymbol{y}$	50	1.0	0.17	0.09	2.41	0.253
4	$\boldsymbol{x}\|\boldsymbol{y}$	20	1.0	1.00	19.47	8.01	0.334
5	$\boldsymbol{x}\|\boldsymbol{y}$	20	1.0	0.25	0.71	3.93	0.272
6	$\boldsymbol{x}\|\boldsymbol{y}$	50	0.5	0.17	0.74	2.10	0.324
7	$\boldsymbol{x}\|\boldsymbol{y}$	50	1.5	0.25	0.20	3.21	0.249
8	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	50	1.0	1.00	33.16	5.03	0.362
9	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	50	1.0	0.25	3.95	2.49	0.287
10	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	50	1.0	0.17	1.40	1.89	0.284
11	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	20	1.0	1.00	31.56	5.83	0.384
12	$\boldsymbol{x}\|\boldsymbol{x}_{obs},\boldsymbol{y}$	20	1.0	0.25	2.47	2.96	0.266

where $d_{\boldsymbol{x}}$ is the separate distance and $\lambda_x = \lambda_y$ is the correlation length. For the case as used in this synthetic example, the key structure parameters boil down to $\theta_{\boldsymbol{x}} = (\sigma_x^2, \lambda_{\boldsymbol{x}})$. The attention is then focused on the effect of *a priori* specifying the spatial correlation length $\lambda_{\boldsymbol{x}}$ (this subsection) and the variance σ_x^2 (the next subsection) on the estimation of model parameters.

Figure 4.4 compares the influence of correlation length on the reproduction of reference fields. The left column indicated by (A) shows the result of $\lambda_{\boldsymbol{x}} = 50$ (correctly specified) while the right one indicated by (B) shows the result of $\lambda_{\boldsymbol{x}} = 20$ (wrongly specified). Both of the $\ln K$ fields (A.1 and B.1) reproduce the main characteristic of the reference $\ln K$ field (Figure 4.2 (A)). However the identified magnitudes are different. The correct correlation length better improves the signal-to-noise ratio. Note that the left image (A.1) has a larger scale of legend. Moreover, the variance of estimated $\ln K$ fields as shown in A.2 is smaller than that of the right one (B.2) which means that the correct $\lambda_{\boldsymbol{x}}$ yields more reliable estimates. The simulated head fields also show the same trend: the one from correct $\lambda_{\boldsymbol{x}}$ (A.3) is closer to the real case (Figure 4.2(C)) than the wrong one (B.3). This numerical experiment shows that a prior $\lambda_{\boldsymbol{x}}$ configuration closer to the reality gives more stable and reliable estimates. However, even though the correlation length is wrongly specified *a priori* which most often happens since the underground reality is never fully observed or is observed but with large errors, the main features of the reference fields are also reproduced which proves the robustness of the proposed method.

Figure 4.5 reveals the effect of correlation length on the performance of MsBMcMC. The sequential plots as shown in (A) and (B) demonstrate the
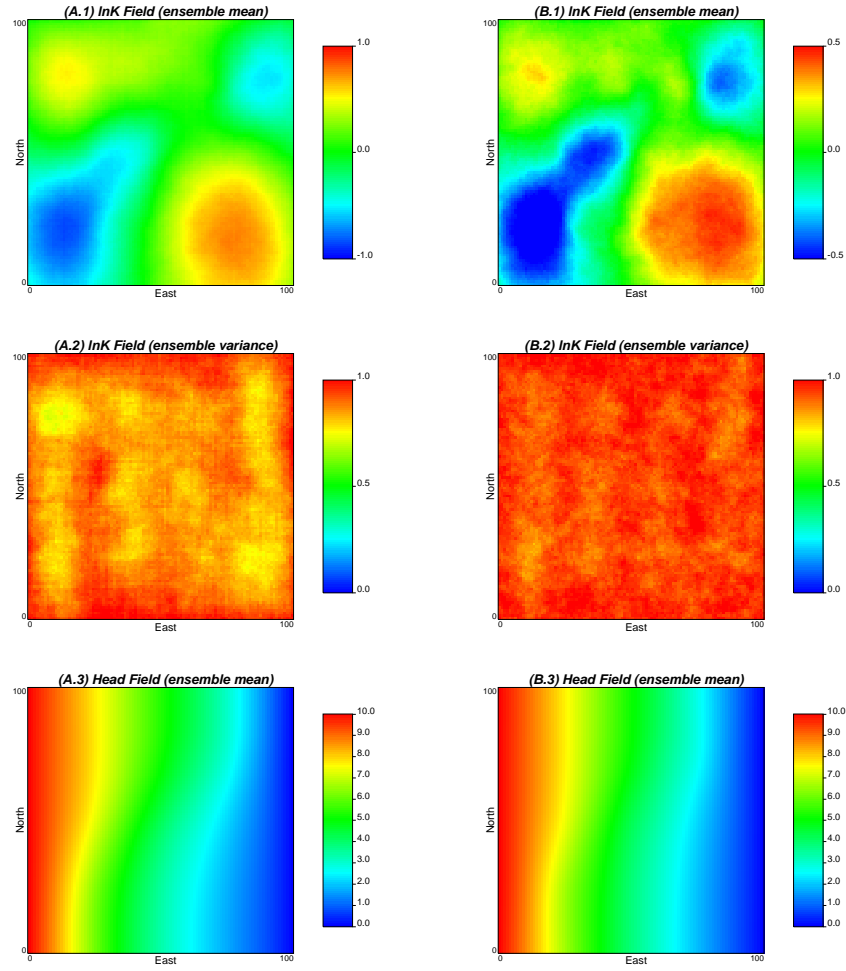
Figure 4.4: *Effect of prior correlation length on the reproduction of reference fields* ($\sigma_x^2 = 1.0, \sigma_y^2 = 0.25, (A)\lambda_{\boldsymbol{x}} = 50, (B)\lambda_{\boldsymbol{x}} = 20$)

squared head mismatch resulted from the correct correlation length configuration ($\lambda_{\boldsymbol{x}} = 50$) and wrong one ($\lambda_{\boldsymbol{x}} = 20$), respectively. Statistically, the correct one has a smaller mean mismatch (3.26 for $\lambda_{\boldsymbol{x}} = 50$ and 3.93 for $\lambda_{\boldsymbol{x}} = 20$) which implies that the correct identification of correlation length produces results more reliable and closer to the reality. Figure 4.5 (C) compares the sequential CUSUM where the correct $\lambda_{\boldsymbol{x}}$ obviously has a slightly smaller excursion size which means the chain mixes a little more rapidly. Moreover, the degree of smoothness does not alter a lot as shown in Figure 4.5 (D) where the hairiness indices of two graphs are very similar (0.243 for $\lambda_{\boldsymbol{x}} = 50$ and 0.272 for $\lambda_{\boldsymbol{x}} = 20$). In summary, correct identification and configuration for correlation length help, but slightly, improve the mixing of the chain and produce results more reliable and closer to the reality.

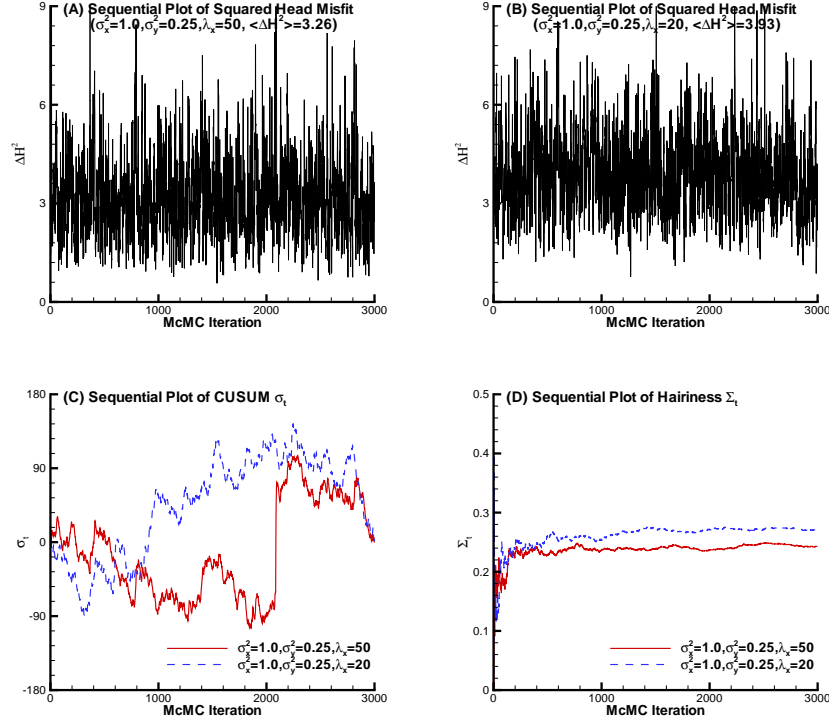Figure 4.5: *Effect of prior correlation length on the performance of MsBMcMC* ($\sigma_x^2 = 1.0, \sigma_y^2 = 0.25, (A)\lambda_{\boldsymbol{x}} = 50, (B)\lambda_{\boldsymbol{x}} = 20$)

The correlation length of model parameters has a certain effect on the acceptance rate which can be seen from Table 4.1. Such influence is more apparent for the case unconditional to the hard data. Compared to basic

configurations in scenario 1 and 2 ($\lambda_{\boldsymbol{x}} = 50$), experiments in scenario 4 and 5 ($\lambda_{\boldsymbol{x}} = 20$) are designed to examine such influence for the cases where the candidates are not conditioned to the hard data. The shorter correlation length obviously enhances the acceptance rate, i.e., from 9.50% to 19.47% for the case $\sigma_y^2 = 1.00$ and from 0.33% to 0.71% for the case $\sigma_y^2 = 0.25$. However, for the case where the candidates are conditioned to the hard data, the acceptance rate almost remains unchanged which can be observed clearly by comparing the scenario 11 (31.56%) to the scenario 8 (33.16%) and the scenario 12 (2.47%) to the scenario 9 (3.95%). The reasoning for this observation is as follows. A longer correlation length increases the dependence of model parameters which makes it less free for the proposal kernel to move around the parameter space that the MsBMcMC has to explore extensively. In other words, if their correlation length is longer the candidates have less degree of freedom, and hence the size of the admissible model set becomes smaller and the acceptance rate decreases correspondingly. However, conditioning to hard data effectively decreases the dimension of candidates which essentially relaxes this rigidity of degree of freedom that a long correlation length strictly sets on. As a result, the acceptance rate increases in nature.

4.4.2   Effects of variances of model parameters

Several conventional means for estimating the variances of model parameters include, (1) collecting the direct measurements of parameters, (2) collecting the indirect geophysical measurements, and (3) analogizing to the geological observations from the outcrop. Although the determination of the variance benefits from such information, the assumed statistics from this are difficult to verify in practice. For this reason, it is of considerate interest to analyze the sensitivity of inverse results on the prior variances of model parameters. As shown in Equation 4.1, the mean spatial distribution $\boldsymbol{\mu}$ is practically unknown *a priori* so that a large variance for $\boldsymbol{C_x}$ is most often specified in order to avoid an unreliable estimate of $\boldsymbol{\mu}$.

Figure 4.6 displays the effect of the prior variance configuration for model parameters on the reproduction of reference fields. The left column indicated by (A) plots the results from a smaller variance specification ($\sigma_x^2 = 0.5$) and the right column indicated by (B) plots the results from a larger variance specification ($\sigma_x^2 = 1.5$). Other relevant results for the base variance ($\sigma_x^2 = 1.0$) are plotted in Figure 4.4 (A) and Figure 4.3 for $\sigma_y^2 = 0.25$ and $\sigma_y^2 = 0.17$, respectively. One can easily find that the *a priori* underestimated variance gives rise to incorrect, at least in visual, estimations for both the ln$K$ field (A.1) and the head field (A.3) that obviously deviate from true cases (Figure 4.2 (A) and (C), respectively) even though the posterior uncertainties are reduced due to condition on head data (A.2) which demonstrates the robustness of the pro-

posed MsBMcMC method. On the other hand, the overestimated variance yields visually correct estimations for both the $\ln K$ field (B.1) and the head field (B.3). Another observation is that such overestimation helps improve the identification of signal in amplitude (B.1) but the posterior uncertainties are also exaggerated (B.2).
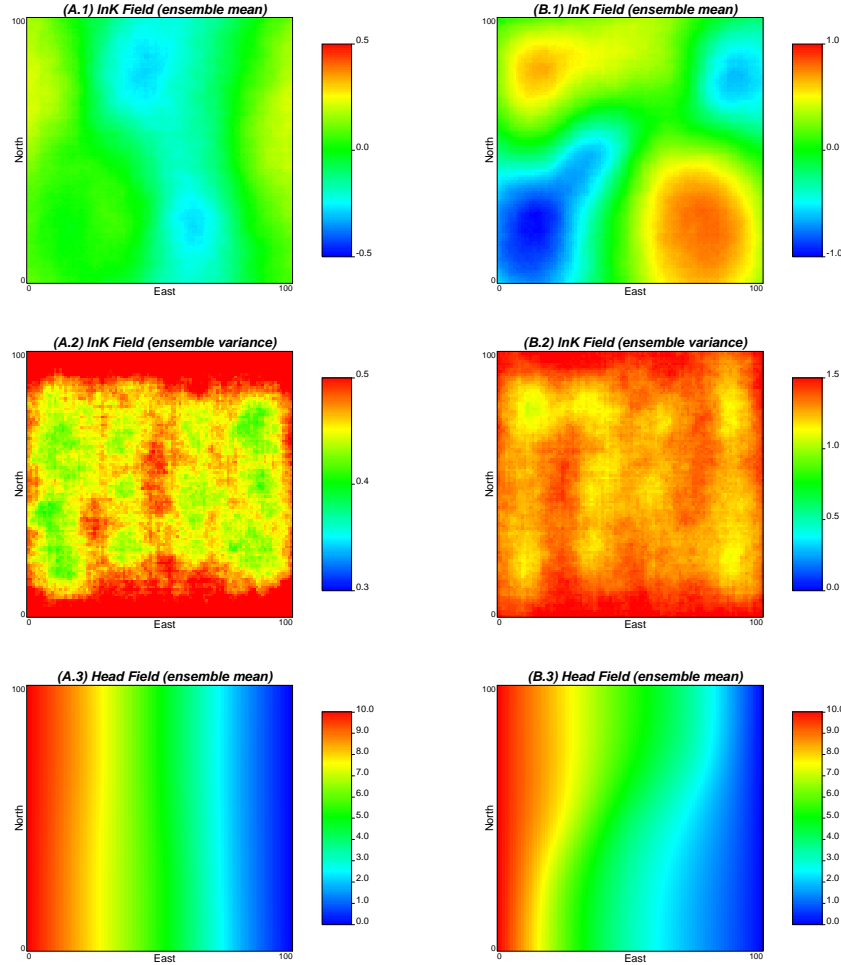
Figure 4.6: *Effect of prior variances (σ_x^2) on the reproduction of reference fields* $(\lambda_{\boldsymbol{x}} = 50, (A)\sigma_x^2 = 0.5, \sigma_y^2 = 0.17, (B)\sigma_x^2 = 1.5, \sigma_y^2 = 0.25)$

Figure 4.7 compares the effect of different variance specifications on the performance of the MsBMcMC method. It is hard to tell the difference of estimated accuracy between the correct variance and the wrong one only from the sequential plot for squared head mismatch (Figure (A) through (D)) but a

smaller variance gives a slightly smaller mismatch. From the sequential plots of CUSUM (E) and hairiness indices (F), one can find a smaller variance seems to mix more rapidly and renders a more stable and reliable estimation.

Prior specification for the variance of model parameters has a significant effect on the acceptance rate as shown in Table 4.1. A small variance enhances the acceptance rate while a large variance lowers the acceptance rate. By comparing the scenario 3 to the scenario 6, one can find that the variance decreases from 1.0 to 0.5 and the acceptance rate increases from 0.09% to 0.74% correspondingly. Similarly, by comparing the scenario 2 to the scenario 7, one can find that the variance increases from 1.0 to 1.5 and the acceptance rate decreases from 0.33% to 0.20% correspondingly. In a certain limit, a larger prior variance obviously decreasing the acceptance rate is due to the larger uncertainty of model parameters while a smaller prior variance increasing the acceptance rate is because the model parameters are relatively deterministic. Another reason is that the smaller variance yields smoother candidates and the fluctuation of their likelihoods is limited to a smaller range, so does the fluctuation of the posterior densities. A natural result is that it increases the possibility of accepting a candidate. However, this conclusion is not always applicable for all the prior variances. For example, if the prior mean is over-confident and the variance is extremely small, the acceptance rate may be small as well since the probability of a candidate matching state data becomes extremely nondeterministic.

4.4.3   Effects of variances of state variables

Carefully examining the Equation 4.3 for the posterior distribution, one can weigh the worth of information provided by the prior assumption for the model (essentially from the direct measurements) and by the posterior observations of state variables. The weight of the prior term is small if the measurement errors of state variables are large while the weight of the likelihood will be small if the prior statistics are uninformative, i.e., the variance is relative large. Due to the worth discrepancies that sources of information hold, different relative weights may be assigned to different types of conditioning data, i.e., the direct measurements of model parameters and indirect observations of state variables. For the sake of retaining the model structure, C_x and μ are generally specified *a priori* before the conditional simulation starts, therefore, a user is more than encouraged to adjust C_y to weigh the state data for the inverse-conditional simulation. The weight assignment of state variables fully depends on the reliability of observations. As the information content of observations of state variables increases, i.e., a decrease of their variances, the output of MsBMcMC reproduces more features of the reference field. Figure 4.8 compares such effect of different variances of head data on the reproduction of reference fields.

Figure 4.7: *Effect of prior variances (σ_x^2) on the performance of MsBMcMC*
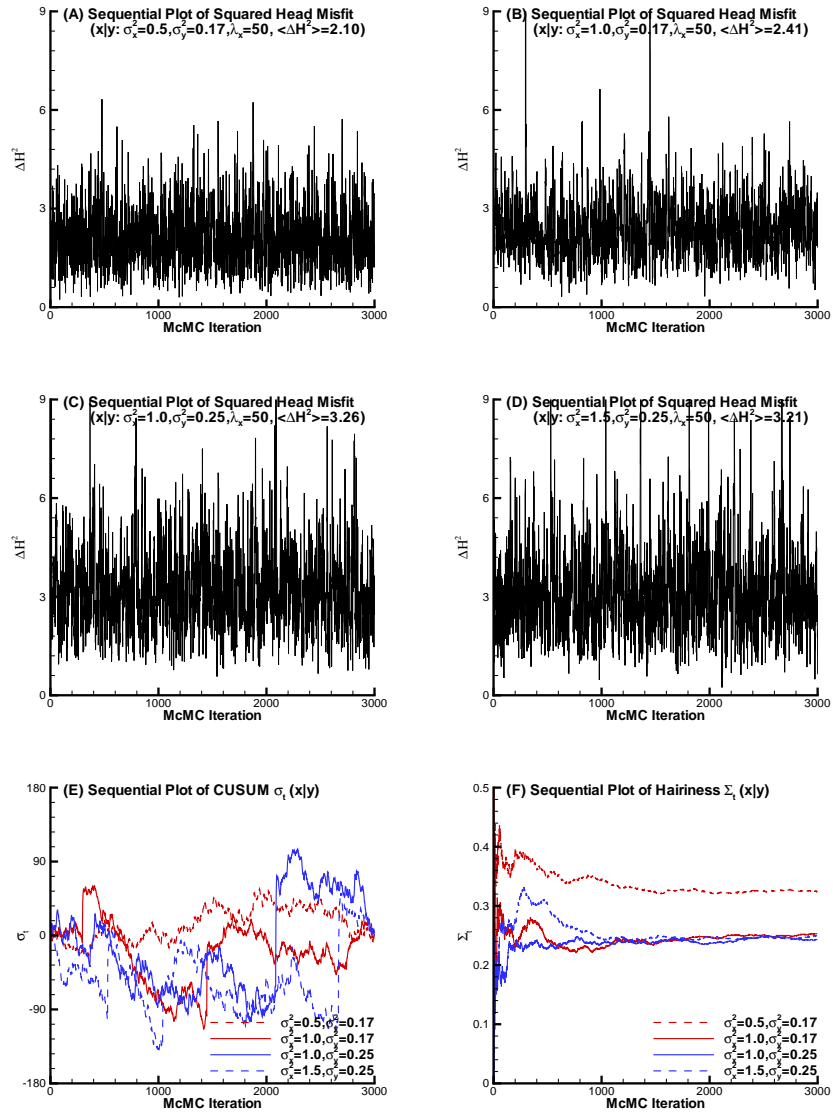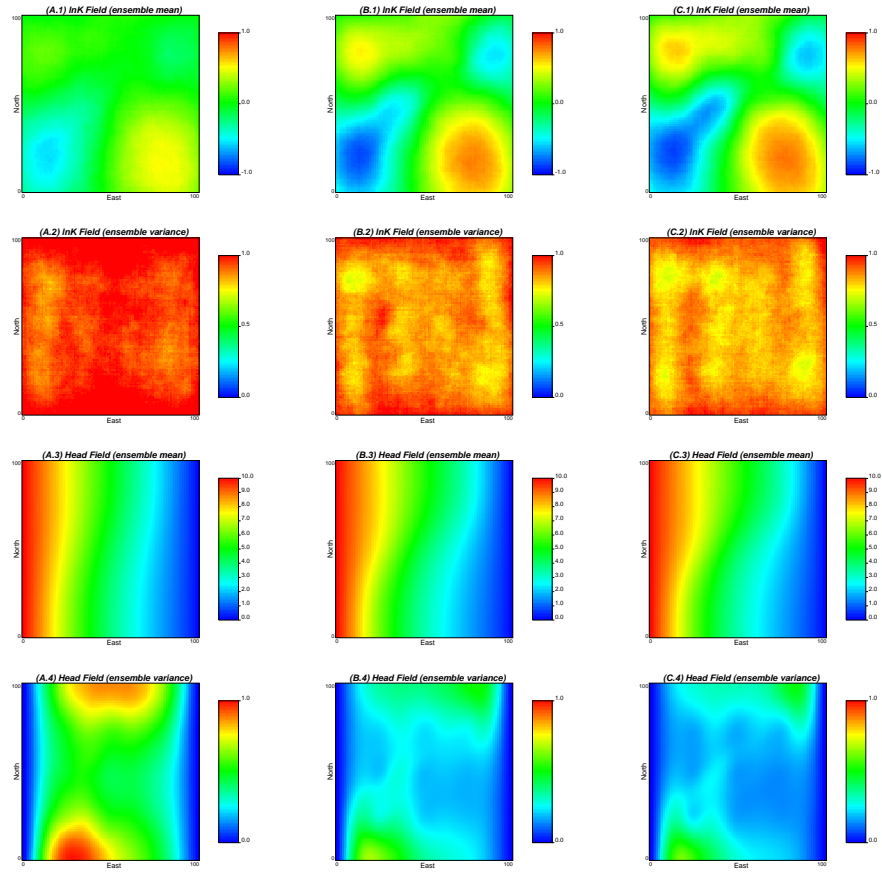
Figure 4.8: *Effect of variances of head data* (σ_y^2) *on the reproduction of reference fields* $(\lambda_{\boldsymbol{x}} = 50, \sigma_x^2 = 1.0, (A)\sigma_y^2 = 1.00, (B)\sigma_y^2 = 0.25), (C)\sigma_y^2 = 0.17)$

The left column indicated by (A) in the figure shows a rather big variance of state variables ($\sigma_y^2 = 1.00$), the middle one indicated by (B) shows a rather moderate variance ($\sigma_y^2 = 0.25$), and the right one indicated by (C) shows a rather small variance ($\sigma_y^2 = 0.17$). One can easily find that the ability of the reproduction of reference fields increases as the variance decreases. The crucial pattern of the reference $\ln K$ field is well uncovered by all of them (from A.1 through C.1) but the mean $\ln K$ field of the small variance is closer to the reality. Notice that the amplitude of signal of (C.1) is obviously higher than the others (A.1 and B.1). The variance of estimated $\ln K$ fields decreases as the decrease of the variance of state variables (from A.2 through C.2) which means that the small variance specification gives more stable estimation. Moreover, the estimated head distribution is closer to the real case (from A.3 through C.3) and more reliable and stable (from A.4 through C.4) as the variance decreases.

Figure 4.9 displays the effect of variances of state data on the performance of MsBMcMC. The sequential plots (A) and (B) show the squared head mismatch resulted from different variance specification, $\sigma_y^2 = 1.00$ and 0.17, respectively. For $\sigma_y^2 = 0.25$, the sequential plot is shown in Figure 4.5 (A). Obviously, as the decrease of variances, the mismatch significantly decreases. The mean squared head misfits are 7.98, 3.26, and 2.41 for $\sigma_y^2 = 1.00$, 0.25, and 0.17, respectively. Figure 4.9 (C) and (D) show the sequential plots of CUSUM and their hairiness indices, respectively. It can be observed that a smaller variance has a smaller excursion size. But at the same time, the hairiness indices also drop down from 0.308 to 0.243 to 0.253 as the variance increases from 1.00 to 0.25 to 0.17. Therefore, the variance of σ_y^2 has a significant influence on the performance of MsBMcMC.

The acceptance rate is prominently affected by the variances of state variables which can be seen clearly from Table 4.1. The increase or decrease of the acceptance rate α is positively proportional to the increase or decrease of the variances of state variables (σ_y^2). For example, from the scenario 1 through 3, the acceptance rate decreases from 9.5% to 0.09% as the variance decreases from 1.00 to 0.17. The same trend is also observed in the case of shorter correlation length (scenario 4 and 5) and in the case conditional to hard data (scenario 8 through 10). The reasoning for such observation is as follows. As the variances of state variables increase, the state data becomes less informative and their constraints on the candidates become less rigorous. More candidates are therefore accepted as the components of the Markov chain. On the contrary, as the variances decrease, the state data are considered as more accurate measurements and their constraints are stricter. Fewer candidates may meet more stringent constraints and the acceptance rate decreases correspondingly since the parameter space remains unchanged.
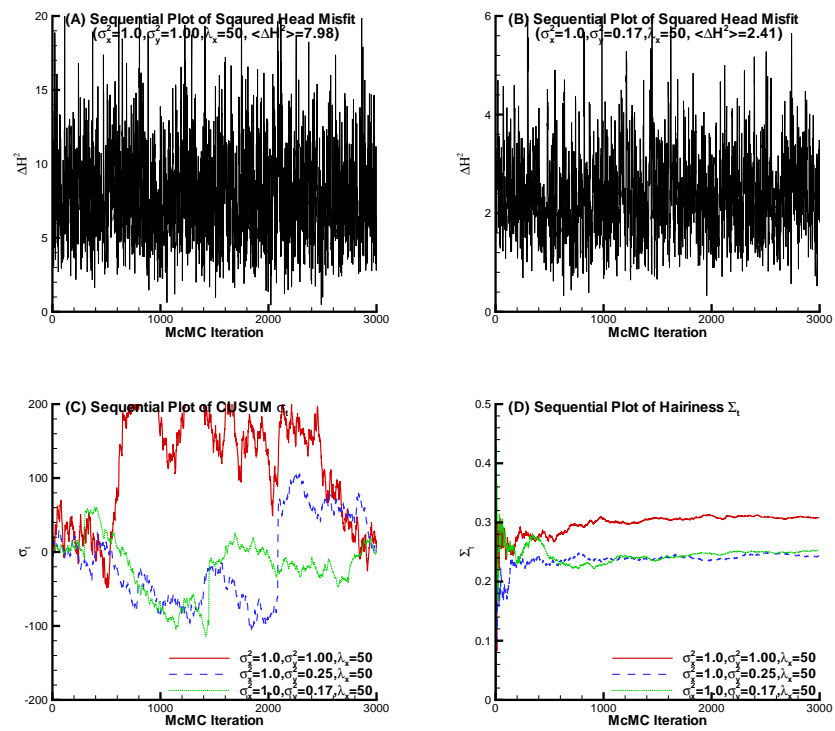
Figure 4.9: *Effect of variances of head data (σ_y^2) on the performance of Ms-BMcMC*

4.4.4   Effects of conditioning on hard data

Conditioning to hard data efficiently decreases the degree of freedom of a stochastic process and hence narrows down the size of the parameter space, i.e., from R^n to R^{n-m}, which adversely helps improve the exploration efficiency of the MsBMcMC method owing to the increase of the weight of the potentially admissible parameter space. In the case of highly correlated random field, such improvement is more significant since the correlated neighborhood also benefits from the conditioning.

One direct result is the enhancement of the acceptance rate which can be easily observed from the Table 4.1. From the scenario 8 through 12 which are the cases conditional to the hard data, the acceptance rates are obviously higher than the unconditional cases (from the scenario 1 through 5, respectively). The enhancement is more obvious for the cases with a small variance of state variables. Note that in the scenario 3 and 10 where the variance is 0.17 the acceptance rate increases from 0.09% to 1.4% by conditioning on nine hard data. It should be pointed out that such big influence of conditioning to the hard data on the acceptance rate may be only applicable to the case of high correlation as in this case study. The effect of the conditioning number of state data and hard data on the estimation is undergoing more investigation. But from the experience of the authors, an effective number of measurements for the most probable estimate is about $l : \lambda_{\boldsymbol{x}}$, i.e., the ratio of the computational domain to the correlation length, specifically 2×2 in this case study.

Figure 4.10 plots an MsBMcMC estimate by conditioning to the hard data. The mean $\ln K$ field (Figure 4.10 (A)) can be viewed as a superimposition of the kriging estimate on the unconditional estimate (Figure 4.3 (A)). The distribution of high and low values basically reflects the main features of the reference field both in location and in magnitude. Figure 4.10 (B) shows the uncertainty of estimated $\ln K$ fields. Since the local values at the nine observation points are fully honored their variances are zero. In the neighborhood of the nine observations the variance decreases obviously. Close to the boundaries the estimate is subject to more uncertainties. Compared to the unconditional case as shown in Figure 4.3 (A) and (B), the kriging estimate seems to dominate the magnitude over the inverse-conditional estimate. The simulated head field (Figure 4.10 (C)) reproduces the reference model slightly better than that of unconditional estimate (see Figure 4.3 (C)). The reliability of such estimate also improves (Figure 4.10 (D)). Note that the variances close to boundaries may not necessarily decrease due to the influence of boundary conditions. But at the center part of computational domain the variances do decrease compared to Figure 4.3 (D).

Figure 4.11 demonstrates the effect of conditioning to hard data on the performance of MsBMcMC. The sequential plot (A) shows the squared head
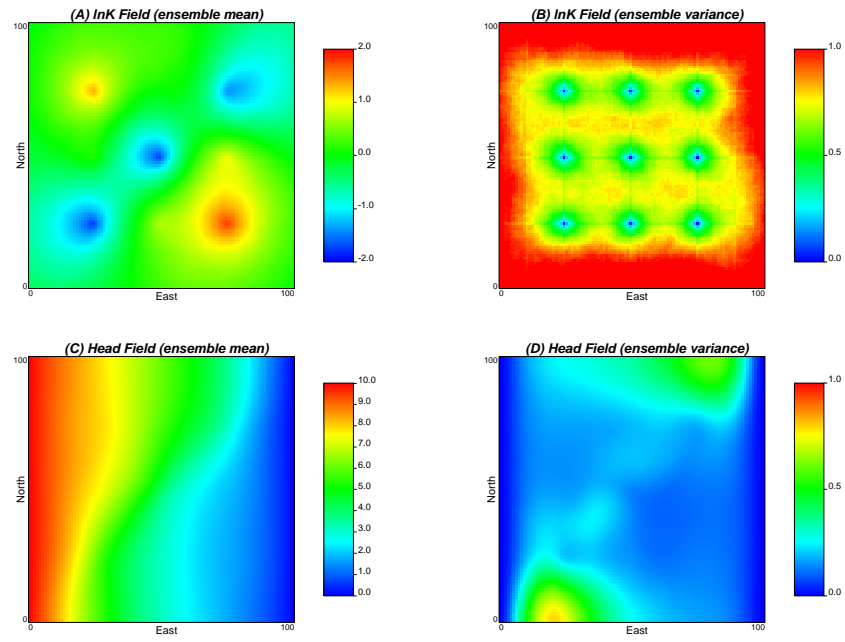
Figure 4.10: *Effect of conditioning to hard data on the reproduction of reference fields* $(\lambda_{\boldsymbol{x}} = 50, \sigma_x^2 = 1.0, \sigma_y^2 = 0.17)$

mismatch for the case unconditional to the hard data while the plot (B) shows the squared head mismatch for the case conditional to the hard data. Obviously, conditioning to hard data makes the estimated mismatch smaller, i.e., from 2.41 to 1.89. The graph (C) plots the CUSUM sequences and the graph (D) plots their hairiness indices. Conditioning to hard makes small the excursion size and also increases the irregularity of CUSUM sequence. The hairiness indices increase from 0.253 to 0.284 owing to conditioning to hard data. Both measures show that conditioning to the hard data helps improve the mixing of the chain.
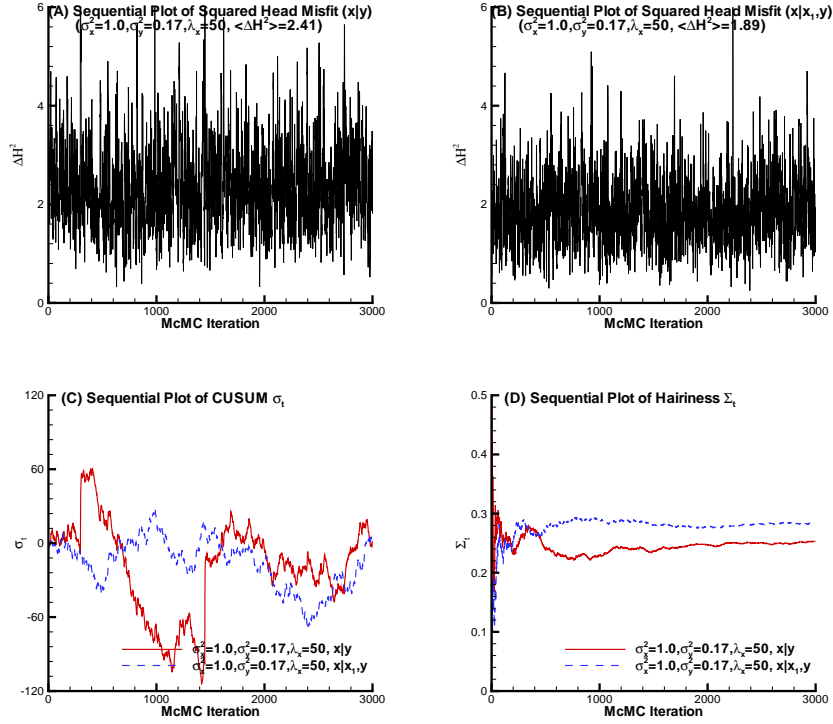
Figure 4.11: *Effect of conditioning to hard data on the performance of MsBM-cMC*

4.4.5   Improve output by pruning

The evaluation of model responses $\xi(\boldsymbol{x})$ is computationally demanding since the number and dimension of random fields are usually large. On the other hand, the MsBMcMC method generally calls for a large number of realizations

to obtain a stable estimate for a stochastic process in order to compensate for the anomalous cases of low probabilities. A way to mitigating this discrepancy is to select only the most probable models subject to the forward response evaluation. It achieves by ranking the geostatistical realizations according to their weights of probability and selecting only those with top scores of ranking. Indeed, when the most probable estimate and its variance are computed from output $i.i.d$ realizations as follows,

$$E[\boldsymbol{x}] = \bar{\boldsymbol{x}} = \int \boldsymbol{x} p(\boldsymbol{x}) d\boldsymbol{x}, \qquad (4.15a)$$

$$Var[\boldsymbol{x}] = \int (\boldsymbol{x} - \bar{\boldsymbol{x}})^2 p(\boldsymbol{x}) d\boldsymbol{x}, \qquad (4.15b)$$

those realizations with low posterior probabilities, i.e., when $p(\boldsymbol{x}) = \pi(\boldsymbol{x}|\boldsymbol{y})$ is small, may be discarded by which the estimate should not be affected too much since their weights are generally small.

The advantage of this post-process can be verified by examining the residual distribution of $\ln K$ field. Actually, one of main results that should be extracted from the head observations by inverse-conditioning is the reproduction of the large-scale spatial trend of $\ln K$ distribution. If the relationship between the spatial trend and the random fluctuation is considered as signal to noise (*McLaughlin and Townley*, 1996), the signal-to-noise ratio defined as such is a performance measure of an algorithm. A simple way is that, by means of subtracting the identified trend (signal) from the reference field (system), the uncorrelated degree of residual distribution (noise) measures the efficiency of the algorithm. The residual distribution tends to behave like white noise if the signal is separated perfectly. Figure 4.12 (B) plots such noise distribution by subtracting the signal identified by 3000 realizations (Figure 4.12 (A)) from the system (Figure 4.2 (A)). The degree of correlation has been efficiently decreased visually in comparison with the original reference model (Figure 4.2 (A)). Notice that the clouds of the high and low values are de-coagulated and the correlation length is significantly shorter which exhibits some trend, though not perfect, to the behaviors of a white noise. The reason for such imperfect is partly due to the underestimation of the amplitude of signal.

However, the post-process of MsBMcMC outputs can efficiently improve the estimate of the amplitude of signal. Figure 4.12 (C) plots the identified mean $\ln K$ distribution by selecting the first 100 realizations with top rankings amongst the original 3000 realizations. One can easily find that the magnitude of (C) is obviously greater than that of (A). Statistical analysis shows that the standard deviation of the former (0.54) is higher than that of the latter (0.45) and the mean remains unchanged (0.03). Figure 4.12 (D) plots the noise distribution by subtracting the identified signal from the system. The de-

aggregation of extremes is in a certain degree enhanced especially around the center of computation domain. As for the estimate of the phase of signal, the design of observation networks is mainly responsible for detecting the locations of the high and lower values.
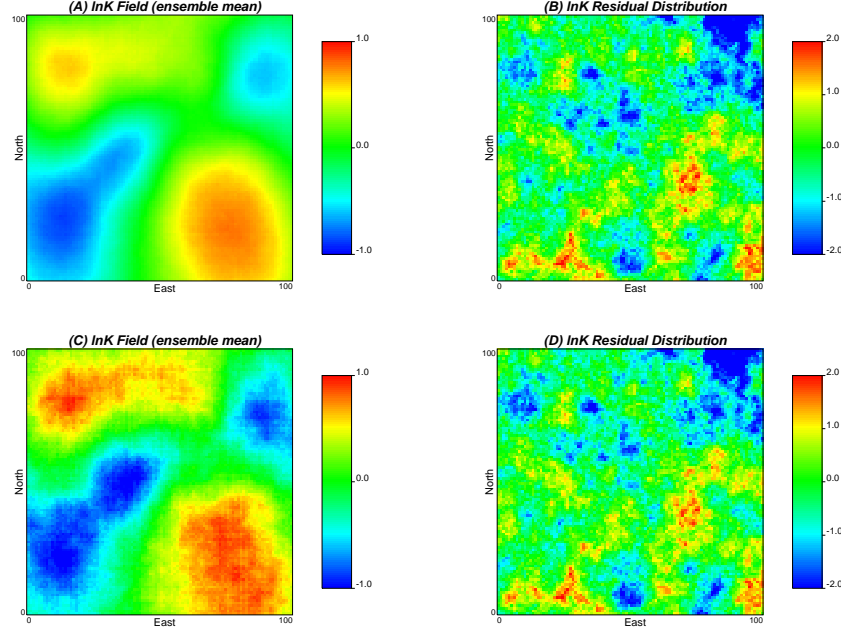
Figure 4.12: *Identified mean* $\ln K$ *fields and residual distributions* ($\lambda_{\boldsymbol{x}} = 50$, $\sigma_x^2 = 1.0$, $\sigma_y^2 = 0.17$). *Note that* (A) *and* (B) *are resulted from 3000 realizations, while* (C) *and* (D) *are based on 100 realizations.*

4.5   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 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 even for a high-dimensional case simply because the spectral decomposition of the covariance matrix is quickly done with aid of the fast Fourier transform. 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. Several important influence factors on the performance of the MsBMcMC have been widely examined on the basis of a synthetic example under the uniform, natural-gradient flow condition. Results show that the parameter configuration for models and the measurement errors of state data have important effects on the reproduction of reference models and computational efficiency.

Unlike the traditional inverse methods, one of the most important features of the proposed method is that the generated realizations have the specified spatial structure and parameter statistics as the user expected. The essential difference between the traditional inverse methods and the presented MsBMcMC method in handling with the model structure problem can be explained under the Bayesian framework. The MAP method, for example, typically assumes a prior model structure and generates an ensemble of initial seed fields conditional to the hard data. Then those realizations are subject to a model calibration procedure in order to inverse-condition on the dependent state data. Classically, a conditional and inverse-conditional realization is constructed by minimizing the objective function,

$$O = \frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{C_x}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}) + \frac{1}{2}(g(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{C_y}^{-1}(g(\boldsymbol{x}) - \boldsymbol{y}). \qquad (4.16)$$

During the procedure of model calibration, the initial seed field following $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C_x})$ is perturbed so as to match the state data and at the same time the parameter statistics and model structure have to be modified correspondingly, i.e., $\boldsymbol{x} \sim N(\boldsymbol{\mu}_m, \boldsymbol{C}_m)$. Indeed, the posterior mean and the posterior covariance matrix of the final individual realizations, in a linear approximation, are,

$$\boldsymbol{\mu}_m = \boldsymbol{\mu} + (\boldsymbol{J}\boldsymbol{C_y}\boldsymbol{J}^T + \boldsymbol{C_x})^{-1}\boldsymbol{J}^T\boldsymbol{C_y}(\boldsymbol{y} - g(\boldsymbol{\mu})), \qquad (4.17\text{a})$$

$$\boldsymbol{C}_m = (\boldsymbol{J}^T\boldsymbol{C_x}^{-1}\boldsymbol{J} + \boldsymbol{C_y}^{-1})^{-1}, \qquad (4.17\text{b})$$

where $\boldsymbol{J} = \frac{\partial O}{\partial \boldsymbol{x}}$ is the Jacobian matrix. The second formula means that the structure of conditional and inverse-conditional realizations, $\boldsymbol{C}_m$, depends not only on the prior specification for model parameters, $\boldsymbol{C_x}$, but also on the posterior observations of state variables, $\boldsymbol{C_y}$. This is the way that the MAP method views the parameter statistics and model structure of conditional realizations, which is also applicable to other sensitivity-based inverse methods, such as, the ML method (*Carrera and Neuman*, 1986), the pilot point method (*RamaRao et al.*, 1995), and the sequential self-calibration method (*Gomez-Hernandez et al.*, 1997). In summary, the modification on parameter statistics and model structure for matching the state data is not a requirement as *a priori* specified but a passive result from the model calibration which is inherent

to the algorithms. This fault is not the prior configuration for models but the posterior assignment for parameter values.

From the viewpoint of the blocking McMC method, however, the parameter statistics and model structure are specified *a priori* and remain unchanged in generating all of candidates, i.e.,

$$\boldsymbol{\mu}_m = \boldsymbol{\mu}, \tag{4.18a}$$

$$\boldsymbol{C}_m = \boldsymbol{C_x}. \tag{4.18b}$$

Then the McMC theory is invoked to select those candidates which are considered to inverse-condition on the state data under the scope of the observation errors. Therefore, the final realizations closely follow the prior specification on parameter statistics and model structure, i.e., $\boldsymbol{x} \sim N(\boldsymbol{\mu}_m, \boldsymbol{C}_m) = N(\boldsymbol{\mu}, \boldsymbol{C_x})$. In essence, the prior statistics for model parameters are independent from the observations on other parameters, including the dependent state variables. However, although the inference on parameter statistics and model structure may, in theory, benefit from the observations of relevant variables, it should be warned that the reliability of such structure analysis is always doubtful because of the inherent sparseness of relevant variables on which the inference making is based. The overconfidence on derivation of parameter statistics and model structure from highly incomplete information is unreliable and thus discouraged. Estimation on parameter statistics and model structure from state data is different than that from hard data. The former is useful but secondary for structure estimation while the latter is required especially for the prior configuration for models.

In addition, the proposed method has several other advantages over classical inverse methods: (1) Unlike those conventional approaches, the MsBM-cMC method is remarkably stable and well-behaved. It fully circumvents the difficulties of ill-posedness and identifiability as faced by the conventional sensitivity-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 widely 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 of the other Monte Carlo methods, the number of realizations to be generated can not be efficiently determined in order to

obtain a convergent result. Second, the efficiency of the proposed MsBMcMC method in exploring parameter space is desired to improve. The main discrepancy lies in the tradeoff between the acceptance rate of candidate models and the accuracy of state data. If the state data are considered to be highly confident, i.e., the measurement errors are extremely small, the exploring efficiency of the Markov chain will decreases drastically since the acceptance rate is dramatically low. Fortunately, this study finds that the high confident measurement errors make no significant difference for the estimation on the model parameters in the synthetic example problem. But this conclusion deserves more validation of real field applications.
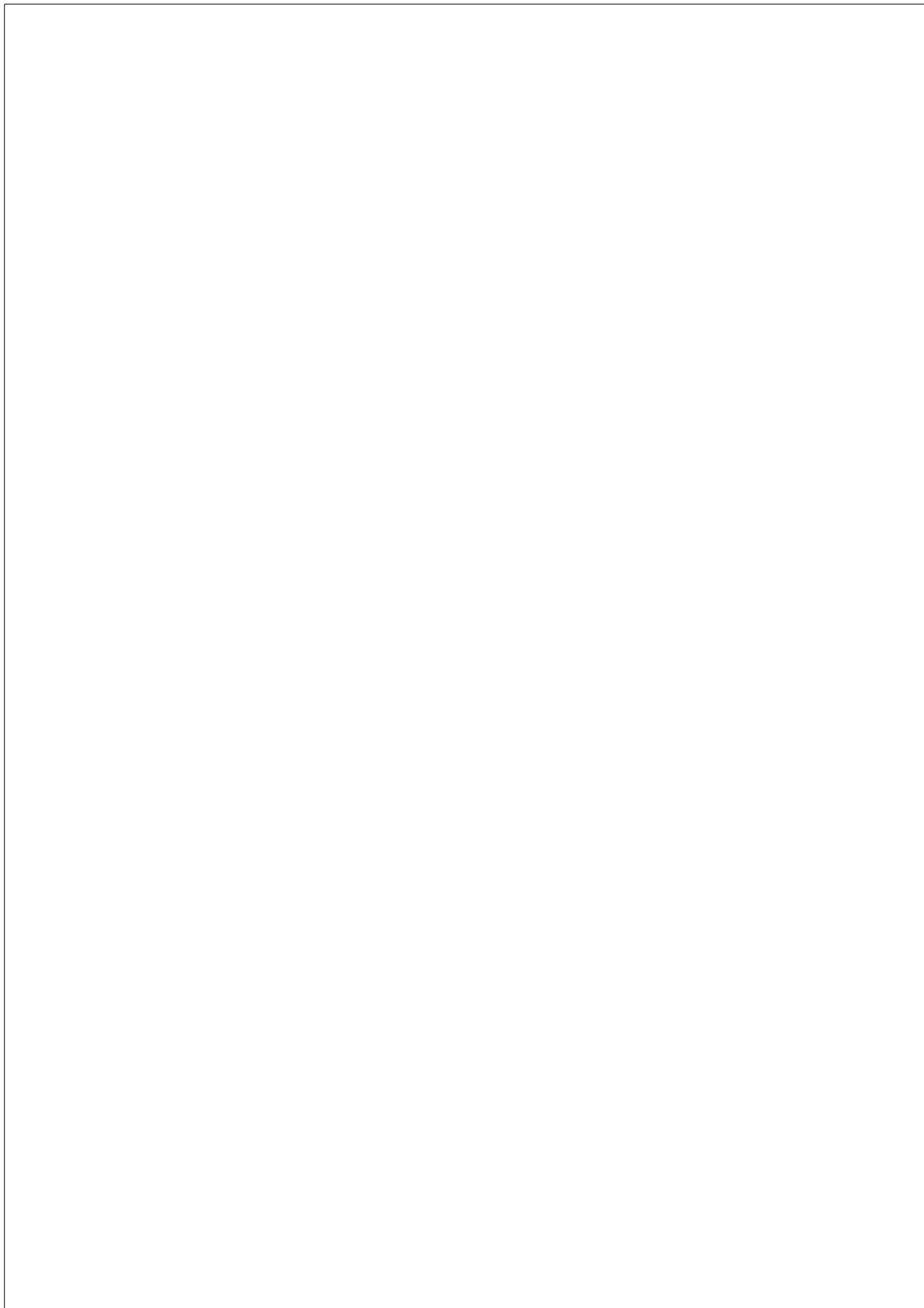
5

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

Abstract

A multi-scale blocking McMC (MsBMcMC) method has been presented in Chapter 4 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 MsBM-cMC 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.

5.1 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, which is well demonstrated in Chapter 4. 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). 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 dis-

tributed concentration data with 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 different knowledge about the flow and transport properties. High connectivity generally results in earlier breakthrough, i.e., an earlier front part of breakthrough curve. 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*, 1997). 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 4 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. It becomes more valuable in integrating nonlinear data since most often there are not sufficient high-resolution, conditional realizations available for the Monte Carlo assessment due to the computational expense of the inverse-conditional simulation. Note that the purpose of quantitative uncertainty analysis is not to reduce uncertainties which can only be achieved by collecting additional effective information. Additional effective information essentially reduces the uncertainty of model. 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.

5.2   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 models generated 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 uncertainties. 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.

5.2.1   Model uncertainty

Although the reference model is well defined and observable in this study, we generally do not know what it is in advance 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 $(I(\boldsymbol{x})_3)$ and the standard deviation of the ensemble average error $(I(\boldsymbol{x})_4)$, which are defined as the L_1-norm and L_2-norm between the simulated models and the mean models, i.e.,

$$I(\boldsymbol{x})_1 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r} - \bar{x}_i|, \qquad (5.1a)$$

$$I(\boldsymbol{x})_2^2 = \|\boldsymbol{x}_{sim} - \bar{\boldsymbol{x}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r} - \bar{x}_i)^2, \qquad (5.1b)$$

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,

$$I(\boldsymbol{x})_3 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |x_{i,r}^{sim} - x_i^{ref}|, \qquad (5.2a)$$

$$I(\boldsymbol{x})_4^2 = \|\boldsymbol{x}_{sim} - \boldsymbol{x}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (x_{i,r}^{sim} - x_i^{ref})^2, \qquad (5.2b)$$

respectively. Note that $\boldsymbol{x}_{ref}$ is the vector of reference attribute values. Obviously, the smaller $I(\boldsymbol{x})_1$ and $I(\boldsymbol{x})_2$ are, the closer to the real model the generated realizations are.

5.2.2   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,

$$I(\boldsymbol{y})_1 = \|\boldsymbol{y}_{sim} - \bar{\boldsymbol{y}}_{sim}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r} - \bar{y}_i|, \qquad (5.3a)$$

$$I(\boldsymbol{y})_2^2 = \|\boldsymbol{y}_{sim} - \bar{\boldsymbol{y}}_{sim}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r} - \bar{y}_i)^2, \qquad (5.3b)$$

$$I(\boldsymbol{y})_3 = \|\boldsymbol{y}_{sim} - \boldsymbol{y}_{ref}\|_1 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} |y_{i,r}^{sim} - y_i^{ref}|, \qquad (5.3c)$$

$$I(\boldsymbol{y})_4^2 = \|\boldsymbol{y}_{sim} - \boldsymbol{y}_{ref}\|_2^2 = \frac{1}{n_{xyz}} \sum_{i=0}^{n_{xyz}-1} \frac{1}{n_r} \sum_{r=0}^{n_r-1} (y_{i,r}^{sim} - y_i^{ref})^2, \qquad (5.3d)$$

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.

5.2.3   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 preserve 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),

$$A_L(x) = \frac{x}{2} \frac{\langle \sigma_t^2(x) \rangle + \langle \sigma_T^2(x) \rangle}{\langle T(x) \rangle^2}, \tag{5.4}$$

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,

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

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.,

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

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,

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

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),

$$A_L(x) = \frac{x}{2} \frac{\sigma_t^2(x)}{m_t^2(x)}, \tag{5.5}$$

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.

5.3   Synthetic Experiments and Results

The stochastic inverse-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.

5.3.1   Experiment configurations

The synthetic example as given in Chapter 4 is employed again to assess the uncertainties and the worth of diverse source data under a natural-gradient flow condition. The reference $\ln K$ model, the reference piezometric head field, and the corresponding conditioning data set are already displayed in Figure 4.2. 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 5.1 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 variations 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.

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 5.2. To ease the computational burden, the relative errors of head and travel time data are assumed to be rather large.

Table 5.1: *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

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 5.2: *Parameter configuration for stochastic simulations*

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

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 5.1 through 5.5, respectively. The importance of hydraulic conductivity, piezometric head, and travel time (moments) on reduction of model uncertainty and prediction uncertainty is analyzed as follows.

5.3.2 Model uncertainty

Figure 5.1 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 4.2(A) as the reference. Note that the legends used are different from each other. As expected, 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 5.2 shows the uncertainties of the estimates given in Figure 5.1. Conditioning on $\ln K$ measurements only makes the uncertainties of estimates 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 5.1 and 5.2 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 minor 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
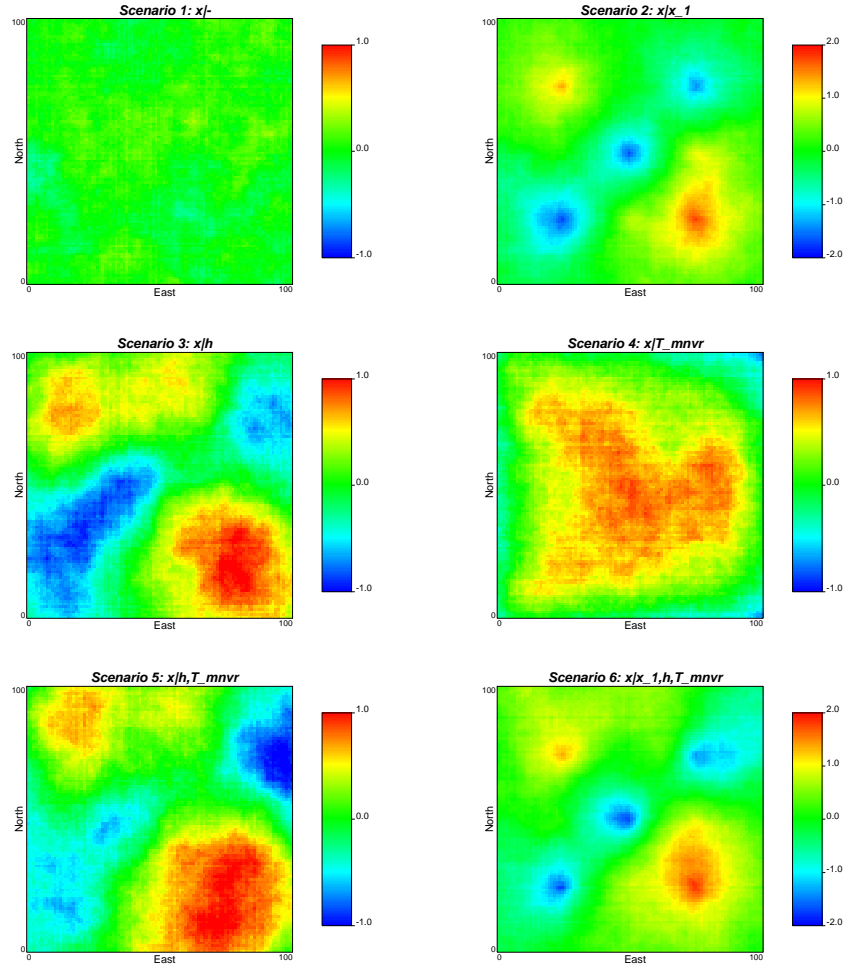
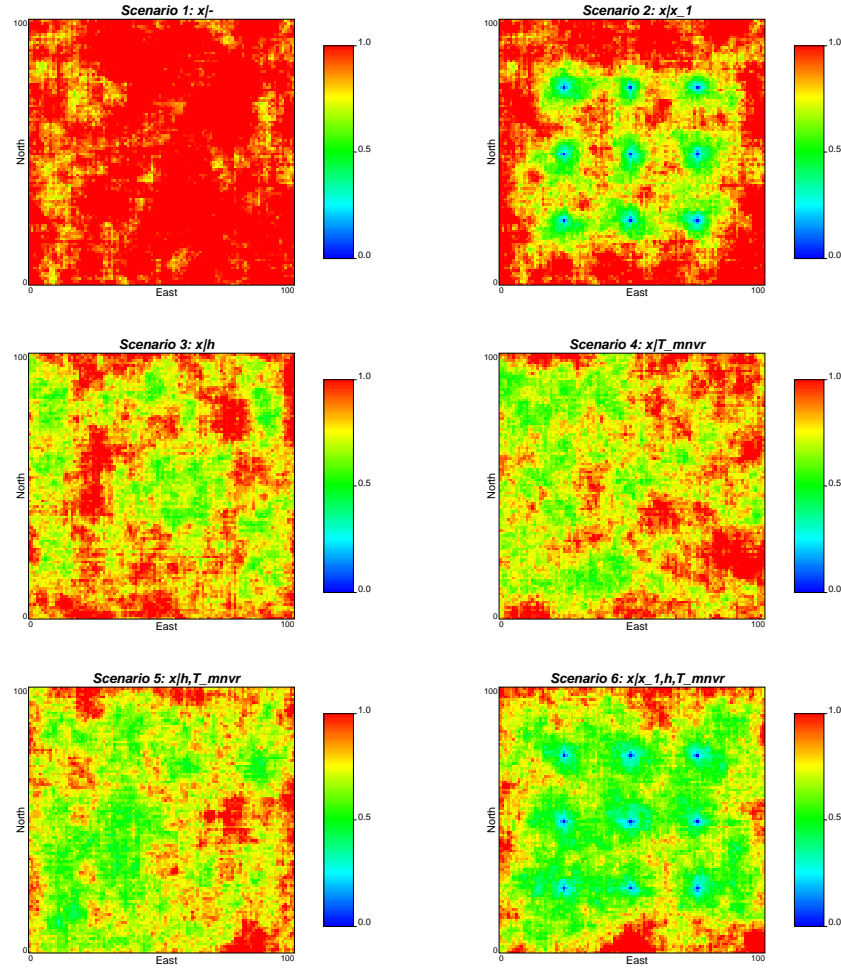Figure 5.1: *The mean* ln*K fields due to conditioning to various source data*

Figure 5.2: *The variances of* ln*K fields due to conditioning to various source data*

$\ln K$ field is highly correlated such that the model responses at the observation points is also "kriged" or propagated to the neighborhood.

Besides, from the Scenario 4 and 5 in Figure 5.2, 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 *Franssen et al.* (2003) contributed this phenomenon to more sensitivity of concentration to flow velocity in the upstream zone. This explanation seems unreasonable since it ignores the fact that the mean hydraulic gradient is uniform over the entire field in their case study. It never means that the velocity in the upstream is necessarily higher than that in the downstream. One reason is possibly due to the measurement error assignment to the travel times during the inverse-conditioning. Since an equal relative weight is imposed on the travel times for both the upstream and and downstream, the upstream part actually is set a smaller error for inverse-conditioning which yields a more accurate result than the downstream.

Table 5.3 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 while the collection of state data helps identify the spatial trend of physical models. From Table 5.3, 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 more 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.

In addition, correct prior configurations for spatial statistics and model structure also play important roles in reducing model uncertainty. Two sets of experiments have already been conducted in Chapter 4 to show the worth

Table 5.3: *Mean absolute error and mean variance of* $\ln K$

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

of model statistics and spatial structure: one is the variance of parameter variables and the other is the spatial correlation length. Comparing to the wrong specifications, the correct cases obviously render the better reproduction of the reference models (see Figure 4.4 and Figure 4.6). Figure 5.3 plots the histograms of $\ln K$ distributions before and after conditioning on diverse source data. 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 (e.g., Figure 1.5 in Chapter 1).

5.3.3 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 5.4 compares the predicted steady-state head distribution. The reference field has been depicted in Figure 4.2(C). 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.
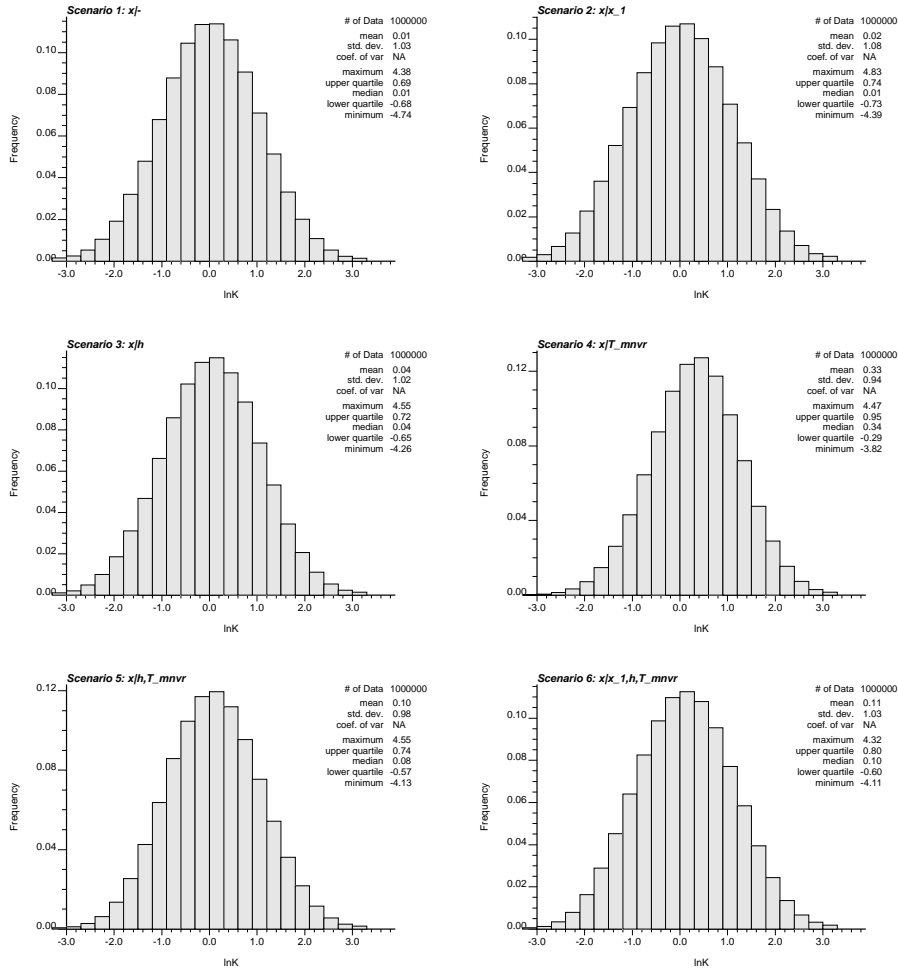
Figure 5.3: *The histogram of* ln*K distribution after conditioning to data from various sources*
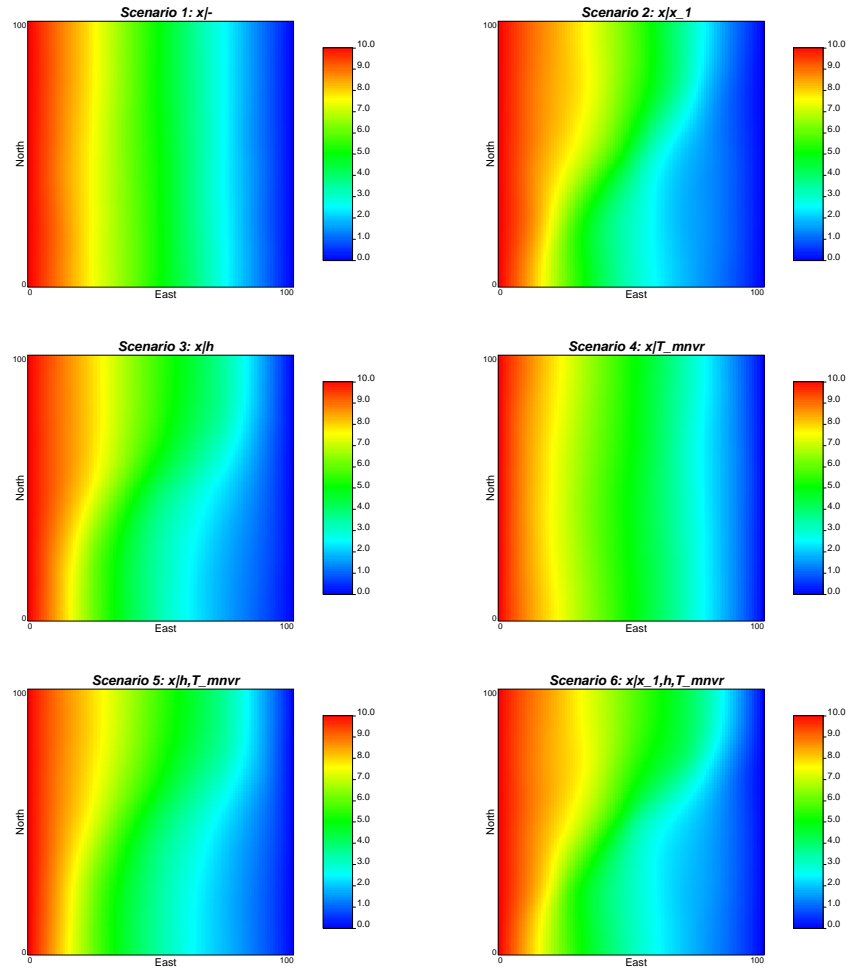
Figure 5.4: *The mean head fields due to conditioning to data from various sources*

Figure 5.5 plots prediction uncertainties of model responses. Conditioning merely on hard data (Scenario 2) and inverse-conditioning only to travel time moment 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.
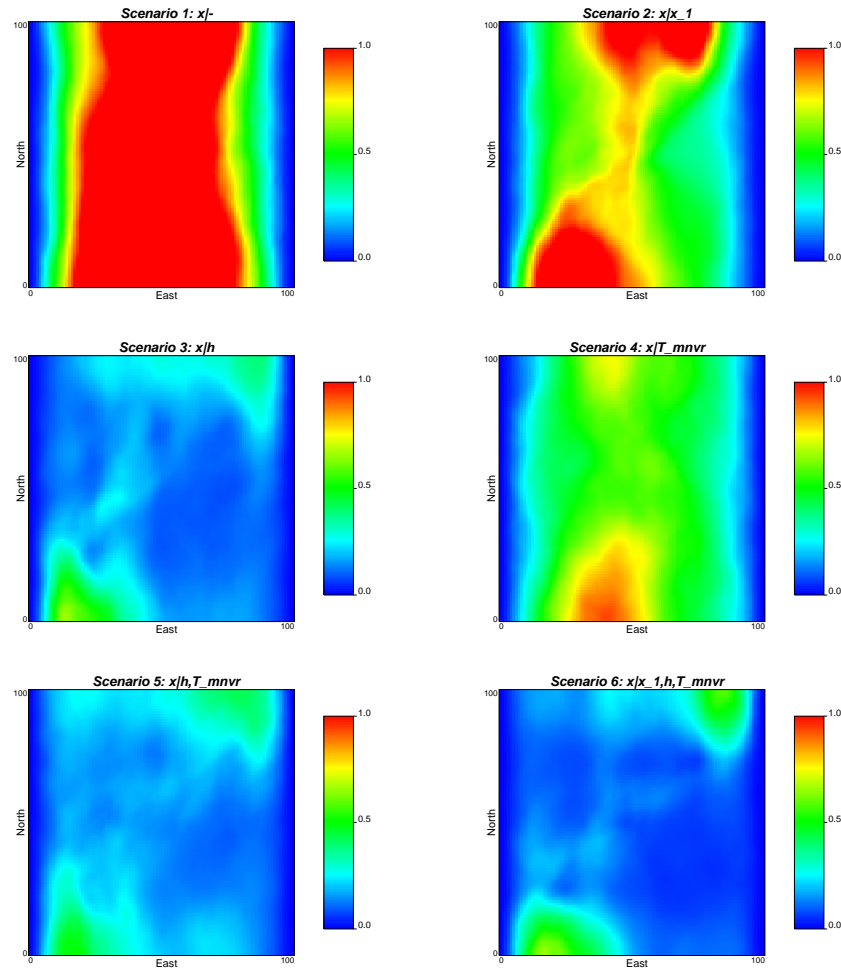
Figure 5.5: *The variances of head fields due to conditioning to data from various sources*

Table 5.4 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 uncer-

tainty 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 5.4: *Mean absolute error and mean variance of predicted head*

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

In addition, from Table 5.3 and Table 5.4, 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 $I(\boldsymbol{h})_2 < I(\boldsymbol{h})_4$.

5.3.4   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 assumption and hence some more general results are expected to be reached.

By setting various control planes in the x direction, the unconditional and conditional macrodispersion coefficients may be computed by Equation 5.4 or 5.5 from the temporal moments of the BTCs. Figure 5.6 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., $x/\lambda_{\boldsymbol{x}} \leq 2$.

Several observations may be drawn from the Figure 5.6. 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*
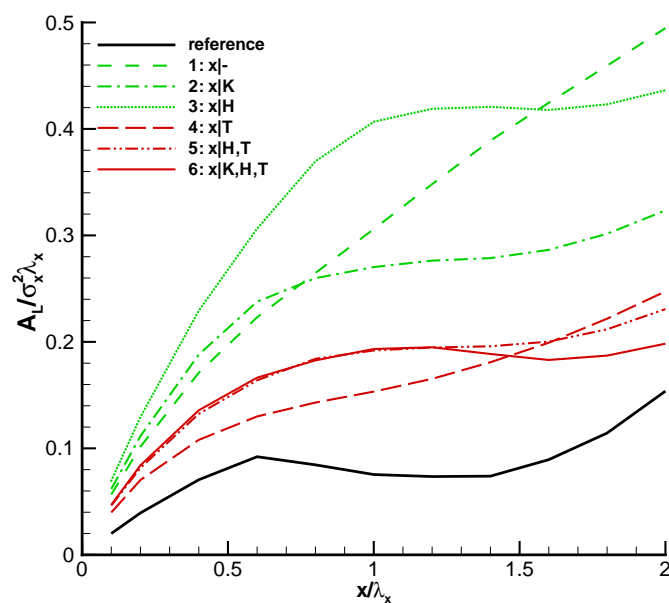
Figure 5.6: *A comparison of simulated macrodispersions due to conditioning to data from various sources*

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 5 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 is 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 5.3, Table 5.4, and Figure 5.6, one can find that conditioning data 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 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 5.6). 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 5.6. But the prediction of hs 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 that a contaminant plume arrives at a location of interest while $t_{95\%}$ estimates the operational time to remove most of contaminants from the aquifer. Figure 5.7 plots the evolution history of contaminants plume's front and tail along the mean plume trajectories before and after conditioning to diverse data.
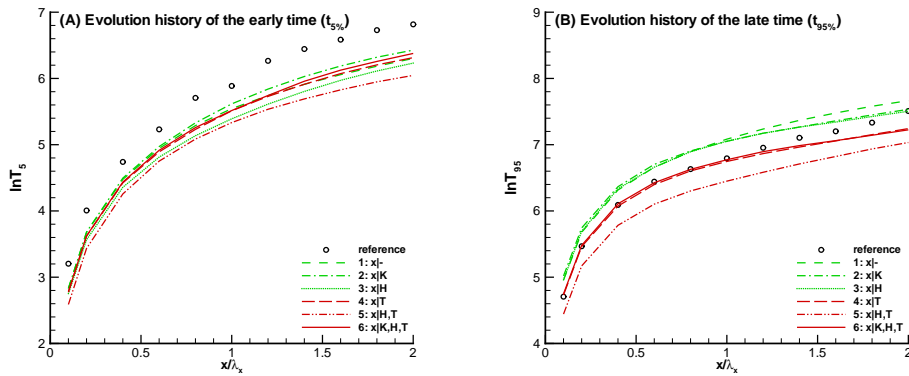
Figure 5.7: *A comparison on evolution history of contaminant plumes (A) at the early time $t_{5\%}$ and (B) at the late time $t_{95\%}$ due to conditioning to data from various sources*

Two observations from Figure 5.7 (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 the early arrival time, i.e., a conservative estimation is obtained. From Figure 5.7 (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, only is 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.

5.4 Conclusions and Discussions

The MsBMcMC method is an useful stochastic conditional and inverse-conditional simulation tool for quantitative uncertainty evaluation. 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 of $\ln K$ 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), which is beyond the scope of this article, most of

algorithms construct the objective function to be optimized 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 error and ignoring the other may not result in an improvement of prediction accuracy (*Warwick*, 1989; *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 error of a response from different sources, e.g., the model structure, the parameter estimation, and the inherent variation of aquifers. The ranking procedure can be achieved by examining the degree of correlation between the predicted response errors and the parameter errors. 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. 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.
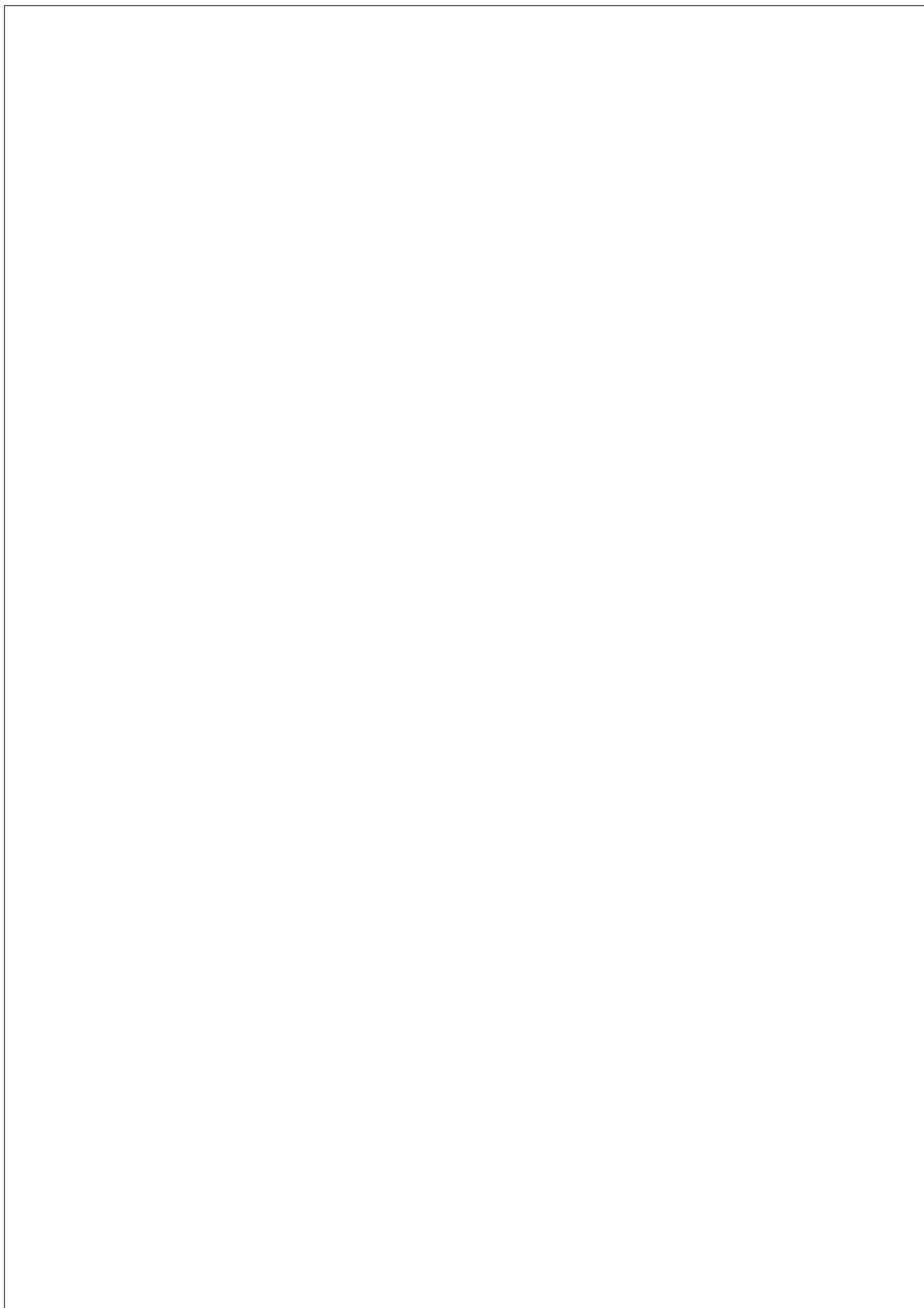
6

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.

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 1 reports some challenges that traditional inverse methods have to face but are hard to overcome.

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 burden 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. Chapter 3 details the proposed method and widely evaluates its performance and influence factors based on a synthetic example. The uncertainty assessment proves the usefulness of the presented method in performing conditional and inverse-conditional stochastic simulations.

Due to the limitation of the LU-decomposition of the covariance matrix, however, the implementation in Chapter 3 fails to cope with a high-dimensional case which is a basic requirement in high-performance reservoir simulation for both petroleum engineering and groundwater communities. The key problem lies in the computational velocity, which is twofold: one is the fast generation of the proposal kernel and the other is the fast computation of the likelihood. An FFT-based spectral conditional simulator, which is known as one of the fastest random field generator, is employed to attack the first problem. To the second problem, 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 4 gives a detailed description on the development of the proposed MsBMcMC method and its performance evaluation.

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 5 with 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.

6.2 Recommendations for Further Research

This study mainly focuses on the feasibility, the correctness and the performance evaluation 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 indirect geophysical measurements**, e.g., seismic data, ground-penetrating radar (GPR), 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 (*Rubin et al.*, 1992; *Hyndman et al.*, 1994; *Copty and Rubin*, 1995; *Hyndman and Gorelick*, 1996).

- **Extend to other parameter identification**, 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, radioactive decay coefficient λ (*Varni and Carrera*, 1998), 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.

- **Compare with a forced-gradient flow case**. The evaluation of the proposed McMC scheme in this study is totally based on synthetic examples under a uniform, natural-gradient flow condition. A radial, forced-gradient flow experiment may produce a distinct result.

- **Compare with other inverse methods**, e.g., real-time SSC, EnKF, MAP, etc. This study presents a stochastic simulation algorithm and its implementation for computationally intensive inverse problems. A special emphasis is put on preserving the spatial statistics and structure for models, but the importance on the uncertainty reduction and propagation has not been fully elaborated. A detailed comparison to other inverse approaches is expected through some real-world case studies.

- **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,

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

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, $\boldsymbol{x}_1 = \boldsymbol{x}_{obs} \subset R^m$, $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2)^T$, and $\boldsymbol{x}_2 \subset R^{n-m}$.

The objective is to draw randomly a large number of equi-probable realizations from the distribution $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{C_x})$. 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 $x \sim N(\mu, C_x)$, where $x, \mu \subset R^n$, and $C_x \subset R^{n \times n}$:

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

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

(3) Calculate $v = Lz \subset R^n$;

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

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., Lz, as in the step 3.

2. Conditional Sampler

The joint distribution of $x = (x_1, x_2)^T$ is,

$$\left(\begin{array}{c} x_1 \\ x_2 \end{array} \right) \sim N \left(\left(\begin{array}{c} \mu_1 \\ \mu_2 \end{array} \right), \left[\begin{array}{cc} C_{11} & C_{12} \\ C_{21} & C_{22} \end{array} \right] \right),$$

where $x_1 \subset R^m$ is the (normalized) conditioning dataset; $x_2 \subset R^{n-m}$ is the (normalized) conditional simulated values; $C_{11} \subset R^{m \times m}$ is the data-to-data covariance matrix; $C_{22} \subset R^{(n-m) \times (n-m)}$ is the unknowns-to-unknowns covariance matrix; and $C_{21} = C_{12}^T$ is the unknowns-to-data covariance matrix, $C_{21} \subset R^{(n-m) \times m}$. It can be shown that the expected value of x_2 is $\mu_2 + C_{21} C_{11}^{-1}(x_1 - \mu_1)$, which is known as the simple kriging estimate, and the covariance matrix of x_2 is $C_{22} - C_{21} C_{11}^{-1} C_{12}$. Therefore, the conditional realizations can be drawn from,

$$x_2 \sim N(\mu^*, C^*),$$

where,

$$\mu^* = \mu_2 + C_{21} C_{11}^{-1}(x_1 - \mu_1),$$

and,

$$C^* = C_{22} + C_{21} C_{11}^{-1} C_{12}.$$

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,

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = LU$$

$$= \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} U_{11} & U_{12} \\ 0 & U_{22} \end{bmatrix} = \begin{bmatrix} L_{11}U_{11} & L_{11}U_{12} \\ L_{21}U_{11} & L_{21}U_{12} + L_{22}U_{22} \end{bmatrix}.$$

Therefore, L_{11}, L_{21}, and L_{22} can be obtained by $C_{11} = L_{11}U_{11}$, $C_{21} = L_{21}U_{11}$, and $C_{22} = L_{21}U_{12} + L_{22}U_{22}$, 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 $\boldsymbol{z} \subset R^n$,

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

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.

128   *APPENDIX A. AN LU-DECOMPOSITION-BASED SAMPLER*

B

An FFT-based Spectral 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,

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

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

129

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,

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

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,

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

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 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.

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

(1) Construct the covariance matrix $\boldsymbol{C}$.

(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 $\boldsymbol{\epsilon} = \boldsymbol{\epsilon}_1 + i\boldsymbol{\epsilon}_2$, where $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2 \sim N(\boldsymbol{0}, \boldsymbol{1})$;

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

(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, s. This procedure can be done by FFT transforming of the sampled covariance sequence, i.e.,

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

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

(4) Randomly draw the phase ϕ by,

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

(5) Build up the Fourier coefficients a, and thus α and β, by,

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

(6) Generate the random field x by inverse FFT transforming a, e.g.,

$$x_k = F^{-1}(a) = \sum_{j=0}^{n-1} a_j e^{2\pi i \frac{k}{n} j}$$

$$= \sum_{j=0}^{n-1} \left[\alpha_j \cos\left(2\pi \frac{k}{n} j\right) + i\beta_j \sin\left(2\pi \frac{k}{n} j\right) \right]; k \in [0, n).$$

(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 FFT 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 $\boldsymbol{x}_K^*$.

C

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,

$$\boldsymbol{\psi}_{t,r} = \left[\begin{array}{c} \boldsymbol{x}_{t,r} \\ \boldsymbol{y}_{t,r} \end{array} \right] ; r \in [0, n_r), \tag{C.1}$$

where $\boldsymbol{x}_t = (\boldsymbol{x}_{t,r} \subset R^{n_{xyz}}; r \in [0, n_r))^T$ 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,

$$\boldsymbol{\psi}_{0,r} = \left[\begin{array}{c} \boldsymbol{x}_{0,r} \\ \boldsymbol{y}_0 \end{array} \right] ; r \in [0, n_r),$$

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 conditional 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,

$$\boldsymbol{\psi} = (\boldsymbol{\psi}_0, \boldsymbol{\psi}_1, ..., \boldsymbol{\psi}_t, ...)^T . \tag{C.2}$$

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,

$$\boldsymbol{\psi}_{1,r}^{sim} = \left[\begin{array}{c} \boldsymbol{x}_{0,r} \\ \boldsymbol{y}_{1,r}^{sim} \end{array} \right] ; r \in [0, n_r).$$

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 as $\boldsymbol{x}_1$ 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) in the Bayesian framework. Starting from the time $t = 1$ when the nonlinear dynamic observations are available, the joint vector is updated by,

$$\boldsymbol{\psi}_{t,r} = \boldsymbol{\psi}_{t,r}^{sim} + \boldsymbol{G}_t \left(\boldsymbol{y}_t^{obs} + \boldsymbol{\epsilon}_t - \boldsymbol{H}_t \boldsymbol{\psi}_{t,r}^{sim} \right) ,$$
$$= \boldsymbol{\psi}_{t,r}^{sim} + \boldsymbol{G}_t \left(\boldsymbol{y}_t^{obs} + \boldsymbol{\epsilon}_t - \boldsymbol{y}_{t,r}^{sim} \right) , \tag{C.3}$$

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,

$$\boldsymbol{G}_t = \boldsymbol{C}_{\boldsymbol{\psi},t}^{sim} \boldsymbol{H}_t^T \left(\boldsymbol{H}_t \boldsymbol{C}_{\boldsymbol{\psi},t}^{sim} \boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t} \right)^{-1} , \tag{C.4}$$

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,

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

$$= \frac{1}{n_r - 1}\sum_{r=0}^{n_r-1}\left(\boldsymbol{\psi}_{t,r}^{sim} - \bar{\boldsymbol{\psi}}_t^{sim}\right)\left(\boldsymbol{\psi}_{t,r}^{sim} - \bar{\boldsymbol{\psi}}_t^{sim}\right)^T, \qquad \text{(C.5)}$$

where $\boldsymbol{\psi}_t^{sim} \subset R^{(n_{xyz}+n_{d,t})\times n_r}$ is a vector which consists of a 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,

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

where $\boldsymbol{0} \subset R^{n_{d,t}\times n_{xyz}}$ and $\boldsymbol{1} \subset R^{n_{d,t}\times n_{d,t}}$; $\boldsymbol{C}_{\boldsymbol{y},t} \subset R^{n_{d,t}\times n_{d,t}}$ 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. The components of $\boldsymbol{C}_{\boldsymbol{y},t}$ are randomly drawn from a normal distribution with a zero mean and a small deviation, e.g., 2×10^{-16}. The introduction of this random noise may also help maintain the matrix $\boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T + \boldsymbol{C}_{\boldsymbol{y},t}$ positive-defined so that it can be directly inverted numerically, which consists of a key part of this method. 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,

$$\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T = \boldsymbol{C}_{\boldsymbol{\psi y},t}^{sim},$$
$$\boldsymbol{H}_t\boldsymbol{C}_{\boldsymbol{\psi},t}^{sim}\boldsymbol{H}_t^T = \boldsymbol{C}_{\boldsymbol{yy},t}^{sim},$$

where $\boldsymbol{y}_{t,r}^{sim}$ is the simulated dynamic response for the realization r; $\boldsymbol{C}_{\boldsymbol{\psi y},t}^{sim}$ denotes the simulated covariance between the joint vector $\boldsymbol{\psi}$ and the response

$\boldsymbol{y}$; and $\boldsymbol{C}_{\boldsymbol{yy},t}^{sim}$ 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 eigenvalue decomposition from the matrix, i.e.,

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

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.

Therefore, the EnKF for history matching is essentially a Monte Carlo 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 spread. The prediction of the estimate and uncertainty is performed by integrating the ensemble of reservoir models.

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 susums, *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 non-aqueous 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. Marño, 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.

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.