Adaptive Error Modelling in MCMC Sampling for Large Scale Inverse Problems

by

T. Cui, C. Fox and M. J. O'Sullivan

Report, Univeristy of Auckland, Faculty of Engineering, no. 687

ISSN 1178-360

Adaptive Error Modelling in MCMC sampling for Large Scale Inverse Problems

T. Cui

The University of Auckland, Auckland, New Zealand.

C. Fox

University of Otago, Dunedin, New Zealand.

M. J. O'Sullivan

The University of Auckland, Auckland, New Zealand.

Summary. We present a new adaptive delayed-acceptance Metropolis-Hastings algorithm (ADAMH) that adapts to the error in a reduced order model to enable efficient sampling from the posterior distribution arising in complex inverse problems. This use of adaptivity differs from existing algorithms that tune proposals of the Metropolis-Hastings algorithm (MH), though ADAMH also implements that strategy. We build on the recent simplified conditions given by Roberts and Rosenthal (2007) to give practical constructions that are provably convergent to the correct target distribution. The main components of ADAMH are the delayed acceptance scheme of Christen and Fox (2005), the enhanced error model introduced by Kaipio and Somersalo (2007) as well as recent advances in adaptive MCMC (Haario et al., 2001; Roberts and Rosenthal, 2007). We developed this algorithm for automatic calibration of large-scale numerical models of geothermal reservoirs. ADAMH shows good computational and statistical efficiencies on measured data sets. This algorithm could allow significant improvement in computational efficiency when implementing sample-based inference in other large-scale inverse problems.

Keywords: adaptive MCMC, delayed acceptance, reduced order models, enhanced error model, geothermal reservoir modelling, inverse problem

1. Introduction

1.1. Overview

We present an adaptive delayed acceptance Metropolis-Hastings algorithm (ADAMH) to enable efficient sampling from the computationally expensive posterior distributions that occur in large scale inverse problems. Given a physical system, the inverse problem commonly arises from reconstructing some quantities of interest from measurable features of the system which indirectly relates to the quantities of interest. Formally, a forward model is used to describe the physical system, and the quantities of interest are parametrized by model parameters. Hence the measurable features of the system can be predicted by the forward model given a set of model parameters. The corresponding inverse problem consists of inferring the model parameters from a set of field measurements of those features.

The unknowns of the interested physical system are usually spatially distributed quantities, e.g. permeability distribution of subsurface fluid transportation model (Higdon et al., 2003; Cui, 2010), resistivity/conductivity field of impedance imaging (Vauhkonen et al., 1999; Andersen et al., 2004; Watzenig and Fox, 2009), and flow field of ocean circulation (McKeague et al., 2005). These spatially distributed parameters can be considered

in the form of images, and hence techniques developed in statistical image analysis, such as Grenander (1967, 1976, 1981); Geman and Geman (1984); Besag (1986); Grenander and Miller (1994); Besag et al. (1995); Hurn et al. (2003), can be applied to solve inverse problems. In particular, we formulate the inverse problem into the Bayesian inferential framework, which offers a coherent and rigorous foundation for solving the inverse problem by accounting various sources of uncertainties in the data modelling process, and incorporating prior information. Based on the posterior distribution, answers to the inverse problem, such as parameter estimation, uncertainty assessment, and model prediction are calculated from the mean, modes, or higher oder moments of statistics of interest over the posterior.

Historically, Bayesian image analysis put a great emphasis on prior modelling of the image, extensive investigations have been carried. These include Gaussian Markov random field models (Besag, 1986; Kuensch, 1987; Besag et al., 1991; Besag and Kooperberg, 1995; Higdon, 1998; Rue and Held, 2005) and Gaussian process models (Cressie, 1993) that have been developed for low-level pixel based representation, intermediate-level modelling that based on continuous random partition of a plane (Arak et al., 1993; Green, 1995; Moller and Plenge Waagepetersen, 1998; Nicholls, 1998; Moller and Skare, 2001), and object based high-level modelling such as deformable template approaches of Amit et al. (1991); Baddeley and van Lieshout (1993); Grenander and Miller (1994); Jain et al. (1998); Rue and Hurn (1999). In comparison with the effort put into prior modelling, the data modelling is a rather neglected area in Bayesian image analysis. However, this is crucial to the credibility of the inference results in physical applications, because constructing the likelihood function requires a forward model that accurately describes the map from the image to the measurements. Some works on using physical models in the Bayesian image analysis can be found in impedance imaging (Nicholls and Fox, 1998; Andersen et al., 2004), ultrasound imaging (Husby et al., 2001), and emission tomography (Green, 1990; Higdon et al., 1997; Weir, 1997). The resulting high dimensional posterior distribution are sampled by Markov chain Monte Carlo (MCMC), and then the estimations can be calculated by Monte Carlo integration over samples.

In real applications of the inverse problem, the forward model usually consists of a set of non-linear governing partial differential equations (PDE) and its numerical implementation. Difficulties on applying MCMC sampling to inverse problems arise from the high dimensionality of the model parameters, and computational demands of numerical simulations of forward models. Unlike most of the problems in Bayesian image analysis, where the high dimensionality is not necessarily an issue because the posterior distributions are trivial and millions of MCMC iterations can be carried out. Each of the forward model simulation of a typical inverse problem requires the order of minutes to hours of CPU time to evaluate. In practice, it is computationally prohibitive to run MCMC for many number of iterations. Hence the feasibility of applying MCMC to solve inverse problems is greatly constrained by the computing time of the forward model. Even though there are few comprehensive studies on posterior sampling of large scale inverse problems, include the work on retrieving ocean circulation field (McKeague et al., 2005), retrieving fields of wind vectors (Cornford et al., 2004), recovering atmosphere gas density (Haario et al., 2004), and characterising subsurface properties of aquifers and petroleum reservoirs (Oliver et al., 1997; Higdon et al., 2002, 2003; Mondal et al., 2010).

The massive scale of computation in each of these examples indicate that considerable improvement of sampling efficiency of MCMC algorithms for inverse problems is necessary. Recently, several investigations on the improvement of sampling efficiency have been carried

out, includes the adaptive Metropolis algorithm (AM) of Haario et al. (2001), the delayed rejection AM algorithm (DRAM) of Haario et al. (2006), the differential evolution Monte Carlo algorithm (DEMC) developed by Ter Braak (2006), and many other adaptive MCMC algorithms (e.g., Atchade and Rosenthal, 2005; Andrieu and Moulines, 2006; Roberts and Rosenthal, 2009). These algorithms focus on optimizing the proposal distributions of MH to efficiently traverse the parameter space. For computationally fast models that require in the order of one minute to simulate, the above algorithms can adequately explore the posterior distribution in the order of one hundred thousands iterations within a week of computing time. However, it is still computationally prohibitive to apply MCMC for computationally expensive models that may requires hours to simulate.

The ADAMH algorithm we presented here employs a computationally fast approximate posterior distribution (also referred to as approximation) to reduce the computing time per iteration of the MCMC sampling. The approximation is based on a reduced order model (ROM) that approximates the behavior of the accurate forward model, and adaptively tuned by ADAMH to accommodate the model reduction error. This use of adaptivity improves the accuracy of the approximation on-line, and differs from existing adaptive MCMC algorithms (Haario et al., 2001; Atchade and Rosenthal, 2005; Andrieu and Moulines, 2006; Roberts and Rosenthal, 2007, 2009) that tune random walk proposals, though this strategy is also implemented by ADAMH.

The ROM plays a central role in numerical analysis on reducing the computational cost of the numerical models, while retain the basic features of the original numerical model. There are many possible ways to construct a ROM, common approaches include coarsening the grid structure of the numerical model (e.g., Christie and Blunt, 2001; Kaipio and Somersalo, 2007), linearizing the forward model (e.g., Christen and Fox, 2005), and projection methods (e.g., Grimme, 1997; Li, 2002). Computer experimental design community have also applied Gaussian process models to construct emulators of the numerical models (Sacks et al., 1989; Welch et al., 1992; Morris et al., 1993; Kennedy and O'Hagan, 2001; Santner et al., 2003; Bayarri et al., 2007; Higdon et al., 2008; Gramacy and Lee, 2008). However, this approach is constrained by the dimensionality of the parameter space, and usually have difficulties in dealing with the extremely high dimensional parameters in inverse problems. In analysing the geothermal reservoir modelling problem, we construct the ROM by coarsening the grid structure used by the forward model, see Section 4.

ROM has been employed by several researchers to perform fast MCMC sampling. For example, Higdon et al. (2003) developed a Metropolis coupled MCMC scheme that simultaneously runs chains for models that have different levels of grid resolution. Information from the faster running coarse formulations speed-up the mixing of the finest scale chain, from which samples are taken. Lieberman et al. (2010) employed a projection based ROM to construct an approximate posterior density on a reduced parameter space, and then samples are drawn from the approximate posterior distribution directly. ADAMH aims at sampling directly from the exact posterior distribution, and uses the ROM in a different way, within the framework of the delayed acceptance Metropolis-Hastings algorithm (DAMH) of Christen and Fox (2005).

In the first step of DAMH, an approximation is used to pre-compute the acceptance probability of proposals as in MH, then a second step evaluation of the exact posterior density only occurs for the proposals accepted in the first step. Here, a modified acceptance probability is used in the second step for ensuring the ergodicity of the Markov chain. This two step MCMC scheme is very similar to the "surrogate transition" method (e.g., Liu, 2001, Section 9.4.3). However, the "surrogate transition" method only allows the use of a state

independent approximate posterior distribution, while one of the most important feature of DAMH is that it can deal with a more general form of approximation that includes both the state-dependent and state-independent cases. The extensive numerical experiments in Cui (2010) show that using a properly designed state-dependent approximation results in a significant improvement in the sampling efficiency compared to state-independent approximations.

The main disadvantage of the ROM is that it usually has a non-negligible discrepancy with the accurate model. In building the approximate posterior distribution, ignorance of the statistics of this model reduction error would result in a biased estimation. In DAMH, this would causes a very low acceptance rate in the second step, and hence a poorly mixed Markov chain. ADAMH implements a modified version of the enhanced error model (EEM) of Kaipio and Somersalo (2007) to estimate the model reduction error, and allows for constructing the approximation that adapts to the model reduction error.

We first validate ADAMH on a 1D homogeneous geothermal reservoir model with synthetic data. Then, ADAMH is applied to the estimation of the heterogeneous and anisotropic permeability distribution and the heterogeneous boundary conditions of a 3D steady state model of a two phase geothermal reservoir. There are about 10^4 parameters in our model, and each model simulation requires about 30 to 50 minutes of CPU time, which makes it virtually impossible for the standard MH algorithm to be applied. ADAMH shows a great enhancement in the sampling efficiency, and achieves a speed-up factor about 7.7 compare to the standard MH. We are able to run 11,200 iterations in about 40 days. The sampling results show good agreement between the estimated temperature profiles and the measured data. We expect that ADAMH will produce significant improvement in computational efficiency when applied to sample-based inference in other large scale inverse problems.

1.2. Outline of this paper

Section 2 provides a brief review of the Bayesian inference for inverse problems and the associated computational issues, including a discussion of DAMH and EEM. Section 3 reviews adaptive MCMC algorithms of Haario et al. (2001); Roberts and Rosenthal (2009), and presents the ADAMH algorithm and several adaptive approximations. Section 4 presents two case studies of ADAMH on calibrating two geothermal reservoir models. Section 5 offers some conclusions and discussions.

2. Bayesian inverse problems and computation

2.1. Bayesian formulation of inverse problems

Suppose a physical system is represented by a forward model $F(\cdot)$, and the unknowns of the physical system are parameterized by model parameters $\mathbf{x}$. Given a set of model parameters, the forward model produces outputs $\mathbf{d} = F(\mathbf{x})$, which represents the measurable features of the system. In solving a inverse problem, we wish to recover unknown $\mathbf{x}$ from measurements $\tilde{\mathbf{d}}$, and to make future prediction about the output of the physical system.

Because the parameters of interest such as permeabilities are usually spatially distributed and highly heterogeneous, and field data are only able to be sparely measured, recovering the parameters from field data is an ill-posed inverse problem (Hadamard, 1902; Kaipio and Somersalo, 2004). Ill-posedness means that there exist a range of feasible parameters that are consistent with the measured data, and hence a range of possible model predictions.

In a Bayesian framework, this consistency is quantified by the posterior distribution which provides the relative probability of a set of parameters being correct. Thus, the assessment of parameters and model predictions can be performed by summarizing information over the posterior distribution.

In oder to formulate the posterior distribution, it is necessary to build the prior model of unknown parameters and the likelihood function. Formulating the prior distribution $\pi(\mathbf{x})$ requires stochastic modelling of the unknowns. This is typically derived from expert knowledge of allowable parameter values, previous measurements, modelling of processes that produce the unknowns, or a combination of these. The parametrization of unknowns is a composite part of prior modelling, since expressing certain types of knowledge is simpler in some representations than others, and solutions that cannot be represented are excluded. For inverse problems with spatially distributed parameters, representations and prior distributions are usefully drawn from spatial statistics developed in Bayesian image analysis. Hurn et al. (2003) provides a comprehensive review about this topic.

In the geothermal reservoir modelling problem we presented here, a non-informative prior is used in the 1D homogeneous model with synthetic data. For the 3D model with heterogeneous and spatially distributed parameters, a low-level pixel based representation and Gaussian Markov random field prior (Rue and Held, 2005) are employed to parametrize the permeabilities. A separate radial basis function (RBF) model is used to represent the unknown boundary conditions.

The likelihood function $L(\tilde{\mathbf{d}} \mid \mathbf{x})$ is constructed by given the probabilistic model of various uncertainties associated with the data modelling process. The commonly used stochastic relationship between the measurements and the model parameters is

$$\tilde{\mathbf{d}} = F(\mathbf{x}) + \mathbf{e}, \tag{1}$$

where the vector $\mathbf{e}$ represents random noise in the data modelling process, and the forward model $F(\mathbf{x})$ is usually deterministic. The noise vector $\mathbf{e}$ probes both the measurement noise and uncertainties that arise from model bias between $F(\mathbf{x})$ and the underlying physical system. Such model bias may be caused by numerical error in computer implementation of the forward model, spatial discretization of the unknown parameters, and inappropriate assumptions in the mathematical model, etc. Many classical literatures of inverse problems consider the forward model is perfect in representing the underlying system, and only assume the existence of the measurement noise. However, when the model bias is more significant than the measurement noise (which is common in practice), ignorance of such model bias would results an imprecise likelihood function, and the inversion results often have biased estimates and over-tighten uncertainty intervals. It is worth mentioning that Kennedy and O'Hagan (2001) offer a framework for analysing the model biases of complex computer code outputs, which requires multiple sets of data are measured from experiments with various controllable model inputs. Unfortunately, there is only one set of data comes from a single experiment in most of the inverse problems, and hence it is unclear how to apply the framework of Kennedy and O'Hagan (2001) in such case. As discussed by Higdon et al. (2003), it may not possible to separate the measurement noise and model bias in case that only single set of data is available. Thus, it is necessary to incorporate the modeller's judgments about the appropriate size and nature of the noise term $\mathbf{e}$.

For the geothermal reservoir models, a Gaussian noise with known variance is used to generate the synthetic data set in the 1D homogeneous test case. In the 3D problem with measured data set, since the data are reasonably sparse and measured in steady state, we

assume that the noise $\mathbf{e}$ follow a zero mean independent and identically distributed Gaussian distribution. The value of the variance is suggested by manual calibration results and trial runs of MCMC sampling, and the posterior analysis shows that this simplified assumption is appropriate. Therefore, the likelihood function used in this research has the form

$$L(\tilde{\mathbf{d}}|\mathbf{x}) \propto \exp\left\{-\frac{1}{2}[F(\mathbf{x}) - \tilde{\mathbf{d}}]^T \mathbf{\Sigma}_\mathbf{e}^{-1}[F(\mathbf{x}) - \tilde{\mathbf{d}}]\right\}, \tag{2}$$

where $\mathbf{\Sigma}_\mathbf{e}$ is the covariance matrix of the noise vector $\mathbf{e}$.

Following the Bayes' rule, the unnormalized posterior distribution is given as the product of likelihood function and prior distribution:

$$\pi(\mathbf{x}|\tilde{\mathbf{d}}) \propto L(\tilde{\mathbf{d}}|\mathbf{x})\pi(\mathbf{x}). \tag{3}$$

2.2. Computation

In principle, the posterior distribution can be explored by MCMC method, the Metropolis-Hastings algorithm (MH) (Metropolis et al., 1953; Hastings, 1970) and its extensions (e.g., Green, 1995) in particular. Then, robust model predictions and uncertainty quantification are calculated as expectations of desired quantities over the posterior distribution by Monte Carlo integration. In recent years, this method has been applied to various field of inverse problems as summarized in Section 1. However, there are several important features about the posterior distribution that would make it difficult to be explored:

(a) The spatially distributed parameters $\mathbf{x}$ are usually high dimensional, from the order of ten to ten thousands, depends on the parametrization and grid resolution.
(b) Evaluation of the posterior distribution is computationally expensive, because simulating the forward model $F(\cdot)$ usually involves computationally demanding numerical schemes such as finite element and finite difference methods.
(c) Since the governing equations of the forward map is usually a set of non-linear PDEs, the resulting posterior may be highly non-linear and has strong spatial interactions.

For example, the 3D geothermal reservoir model presented in Section 4 has about ten thousands parameters, each model evaluation takes about 30 to 50 minutes CPU time, and the governing PDEs is highly non-linear (phase changing and table look-up are involved).

These features render that there are two crucial issues in applying MCMC method to practical inverse problem, includes: (i) designing proposal distributions that is able to efficiently traverse the support of the posterior, and (ii) lowering the computational cost per iteration of MH to make the MCMC sampling being computationally feasible. Formally, the former is a problem on improving the *statistical efficiency*, i.e., using less number of MCMC iterations to generate statistically independent samples; and the latter requires enhancement in the *computational efficiency*, i.e., using less CPU time to generate statistically independent samples.

By using adaptivity or multiple chains to optimize the scale and orientation of the proposal distribution, recent algorithms such as AM, DRAM, the adaptive Metropolis within Gibbs algorithm (AMWG) of Roberts and Rosenthal (2009), and DEMC demonstrate statistical efficiencies in several high dimensional sampling problems (Haario et al., 2004; Turro et al., 2007; Vrugt et al., 2008). Also, these algorithms avoid the expensive (in terms of computing and human input) tuning process by employing adaptivity or multiple chains.

However it is still infeasible to apply these algorithms directly to computationally demanding models that requires hours to simulate. To improve the computational efficiency, the ADAMH algorithm we present here uses the idea of delayed acceptance (Christen and Fox, 2005) to reduce the computational cost per iteration.

2.3. Delayed acceptance Metropolis-Hastings algorithm

Suppose we have a computationally expensive target distribution $\pi(\cdot)$, and a computationally fast approximation $\pi^*_{\mathbf{x}}(\cdot)$ (can be state dependent or state independent) of $\pi(\cdot)$. Consider a proposal distribution $q(\mathbf{x}, \mathbf{y})$. DAMH first "corrects" the proposal with the approximation $\pi^*_{\mathbf{x}}(\mathbf{y})$ to create a modified proposal distribution $q^*(\mathbf{x}, \mathbf{y})$, to be used in a standard MH. If the modified proposal has a high acceptance rate, DAMH gains computational efficiency by avoiding calculating $\pi(\mathbf{y})$ when proposals are rejected by $\pi^*_{\mathbf{x}}(\mathbf{y})$. Following the description of of Christen and Fox (2005), one iteration of the process is given by:

ALGORITHM 1. *Delayed acceptance Metropolis-Hastings*
At step n, given $\mathbf{X}_n = \mathbf{x}$, then $\mathbf{X}_{n+1}$ is determined in the following way:

(a) Generate a proposal $\mathbf{y}$ from the proposal distribution $q(\mathbf{x}, \cdot)$.
(b) Let

$$\alpha(\mathbf{x}, \mathbf{y}) = \min\left[1, \frac{\pi^*_{\mathbf{x}}(\mathbf{y})q(\mathbf{y}, \mathbf{x})}{\pi^*_{\mathbf{x}}(\mathbf{x})q(\mathbf{x}, \mathbf{y})}\right].$$

With probability $\alpha(\mathbf{x}, \mathbf{y})$, accept $\mathbf{y}$ to be used as a proposal for the standard Metropolis–Hastings algorithm. Otherwise use $\mathbf{y} = \mathbf{x}$ as a proposal. The actual proposal distribution used is

$$q^*(\mathbf{x}, \mathbf{y}) = \alpha(\mathbf{x}, \mathbf{y})q(\mathbf{x}, \mathbf{y}) + (1 - r(\mathbf{x}))\delta_{\mathbf{x}}(\mathbf{y})$$

where $r(\mathbf{x}) = \int_{\mathcal{X}} \alpha(\mathbf{x}, \mathbf{y})q(\mathbf{x}, \mathbf{y})\mathbf{dy}$ and $\delta_{\mathbf{x}}(\cdot)$ denotes the Dirac mass at $\mathbf{x}$.
(c) Let

$$\beta(\mathbf{x}, \mathbf{y}) = \min\left[1, \frac{\pi(\mathbf{y})q^*(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})q^*(\mathbf{x}, \mathbf{y})}\right].$$

With probability $\beta(\mathbf{x}, \mathbf{y})$ accept $\mathbf{y}$ setting $\mathbf{X}_{n+1} = \mathbf{y}$. Otherwise reject $\mathbf{y}$ by setting $\mathbf{X}_{n+1} = \mathbf{x}$.

For state dependent approximation, we can assume that the approximate posterior has the same density as the exact target distribution at the "current" state of the chain, i.e., $\pi^*_{\mathbf{x}}(\mathbf{x}) = \pi(\mathbf{x})$. Hence, the second step acceptance probability can be simplified by

$$\beta(\mathbf{x}, \mathbf{y}) = \min\left[1, \frac{\min\{\pi(\mathbf{y})q(\mathbf{y}, \mathbf{x}), \pi^*_{\mathbf{y}}(\mathbf{x})q(\mathbf{x}, \mathbf{y})\}}{\min\{\pi(\mathbf{x})q(\mathbf{x}, \mathbf{y}), \pi^*_{\mathbf{x}}(\mathbf{y})q(\mathbf{y}, \mathbf{x})\}}\right].$$

If the approximation is state-independent, $\pi^*(\cdot)$ is used instead of $\pi^*_{\mathbf{x}}(\cdot)$, and the second step acceptance probability has the following form

$$\beta(\mathbf{x}, \mathbf{y}) = \min\left[1, \frac{\pi(\mathbf{y})}{\pi^*(\mathbf{y})}\frac{\pi^*(\mathbf{x})}{\pi(\mathbf{x})}\right],$$

which is exactly the "surrogate transition method" (Liu, 2001). However, the surrogate transition does not include a state-dependent approximation, which is an important aspect

of DAMH since that case commonly occurs in efficient construction of ROM to nonlinear problems.

As cited in Theorem 2 of Christen and Fox (2005), both of statistical and computational efficiencies of DAMH depend on the quality of the approximate target distribution. A "good" approximation will produce $\beta(\mathbf{x}, \mathbf{y}) \approx 1$, and hence the transaction probability from $\mathbf{x}$ to $\mathbf{y}$ of the delayed acceptance scheme is close to of the standard MH. One obvious way to construct an approximation is to directly replace the forward model by a ROM in the likelihood function. Since a state independent coarse model is used as the ROM in this research, we have the following approximation:

Approximation 1. *Approximation built by directly using ROM. A state independent ROM is used, and the approximate posterior distribution has the form of*

$$\pi^*(\mathbf{x}|\tilde{\mathbf{d}}) \propto \exp\left\{-\frac{1}{2}[F^*(\mathbf{x}) - \tilde{\mathbf{d}}]^T \mathbf{\Sigma}_{\mathbf{e}}^{-1}[F^*(\mathbf{x}) - \tilde{\mathbf{d}}]\right\} \pi(\mathbf{x}). \tag{4}$$

This approximation can be generalized to state dependent case when a state dependent ROM $F^*_{\mathbf{x}}(\mathbf{y})$ is used. To make consistency notations, the terms "coarse model" and "ROM" will be refer to as state independent ROM hereinafter. The state dependent ROM will be addressed specifically.

The model reduction error of the ROM is usually significant for complex forward models. This has not been counted by the above approximation, and would induces a very low acceptance rate in the second step of DAMH. Efendiev et al. (2005) and the numerical experiments in Section 4 show that the second step acceptance rate is only about 20% for approximation 1, which results in a very poorly mixing chain. It always possible to improve the accuracy of the ROM by reduce the level of model reduction, e.g., using a coarse model with grid resolution close to the forward model. However, this would result in an increasing computational cost in evaluating the approximation. The EEM introduced in next section provides an alternative way to improve the accuracy of the approximation by including the statsitics of the model reduction error. This approach uses the same ROM, and does not incur additional computational cost.

2.4. Enhanced error model

Kaipio and Somersalo (2007) suggest that, by employing an estimation of the model reduction error between the low precision ROM and the forward model, it is possible to make an accurate approximate posterior distribution based on a ROM. Suppose that, we have a ROM $F^*(\cdot)$, then Equation (1) can be expressed as

$$\begin{aligned} \tilde{\mathbf{d}} &= F^*(\mathbf{x}) + [F(\mathbf{x}) - F^*(\mathbf{x})] + \mathbf{e} \\ &= F^*(\mathbf{x}) + B(\mathbf{x}) + \mathbf{e}. \end{aligned} \tag{5}$$

Where $B(\mathbf{x})$ is the change in the model reduction error between the forward model and the ROM.

The enhanced error model of Kaipio and Somersalo (2007) assumes that the model reduction error is independent of the model parameters and normally distributed, then Equation (5) is reduced to

$$\tilde{\mathbf{d}} = F^*(\mathbf{x}) + B + \mathbf{e},$$

where $B \sim N(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B)$. To construct the EEM, a set of L random samples is drawn from the prior distribution $\mathbf{x}_i \sim \pi(\mathbf{x}), i = 1, \cdots, L$, for each of $\mathbf{x}_i$. Then, $\boldsymbol{\mu}_B$ and $\boldsymbol{\Sigma}_B$ are empirically estimated from the differences of the model outputs of the forward model and the ROM, $\{F(\mathbf{x}_i) - F^*(\mathbf{x}_i)\}_{i=1}^L$.

The EEM and a ROM can be used to give the following approximation to the exact likelihood function (2), and hence the approximate unnormalized posterior distribution (3), which have the form of:

Approximation 2. *EEM built over the prior distribution. A ROM and the EEM are used to construct the following approximate posterior distribution*

$$\pi^*(\mathbf{x}|\tilde{\mathbf{d}}) \propto \exp\left\{-\frac{1}{2}[F^*(\mathbf{x}) + \boldsymbol{\mu}_B - \tilde{\mathbf{d}}]^T(\boldsymbol{\Sigma}_B + \boldsymbol{\Sigma}_\mathbf{e})^{-1}[F^*(\mathbf{x}) + \boldsymbol{\mu}_B - \tilde{\mathbf{d}}]\right\}\pi(\mathbf{x}). \quad (6)$$

This approximation has been employed by Arridge et al. (2006) to analyse an optical diffusion problem, and the MAP estimate shows a reliable reconstruction result. More interestingly, the marginal density of the EEM is very close to the exact marginal posterior density, while the marginal density of Approximation 1 has almost zero probability in the mode of the exact marginal posterior density. This suggest that, the approximate posterior distribution (6) provides a potential approximation for DAMH.

Note that the EEM is constructed *a priori* to the parameter reconstruction. Because the prior distribution may have a quite different structure compared to the posterior distribution. When the EEM is used to estimate the parameters from a particular measured data set, it may not be able to accurately capture the differences between the outputs of the forward model outputs and ROM that are compatible with data. To make a more accurate estimation of the EEM and hence a better approximate posterior distribution, we integrate the EEM into ADAMH as an approximation used in the first step of the algorithm. This allows for estimating the *a posteriori* model reduction error adaptively through MCMC sampling.

3. Adaptive delayed acceptance Metropolis-Hastings algorithm

In practice, using a MCMC algorithm usually requires adjusting the variables in the transition kernel to achieve statistical efficiency, e.g. the scale variables in the random walk proposals of MH. If there exists many variables to adjust or the target distribution is computationally expensive to evaluate, this tuning process can be very time consuming in terms of human input and computing time. By summarizing the statistical information of the samples previously drawn from a Markov chain, the adaptive MCMC algorithms (Haario et al., 2001; Atchade and Rosenthal, 2005; Andrieu and Moulines, 2006; Roberts and Rosenthal, 2007, 2009) provide possibilities for automatically adjusting those variable.

To allow for the construction of adaptive MCMC algorithms under more general and simpler conditions, Roberts and Rosenthal (2007) proved ergodicity theorems of adaptive MCMC algorithms under simpler conditions, namely simultaneously ergodicity and diminishing adaptation. Based on these regularity conditions, we design ADAMH that does not only adaptively adjust the form of the first step proposal distribution, but also uses the empirical estimation of the model reduction error from the sampling history to improve the accuracy of the approximate posterior distribution. Before discussing ADAMH, we first review two adaptive MCMC algorithms: the adaptive Metropolis algorithm (AM) (Haario

et al., 2001) and the adaptive Metropolis-within-Gibbs algorithm (AMWG) (Roberts and Rosenthal, 2009).

3.1. Examples of adaptive MCMC algorithms

Let $\mathbf{x}$ and $\mathcal{X}$ denote the current state and parameter space, respectively. Assume that the parameter space $\mathcal{X} \in \mathcal{R}^d$ is a subspace of $\mathcal{R}^d$ with compact support. AM is a Metropolis algorithm with Gaussian proposal $q(\mathbf{x}, \cdot) = N(\mathbf{x}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is estimated adaptively from the past history of the chain. Suppose we have a target distribution $\pi(\cdot)$, AM is given by:

ALGORITHM 2. *Adaptive Metropolis*
At step n, given $\mathbf{X}_n = \mathbf{x}$, then $\mathbf{X}_{n+1}$ is determined in the following way:

(a) Propose new state $\mathbf{y}$ from the proposal

$$q_n(\mathbf{x}, \cdot) = \begin{cases} N(\mathbf{x}, \frac{0.1^2}{d}\mathbf{I}_d) & n \leq 2d \\ (1-\beta)N(\mathbf{x}, \frac{2.38^2}{d}\mathbf{\Sigma}_n) + \beta N(\mathbf{x}, \frac{0.1^2}{d}\mathbf{I}_d) & n > 2d \end{cases},$$

where Σ_n is the empirical estimation of the covariance structure of the target distribution up to step n, and β is a small positive constant.

(b) With probability $\min\left[1, \frac{\pi(\mathbf{y})}{\pi(\mathbf{x})}\right]$, set $\mathbf{X}_{n+1} = \mathbf{y}$, otherwise $\mathbf{X}_{n+1} = \mathbf{x}$.

The motivation of AM is that the proposal distribution $N(\mathbf{x}, \frac{2.38^2}{d}\mathbf{\Sigma}_n)$ is optimal in a particular large dimensional context (Roberts et al., 1997; Roberts and Rosenthal, 2001). The mixer with the non-adaptive Gaussian distribution, $N(\mathbf{x}, \frac{0.1^2}{d}\mathbf{I}_d)$ is introduced to prevent the algorithm from being stuck at problematic value of $\mathbf{\Sigma}_n$.

Unlike AM tune the random walk proposal by calculating empirical estimates of past samples, AMWG uses a different idea of adaptation by adjusting the proposal variable to match the target acceptance rate. This algorithm uses a Gaussian distribution $N(x_i, \gamma_i^2)$ component-wise as the proposal, which is defined as:

ALGORITHM 3. *Adaptive Metropolis-within-Gibbs*
At step n, given $\mathbf{X}_n = \mathbf{x}$, then $\mathbf{X}_{n+1}$ is determined in the following way:

(a) Initialize $\mathbf{y} = \mathbf{x}$, for all $i = 1 \ldots, d$:

- *we draw z from the proposal distribution $q_{\gamma_i}(x_i, \cdot) = N(x_i, \gamma_i^2)$,*
- *with probability $\min\left[1, \frac{\pi(y_1, \ldots, z, \ldots, y_d)}{\pi(\mathbf{y})}\right]$, set $y_i = z$, otherwise y_i unchanged.*

(b) Then $\mathbf{X}_{n+1} = \mathbf{y}$ after updating all the components of the parameter $\mathbf{y}$.

(c) For a pre-specified batch number N, if $n \bmod N = 0$, for all $i = 1 \ldots, d$, the acceptance rate $\hat{\alpha}_i$ from the past N updates on ith component is calculated. Then, if $\hat{\alpha}_i > 0.44$, $\gamma_i = \gamma_i + \exp(\delta)$, otherwise $\gamma_i = \gamma_i - \exp(\delta)$. Here $\delta = \min[0.01, (\frac{n}{N})^{-1/2}]$.

AMWG attempts to vary the scale of the proposal to match the acceptance rate of 0.44 for all the elements of the parameter, where the acceptance rates of 0.44 is optimal for one-dimensional proposals in certain settings, see Roberts et al. (1997); Roberts and Rosenthal (2001).

3.2. Adaptive delayed acceptance Metropolis-Hastings algorithm

To allow estimation of the model reduction error from the posterior distribution adaptively through running ADAMH, we have the following adaptive approximation:

APPROXIMATION 3. *EEM built over the posterior distribution. A ROM is used, and the approximate posterior distribution at step n has the form of*

$$\pi^*_{\boldsymbol{\xi}}(\mathbf{x}|\tilde{\mathbf{d}}) \propto \exp\left\{-\frac{1}{2}[F^*(\mathbf{x}) + \boldsymbol{\mu}_{B,n} - \tilde{\mathbf{d}}]^T(\boldsymbol{\Sigma}_{B,n} + \boldsymbol{\Sigma}_{\mathbf{e}})^{-1}[F^*(\mathbf{x}) + \boldsymbol{\mu}_{B,n} - \tilde{\mathbf{d}}]\right\}\pi(\mathbf{x}). \quad (7)$$

Where $\boldsymbol{\xi} = (\boldsymbol{\mu}_{B,n}, \boldsymbol{\Sigma}_{B,n})$ represents the mean and covariance matrix of model reduction error after n steps of adaptive updating. For an **accepted** state $\mathbf{x}_n$ at time n, the difference between the forward model and the ROM, $B_n = F(\mathbf{x}_n) - F^*(\mathbf{x}_n)$, can be computed virtually no cost compared to the evaluation time of the forward models. Therefore, $(\boldsymbol{\mu}_{B,n}, \boldsymbol{\Sigma}_{B,n})$ is updated to $(\boldsymbol{\mu}_{B,n+1}, \boldsymbol{\Sigma}_{B,n+1})$ by:

$$\begin{aligned}
\boldsymbol{\mu}_{B,n+1} &= \frac{1}{n+1}[n\boldsymbol{\mu}_{B,n} + B_{n+1}], \\
\boldsymbol{\Sigma}_{B,n+1} &= \frac{1}{n}[(n-1)\boldsymbol{\Sigma}_{B,n} + B_{n+1}{B_{n+1}}^T - n\boldsymbol{\mu}_{B,n+1}{\boldsymbol{\mu}_{B,n+1}}^T].
\end{aligned} \quad (8)$$

By Combining the adaptive approximation (e.g., Approximation 3), adaptive MCMC algorithms and DAMH together, we have ADAMH. To simplify the notation, let $\pi(\cdot) = \pi(\cdot|\tilde{\mathbf{d}})$ be the exact posterior distribution based on the forward model. To give a more general type of approximation than the state-independent Approximation 3, we let $\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot) = \pi^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot|\tilde{\mathbf{d}})$ denotes the state dependent approximation to the exact posterior distribution, where $\boldsymbol{\xi}$ is the variable used in the approximate posterior distribution, e.g., Equation (7) in Approximation 3. Let $q_{\boldsymbol{\gamma}}(\mathbf{x}, \cdot)$ represents the adaptive proposal distribution used in the first step, where $\boldsymbol{\gamma}$ is the variables in the proposals, e.g., the covariance matrix of AM. Thus, ADAMH is given as:

ALGORITHM 4. *Adaptive delayed acceptance Metropolis-Hastings*
At step n, suppose we have $\mathbf{X}_n = \mathbf{x}$*, then* $\mathbf{X}_{n+1}$ *is determined in the following way:*

(a) Generate a proposal $\mathbf{y}$ *from the adaptive proposal distribution* $q_{\boldsymbol{\gamma}}(\mathbf{x}, \cdot)$.
(b) Let

$$\alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) = \min\left[1, \frac{\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y})q_{\boldsymbol{\gamma}}(\mathbf{y}, \mathbf{x})}{\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{x})q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{y})}\right].$$

With probability $\alpha(\mathbf{x}, \mathbf{y})$*, accept* $\mathbf{y}$ *to be used as a proposal for the standard MH. Otherwise use* $\mathbf{y} = \mathbf{x}$ *as a proposal. The actual proposal distribution used is*

$$q^*_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) = \alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{y}) + (1 - r_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}))\delta_{\mathbf{x}}(\mathbf{y}),$$

where $r_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}) = \int_{\mathcal{X}} \alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{y})\mathbf{dy}$ *and* $\delta_{\mathbf{x}}(\cdot)$ *denotes the Dirac mass at* $\mathbf{x}$.
(c) Let

$$\beta_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) = \min\left[1, \frac{\pi(\mathbf{y})q^*_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{y}, \mathbf{x})}{\pi(\mathbf{x})q^*_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})}\right].$$

With probability $\beta_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})$ *accept* $\mathbf{y}$ *setting* $\mathbf{X}_{n+1} = \mathbf{y}$*. Otherwise reject* $\mathbf{y}$ *setting* $\mathbf{X}_{n+1} = \mathbf{x}$.

(d) Update the approximation $\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot)$.
(e) Update the adaptive proposal $q_{\boldsymbol{\gamma}}(\mathbf{x},\cdot)$.

In this algorithm, the proposal $q_{\boldsymbol{\gamma}}(\mathbf{x},\cdot)$ in step (a) and its adaptation in step (e) may have the form of any of the adaptive algorithms such as AM and AMWG. When Approximation 3 is used in step (b), for each state $\mathbf{X}_n = \mathbf{x}$ in the Markov chain, a variable B_n is required to represent the difference between the forward model and the ROM, if a candidate state $\mathbf{y}$ is accepted in step (c), B_{n+1} is updated to $F(\mathbf{y}) - F^*(\mathbf{y})$, otherwise $B_{n+1} = B_n$. Then, the EEM is updated in step (d) as in the updating rule (8).

3.3. Ergodicity conditions and theorem

We follow the notation in Roberts and Rosenthal (2007) to formalise ADAMH. Suppose $\pi(\cdot)$ is a fixed target distribution, defined on state space $\mathcal{X}$ with σ-algebra $\mathcal{B}(\mathcal{X})$, e.g., the exact posterior distribution. Let $\{K_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma}\in\mathcal{Y}}$ be a family of Markov chain kernels (associated with standard MH) on $\mathcal{X}$, and each has $\pi(\cdot)$ as the unique stationary distribution for all $\boldsymbol{\gamma} \in \mathcal{Y}$. Let $\{\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot)\}_{\boldsymbol{\xi}\in\mathcal{E}}$ be a family of state-dependent approximations to the exact target distribution $\pi(\cdot)$ for all $\boldsymbol{\xi} \in \mathcal{E}$, e.g., Approximation 3. The variable $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ are referred as adaptation indices, e.g. the $\boldsymbol{\xi} = (\boldsymbol{\mu}_{B,n}, \boldsymbol{\Sigma}_{B,n})$ in Approximation 3 and the proposal variables $\gamma_i, i = 1,\ldots,d$ in AMWG.

At each time n, ADAMH updates $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ by $\mathcal{Y}$-valued random variable $\boldsymbol{\Gamma}_n$ and $\mathcal{E}$-valued random variable $\boldsymbol{\Xi}_n$, respectively. The transition kernel of ADAMH is denoted by $\{K_{\boldsymbol{\gamma},\boldsymbol{\xi}}\}_{\boldsymbol{\gamma}\in\mathcal{Y},\boldsymbol{\xi}\in\mathcal{E}}$. By using the conditions stated in Theorem 1 and Corollary 5 in Roberts and Rosenthal (2007). We prove the following theorem that, the ergodicity of ADAMH can be achieved by imposing certain regularity conditions.

THEOREM 1. *Consider an adaptive delayed acceptance Metropolis-Hastings algorithm, with target distribution* $\pi(\cdot)$ *defined on a state space* $\mathcal{X}$*, with first step proposal adaptation index* $\mathcal{Y}$ *and approximation adaptation index* $\mathcal{E}$.

Suppose for each $\boldsymbol{\gamma} \in \mathcal{Y}$, $K_{\boldsymbol{\gamma}}$ *represents a standard Metropolis-Hastings algorithm with proposal kernel* $Q_{\boldsymbol{\gamma}}(\mathbf{x},\mathbf{dy}) = q_{\boldsymbol{\gamma}}(\mathbf{x},\mathbf{dy})\lambda(\mathbf{dy})$ *having a density* $q_{\boldsymbol{\gamma}}(,\cdot)$ *with respect to some finite reference measure* $\lambda(\cdot)$*, with corresponding density* h *for* $\pi(\cdot)$ *so that* $\pi(\mathbf{dy}) = h(\mathbf{y})\lambda(\mathbf{dy})$. *Similarly, for each* $\boldsymbol{\xi} \in \mathcal{E}$*, the state-dependent approximation* $\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot)$ *has density* $h^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot)$ *such that,* $\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{dy}) = h^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y})\lambda(\mathbf{dy})$. *Let* $K_{\boldsymbol{\gamma},\boldsymbol{\xi}}$ *denotes the transition kernel of an adaptive delayed acceptance Metropolis-Hastings algorithm with a given first step transition kernel* $K_{\boldsymbol{\gamma}}$ *and a given approximation* $\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot)$. *Suppose further the following conditions hold:*

(a) The state space $\mathcal{X}$ *is compact with respect to some topology.*
(b) Each transition kernel $K_{\boldsymbol{\gamma}}$ *is ergodic for* $\pi(\cdot)$.
(c) For all $\boldsymbol{\gamma} \in \mathcal{Y}$, $q_{\boldsymbol{\gamma}}(\mathbf{x},\cdot)$ *is uniformly bounded. For each fixed* $\mathbf{y} \in \mathcal{X}$*, the mapping* $(\mathbf{x},\boldsymbol{\gamma}) \to q_{\boldsymbol{\gamma}}(\mathbf{x},\mathbf{y})$ *is continuous with respect to some product metric space topology, with respect to which* $\mathcal{X}\times\mathcal{Y}$ *is compact.*
(d) For given $\boldsymbol{\gamma} \in \mathcal{Y}$ *and* $\boldsymbol{\xi} \in \mathcal{E}$, $K_{\boldsymbol{\gamma}}(\mathbf{x},\mathbf{y}) > 0$ *implies* $\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y}) > 0$.
(e) For each fixed $\mathbf{y} \in \mathcal{X}$*, the map* $(\mathbf{x},\boldsymbol{\xi}) \to h^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y})$ *is continuous with respect to some product metric space topology, with respect to which* $\mathcal{X}\times\mathcal{E}$ *is compact.*
(f) ***Diminishing adaptation****. The amount of adaptation in the transition kernel vanishes as* $n \to \infty$*, i.e.* $\lim_{n\to\infty} \|K_{\boldsymbol{\Gamma}_{n+1},\boldsymbol{\Xi}_{n+1}}(\mathbf{x},\cdot) - K_{\boldsymbol{\Gamma}_n,\boldsymbol{\Xi}n}(\mathbf{x},\cdot)\|_{TV} = 0$ *in probability, where* $\|\mu(\cdot) - \nu(\cdot)\|_{TV} = \sup_{\mathbf{A}\in\mathcal{B}(\mathcal{X})} \|\mu(\mathbf{A}) - \nu(\mathbf{A})\|$ *is the total variational norm.*

Then ADAMH is ergodic.

Proof: In order to prove this theorem, the elements of Theorem 1 of Roberts and Rosenthal (2007) are used along with the method of proof used in Corollary 5 (ergodicity of the single step Metropolis-Hastings algorithm with adaptive proposal) of Roberts and Rosenthal (2007).

By Theorem 1 of Christen and Fox (2005), Condition (b) and (d) imply that the transition kernel $K_{\boldsymbol{\gamma},\boldsymbol{\xi}}$ is ergodic for given $\boldsymbol{\gamma} \in \mathcal{Y}$ and $\boldsymbol{\xi} \in \mathcal{E}$. The effective proposal density in the second step of ADAMH can be written as

$$q^*_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{dy}) = \alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{y}) \lambda(\mathbf{dy}) + [1 - r_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x})] \delta_{\mathbf{x}}(\mathbf{dy}),$$

where $\alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})$ is the first step acceptance probability, in the form

$$\alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) = \min \left[1, \frac{h^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y}) q_{\boldsymbol{\gamma}}(\mathbf{y}, \mathbf{x})}{h^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{x}) q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{y})} \right],$$

and $r_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x})$ is the probability of accepting a proposal from $\mathbf{x}$, which is given by:

$$r_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}) = \int_{\mathcal{X}} \alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{y}) \lambda(\mathbf{dy}).$$

Since $\alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})$ is jointly continuous in $\mathbf{x}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\xi}$ by Condition (c) and (d), $r_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x})$ is a jointly continuous function of $(\mathbf{x}, \boldsymbol{\gamma}, \boldsymbol{\xi}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{E}$ by the Bounded Convergence Theorem. We can assume without loss of generality that $r_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}) = 1$ whenever $\lambda(\mathbf{x}) > 0$, i.e., that $\delta_{\mathbf{x}}(\cdot)$ and $\pi(\cdot)$ are orthogonal measures. Furthermore, the probability of being able to accept a proposal in the first step $f_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) = \alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{y})$ is uniformly bounded and jointly continuous in $\mathbf{x}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$.

Therefore, the overall probability of accepting a proposal from $\mathbf{x}$ in the an adaptive delayed acceptance Metropolis-Hastings algorithm is given by:

$$\rho_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}) = \int_{\mathcal{X}} \beta_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) f_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) \lambda(\mathbf{dy}),$$

where $\beta_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})$ is the second step acceptance rate, in the form of

$$\beta_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y}) = \min \left[1, \frac{h(\mathbf{y}) f_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{y}, \mathbf{x})}{h(\mathbf{x}) f_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{y})} \right].$$

Using a similar argument to that above, $\rho_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x})$ is a jointly continuous function of $(\mathbf{x}, \boldsymbol{\gamma}, \boldsymbol{\xi}) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{E}$, and the transition kernel of the ADAMH can be decomposed as

$$K_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{dz}) = g_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{z}) \lambda(\mathbf{dz}) + [1 - \rho_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x})] \delta_{\mathbf{x}}(\mathbf{dz}),$$

where $g_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{z}) = q_{\boldsymbol{\gamma}}(\mathbf{x}, \mathbf{z}) \alpha_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{z}) \beta_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{z})$ is jointly continuous in $\mathbf{x}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$.

Now we can exactly follow the proof of Corollary 5 of Roberts and Rosenthal (2007). By iterating $K_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{dz})$, we have the n-step transition law

$$K^n_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{dz}) = g^n_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}, \mathbf{z}) \lambda(\mathbf{dz}) + [1 - \rho_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x})]^n \delta_{\mathbf{x}}(\mathbf{dz}).$$

Again, we can assume without loss of generality that $\rho_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x}) = 1$ whenever $\lambda(\mathbf{x}) > 0$. It then follows that:

$$\|K^n_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x},\cdot) - \pi(\cdot)\|_{TV} = \frac{1}{2}\int_{\mathcal{X}} |g^n_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x},\mathbf{z}) - h(\mathbf{z})|\lambda(\mathbf{dz}) + [1 - \rho_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x})]^n.$$

This quantity is jointly continuous in $\mathbf{x}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$ by the Bounded Convergence Theorem. Moreover, it converges to zero as $n \to \infty$ for each fixed $\mathbf{x}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$. Hence, by compactness, the convergence is uniform in $\mathbf{x}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\xi}$. Thus, the simultaneous uniform ergodicity condition in Theorem 1 of Roberts and Rosenthal (2007) is satisfied. Then, also the transition kernel $K^n_{\boldsymbol{\gamma},\boldsymbol{\xi}}(\mathbf{x},\cdot)$ satisfies the diminishing adaptation. Therefore, all the conditions in Theorem 1 of Roberts and Rosenthal (2007) are satisfied, and the result follows. ■

The strong conditions imposed in Theorem 1 can be relaxed further by using more specialized arguments for specific algorithms, e.g., Saksman and Vihola (2010); Vihola (2011). However, these conditions are sufficient to design usable adaptive approximations in ADAMH for practical purpose. Given an adaptive proposal density $q_{\boldsymbol{\gamma}}(\mathbf{x},\cdot)$ satisfies Condition (b) and (c) of Theorem 1, the following corollary prove that Approximation 3 satisfies Condition (d) and (e) of Theorem 1, and hence ADAMH using Approximation 3 is ergodic.

COROLLARY 1. *Suppose that the class of proposal kernels used in the first step of ADAMH, $\{Q_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma}\in\mathcal{Y}}$, satisfy the diminishing adaptation condition and Condition (b) and (c) of Theorem 1, ADAMH with Approximation 3 is ergodic.*

Proof: Without loss of generality, we assume that the parameter space $\mathcal{X}$ is compact, as we can always define some bounds for the input parameter to the computer model. Since any continuous image of a compact set is compact, for any continuous mathematical model, the image is compact. When the continuous map is solved numerically by finite element methods or integrated finite difference methods, the computer model is necessarily continuous since the stiffness matrix is non-singular. This implies that the outputs of the forward model $F(\cdot)$ and its ROM $F^*(\cdot)$ are compact, and hence the model reduction error $B(\cdot) = F(\cdot) - F^*(\cdot)$ is compact as well. Thus, the mean and covariance of the model reduction error is bounded, i.e., $\|\boldsymbol{\mu}_{B,n}\|_\infty < C_1$ and $0 \leq \boldsymbol{\Sigma}_{B,n} < C_2\mathbf{I}$ for some $C_1, C_2 > 0$. This guarantees that the $\mathcal{E}$ is compact, and hence $\mathcal{X} \times \mathcal{E}$ is compact.

The approximate posterior distribution (7) is always continuous except when $\boldsymbol{\Sigma}_{B,n} + \boldsymbol{\Sigma}_{\mathbf{e}}$ is singular. We can show that $\boldsymbol{\Sigma}_{B,n} + \boldsymbol{\Sigma}_{\mathbf{e}}$ is bounded away from zero as follows. The positive definite observation noise covariance $\boldsymbol{\Sigma}_{\mathbf{e}}$ ensures that, there exists $C_4 > C_3 > 0$ such that $C_3\mathbf{I} \leq \boldsymbol{\Sigma}_{B,n} + \boldsymbol{\Sigma}_{\mathbf{e}} < C_4\mathbf{I}$, and hence the approximate posterior distribution defined by the approximate posterior distribution (7) and an appropriate prior distribution has the continuous and bounded density mapping $(\mathbf{x}, \boldsymbol{\xi}) \to h^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y})$. Therefore, Condition (d) and (e) in Theorem 1 is satisfied.

$\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\cdot)$ give in the updating rule (8) is easy to satisfy the diminishing adaptation condition. Because the empirical estimates change about $O(1/n)$ at the nth iteration, the mean and covariance matrix trend to constant expected values over $\pi(\cdot)$. Together with the diminishing adaptation assumed for $\{Q_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma}\in\mathcal{Y}}$, the diminishing adaptation condition is satisfied, and thus Theorem 1 applies. ■

3.4. State dependent approximations

One significant advantage of Approximation 3 is that it has more accurate mean estimation and smaller variance than Approximation 2. Another advantage is that this adaptive approach does not require any pre-computation to give an estimation of EEM before setting

up an MCMC simulation. Since the model reduction error commonly has different structures in different regions of the state space, a local estimation of $B(\mathbf{x})$ can be used instead of the global mean. Suppose that, the model reduction error $B(\mathbf{x})$ is known for some point $\mathbf{x} \in \mathcal{X}$. We assume for the points around $\mathbf{x}$, say $\mathbf{y} \in \{\|\mathbf{y}-\mathbf{x}\| \leq \delta : \mathbf{y} \in \mathcal{X}\}$ for given $\delta > 0$, the model reduction error $B(\mathbf{y})$ is close to $B(\mathbf{x})$, i.e., $\|B(\mathbf{y}) - B(\mathbf{x})\| \leq C\delta^P$ for some $C > 0$ and $P > 1$. Let us define a state-dependent ROM based on the zeroth order correction to the ROM, $F^*_{\mathbf{x}}(\mathbf{y}) = F^*(\mathbf{y}) + [F(\mathbf{x}) - F^*(\mathbf{x})]$, a non-adaptive version of approximation is given:

Approximation 4. *Local approximation. Suppose at step n, the Markov chain has state* $\mathbf{X}_n = \mathbf{x}$*, a state-dependent ROM* $F^*_{\mathbf{x}}(\cdot)$ *can be given as the sum of model reduction error at* $\mathbf{X}_n = \mathbf{x}$ *and the ROM* $F^*(\cdot)$ *to the forward model* $F(\cdot)$*. For a proposed state* $\mathbf{y} \sim q(\mathbf{x}, \cdot)$*, it is given by:*

$$F^*_{\mathbf{x}}(\mathbf{y}) = F^*(\mathbf{y}) + [F(\mathbf{x}) - F^*(\mathbf{x})]. \tag{9}$$

The approximate posterior distribution has the form of

$$\pi^*_{\mathbf{x}}(\mathbf{y}|\tilde{\mathbf{d}}) \propto \exp\left\{-\frac{1}{2}[F^*_{\mathbf{x}}(\mathbf{y}) - \tilde{\mathbf{d}}]^T \boldsymbol{\Sigma}_{\mathbf{e}}^{-1}[F^*_{\mathbf{x}}(\mathbf{y}) - \tilde{\mathbf{d}}]\right\}\pi(\mathbf{y}). \tag{10}$$

It is worth mentioning that the above state-dependent ROM is just an example based solely on a state independent ROM and a zeroth order local correction. Whenever it is possible, higher oder derivatives can be used to give a more accurate state-dependent ROM. The error structure of the above state-dependent ROM can also be estimated by employing the EEM, Approximation 3 and 4 can be combined to give a more sophisticated approximation:

Approximation 5. *EEM built over the posterior distribution with state-dependent ROM. Suppose at step n of the Markov chain, we have* $\mathbf{X}_n = \mathbf{x}$ *and a proposed state* $\mathbf{y} \sim q(\mathbf{x}, \cdot)$*. Let* $\boldsymbol{\xi} = (\boldsymbol{\mu}_{B,n}, \boldsymbol{\Sigma}_{B,n})$ *represents the mean and covariance matrix of EEM of the state-dependent ROM after n steps of adaptively updating. The state-dependent approximate posterior distribution has the form of*

$$\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y}|\tilde{\mathbf{d}}) \propto \exp\left\{-\frac{1}{2}[F^*_{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\mu}_{B,n} - \tilde{\mathbf{d}}]^T(\boldsymbol{\Sigma}_{B,n} + \boldsymbol{\Sigma}_{\mathbf{e}})^{-1}[F^*_{\mathbf{x}}(\mathbf{y}) + \boldsymbol{\mu}_{B,n} - \tilde{\mathbf{d}}]\right\}\pi(\mathbf{y}). \tag{11}$$

When the above zeroth order correction (9) is used, the model reduction error is given by $B_{\mathbf{x}}(\mathbf{y}) = F(\mathbf{y}) - F^*_{\mathbf{x}}(\mathbf{y})$ for an *accepted* state $\mathbf{y}$, otherwise $B_{\mathbf{x}}(\mathbf{y}) = \mathbf{0}$.

Given a current state $\mathbf{x}$, the mean and covariance are only estimated from the accepted state $\mathbf{y}$. Hence, the mean and covariance of the EEM in Approximation 5 with ROM (9) are actually estimations of $\mathrm{E}_\pi[\int_{\mathcal{X}} B_{\mathbf{x}}(\mathbf{y})K(\mathbf{x},\mathbf{y})d\mathbf{y}]$ and $\mathrm{Cov}_\pi[\int_{\mathcal{X}} B_{\mathbf{x}}(\mathbf{y})K(\mathbf{x},\mathbf{y})d\mathbf{y}]$, respectively. Where the transition kernel $K(\mathbf{x},\mathbf{y})$ gives the transaction probability from $\mathbf{x}$ to a candidate $\mathbf{y}$. In general, when the distance between two states $\mathbf{x}$ and $\mathbf{y}$ increases, the model reduction error $B_{\mathbf{x}}(\mathbf{y})$ increases, while the transaction probability $K(\mathbf{x},\mathbf{y})$ decreases. $\mathrm{E}_\pi[\int_{\mathcal{X}} B_{\mathbf{x}}(\mathbf{y})K(\mathbf{x},\mathbf{y})d\mathbf{y}]$ can be written as:

$$\int_{\mathcal{X}}\int_{\mathcal{X}} [F(\mathbf{y}) - F^*(\mathbf{y})]\pi(\mathbf{x})K(\mathbf{x},\mathbf{y})d\mathbf{y}d\mathbf{x} - \int_{\mathcal{X}}\int_{\mathcal{X}} [F(\mathbf{x}) - F^*(\mathbf{x})]\pi(\mathbf{x})K(\mathbf{x},\mathbf{y})d\mathbf{y}d\mathbf{x}.$$

From the detailed balance condition, $\pi(\mathbf{x})K(\mathbf{x},\mathbf{y}) = \pi(\mathbf{y})K(\mathbf{y},\mathbf{x})$, we have:

$$\int_{\mathcal{X}}\int_{\mathcal{X}} [F(\mathbf{y}) - F^*(\mathbf{y})]\pi(\mathbf{y})K(\mathbf{y},\mathbf{x})d\mathbf{y}d\mathbf{x} - \int_{\mathcal{X}}\int_{\mathcal{X}} [F(\mathbf{x}) - F^*(\mathbf{x})]\pi(\mathbf{x})K(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y},$$

which has value zero. However, the covariance of the model reduction error $B_{\mathbf{x}}(\mathbf{y})$ is non-zero, and can be computed adaptively by:

$$\mathbf{\Sigma}_{B,n} = \frac{1}{n-1}\left[(n-2)\mathbf{\Sigma}_{B,n-1} + B_{\mathbf{x}_{n-1}}(\mathbf{x}_n)B_{\mathbf{x}_{n-1}}(\mathbf{x}_n)^T\right] \tag{12}$$

Hence, the approximate posterior distribution (11) can be rewritten in the following form:

$$\pi^*_{\boldsymbol{\xi},\mathbf{x}}(\mathbf{y}|\tilde{\mathbf{d}}) \propto \exp\left\{-\frac{1}{2}[F^*_{\mathbf{x}}(\mathbf{y}) - \tilde{\mathbf{d}}]^T(\mathbf{\Sigma}_{B,n} + \mathbf{\Sigma}_{\mathbf{e}})^{-1}[F^*_{\mathbf{x}}(\mathbf{y}) - \tilde{\mathbf{d}}]\right\}\pi(\mathbf{x}). \tag{13}$$

The Approximation 4 can be considered as a special case of Approximation 5 with zero covariance. Given an adaptive proposal distribution $q_{\boldsymbol{\gamma}}(\mathbf{x},\cdot)$ that satisfies the Condition 2, then following the proof of Corollary 1, the ergodicity of ADAMH with Approximation 5 can be shown.

COROLLARY 2. *Suppose that the class of proposal kernels used in the first step of ADAMH, $\{Q_{\boldsymbol{\gamma}}\}_{\boldsymbol{\gamma}\in\mathcal{Y}}$, satisfy the diminishing adaptation condition and Conditions (b) and (c) of Theorem 1, ADAMH with Approximation 5 is ergodic.*

Proof: See proof of Corollary 1. ■

Compare Approximation 5 to Approximation 3, the use of a state dependent approximation is able to give a more accurate approximation is the jump size in the first step proposal is small. However, this may not hold for large jump size. Another practical advantage give by Approximation 5 with ROM (9) is that it does not require as many burn-in steps as Approximation 3 to achieve good second step acceptance rate. In practice, we found that Approximation 5 with ROM (9) could archive good second step acceptance rate without any burn-in steps, while Approximation 3 need few thousands or even more burn-in steps to build a reasonably good approximation. This mainly because of the local structure used in Approximation 5 does not require any computation from the sample history, which is compulsory for the EEM used in Approximation 3. However, Approximation 5 does require burn-in steps to build the covariance matrix, which also improves the second step acceptance rate. This is observed in several computing experiments in Cui (2010), and further study is necessary to understand how the covariance matrix affect the statistical efficiency of Approximation 5.

3.5. Performance estimates

To estimate the speed-up factor of ADAMH, let the computing time of the approximate target density and the exact target density be t^* and t, respectively. Suppose that the average first step acceptance rate of ADAMH is about $\hat{\alpha}$. Let the integrated autocorrelation time (IACT) of some statistics h be τ^*. Consider the same target distribution also being sampled by the standard MH, with the IACT of h denoted by τ. The number of effectively independent samples generated in n iteration by the standard MH and ADAMH are $n/(2\tau)$ and $n/(2\tau^*)$ respectively. The time required to run n iteration for the standard MH and ADAMH are nt and $n(1-\hat{\alpha})t^* + n\hat{\alpha}(t^*+t)$ (or $n(1-\hat{\alpha})t^* + n\hat{\alpha}(2t^*+t)$ if calculating the second step acceptance rate requires backward evaluation of the approximation $\pi^*_{\mathbf{y}}(\mathbf{x})$, e.g. with the linearisation in Christen and Fox (2005)). Hence the speed-up factor of

ADAMH compared to the standard MH can be estimated as the ratio of computing time per effectively independent sample, which has the form of

$$\frac{\tau}{\tau^*}\frac{1}{\hat{\alpha}+\frac{t^*}{t}}, \tag{14}$$

or

$$\frac{\tau}{\tau^*}\frac{1}{\hat{\alpha}+(1+\hat{\alpha})\frac{t^*}{t}}, \tag{15}$$

for the case which requires backward evaluation of the approximation. Here $\tau/\tau^* < 1$ is a measure of the decrease of the statistical efficiency of ADAMH. ADAMH is always statistically less efficient than the standard MH (Christen and Fox, 2005), because it requires more iterations to generate a statistically independent sample than the standard MH. This fact follows directly from the condition $\beta_n(\mathbf{x}, \mathbf{x}') \leq 1$. This suggest that there are two important issues in implementing ADAMH: (i) designing efficient moves in the first step, which optimize the statistical efficiency in the sense of running a standard MH; and (ii) maintaining the statistical efficiency of ADAMH, which requires a "good" approximate posterior distribution that has transaction probability from $\mathbf{x}$ to $\mathbf{y}$ close to the standard MH, i.e., $\beta_n(\mathbf{x}, \mathbf{y}) \approx 1$, and hence $\tau/\tau^* \approx 1$.

The speed-up factor (14) and (15) suggest that we can always lower the first step acceptance rate in ADAMH, and hence the overall acceptance rate, thus lowering the computing time per independent sample. However, the statistical efficiency of the algorithm drops as the acceptance rate approaches 0 (Roberts et al., 1997; Roberts and Rosenthal, 2001). There is a trade off in selecting the appropriate acceptance rate. Roberts et al. (1997); Roberts and Rosenthal (2001) suggested an optimal acceptance rate of about 0.23 for high dimensional parameters, but they also suggest that any acceptance rate in the rage of $[0.1, 0.6]$ gives good statistical efficiency. The numerical experiments based on geothermal reservoir models in (Cui, 2010) shows that an acceptance rate of 0.1 produces optimal statistical efficiency, and hence we adopt it for the higher dimensional problem in Section 4.

4. Applications in geothermal reservoir models

We apply ADAMH to two sampling problems of geothermal reservoir models here. First, an one dimensional radial symmetry model of the feedzone of a geothermal reservoir with synthetic data is used to validate the algorithm. Then ADAMH is applied to sample a 3D model with measured data. We start this section by reviewing the governing equations and its numerical simulator of the geothermal reservoir.

4.1. Data simulation

Multiphase non-isothermal flow in a geothermal reservoir can be simulated by the numerical package TOUGH2 (Pruess, 1991). In TOUGH2, two phase flow (water and vapour) is modelled by general mass balance and energy balance equations. For an arbitrary subdomain Ω with bounding surface $\partial\Omega$, the balance equations can be written in the form

$$\frac{d}{dt}\int_\Omega M_\alpha \, dV = \int_{\partial\Omega} Q_\alpha \cdot \hat{n} \, d\Gamma + \int_\Omega q_\alpha \, dV, \tag{16}$$

where Ω is the control volume and $\partial\Omega$ is the boundary of the control volume. The accumulation term q_α represents the mass (q_m) and heat q_e sources or sinks in the control volume,

and Q_α denotes the mass (Q_m) or energy (Q_e) flux through the boundary of the control volume. The mass and energy within the control volume are represented by M_m and M_e, respectively. Here, M_m and M_e are defined by

$$\begin{aligned} M_\mathrm{m} &= \phi\,(\rho_\mathrm{l} S_\mathrm{l} + \rho_\mathrm{v} S_\mathrm{v}), \\ M_\mathrm{e} &= (1-\phi)\,\rho_\mathrm{r} c_\mathrm{r} T + \phi\,(\rho_\mathrm{l} u_\mathrm{l} S_\mathrm{l} + \rho_\mathrm{v} u_\mathrm{l} S_\mathrm{v}). \end{aligned}$$

Here ϕ is porosity, ρ is density, S is saturation, u is specific internal energy, c is specific heat and T is the temperature. The subscripts l, v and r indicate the quantities pertaining to liquid, vapour and rock, respectively. Note that: $S_\mathrm{l} + S_\mathrm{v} = 1$.

The mass flux term Q_m is given as a sum over liquid and vapour phase:

$$Q_\mathrm{m} = \sum_{\beta=\mathrm{l,v}} \frac{k k_{\mathrm{r}\beta}}{\nu_\beta}(\nabla p - \rho_\beta \mathbf{g}). \tag{17}$$

Here k is a diagonal second order permeability tensor in 3-dimensions, $\mathbf{g}$ represents the acceleration due to gravity, ∇p is the pressure gradient acting on the fluid flow, and ν_β is the kinematic viscosities. Relative permeability $k_{r\beta}$ is introduced to account for the interference between liquid and vapour phases as they move through the rock matrix in the geothermal reservoir. There are several empirically derived curves available to model $k_{\mathrm{r}\beta}$ as functions of vapour saturation S_v. We use the van Genuchten-Mualem model (van Genuchten, 1980):

$$(k_\mathrm{rl}, k_\mathrm{rv}) = f_\mathrm{vGM}(m, S_\mathrm{rl}, S_\mathrm{ls})$$

where m, S_rl, and S_ls are hyper-parameters. This model has restrictions $0 \leq k_\mathrm{rl}, k_\mathrm{rv} \leq 1$, $S_\mathrm{rl} + S_\mathrm{rv} < 1$, and $S_\mathrm{rl} < S_\mathrm{ls}$.

Energy is carried by the movement of steam and water, and by thermal conduction. Hence the energy flux Q_e is given by

$$Q_\mathrm{e} = \sum_{\beta=\mathrm{l,v}} \frac{k k_{\mathrm{r}\beta}}{\nu_\beta}(\nabla p - \rho_\beta \mathbf{g}) h_\beta - K \nabla T. \tag{18}$$

Here h denotes specific enthalpy and K is the thermal conductivity in a saturated medium.

In the system of equations (16) - (18), pressure p and temperature T (or vapor saturation S_v for two phase flow) are spatially distributed quantities that represent the state of the system. Permeability k, relative permeabilities k_{rl} and k_{rv} and porosity ϕ are the parameters of interest. The thermolphysical properties of liquid water and steam (such as density, viscosity, specific internal energy, specific heat and specific enthalpy) are calculated using steam table equations given by the International Formulation Committee (1967).

Spatial discretization of (16) is based on an integrated finite difference or finite volume method, which is implemented in the existing Fortran code TOUGH2. To guarantee the numerical stability of simulation, TOUGH2 uses a fully implicit method for numerical integration in time, and upstream weighting for calculating the velocity of fluid movement between adjacent blocks (in (17) and (18)). For each time step, the Newton-Raphson method is used to solve the resulting system of non-linear difference equations. TOUGH2 uses a preconditioned iterative sparse matrix solver for solving the linear equations at each Newton-Raphson iteration (see Pruess, 1991).

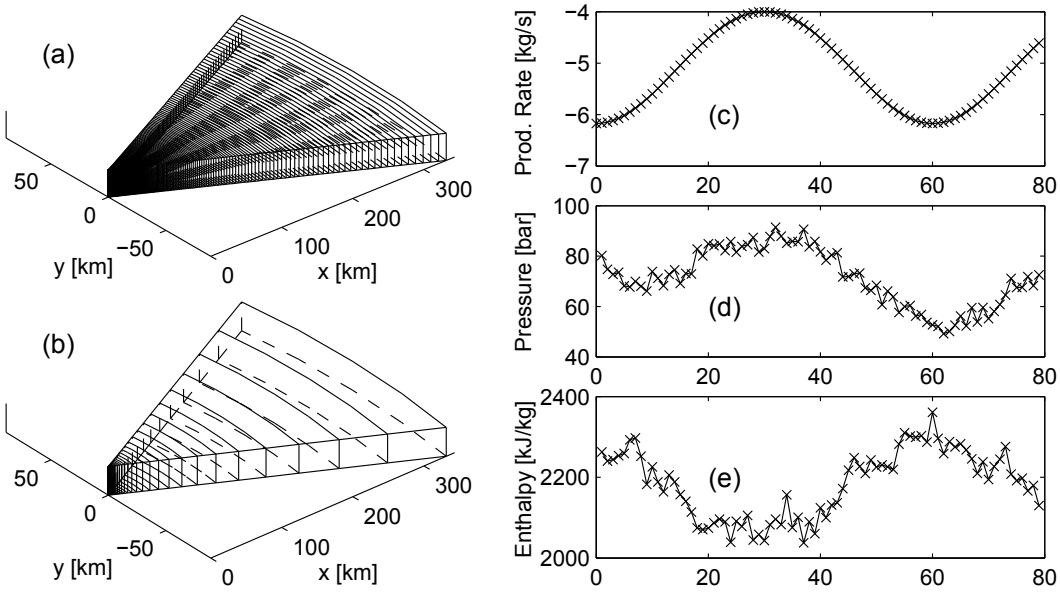

Fig. 1. Finite volume grids used for well discharge test analysis and data sets used for well discharge test. (a): the forward model (640 blocks), (b): the ROM (80 blocks), (c): the production rate (kg/seconds), (d): the pressure (bar), and (e): the flowing enthalpy (kJ/kg).

4.2. Well discharge test analysis

4.2.1. Model and parameter

In geothermal reservoir modelling, well discharge analysis are usually used to interpret the local properties of the reservoir from pressure and enthalpy data measured during a short period of field production. Based on assumption that all flows into the well come through a single layer feedzone, An one dimensional radial symmetry forward model with 640 blcoks is built to infer the near-well properties of the reservoir, as shown in the plot (a) of Figure 1. A high resolution grid is used immediately outside the wellbore and the thickness increases exponentially outside this region, where the wellbores is located in the centre of the models. The ROM is build by coarsening the grid of the forward model to a coarse grid with 80 blocks, as shown in the plot (b) of Figure 1. The CPU time of evaluating the forward model and the ROM are 2.60 and 0.29 seconds on a DELL T3400 workstation, where the computing time of the ROM is about 11% of the forward model. Since the computing time of these models are sensitive to the input parameters, they are measured from 1,000 simulations with different set of parameters.

The parameters of interest are the porosity, permeability, initial conditions and the hyperparameters in the van Genuchten-Mualem relative permeability model, as well as the initial vapour saturation (S_{v0}) and initial pressure (p_0) that are used to represent the initial thermodynamics state of the two-phase system. These make up the seven unknown parameters for the data simulation:

$$\mathbf{x} = (\phi, \log_{10}(k), p_0, S_{\mathrm{v0}}, m, S_{\mathrm{rl}}, S_{\mathrm{ls}})\,.$$

Note that the permeability k is represented on a base 10 logarithmic scale.

Table 1. Performance summary of various approximations.

Approximation	No EEM	EEM *a priori*	EEM *a posteriori*	local EEM *a posteriori*
Acc. Rate (2nd)	0.17	0.47	0.80	0.97
IACT	-	-	147.09	145.87
Avr. Jump Dist.	-	-	0.0123	0.0155

To test ADAMH, a set of synthetic data is generated over 80 days with production rates varying smoothly from about 4 kg/seconds to about 6 kg/seconds. The forward model and the model parameters $\mathbf{x} = (0.12, 1.5 \cdot 10^{-15}, 120, 0.1, 0.65, 0.25, 0.91)$ are employed to generate the synthetic data. The production rate and the noise corrupted pressure and flowing enthalpy response are shown in plots (c)-(e) in Figure 1, where the noise follow an i.i.d. Gaussian distribution with standard deviations $\sigma_{\mathrm{p}} = 3$ bar for pressure and $\sigma_{\mathrm{h}} = 30$ kJ/kg for the flowing enthalpy. This gives

$$\mathbf{\Sigma_e} = \begin{pmatrix} \sigma_{\mathrm{h}}^2 \mathbf{I}_n & 0 \\ 0 & \sigma_{\mathrm{p}}^2 \mathbf{I}_n \end{pmatrix}$$

in the likelihood function (2), and observed data $\tilde{\mathbf{d}} = (\tilde{\mathbf{d}}_{\mathrm{h}}, \tilde{\mathbf{d}}_{\mathrm{p}})^T$ comprises the observed flowing enthalpy and pressure, and n is number of measurements.

4.2.2. MCMC sampling

Fixed single-site update proposals are used to test ADAMH with Approximation 1, 2, 3 and 5. The proposals give an first step acceptance rate of 13.4% in ADAMH. We use IACT of the log-likelihood function and average jump distance to benchmark the performance of various approximations, these are summarized in Table 4.2.2. Since the Approximation 1 and 2 cannot produces sufficient mixing after 2×10^5 iterations, these chains are terminated, and only the acceptance rates in the second step are recorded. For Approximation 3 and 5, the chains are run for 2×10^6 iterations, and first 5×10^5 are discarded as burn-in steps. The output traces of log-likelihood functions and estimations of IACTs for Approximation 3 and 5 are shown in Figure 2.

Firstly, we run ADAMH without using EEM (approximate posterior (1)), the resulting Markov chain has a second step acceptance rate of about 17%, and demonstrates very poor mixing, i.e., IACT of the log-likelihood function cannot be estimated through generated samples after a large number of steps. After introducing the EEM in the prior (Approximation 2), the second step acceptance rate increases to 47%. However, this Markov chain still does not achieve desirable mixing. ADAMH with Approximation 3 significantly increases the mixing, with an estimated second step acceptance of 80%, and IACT of the log-likelihood function is 147.09. Approximation 5 with ROM (9) enhances the performance further, it achieves a second step acceptance rate of 97%, and a similar IACT of the log-likelihood function is 145.87. The method and MATLAB code presented in Wolff (2004) are employed to estimate the IACT. Since IACT cannot be accurately estimated from noise data, we also use the average jump distance to compare the performance of Approximation 3 and 5. Approximation 5 produces an average jump distance about 0.0155, which is 26% larger than 0.0123 of Approximation 3.

To estimate the speed-up factor of ADAMH, because the decrease of the statistical efficiency τ/τ^* in (14) cannot be computed as sampling by standard MH is computationally prohibitive, we can only evaluate it upper bound. In this test example, the upper bound of

the speed-up factor is about 4.1. Given the second step acceptance rates of Approximation 3 and 5 are more than 80%, the decrease of the statistical efficiency should not be significant. Especially for Approximation 5, the 97% accpetance rate in the second step suggest that the statistical efficiency of ADAMH should close to standard MH in this case.

The histograms of the model predictions computed on several different time points are given in Figure 3. We can observe that the predictions at these observation times follow uni-modal distributions that close to Gaussian. The model predictions and the 95% credible intervals over an 80-day period are shown in Figure 3, with pressure on the left and enthalpy on the right. For both predictions, the means follow the observed data reasonably well.

The histograms of the marginal distributions of the parameter $\mathbf{x}$ (see first two rows of Figure 5) show skewness in porosity and two of the hyperparameters of the van Genuchten-Mualem relative permeability model (m and S_{rl}). The number of bins in the histograms are chosen according to the Freedman-Diaconis rule (Freedman and Diaconis, 1981). The scatter plots between parameters show strong negative correlations between the permeability (on base 10 logarithmic scale) and the initial pressure, see the first plot of last row of Figure 5. There also exists strong negative correlations between the initial saturation and one of the hyperparameters of the van Genuchten-Mualem (S_{ls}), see the second plot of last row of Figure 5.

4.3. Natural state modelling

One important step in setting up a large scale geothermal reservoir model is the natural state modelling, in which the aim is the estimation of the large scale permeability structure and boundary conditions (Grant et al., 1982; O'Sullivan, 1985). Because the horizontal permeabilities are only able to be estimated locally by expensive well test analysis (in terms of investment on drilling new production/injection wells), and there is no way to infer the vertical permeability directly. The large scale permeability structure are inferred from steady state temperature distributions indicating the movement of hot fluid, which is directly influenced by the permeability structure and boundary condition. Since the porosities do not affect the solution of the set of equations (16) - (18), and the relative permeabilities only have minor impact on the solution, these parameters are not considered in the natural state modelling.

The temperature measurements are presented in Figure 9, manual calibration results and trial runs of MCMC sampling suggest that the model mis-fit has standard deviation $\sigma_T = 7.5\,^{\circ}\text{C}$. Thus, $\mathbf{\Sigma_e} = \sigma_{\text{T}}^2 \mathbf{I}_n$ is used in the likelihood function (2), where n is the number of observations.

The geological setting of geothermal reservoir model we demonstrate here is summarized in Cui (2010). The model covers a volume of 12.0 km by 14.4 km extending down to 3050 meters below the sea level. Relatively large blocks were used in the outside of the model and then they were progressively refined near the wells to achieve a well-by-well allocation to the blocks. The 3D structure of the forward model has $26,005$ blocks, which is shown in plot (a) of Figure 6, where the blue lines in the middle of the grid are wells drilled into the reservoir. To speed up the computation, a ROM based on a coarse grid with $3,335$ blocks is constructed by combining adjacent blocks in the x, y and z directions of the forward model (see plot (b) of Figure 6). A coarser level of grid resolution is not used here because further coarsening of the grid structure would produce a model that cannot reproduce the convective plume in the reservoir. Each simulation of the forward model takes about 30 to 50 minutes CPU time on a DELL T3400 workstation, and the computing time for the

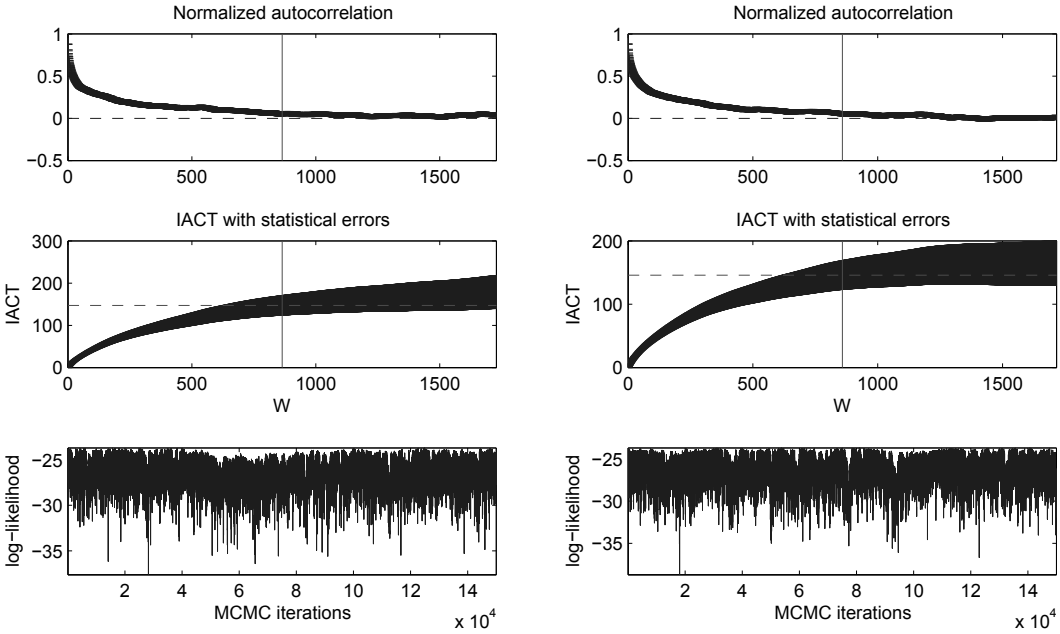

Fig. 2. IACT of the log-likelihood function, ADAMH with Approximation 3 (left column) and 5 (right column). Top row: Normalized autocorrelation, middle row: estimated IACT, bottom row: traces of log-likelihood.

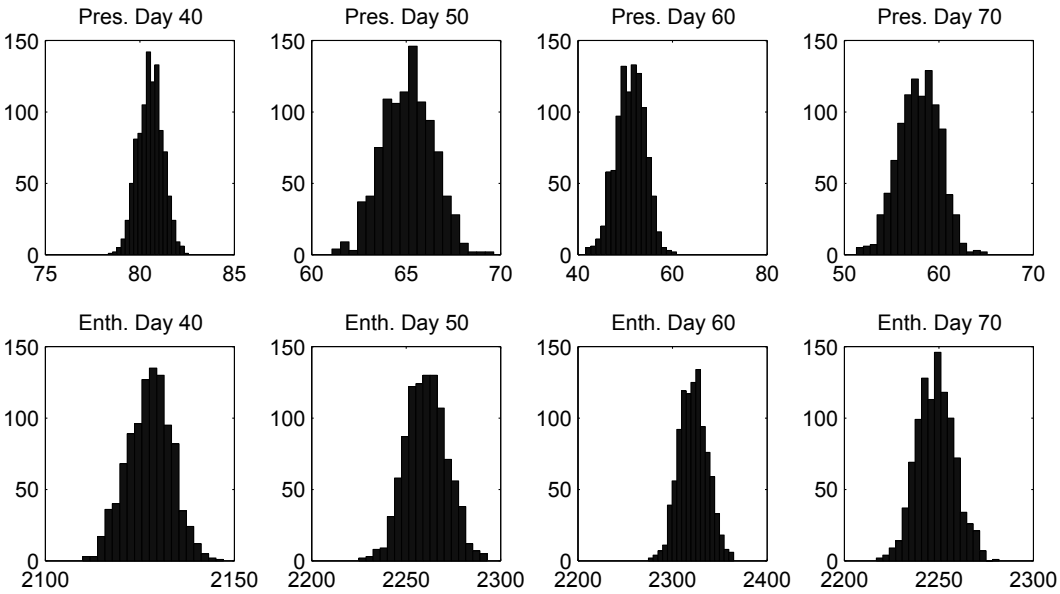

Fig. 3. Histograms of the predictive density, at day 40, 50, 60 and 70. Top row: Pressure, bottom row: enthalpy.

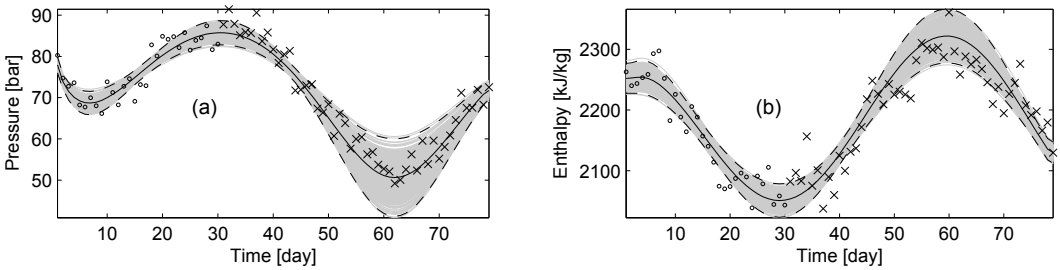

Fig. 4. Predictions for the synthesis data set. (a): pressure, (b): flowing enthalpy. The circles and the crosses are the training data and validation data, respectively; the solid line and dashed lines are the mean prediction and 95% credible interval, respectively; and the shaded lines represent the predictions made by samples.

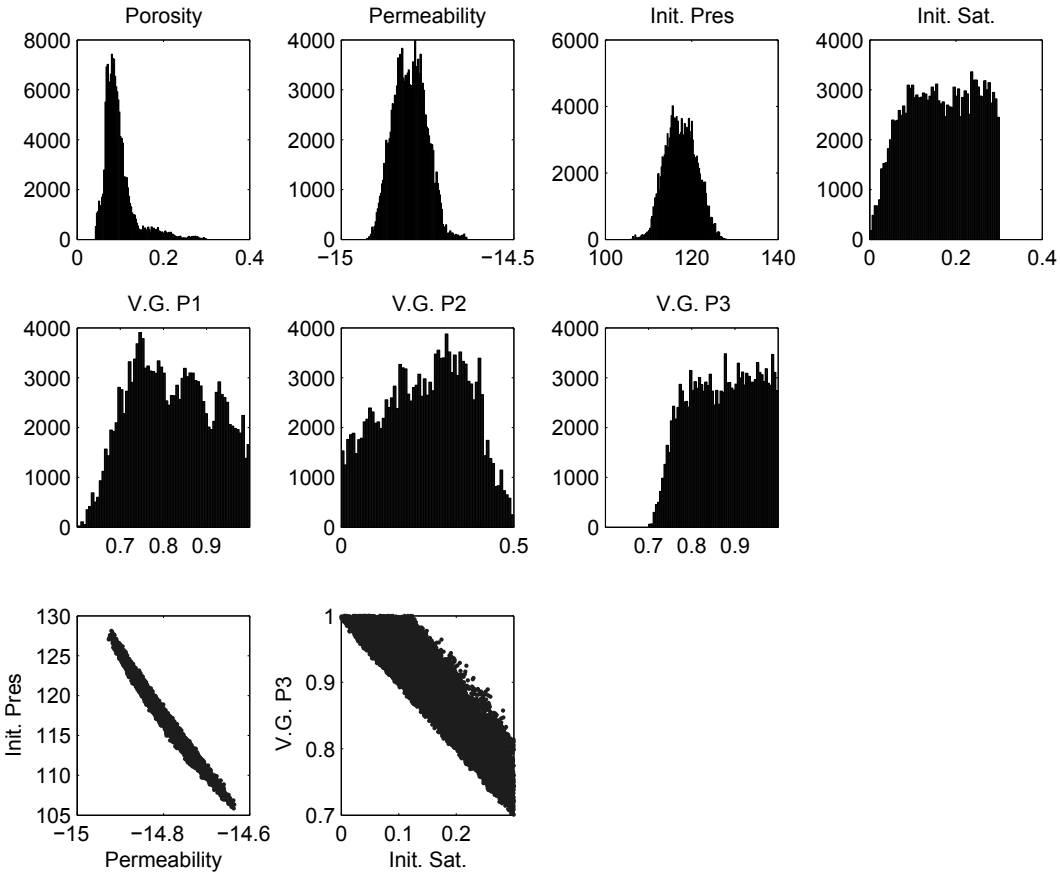

Fig. 5. Histograms of the marginal distributions and scatter plots between parameters, synthesis data set.

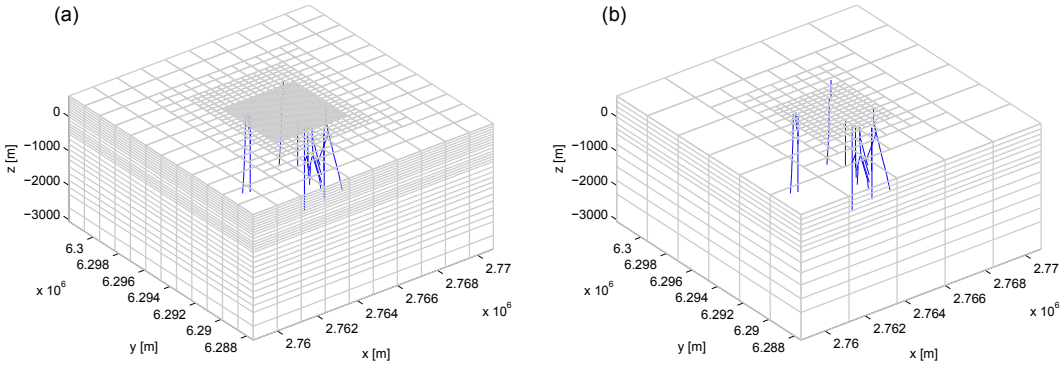

Fig. 6. The fine grid (left) and the coarse grid (right) used for natural state modelling.

ROM is about 1 to 1.5 minutes (roughly 3% of the forward model). The computing time for these models is sensitive to the input parameters.

4.3.1. Prior modelling of permeabilities

Previously, the rock structure of the model was pre-assigned by a geologist, and each rock type covers a range of grid blocks. From manual calibration of the model, we found that the permeability may not have a very good correspondence with the type of rock, and it has a large variation depending on the location. To capture these variabilities, we employ a low-level pixel based representation with the same resolution as the coarse grid used by the ROM. Therefore, the permeabilities in x y, and z directions are represented by $N = 3,335$ dimensional vectors $\mathbf{k}^{(\zeta)} = [\log_{10}(k_1^{(\zeta)}), \ldots, \log_{10}(k_N^{(\zeta)})], \zeta \in \{x, y, z\}$.

Because there is no point measurement and other direct survey (e.g., seismic data) of the permeabilities available for this reservoir, we assume the permeabilities in x, y and z directions are uncorrelated. A first order Gaussian Markov random field (GMRF) model (Rue and Held, 2005) is employed to formulate the prior distribution for each of the x, y, and z direction of permeabilities. Let $i \sim j$ denote that two blocks i and j are adjacent to each other. The prior has the form of

$$\pi(\mathbf{k}^{(\zeta)}) \propto \exp\left\{-\delta_\zeta \sum_{i\sim j} \omega_{ij} \left[\log_{10}(k_i^{(\zeta)}) - \log_{10}(k_j^{(\zeta)})\right]^2\right\}, \tag{19}$$

where ω_{ij} is the inverse Euclidean distance between the two adjacent block centers, and δ_ζ is a hyperparameter controls the smoothing (has different meaning with the δ in Section 3). We take the hyperparameter $\delta_\zeta = 0.5$ for all $\zeta \in \{x, y, z\}$, a value suggested by the trial runs of MCMC. With this setting, the posterior distribution is still dominated by the likelihood function, and hence the model parameters are determined by the data. The prior provides a minor constraint to enforce the condition that the permeabilities are smooth.

4.3.2. Prior modelling of boundary conditions

The top of the model is assumed to be "open", which allows the model to have direct connection with the atmosphere. Atmosphere pressure and temperature are used as the

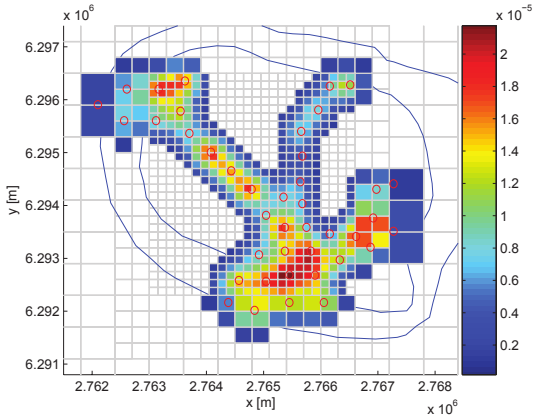

Fig. 7. Location of control points of the RBF model for mass input distribution (red circles) and one realization of the distribution of hot water injection at the base of the model.

boundary conditions at the top of the model. The model covers a sufficiently large area so that the flows through the side of the boundary are negligible in the natural state modeling, and hence the sides of the model are treated as no-flow boundaries.

At the base of the model, a distribution of very hot water is injected to represent the upflow from depth, which also has to be estimated. A radial basis function (RBF) with kernel function

$$C(\mathbf{s}, \mathbf{s}'; r) = \exp\left[-\left(\frac{\|\mathbf{s} - \mathbf{s}'\|}{r}\right)^2\right]$$

is used to represent the distribution of the injected hot water. Where $\mathbf{s} = (x, y)$ denotes the location of the control points. Because the distribution of the injected hot water is unknown, we use a set of evenly spaced control points, with fixed locations to parametrize the distribution. These control points spans a sufficiently large area to cover all the possible locations of injected hot water as indicated by the location and orientation of known faults in the system (M. J. O'Sullivan, unpublished report, 2006). The hyperparameter r is set to have a fixed value of 300 meters, which is empirically adjusted so that each control points has a sufficiently large coverage area, and the overall distribution does not have any spikes. The set of M control points $\mathbf{S} = \{\mathbf{s}_i\}_{i=1}^{M}$ and one realization of the mass input distribution are shown in Figure 7. The blue circles around the control points are the resistivity boundaries obtained from geological survey, which indicate the hot region of the reservoir.

In a full Bayesian way, a hierarchical posterior should be used to model the hyperparameter r in the RBF function, the standard deviation σ_T in the likelihood function, and the hyperparameter δ_ζ used in the prior model of permeabilities. Then, MCMC sampling could be used to either estimate or marginalize these parameters. However, correlation between these parameter and other model parameters could make this process computationally infeasible, because the resulting hierarchical posterior is usually difficult to sample from, and hence requires a large number of iterations in MCMC sampling. Since a fully Bayesian treatment of these parameters is still limited by the computational power available, instead in this research, we use empirical estimates of these quantities that are deduced from either expert judgment or previous trial runs of MCMC. The limited computational power is

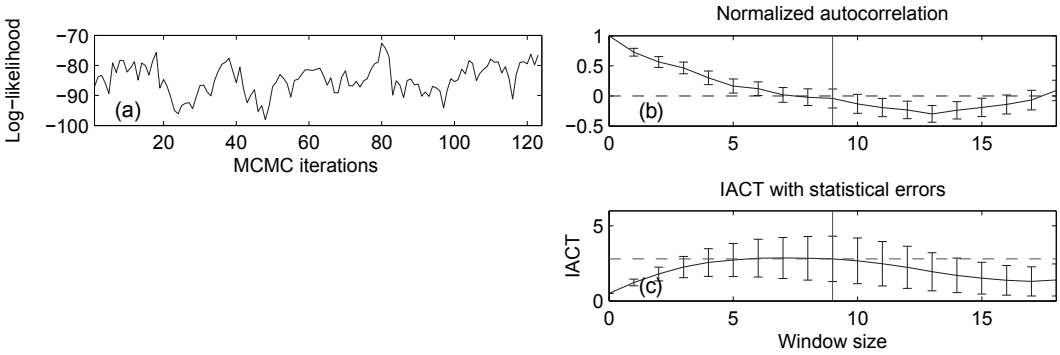

Fig. 8. Trace plot of the log-likelihood function and the integrated autocorrelation time, after discarding burn-in steps. (a): Trace plot of the log-likelihood function, after discarding burn-in steps. (b): Normalized autocorrelation plot of the log-likelihood function. (c): Estimated integrated autocorrelation time with statistical error (Wolff, 2004).

allocated to exploring the model parameters.

By assigning weights $\mathbf{w} = \{w_i\}_{i=1}^{M}$ to each of the control points, the distribution of the water injection (with unit $[\mathrm{kg}/(\mathrm{m}^2 \cdot \mathrm{s})]$) for a given point $\mathbf{s}'$ is

$$f(\mathbf{s}' \mid \mathbf{w}, \mathbf{S}, r) = \sum_{i=1}^{M} w_i C(\mathbf{s}_i, \mathbf{s}'; r). \tag{20}$$

The total amount of mass is a fixed value about 105kg/s, which is deduced from the surface flow data of the reservoir. Integrating Equation (20) gives the constraint on $\mathbf{w}$, which is $\sum_{i=1}^{M} w_i = 3.7136 \cdot 10^4$. Also, all the weights w_i must be non-negative to ensure that the distribution of mass injection (20) is non-negative.

4.3.3. MCMC sampling

Overall, we have the parameters $\mathbf{x} = \left\{\mathbf{k}^{(x)}, \mathbf{k}^{(y)}, \mathbf{k}^{(z)}, \mathbf{w}\right\}$ to be estimated. The permeabilities are separated into 68 groups, and a modified adaptive Metropolis proposal are used for each of these groups. The weights of the distribution of the water injection $\mathbf{w}$ are adjusted separately by an adaptive reversible jump move. These proposals produce a first step acceptance rate of 0.1 in ADAMH, for details of these proposals see Cui (2010). Since the forward model is computationally very demanding, we first simulate the chain with the ROM only for about 200 sweeps of updates (each sweep of updates consists of 69 iterations to update all the elements of the parameter) to get through the initial burn-in period. Then, we start ADAMH with Approximation 5 and the ROM (9) to sample from the posterior distribution based on the forward model. Note that the adaptation of the proposals are started at the beginning of the sampling of the posterior distribution based solely on the ROM.

We are able to sample the posterior distribution for about $11,200$ iterations in 40 days, and ADAMH achieves about 74% accpetance rate in the second step. The estimated upper bound of the speed-up factor is about 7.7. The trace plot of the log-likelihood function and estimation of its IACT are shown in Figure 8, where the first 40 sweeps of updates are discarded as burn-in steps. It was obtained by using the MATLAB code provided by

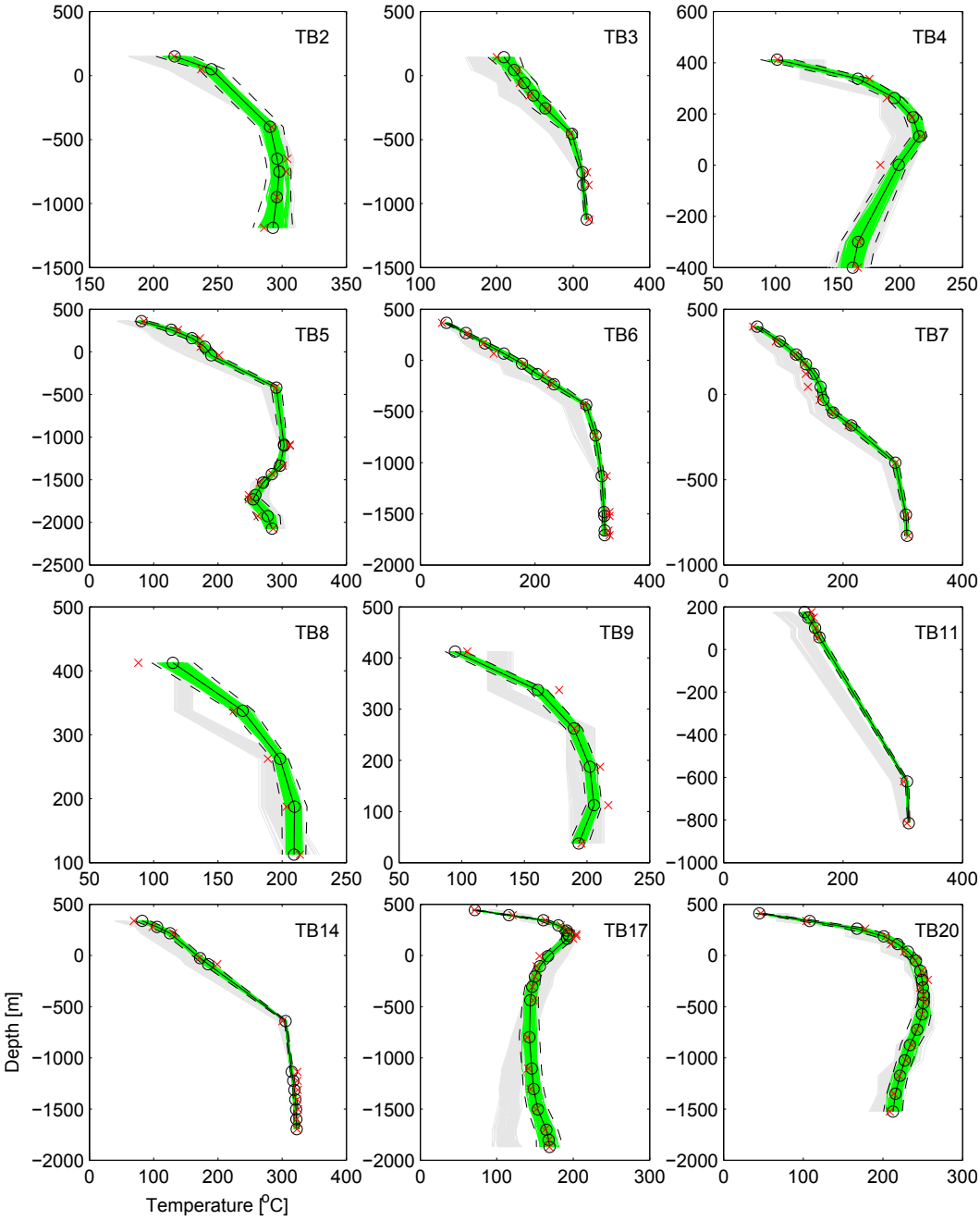

Fig. 9. Comparison of estimated temperatures and measured data. The solid black lines are the estimated mean temperatures, dashed black lines are the 95% percent credible interval, and measured data are shown as red crosses. The green and gray lines represent outputs of the foaward model and the ROM various realizations, respectively.

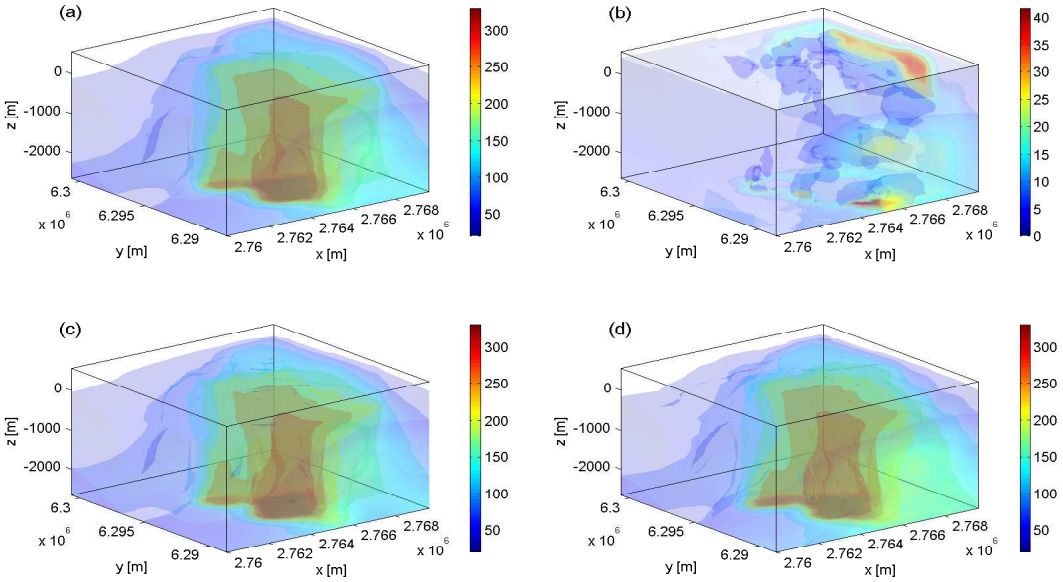

Fig. 10. Distributions of model temperatures, unit is ^oC. (a): the mean realizations, (b): standard deviations of realizations, (c): one realization from the Markov chain, and (d): another realization from the Markov chain.

Wolff (2004). The IACT the log-likelihood function is about 2.80, which is only a rough estimation because the chain has not been running long enough. Thus, we roughly have one independent sample for every 5.6 sweeps of updates.

The mean and standard deviation of the temperature profiles are estimated as sample averages over model realizations. We compare these estimations with the measured data in Figure 9. The solid black lines are the estimated mean temperatures, dashed black lines are the 95% percent credible interval, and measured data are shown as red crosses. The green and gray lines represent outputs of the foaward model and the ROM various realizations, respectively. Figure 9 shows that the forward model and the ROM produce significant different temperature profiles. We can observe that there exist structure error between the forward model outputs and the ROM outputs, and the forward model is hotter than the ROM in average. The mean model reduction error at each of the measurement position span a range of $[-33.64, 54.98]$, and hence the noise level of these model reduction errors are more significant than the zero mean normal distribution with standard deviation $\sigma_e = 7.5\,^{\circ}\mathrm{C}$ used in the likelihood function. Hence, the statistical modeling of the model reduction error in Section 3 is essential for the efficient use of the ROM in this research.

The temperature distributions are shown in Figure 10, where (a) shows the mean temperature distribution, the standard deviation of the temperature distribution is shown in (b), and the temperature distributions of the two realizations from the Markov chain (one from the middle and another from the end) are presented in (c) and (d). From these plots, we can observe that there is one hot plume in the model, and the top and the bottom of this hot plume have larger variations in temperature than the rest of the model.

Most of the estimated model temperatures agree with the measured data very well,

the root mean square error is in the range of $[7.11, 8.66]$, with mean value of 7.22. The assumption that the residuals follow a Gaussian distribution with standard deviation $\sigma_e = 7.5\,^{\circ}\mathrm{C}$ is validated by the normal quantile plot and the cumulative density plot of the residuals (see Figure 11). The normal quantile plot shows that the residuals are symmetric and roughly follow the Gaussian distribution in the interval $[-10, 10]$, where most of them are located. Outside of this range, the residuals are slightly skewed (light tailed below -10 and heavy tailed above 10). The same result is suggested by the cumulative density plot. Overall, these plots show that the normality assumption is reasonable.

We also present the mean estimations, standard deviations and two realizations of the distribution of mass input and permeability distribution in the x, y, and z directions in Figure 12 - 15. In these figures, (a) is the mean estimation, (b) is the standard deviation, and two realizations are shown in (c) and (d).

The mean distribution of the mass input shows that most of the injected hot water occurs at the bottom of the inner resistivity boundary. This corresponds to the opinion of geologists that the major source of deep hot water occurs at the intersection of two geological faults in this area. The standard deviation and two realizations of this distribution suggest that the variation among the samples is very small.

The reasonably large variations in the reconstructed permeability distributions suggest that there exists ambiguities in the permeability distributions, and this may be caused by the sparsely measured data. This effect is more significant in the region close to the boundary of the reservoir, where no measured data are available. We can either give more accurate quantifications to these ambiguities by running the chain for a very long time, or remove these ambiguities by imposing a stronger prior distribution or using different parameterizations of the permeability distributions.

5. Discussion

We developed an adaptive delayed acceptance Metropolis-Hastings algorithm for efficient sampling the computationally demanding posterior distribution occurs in inverse problems. ADAMH improves the computational efficiency of the standard MH algorithm by employing an approximate posterior distribution based on a reduced order model. Components of the delayed acceptance MH algorithm (Christen and Fox, 2005), the enhanced error model of Kaipio and Somersalo (2007), and adaptive MCMC (Haario et al., 2001; Roberts and Rosenthal, 2007) are used to construct ADAMH. By using the regularity conditions provided in Section 3, several adaptive approximate posterior distributions are designed.

As shown in the first example, ADAMH improves the accuracy of the approximation significantly without extra computing time by using EEM and adaptive sampling, and hence improve the computational efficiency of sampling. The state dependent approximation 5 with the ROM (9) have the best performance over all approximations. ADAMH with Approximation 5 is used to sample the posterior distribution of the large scale 3D geothermal reservoir model in the second example. We are able to run the MCMC sampling for about $11,200$ iterations in about 40 days, which would takes about 10 months for the standard MH algorithm. The significant speed-up offered by ADAMH would allow MCMC method to be applied to inverse problems that on the limitation of the computing power.

The ROM we used here is based on coarsening the grid structure of the forward model. Even through the ROM (9) is state dependent by adding a local correction term, the overall structure of the ROM is fixed during the sampling. Based on the conditions supplied in

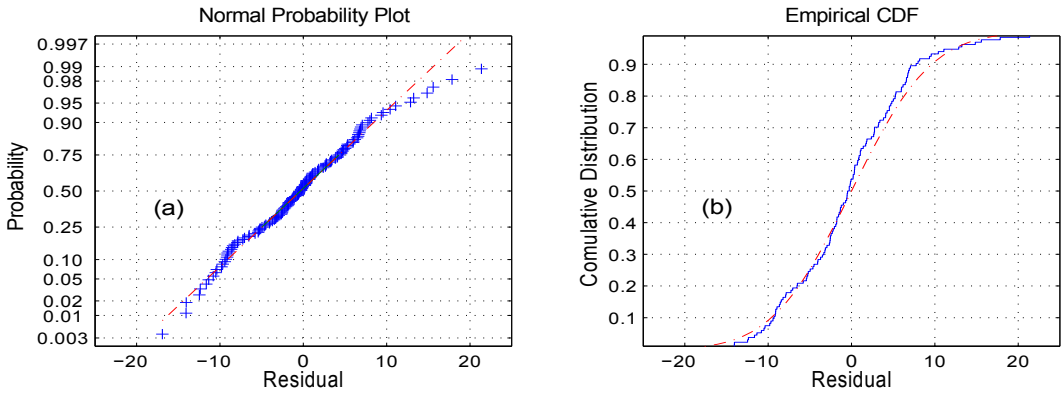

Fig. 11. (a): Normal quantile plot of the residuals. (b): Cumulative density plot of the residuals, the solid line is the estimated cumulative density function, and the dashed line indicates the Gaussian distribution used in the assumption.

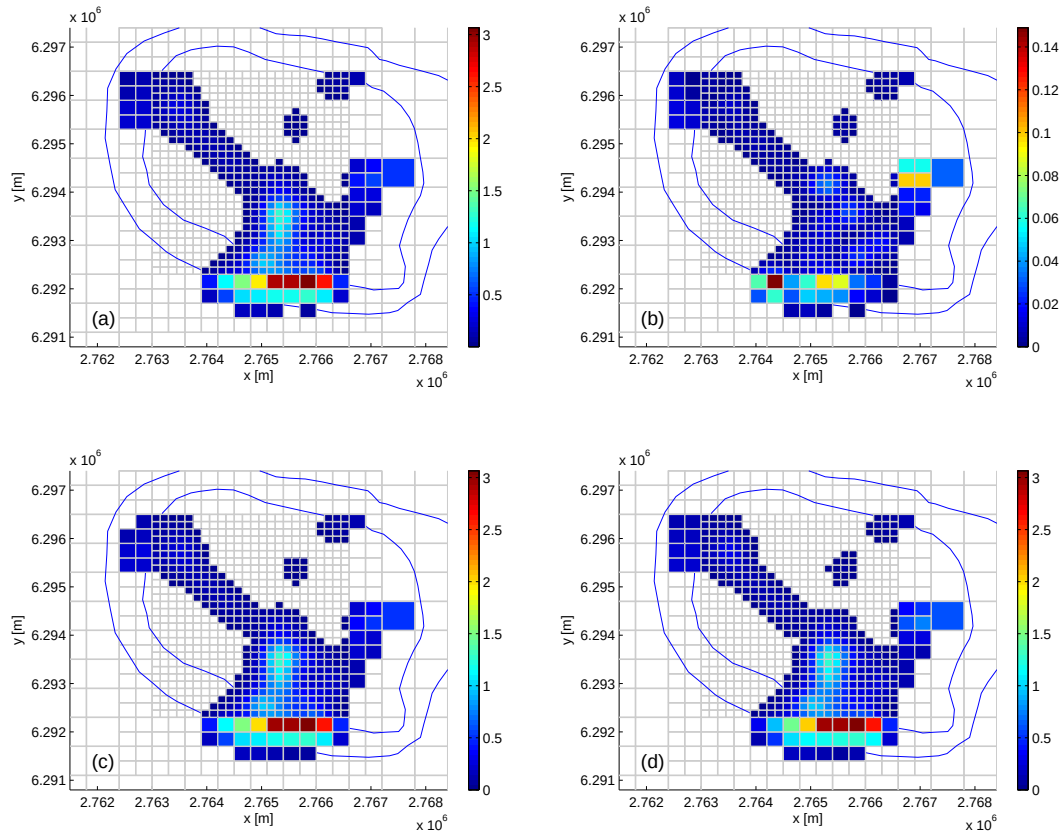

Fig. 12. Distributions of mass input at the bottom of the model, unit in kg/s. (a): the mean realizations, (b): standard deviations of realizations, (c): one realization from the Markov chain, and (d): another realization from the Markov chain.

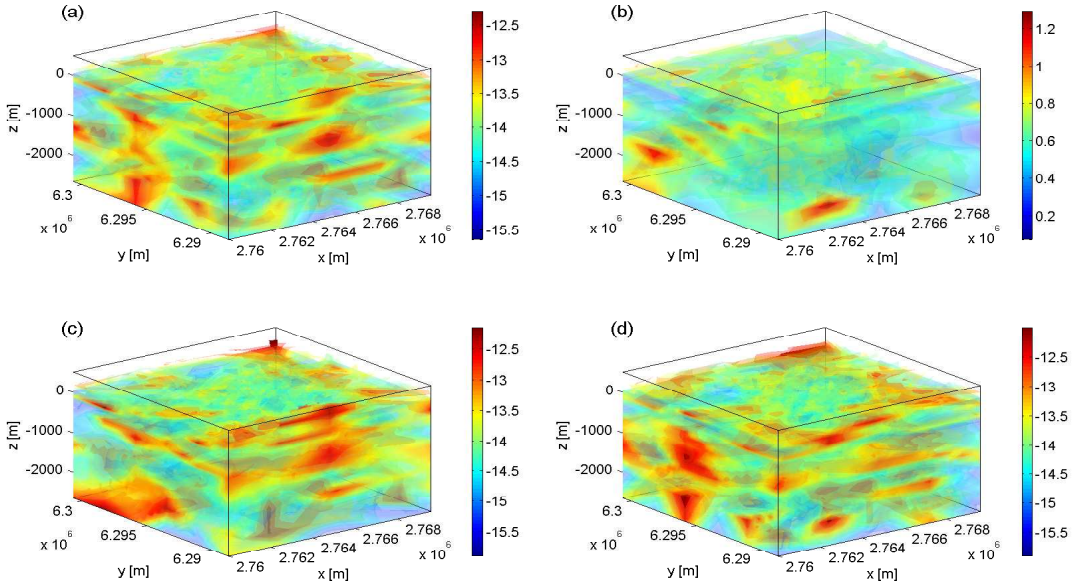

Fig. 13. The permeability distribution on x direction, in base 10 logarithmic scale. (a): the mean realizations, (b): standard deviations of realizations, (c): one realization from the Markov chain, and (d): another realization from the Markov chain.

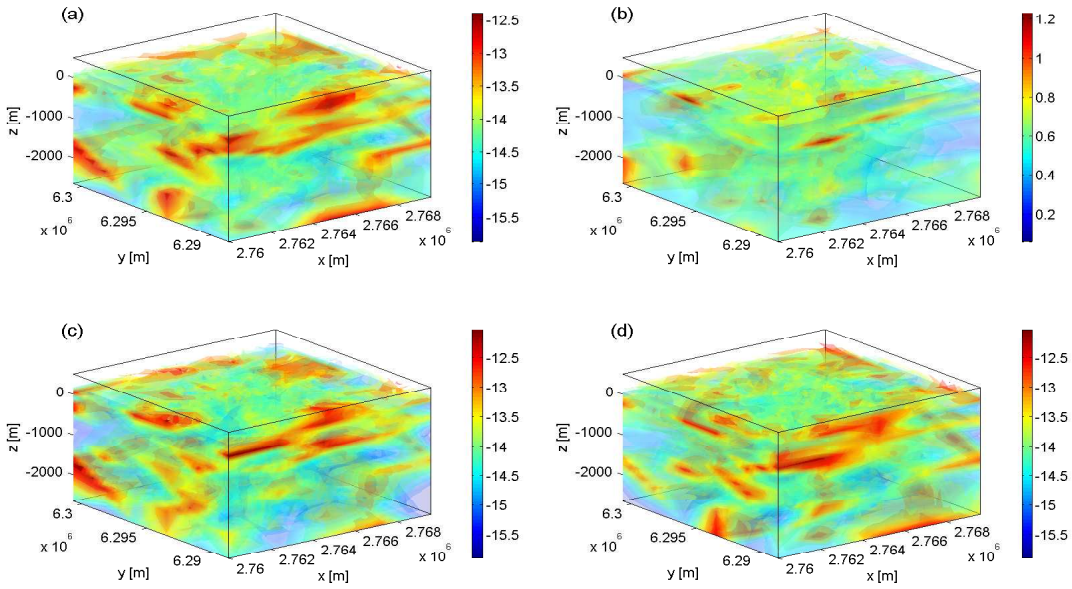

Fig. 14. The permeability distribution on y direction, in base 10 logarithmic scale. (a): the mean realizations, (b): standard deviations of realizations, (c): one realization from the Markov chain, and (d): another realization from the Markov chain.

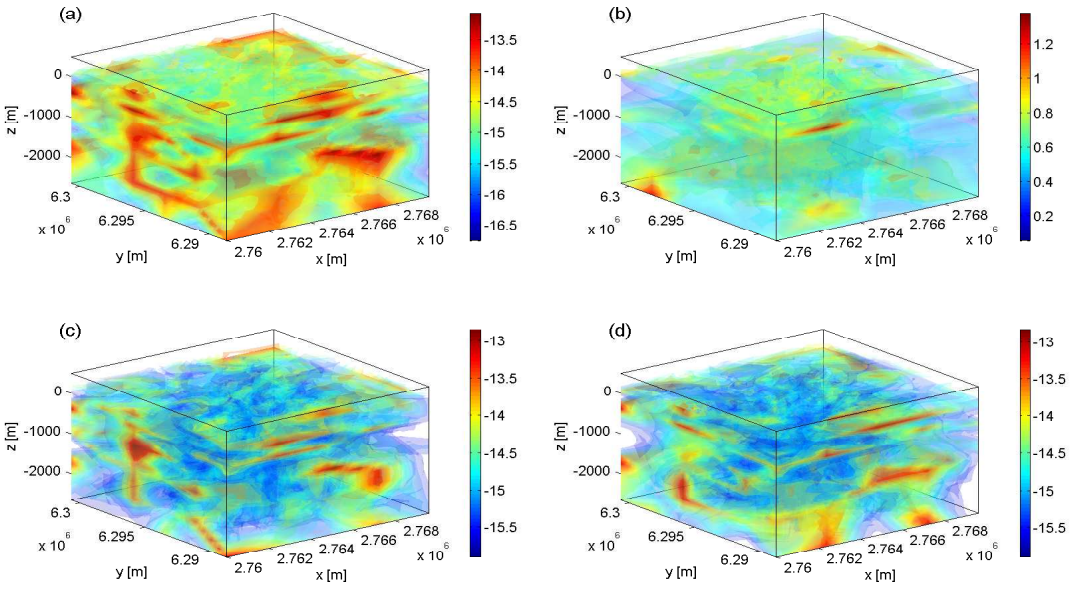

Fig. 15. The permeability distribution on z direction, in base 10 logarithmic scale. (a): the mean realizations, (b): standard deviations of realizations, (c): one realization from the Markov chain, and (d): another realization from the Markov chain.

Theorem 1, other more sophisticated model reduction techniques such as projection based methods (Grimme, 1997; Li, 2002) can be used to construct ROMs that adapts to the local structure of the forward model. Further investigations are necessary to integrate these techniques into ADAMH.

References

Amit, Y., U. Grenander, and M. Piccioni (1991). Structural image restoration through deformable templates. *Journal of the American Statistical Association 86(414)*, 376–387.

Andersen, K. E., S. P. Brooks, and M. B. Hansen (2004). Bayesian inversion of geoelectrical resistivity data. *Journal of the Royal Statistical Society: Series B 65*, 619–644.

Andrieu, C. and E. Moulines (2006). On the ergodicity properties of some adaptive Markov chain Monte Carlo algorithms. *Annals of Applied Probability 16(1)*, 462–505.

Arak, T., P. Clifford, and D. Surgailis (1993). Point based polygonal models for random graphs. *Advances in Applied Probabiity 25*, 348–372.

Arridge, S. R., J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen (2006). Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems 22*, 175–195.

Atchade, Y. F. and J. S. Rosenthal (2005). On adaptive Markov chain Monte Carlo algorithms. *Bernoulli 11*, 815–828.

Baddeley, A. J. and M. N. M. van Lieshout (1993). Stochastic geometry models in high-level vision. In K. V. Mardia (Ed.), *Statistics and Images*, pp. 233–258. Carfax, Abingdon.

Bayarri, M. J., J. O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R. Parthasarathy, R. Paulo, J. Sacks, and D. Walsh (2007). Computer model validation with function output. *Annals of Statistics 35(5)*, 1874–1906.

Besag, J. E. (1986). On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B 48*, 259–302.

Besag, J. E. and C. Kooperberg (1995). On conditional and intrinsic autoregressions. *Biometrika 82(4)*, 733–746.

Besag, J. E., J. York, and A. Mollie (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics 43*, 1–59.

Besag, J. E..and Green, P. J., D. Higdon, and K. Mengersen (1995). Bayesian computation and stochastic systems. *Statistical Science 10*, 3–41.

Christen, J. A. and C. Fox (2005). MCMC using an approximation. *Journal of Computational and Graphical Statistics 14(4)*, 795–810.

Christie, M. A. and M. J. Blunt (2001). Tenth spe comparative solution project: A comparison of upscaling techniques. *SPE Reservoir Engineering and Evaluation 4*, 308–317.

Cornford, D., L. Csat, D. J. Evans, and M. Opper (2004). Bayesian analysis of the scatterometer wind retrieval inverse problem: Some new approaches. *Journal of the Royal Statistical Society. Series B 66(3)*, 609–652.

Cressie, N. A. C. (1993). *Statistics for Spatial Data, Revised Edition*. Wiley-Interscience.

Cui, T. (2010). *Bayesian Calibration of Geothermal Reservoir Models via Markov Chain Monte Carlo*. Ph. D. thesis, The University of Auckland.

Efendiev, Y., A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick (2005). An efficient two-stage Markov chain Monte Carlo method for dynamic data integration. *Water Resources Research 41*.

Freedman, D. and P. Diaconis (1981). On the histogram as a density estimator: l_2 theory. *Probability Theory and Related Fields 57(4)*, 453–476.

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.

Gramacy, R. B. and H. K. H. Lee (2008). Bayesian treed gaussian process models with an application to computer modeling. *Journal of the American Statistical Association 103(483)*, 1119–1130.

Grant, M. A., I. G. Donaldson, and P. F. Bixley (1982). *Geothermal Reservoir Engineering*. Academic Press.

Green, P. J. (1990). Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging 9(1)*, 84–93.

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Grenander, U. (1967). *Lectures in Pattern Theory. Volume 1: Pattern Synthesis*, Volume 24 of *Applied Mathematical Sciences*. Berlin: Springer-Verlag.

Grenander, U. (1976). *Lectures in Pattern Theory. Volume 2: Pattern Analysis*, Volume 24 of *Applied Mathematical Sciences*. Berlin: Springer-Verlag.

Grenander, U. (1981). *Lectures in Pattern Theory. Volume 3: Regular Structures*, Volume 33 of *Applied Mathematical Sciences*. Berlin: Springer-Verlag.

Grenander, U. and M. I. Miller (1994). Representations of knowledge in complex systems. *Journal of the Royal Statistical Society. Series B 56(4)*, 549–603.

Grimme, E. J. (1997). *Krylov projection methods for model. reduction*. Ph. D. thesis, University of Illinois Urbana Champaign.

Haario, H., M. Laine, M. Lehtinen, Saksman, E., and J. Tamminen (2004). Markov chain Monte Carlo methods for high dimensional inversion in remote sensing. *Journal of the Royal Statistical Society: Series B 66*, 591–608.

Haario, H., M. Laine, M. Lehtinen, E. Saksman, and J. Tamminen (2004). Markov chain Monte Carlo Methods for high dimensional inversion in remote sensing. *Journal of the Royal Statistical Society. Series B 66(3)*, 591–607.

Haario, H., M. Laine, A. Mira, and E. Saksman (2006). Dram: Efficient adaptive MCMC. *Statistics and Computing 16*, 339–354.

Haario, H., E. Saksman, and J. Tamminen (2001). An adaptive Metropolis algorithm. *Bernoulli 7*, 223–242.

Hadamard, J. (1902). Sur les problemes aux derivees partielles et leur signification physique. [on the problems about partial derivatives and their physical significance]. *Princeton University Bulletin 13*, 49–52.

Hastings, W. (1970). Monte Carlo sampling using Markov chains and their applications. *Biometrika 57*, 97–109.

Higdon, D. (1998). A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics 5(2)*, 173–190.

Higdon, D., J. E. Bowsher, V. E. Johnson, T. G. Turkington, D. R. Gilland, and R. J. Jaszczak (1997). Fully bayesian estimation of gibbs hyperparameters for emission computed tomography data. *IEEE Transactions on Medical Imaging 16(5)*, 516–526.

Higdon, D., J. Gattiker, B. Williams, and M. Rightley (2008). Computer model calibration using high dimensional output. *Journal of the American Statistical Association 103(482)*, 570–583.

Higdon, D., H. Lee, and Z. B. (2002). A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information. *IEEE Transactions on Signal Processing 50(2)*, 389–399.

Higdon, D., H. Lee, and C. Holloman (2003). Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (Eds.), *Bayesian Statistics 7*, pp. 181–197. Oxford University Press.

Hurn, M. A., O. Husby, and H. Rue (2003). Advances in Bayesian image analysis. In P. J. Green, N. L. Hjort, and S. Richardson (Eds.), *Highly Structured Stochastic Systems*, pp. 301–322. Oxford University Press.

Husby, O., T. Lie, T. Lango, J. Hokland, and H. Rue (2001). Bayesian 2-D deconvolution: a model for diffuse ultrasound scattering. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 48(1)*, 121–130.

International Formulation Committee (1967). *A Formulation of the Thermodynamic Properties of Ordinary Water Substance*. IFC Secretariat.

Jain, A. K., Y. Zhong, and M. P. Dubuisson-Jolly (1998). Deformable template models: A review. *Signal Processing 71(2)*, 109–129.

Kaipio, J. and E. Somersalo (2004). *Statistical and Computational Inverse Problems*. Springer-Verlag.

Kaipio, J. P. and E. Somersalo (2007). Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics 198(2)*, 493–504.

Kennedy, M. C. and A. O'Hagan (2001). Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society: Series B 63*, 425–464.

Kuensch, H. R. (1987). Intrinsic autoregressions and related models on the two-dimensional lattice. *Biometrika 74*, 517–524.

Li, J. R. (2002). *Model reduction of large linear systems via low rank system Gramians*. Ph. D. thesis, MIT.

Lieberman, C., K. Willcox, and O. Ghattas (2010). Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing 32(5)*, 2523–2542.

Liu, J. S. (2001). *Monte Carlo Strategies in Scientific Computing*. Springer-Verlag.

McKeague, I. W., Nicholls, G. K., K. Speer, and R. Herbei (2005). Statistical inversion of South Atlantic circulation in an abyssal neutral density layer. *Journal of Marine Research 63*(4), in press.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *Journal of chemical physics 21*, 1087–1092.

Moller, J. and R. P. Plenge Waagepetersen (1998). Markov connected component fields. *Advances in Applied Probabiity 30(1)*, 1–35.

Moller, J. and O. Skare (2001). Coloured voronoi tessellations for bayesian image analysis and reservoir modelling. *Statistical Modelling 1*, 213–232.

Mondal, A., Y. Efendiev, B. Mallick, and A. Datta-Gupta (2010). Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump markov chain monte carlo methods. *Advances in Water Resources 33(3)*, 241–256.

Morris, M. D., T. J. Mitchell, and D. Ylvisaker (1993). Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction. *Technometrics 35*, 243–255.

Nicholls, G. K. (1998). Bayesian image analysis with Markov chain Monte Carlo and coloured continuum triangulation mosaics. *Journal of the Royal Statistical Society: Series B 60*, 643–659.

Nicholls, G. K. and C. Fox (1998). Prior modelling and posterior sampling in impedance imaging. In *Bayesian Inference for Inverse Problems*, Volume 3459, pp. 116–127. SPIE.

Oliver, D. S., L. B. Cunha, and A. C. Reynolds (1997). Markov chain Monte Carlo methods for conditioning a permeability field to pressure data. *Mathematical Geology 29(1)*, 61–91.

O'Sullivan, M. J. (1985). Geothermal reservoir simulation. *International Journal of Energy Research 9(3)*, 319–332.

Pruess, K. (1991). *TOUGH2 - A General-Purpose Numerical Simulator for Multiphase Fluid and Heat Flow*. Berkeley, California: Lawrence Berkeley National Laboratory.

Roberts, G. O., A. Gelman, and W. R. Gilks (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *Annals of Applied Probability 7*, 110–120.

Roberts, G. O. and J. S. Rosenthal (2001). Optimal scaling for various Metropolis-Hastings algorithms. *Statistical Science 16*, 351–367.

Roberts, G. O. and J. S. Rosenthal (2007). Coupling and ergodicity of adaptive MCMC. *Journal of Applied Probability 44*, 458–475.

Roberts, G. O. and J. S. Rosenthal (2009, Jun). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics 18(2)*, 349–367.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*. Chapman & Hall.

Rue, H. and M. A. Hurn (1999). Bayesian object identification. *Biometrika 86(3)*, 649–660.

Sacks, J., W. J. Welch, T. J. Mitchell, and H. P. Wynn (1989). Design and analysis of computer experiments (c/r: p423-435). *Statistical Science 4*, 409–423.

Saksman, E. and M. Vihola (2010). On the ergodicity of the adaptive metropolis algorithm on unbounded domains. *Annals of Applied Probability 20(6)*, 2178–2203.

Santner, T., B. Williams, and W. Notz (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag.

Ter Braak, C. J. F. (2006). A Markov chain Monte Carlo version of the genetic algorithm differential evolution: easy Bayesian computing for real parameter spaces. *Statistics and Computing 16(3)*, 239–249.

Turro, E., N. Bochkina, A. M. K. Hein, and S. Richardson (2007). Bgx: a bioconductor package for the bayesian integrated analysis of affymetrix genechips. *BMC Bioinformatics 8(439)*.

van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Science Society of America Journal 44*, 892–898.

Vauhkonen, P., M. Vauhkonen, T. Savolainen, and J. Kaipio (1999). Three-dimensional electrical impedance tomography based on the complete electrode model. *IEEE Transactions on Biomedical Engineering 46*, 1150–1160.

Vihola, M. (2011). Can the adaptive metropolis algorithm collapse without the covariance lower bound? *Electronic Journal of Probability 16*, 45–75.

Vrugt, J. A., C. J. F. ter Braak, M. P. Clark, J. M. Hyman, and B. A. Robinson (2008). Treatment of input uncertainty in hydrologic modeling: doing hydrology backward with Markov chain Monte Carlo simulation. *Water Resource Research 44*.

Watzenig, D. and C. Fox (2009). A review of statistical modelling and inference for electrical capacitance tomography. *Measurement Science and Technology 20(5)*.

Weir, I. S. (1997). Fully bayesian reconstructions from single-photon emission computed tomography data. *Journal of the American Statistical Association 92*, 49–60.

Welch, W. J., R. J. Buck, J. Sacks, H. P. Wynn, T. J. Mitchell, and M. D. Morris (1992). Screening, predicting, and computer experiments. *Technometrics 34*, 15–25.

Wolff, U. (2004). Monte Carlo errors with less errors. *Computer Physics Communications 156(2)*, 143–153.