

THE USE OF TECHNIQUES WITHIN THE BAYESIAN FRAMEWORK OF STATISTICS FOR THE SOLUTION OF INVERSE PROBLEMS

Helcio R. B. Orlande

Department of Mechanical Engineering

Escola Politécnica/COPPE

Federal University of Rio de Janeiro, UFRJ

Rio de Janeiro, RJ, Brazil

helcio@mecanica.coppe.ufrj.br



OUTLINE

-
- **Introduction**
 - **General considerations**
 - **Bayesian framework**
 - **Maximum a Posteriori Objective Function**
 - **Markov Chain Monte Carlo (MCMC) methods**
 - **Practical issues for MCMC methods**
 - **Reduction of the computational time for MCMC methods**
 - **Conclusions**

INTRODUCTION

Inverse heat transfer problems deal with the estimation of unknown quantities appearing in the mathematical formulation of physical processes in thermal sciences, by using measurements of temperature, heat flux, radiation intensities, etc.

INTRODUCTION

The term Bayesian is commonly used to refer to techniques for the solution of inverse problems that fall within the framework of statistics developed by the Presbyterian minister Rev. Thomas Bayes (1702 - 1761) [1].

P H I L O S O P H I C A L
T R A N S A C T I O N S :

An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.

Mr. Bayes and Mr. Price

Phil. Trans. 1763 **53**, 370-418, published 1 January 1763

P R O B L E M.

Given the number of times in which an unknown event has happened and failed: *Required* the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named.

INTRODUCTION

- It is attributed to Laplace the mathematical formulation that is known today as Bayes' theorem [3].
- The term Bayesian was first used by R. A. Fisher, but in a pejorative context. Although born more than 120 years after the death of Bayes, Fisher was Bayes biggest intellectual rival [3]. The major issue by Fisher against Bayes and Laplace was that they used the concept of a prior probability, which represents the information about an unknown quantity before the measured data is available [3].
- Fisher's theory relies solely on the measured data and on modelling of their associated uncertainty, aiming at unbiased inference and/or decision; therefore, it is usually referred to as the frequentist framework for statistics [1,3,4].
- On the other hand, within the Bayesian framework, credit is also given to previous beliefs, in addition to that given to the measured data. Such previous information can even be qualitative, but need to be represented in terms of a probability distribution function, and regrettably induce bias in the results [1,3,4].

INTRODUCTION

- Although not always considered in such a way, the solution of inverse problems can be formulated in terms of statistical inference [5]. Statistical inference refers to the process of drawing conclusions or making predictions based on limited information, beyond the immediate data that is available [4].
- This is exactly what is aimed with the solution of inverse problems.
- There are many techniques for the solution of inverse problems, but the most general ones are usually related to the minimization of an objective function that involves the difference between measured and estimated responses of the physical problem [5-27].
- If the objective function is derived based on statistical hypotheses for the measurement errors and unknown parameters/functions, the minimization procedure can be related to statistical inference, thus resulting in point estimates for the unknowns that allow for estimations of their associated uncertainties [5,8].
- Unfortunately, such is generally not the case, in special when the objective function is penalized with regularization terms.

INTRODUCTION

- The major source for the solution of inverse problems within the Bayesian framework is the book by Kaipio and Somersalo [5].
- The reader is referred to the book by Gamerman and Lopes [28] for deeper details about Markov Chain Monte Carlo methods and to the books by Lee [1] and Winkler [4] for fundamental material on Bayesian statistics.

GENERAL CONSIDERATIONS

Consider the mathematical formulation of a heat transfer problem, which, for instance, can be linear or non-linear, one or multi-dimensional, involve one single or coupled heat transfer modes, etc.

We denote the **vector of parameters** appearing in such formulation as:

$$\mathbf{P}^T = [P_1, P_2, \dots, P_N]$$

where **N is the number of parameters**

- These parameters can possibly be thermal conductivity components, heat transfer coefficients, heat sources, boundary heat fluxes, etc.
- They can represent constant values of such quantities, or the parameters of the representation of a function in terms of known basis functions.

GENERAL CONSIDERATIONS

The unknown function $g_p(t)$ is approximated as:

$$g_p(t) = \sum_{j=1}^N P_j C_j(t)$$

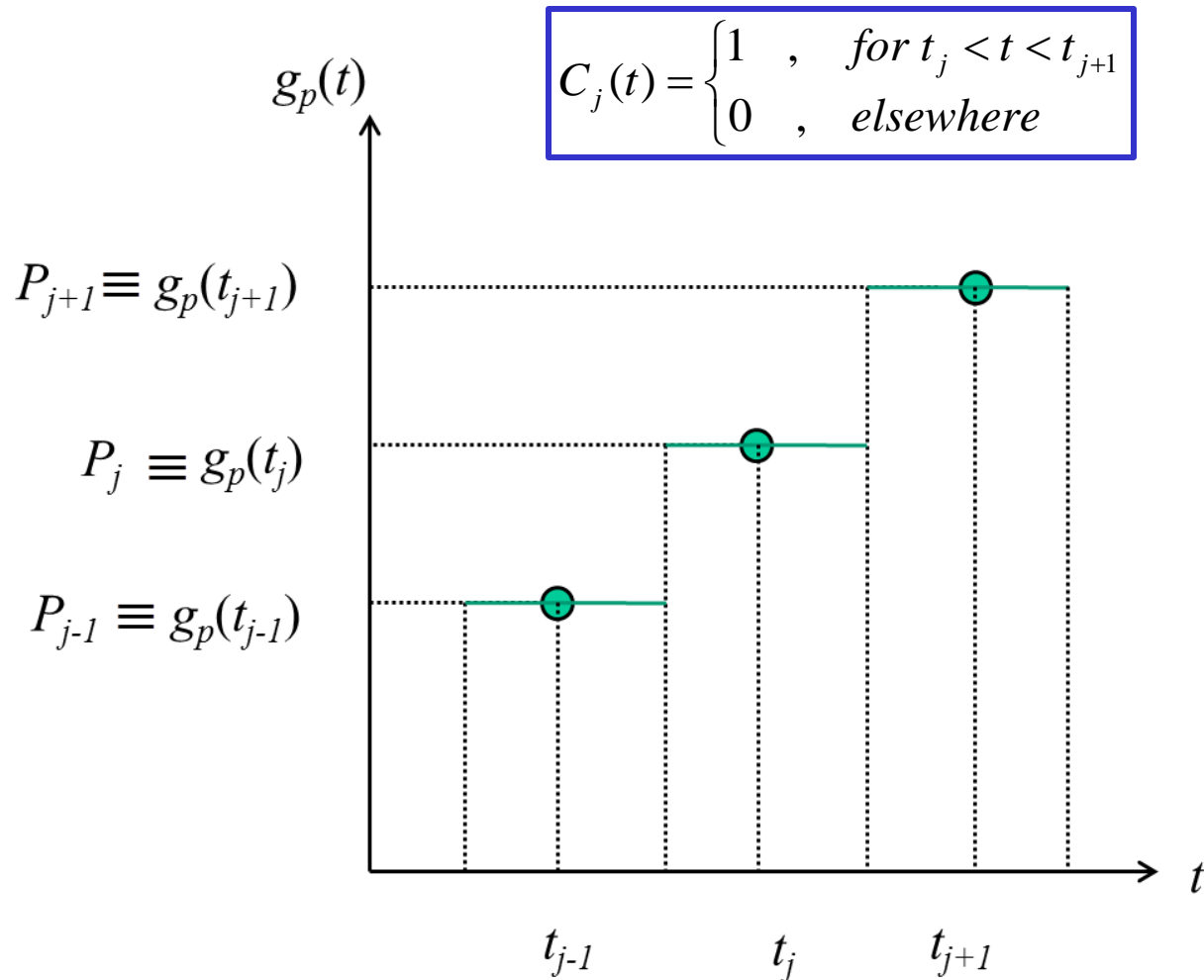
where:

$C_j(t)$ are *known LI basis functions*

N is the *number of basis functions* used in
the approximation
(known for the analysis)

P_j are the *unknown parameters*

GENERAL CONSIDERATIONS



GENERAL CONSIDERATIONS

Consider also that transient measurements are available within the medium, or at its surface, where the heat transfer processes are being mathematically formulated.

The vector containing the **measurements** is written as:

$$\mathbf{Y}^T = (\vec{Y}_1, \vec{Y}_2, \dots, \vec{Y}_I)$$

$$\vec{Y}_i = (Y_{i1}, Y_{i2}, \dots, Y_{iM})$$

$M = \#$ of sensors

$I = \#$ of transient measurements per sensor

} $D = MI = \#$ of measurements

- The measured data are not limited to temperatures, but could also include heat fluxes, radiation intensities, etc.

GENERAL CONSIDERATIONS

Hypotheses:

- The experimental errors are additive, with zero mean and normally distributed.

$$\mathbf{Y} = \mathbf{T}(\mathbf{P}) + \boldsymbol{\varepsilon}$$

- The statistical parameters (covariance matrix \mathbf{W}) describing the errors are known.
- There are no errors in the independent variables.
- The measurement errors are independent of the parameters \mathbf{P}

Likelihood Function

$$\pi(\boldsymbol{\varepsilon}) = (2\pi)^{-D/2} |\mathbf{W}|^{-1/2} \exp \left\{ -\frac{1}{2} [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})] \right\}$$

where $\mathbf{T}(\mathbf{P})$ is the solution of the direct (forward) problem.

GENERAL CONSIDERATIONS

Likelihood Function

$$\pi(\mathbf{Y}|\mathbf{P}) = (2\pi)^{-D/2} |\mathbf{W}|^{-1/2} \exp \left\{ -\frac{1}{2} [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})] \right\}$$

- The likelihood function gives the relative probability density of different measurement outcomes \mathbf{Y} with a fixed \mathbf{P} .
- A very common approach for the solution of inverse problems, dealing with the estimation of the parameters \mathbf{P} with the measurements \mathbf{Y} , is to maximize the likelihood probability density.

Maximum Likelihood Objective Function

$$S_{ML}(\mathbf{P}) = [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})]$$

GENERAL CONSIDERATIONS

Maximum Likelihood Objective Function

$$S_{ML}(\mathbf{P}) = [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})]$$

For uncorrelated measurements:

$$\mathbf{W} = \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \ddots & \\ 0 & & & \sigma_I^2 \end{bmatrix}$$

Constant variances σ^2 :

$$\mathbf{W} = \sigma^2 \mathbf{I}$$

Least-Squares

$$S_{OLS}(\mathbf{P}) = [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T [\mathbf{Y} - \mathbf{T}(\mathbf{P})]$$

GENERAL CONSIDERATIONS

Remark: If the inverse heat transfer problem involves the estimation of only few unknown parameters from many measurements, the minimization of the maximum likelihood objective function can be stable. However, if the inverse problem involves the estimation of a large number of parameters, such as the recovery of the unknown transient strength of the heat source term $g_p(t_i)$ at times $t_i, i=1, \dots, I$, excursion and oscillation of the solution may occur. In this case, *regularization* (or stabilization) techniques are required.

GENERAL CONSIDERATIONS

First-order regularization Tikhonov's Regularization

$$S(\mathbf{P}) = [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})] + \alpha \|\mathbf{D}\mathbf{P}\|_{L_2}^2$$

where: $\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \dots & 0 \\ & & \ddots & & \vdots \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix}$

$\alpha (> 0)$ is the first-order regularization parameter

To select α :

- Discrepancy Principle: $|Y_i - T_i| \approx \sigma_i$
- L curve
- Generalized Cross Validation

BAYESIAN FRAMEWORK

- Classical regularization methods are not based on the modeling of prior information and related uncertainty about the unknown parameters or functions.
- Although very popular and useful in many situations, the minimization of the maximum likelihood objective function is a non-Bayesian estimator. A Bayesian estimator is basically concerned with the analysis of the **posterior probability density**, which is the conditional probability of the parameters \mathbf{P} given the measurements \mathbf{Y} .

BAYESIAN FRAMEWORK

For the solution of inverse problems within the Bayesian framework, **all variables included in the mathematical formulation of the physical problem are modelled as random variables**. Techniques for the solution of inverse problems within the Bayesian framework can be summarized in the following steps [5]:

- 1. Based on all information available for the parameters \mathbf{P} before the measured data \mathbf{Y} is taken, select a probability distribution function, $\pi(\mathbf{P})$, that appropriately represents the prior information.**
- 2. Select the likelihood function, $\pi(\mathbf{Y}|\mathbf{P})$, that appropriately models the measurement errors** and involves a relation between the observations and the mathematical model of the physical problem under picture (see, for example, equation 6.b).
- 3. Develop methods to explore the posterior density function**, which is the conditional probability distribution of the unknown parameters given the measurements, $\pi(\mathbf{P}|\mathbf{Y})$.

BAYESIAN FRAMEWORK

The formal mechanism to combine the new information (measurements) with the previously available information (prior) is known as the Bayes' theorem [5,8,20,22,25-28]. Let \mathbf{P} and \mathbf{Y} be continuous random variables. Then, we can write [4]:

$$\pi(\mathbf{P}|\mathbf{Y}) = \frac{\pi(\mathbf{P}, \mathbf{Y})}{\pi(\mathbf{Y})} \quad (11.a)$$

that is, the conditional density of the random variable \mathbf{P} given a value of the random variable \mathbf{Y} is the joint density of \mathbf{P} and \mathbf{Y} divided by the marginal density of \mathbf{Y} , where

$$\pi(\mathbf{Y}) = \int_{R^N} \pi(\mathbf{P}, \mathbf{Y}) d\mathbf{P} \quad (11.b)$$

The joint density $\pi(\mathbf{P}, \mathbf{Y})$ is not generally known, but it can be written in terms of the likelihood and the prior as [4]

$$\pi(\mathbf{P}, \mathbf{Y}) = \pi(\mathbf{Y}|\mathbf{P})\pi(\mathbf{P}) \quad (12)$$

By substituting (12) into (11.a) we then obtain Bayes' theorem, which is given by

BAYESIAN FRAMEWORK

BAYES' FORMULA

$$\pi_{posterior}(\mathbf{P}) = \pi(\mathbf{P}|\mathbf{Y}) = \frac{\pi_{prior}(\mathbf{P})\pi(\mathbf{Y}|\mathbf{P})}{\pi(\mathbf{Y})}$$

Where: $\pi_{posterior}(\mathbf{P})$ = posterior probability density (conditional probability of the parameters \mathbf{P} given the measurements \mathbf{Y})

$\pi_{prior}(\mathbf{P})$ = prior density (information about the parameters prior to the measurements)

$\pi(\mathbf{Y}|\mathbf{P})$ = likelihood function (expresses the likelihood of different measurement outcomes \mathbf{Y} with \mathbf{P} given)

$\pi(\mathbf{Y})$ = probability density of the measurements (normalizing constant)

$$posterior \propto prior \times likelihood$$

MAXIMUM A POSTERIORI

Hypotheses:

- The errors are additive, with zero mean and normally distributed.
- The statistical parameters describing the errors are known.
- There are no errors in the independent variables.
- **P** is independent of **Y**.
- **P** is Gaussian with known mean $\boldsymbol{\mu}$ and known covariance matrix **V**.

Gaussian Prior

$$\pi(\mathbf{P}) = (2\pi)^{-N/2} |\mathbf{V}|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{P} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{P} - \boldsymbol{\mu}) \right]$$

MAXIMUM A POSTERIORI

Likelihood Function

$$\pi(\mathbf{Y}|\mathbf{P}) = (2\pi)^{-D/2} |\mathbf{W}|^{-1/2} \exp \left\{ -\frac{1}{2} [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})] \right\}$$

Posterior Density

$$\ln [\pi(\mathbf{P} | \mathbf{Y})] \propto -\frac{1}{2} [(D + N) \ln 2\pi + \ln |\mathbf{W}| + \ln |\mathbf{V}| + S_{MAP}(\mathbf{P})]$$

Maximum a Posteriori Objective Function

$$S_{MAP}(\mathbf{P}) = [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})] + (\boldsymbol{\mu} - \mathbf{P})^T \mathbf{V}^{-1} (\boldsymbol{\mu} - \mathbf{P})$$

MAXIMUM A POSTERIORI

$$-2 \mathbf{J}^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})] - 2\mathbf{V}^{-1} [\boldsymbol{\mu} - \mathbf{P}] = 0$$

Linear Problems: \mathbf{J} does not depend on \mathbf{P} \Rightarrow $\mathbf{T}(\mathbf{P}) = \mathbf{J}\mathbf{P}$

$$\mathbf{P} = [\mathbf{J}^T \mathbf{W}^{-1} \mathbf{J} + \mathbf{V}^{-1}]^{-1} [\mathbf{J}^T \mathbf{W}^{-1} \mathbf{Y} + \mathbf{V}^{-1} \boldsymbol{\mu}]$$

Nonlinear Problems: $\mathbf{J} \equiv \mathbf{J}(\mathbf{P})$ \Rightarrow $\mathbf{T}(\mathbf{P}) = \mathbf{T}(\mathbf{P}^k) + \mathbf{J}^k (\mathbf{P} - \mathbf{P}^k)$

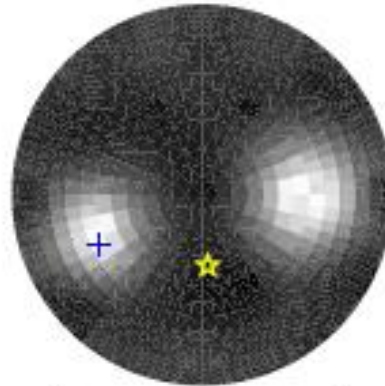
$$\mathbf{P}^{k+1} = \mathbf{P}^k + [\mathbf{J}^T \mathbf{W}^{-1} \mathbf{J} + \mathbf{V}^{-1}]^{-1} \{ \mathbf{J}^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P}^k)] + \mathbf{V}^{-1} (\boldsymbol{\mu} - \mathbf{P}^k) \}$$

$$\text{cov}(\mathbf{P}) = (\mathbf{J}^T \mathbf{W}^{-1} \mathbf{J} + \mathbf{V}^{-1})^{-1}$$

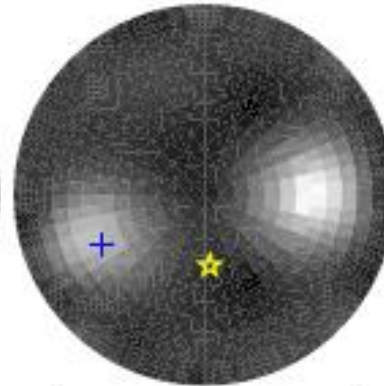
MAXIMUM A POSTERIORI

Example: Thermal Tomography (Gaussian and smooth priors - MAP)

Thermal Conductivity



Volumetric Heat Capacity



Markov Chain Monte Carlo (MCMC) methods

Moreover, whereas the computation of the MAP estimate is an optimization problem, that is,

$$\mathbf{P}_{MAP} = \arg \max_{\mathbf{P} \in R^N} \pi(\mathbf{P} | \mathbf{Y}) \quad (19)$$

other point and confidence estimates from the posterior distribution typically require numerical integration. For example, one common point estimate is the conditional mean defined as [5]:

$$\mathbf{P}_{CM} = E(\mathbf{P}) = \int_{R^N} \mathbf{P} \pi(\mathbf{P} | \mathbf{Y}) d\mathbf{P} \quad (20)$$

where $E(\cdot)$ denotes the expected value. In general, the dimension N of the parameter space is large enough to make the numerical integration in equation (20) impractical. Besides that, the computation of the normalizing constant in the denominator of $\pi(\mathbf{P} | \mathbf{Y})$ (see equations 11-13) already constitutes a challenging problem by itself.

Markov Chain Monte Carlo (MCMC) methods

For those cases that the posterior is not analytical and/or numerical integrations required for estimates are not practical, Markov Chain Monte Carlo (MCMC) methods can provide a solution of the inverse problem, so that inference on the posterior probability becomes inference on its samples [1,4,5,20,22,25-28]. For example, the Monte Carlo integration of equation (20) can be approximated by [5]:

$$\mathbf{P}_{CM} = E(\mathbf{P}) = \int_{R^N} \mathbf{P} \pi(\mathbf{P} | \mathbf{Y}) d\mathbf{P} \approx \frac{1}{n} \sum_{t=1}^n \mathbf{P}^{(t)} \quad (21)$$

where $\mathbf{P}^{(t)}$, for $t = 1, \dots, n$, are samples from $\pi(\mathbf{P} | \mathbf{Y})$. Markov Chain Monte Carlo methods are used to obtain such samples.

Markov Chain Monte Carlo (MCMC) methods

- Draw samples of all possible \mathbf{P} 's, each sample with probability $\pi(\mathbf{P}|\mathbf{Y})$.
- Get a set $\Theta = \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n\}$ of samples distributed like the posterior distribution.
- Inference on $\pi(\mathbf{P}|\mathbf{Y})$ becomes inference on $\Theta = \{\mathbf{P}_1, \mathbf{P}_2, \dots, \mathbf{P}_n\}$, for example the mean of the samples in Θ give us an estimation of the average values of $\pi(\mathbf{P}|\mathbf{Y})$.
- **Metropolis-Hastings Algorithm** and Gibbs Sampler
- Very time consuming.

Markov Chain Monte Carlo (MCMC) methods

Markov Chains

A Markov chain is a stochastic process that, given the present state, past and future states are independent.

The stochastic process is a Markov chain if it satisfies the Markov condition [1,4,5,20,22,25-28]:

$$q(\mathbf{P}^{t+1} = \mathbf{y} \mid \mathbf{P}^t = \mathbf{x}, \mathbf{P}^{t-1} = \mathbf{x}^{t-1}, \dots, \mathbf{P}^0 = \mathbf{x}^0) = q(\mathbf{P}^{t+1} = \mathbf{y} \mid \mathbf{P}^t = \mathbf{x}) \text{ for all } \mathbf{y}, \mathbf{x}, \mathbf{x}^{t-1}, \dots, \mathbf{x}^0 \in S \quad (22)$$

where q is a transition probability. Some concepts regarding Markov chains are now presented. The reader shall consult references [1,4,5,20,22,25-28] for further details.

Markov Chain Monte Carlo (MCMC) methods

Markov Chains

A distribution p^* is said to be a *stationary distribution* of a chain if, once the chain is in p^* , it stays in this distribution. Suppose now that $p^{(t)} \rightarrow p^*$ as $t \rightarrow \infty$ for any $p^{(0)}$, where $p^{(t)}$ is the distribution at state t of the chain. Then, p^* is the *equilibrium distribution* of the Markov chain and the chain is said to be *ergodic*.

Consider the sequence of states $\mathbf{x} \rightarrow \mathbf{k}_1 \rightarrow \mathbf{k}_2 \rightarrow \dots \rightarrow \mathbf{k}_t \rightarrow \mathbf{y}$ so that the transition probabilities $q(\mathbf{k}_1 | \mathbf{x}) \neq 0$, $q(\mathbf{k}_2 | \mathbf{k}_1) \neq 0$, \dots , $q(\mathbf{y} | \mathbf{k}_t) \neq 0$. Then, there is a sequence of states from \mathbf{x} to \mathbf{y} with a nonzero probability of occurring in the Markov chain. It is said that \mathbf{x} and \mathbf{y} communicate. If \mathbf{y} and \mathbf{x} also communicate through nonzero transition probabilities, it is said that these two states intercommunicate. If all states in S intercommunicate, then the state space is said to be *irreducible* under q . A Markov chain is *reversible* if $p(\mathbf{x})q(\mathbf{y} | \mathbf{x}) = p(\mathbf{y})q(\mathbf{x} | \mathbf{y})$.

The period of a state \mathbf{x} , denoted by $d_{\mathbf{x}}$, is the largest common divisor of the set $\{m \geq 1 : p^{(m)}(\mathbf{x}, \mathbf{x}) > 0\}$. A state \mathbf{x} is aperiodic if $d_{\mathbf{x}} = 1$. A chain is *aperiodic* if all of its states are aperiodic.

Markov Chain Monte Carlo (MCMC) methods

Markov Chains

*Let p be a given probability distribution. The Markov chain simulated by the **Metropolis-Hastings algorithm** is reversible with respect to p . If it is also irreducible and aperiodic, then it defines an ergodic Markov chain with unique equilibrium distribution p .*

Markov Chain Monte Carlo (MCMC) methods

- **Parameters with linearly-dependent sensitivity coefficients generally result on periodic and correlated chains and an equilibrium distribution is not reached.**
- Similarly to classical methods of parameter estimation, where the sensitivity coefficients directly influence the topology of the objective function based on the likelihood and a global minimum might not exist, such coefficients directly influence the posterior distribution, which is now sought via the implementation of a Markov chain.
- **The sensitivity coefficients need also to be carefully examined if the solution of the inverse parameter estimation problem is to be obtained within the Bayesian framework.**

Markov Chain Monte Carlo (MCMC) methods

- In classical methods, **parameters with small and linearly dependent sensitivity coefficients** are usually deterministically fixed, based on values known from previous experience and/or literature. **In approaches within the Bayesian framework, uncertainties on such kind of parameters can be appropriately taken into account through their prior distribution functions.**
- On the other hand, the analysis of the sensitivity coefficients reveals that **parameters with small and/or linearly dependent sensitivity coefficients require informative prior distributions for the success of the estimation procedure.**

Metropolis-Hastings Algorithm

THE JOURNAL OF CHEMICAL PHYSICS

VOLUME 21, NUMBER 6

JUNE, 1953

Equation of State Calculations by Fast Computing Machines

NICHOLAS METROPOLIS, ARIANNA W. ROSENBLUTH, MARSHALL N. ROSENBLUTH, AND AUGUSTA H. TELLER,
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

AND

EDWARD TELLER,* *Department of Physics, University of Chicago, Chicago, Illinois*

(Received March 6, 1953)

A general method, suitable for fast computing machines, for investigating such properties as equations of state for substances consisting of interacting individual molecules is described. The method consists of a modified Monte Carlo integration over configuration space. Results for the two-dimensional rigid-sphere system have been obtained on the Los Alamos MANIAC and are presented here. These results are compared to the free volume equation of state and to a four-term virial coefficient expansion.

Metropolis-Hastings Algorithm

Biometrika (1970), 57, 1, p. 97

97

Printed in Great Britain

Monte Carlo sampling methods using Markov chains and their applications

BY W. K. HASTINGS

University of Toronto

SUMMARY

A generalization of the sampling method introduced by Metropolis *et al.* (1953) is presented along with an exposition of the relevant theory, techniques of application and methods and difficulties of assessing the error in Monte Carlo estimates. Examples of the methods, including the generation of random orthogonal matrices and potential applications of the methods to numerical problems arising in statistics, are discussed.

Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm draws samples from a candidate density, such as in acceptance-rejection sampling [1]. The acceptance-rejection method is used to generate samples from a density $p(\mathbf{P}) = \tilde{p}(\mathbf{P}) / K$, where the normalizing constant K might be unknown, such as in the posterior distribution given by equations (13.a,b). Instead of sampling from $p(\mathbf{P})$, assume that there exists a candidate density $h(\mathbf{P})$ that is easy to simulate samples from, where $\tilde{p}(\mathbf{P}) \leq c h(\mathbf{P})$ and c is a constant. The following steps are then used to obtain a random variable $\hat{\mathbf{P}}$ from density $p(\mathbf{P})$ with the acceptance-rejection method [1]:

1. Generate a random variable \mathbf{P}^* from the density $h(\mathbf{P})$;
2. Generate a random value $U \sim U(0,1)$, which is uniformly distributed in $(0,1)$;
3. If $U \leq \tilde{p}(\mathbf{P}) / c h(\mathbf{P})$, let $\hat{\mathbf{P}} = \mathbf{P}^*$. Otherwise, return to step 1.

Markov Chain Monte Carlo (MCMC) methods

Metropolis-Hastings Algorithm

The implementation of the Metropolis-Hastings algorithm starts with the selection of a candidate or proposal distribution $q(\mathbf{P}^* | \mathbf{P}^{(t)})$, which is used to draw a new candidate state \mathbf{P}^* , given the current state $\mathbf{P}^{(t)}$ of the Markov chain. Remember that, for the solution of the inverse problem within the Bayesian framework, one aims at simulating the posterior distribution $\pi_{posterior}(\mathbf{P}) \propto \pi(\mathbf{Y} | \mathbf{P})\pi(\mathbf{P})$ (see equation 13.b). Hence, the balance (reversibility) condition of the Markov chain of interest is given by:

- $$\pi_{posterior}(\mathbf{P}^{(t)})q(\mathbf{P}^* | \mathbf{P}^{(t)}) = \pi_{posterior}(\mathbf{P}^*)q(\mathbf{P}^{(t)} | \mathbf{P}^*) \quad (24)$$

Metropolis-Hastings Algorithm

In order to avoid eventual cases that $\pi_{posterior}(\mathbf{P}^{(t)})q(\mathbf{P}^* | \mathbf{P}^{(t)}) > \pi_{posterior}(\mathbf{P}^*)q(\mathbf{P}^{(t)} | \mathbf{P}^*)$, that is, the process moves from $\mathbf{P}^{(t)}$ to \mathbf{P}^* more often than the reverse, a probability $\alpha(\mathbf{P}^* | \mathbf{P}^{(t)})$ is introduced in equation (24), so that [1]:

$$\pi_{posterior}(\mathbf{P}^{(t)})q(\mathbf{P}^* | \mathbf{P}^{(t)})\alpha(\mathbf{P}^* | \mathbf{P}^{(t)}) = \pi_{posterior}(\mathbf{P}^*)q(\mathbf{P}^{(t)} | \mathbf{P}^*) \quad (25)$$

Therefore,

$$\alpha(\mathbf{P}^* | \mathbf{P}^{(t)}) = \min \left[1, \frac{\pi_{posterior}(\mathbf{P}^*)q(\mathbf{P}^{(t)} | \mathbf{P}^*)}{\pi_{posterior}(\mathbf{P}^{(t)})q(\mathbf{P}^* | \mathbf{P}^{(t)})} \right] \quad (26)$$

Metropolis-Hastings Algorithm

1. Sample a *Candidate Point* \mathbf{P}^* from a proposal distribution $q(\mathbf{P}^* | \mathbf{P}^{(t-1)})$.
2. Calculate the acceptance factor:

$$\alpha = \min \left[1, \frac{\pi(\mathbf{P}^* | \mathbf{Y}) q(\mathbf{P}^{(t-1)} | \mathbf{P}^*)}{\pi(\mathbf{P}^{(t-1)} | \mathbf{Y}) q(\mathbf{P}^* | \mathbf{P}^{(t-1)})} \right]$$

3. Generate a random value U that is uniformly distributed on $(0,1)$.
4. If $U < \alpha$, set $\mathbf{P}^{(t)} = \mathbf{P}^*$. Otherwise, set $\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)}$.
5. Return to step 1.

Proposal Distribution

(i) Random Walk: In this case $\mathbf{P}^* = \mathbf{P}^{(t)} + \Psi$, where Ψ is a vector of random variables with distribution $q_1(\Psi)$. Therefore, $q(\mathbf{P}^* | \mathbf{P}^{(t)}) = q_1(\Psi)$. If the proposal distribution is symmetric, that is, $q_1(\Psi) = q_1(-\Psi)$ or $q(\mathbf{P}^* | \mathbf{P}^{(t)}) = q(\mathbf{P}^{(t)} | \mathbf{P}^*)$, equation (26) reduces to

$$\alpha(\mathbf{P}^* | \mathbf{P}^{(t)}) = \min \left[1, \frac{\pi_{posterior}(\mathbf{P}^*)}{\pi_{posterior}(\mathbf{P}^{(t)})} \right] \quad (27)$$

Thus, for this choice of the proposal density, equation (27) shows that in step 5 of the Metropolis-Hastings algorithm, the candidate point \mathbf{P}^* is always accepted if the move leads to a region of higher posterior probability. Furthermore, the candidate point can also be accepted if $\pi_{posterior}(\mathbf{P}^*) < \pi_{posterior}(\mathbf{P}^{(t)})$ with probability $\alpha(\mathbf{P}^* | \mathbf{P}^{(t)})$, thus allowing that the state space be highly explored.

Proposal Distribution

Uniform and Gaussian distributions are commonly used for $q_1(\boldsymbol{\psi})$. Consider one single component P_j of the vector \mathbf{P} . For the uniform random walk proposal one can write:

$$P_j^* = P_j^{(t)} + w_j(2r - 1) \quad (28.a)$$

where r is a random number with uniform distribution in $(0,1)$, that is, $r \sim U(0,1)$, while w_j is the maximum variation for the parameter P_j .

For the Gaussian random walk proposal we have

$$P_j^* = P_j^{(t)} + w_j \quad (28.b)$$

where now w_j is a Gaussian random number with zero mean and standard deviation s_j .

Proposal Distribution

(ii) Independent Move: This choice for the proposal density is of the kind $q(\mathbf{P}^* | \mathbf{P}^{(t)}) = q_2(\mathbf{P}^*)$, that is, it does not depend on the current state $\mathbf{P}^{(t)}$. In this case, the proposal density $q(\mathbf{P}^* | \mathbf{P}^{(t)})$ can be conveniently selected as the prior density $\pi(\mathbf{P}^*)$. Then, by also utilizing equation (13.b), equation (26) is rewritten as

$$\alpha(\mathbf{P}^* | \mathbf{P}^{(t)}) = \min \left[1, \frac{\pi(\mathbf{Y} | \mathbf{P}^*)\pi(\mathbf{P}^*)}{\pi(\mathbf{Y} | \mathbf{P}^{(t)})\pi(\mathbf{P}^{(t)})} \frac{\pi(\mathbf{P}^{(t)})}{\pi(\mathbf{P}^*)} \right] \quad (29.a)$$

Hence, the Metropolis-Hastings ratio is given by the ratio of the likelihoods, that is,

$$\alpha(\mathbf{P}^* | \mathbf{P}^{(t)}) = \min \left[1, \frac{\pi(\mathbf{Y} | \mathbf{P}^*)}{\pi(\mathbf{Y} | \mathbf{P}^{(t)})} \right] \quad (29.b)$$

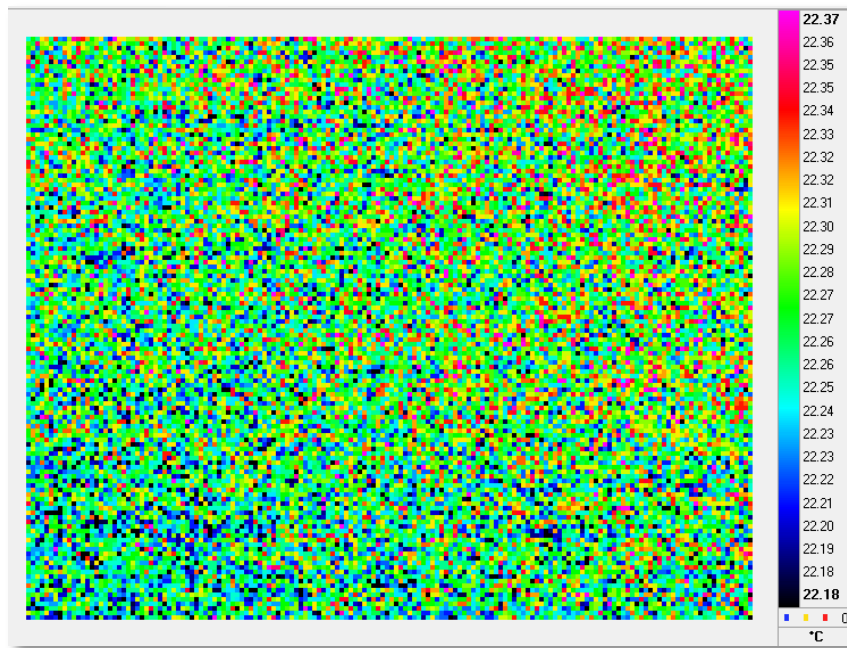
Practical Issues for MCMC Methods

Likelihood

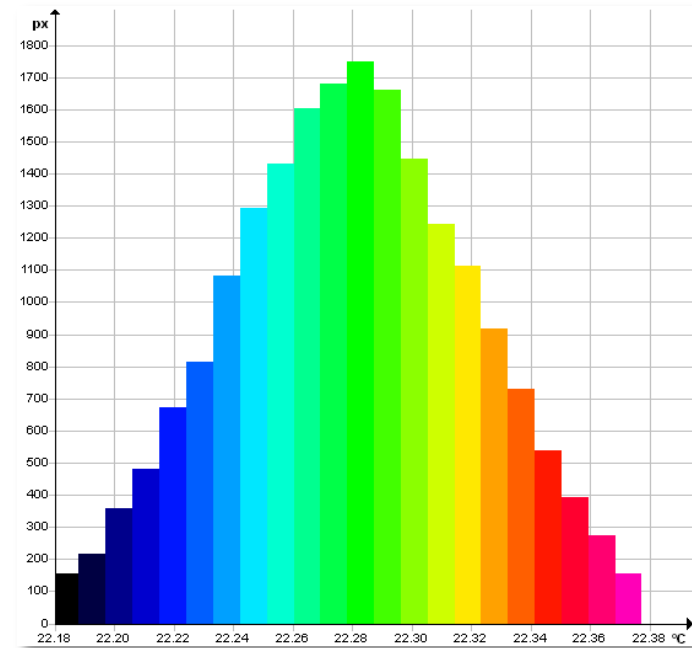
- The likelihood function involves the solution of the mathematical formulation of the physical problem under analysis, that is, the solution of the direct or forward model, as well as the measurements and their related uncertainties.
- Measurement errors are modelled after the calibration of sensors and instruments used to collect the experimental data.
- **Gaussian distributions are in general appropriate for temperature measurements taken with thermocouples and infrared cameras.**
- **Be careful! Measurements are not always Gaussian!**

Practical Issues for MCMC Methods

Likelihood



(a)



(b)

Figure 2. (a) Thermal image with an infrared camera of an isothermal plate;
(b) Histogram of the temperature measurements [33].

Priors

A Gaussian prior was also considered in section 4, given by equation (14) for a multivariate case, with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{V} , denoted as $\mathbf{P} \sim N(\boldsymbol{\mu}, \mathbf{V})$. For one single parameter P_j , a *Gaussian prior* with mean μ_j and variance σ_j^2 , $P_j \sim N(\mu_j, \sigma_j^2)$, is given by

- $$\pi(P_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{1}{2} \frac{(P_j - \mu_j)^2}{\sigma_j^2}\right] \text{ in } -\infty < P_j < \infty \quad (30)$$

Random variables modelled by the Gaussian prior have support in R ; hence, may assume negative values, although this might happen with small probabilities depending on the values of μ_j and σ_j^2 . On the other hand, several physical parameters only allow positive values, such as, for example, thermal conductivity, specific heat and thermal diffusivity.

Practical Issues for MCMC Methods

A very simple prior that allows lower and upper bounds for the parameter values is the *Uniform distribution* $P_j \sim U(a, b)$ given by

$$\pi(P_j) = \begin{cases} \frac{1}{(b-a)} & , \quad a < P_j < b \\ 0 & , \quad elsewhere \end{cases} \quad (31)$$

Mean and variance for the uniform distribution are given by $\frac{1}{2}(a+b)$ and $\frac{1}{12}(b-a)^2$, respectively. In the uniform distribution, any value in $a < P_j < b$ is equally probable. If in this interval values around a known mean are more likely to occur than elsewhere, like in a Gaussian distribution, but the probability density is zero in $P_j \leq a$ and $P_j \geq b$, one possible prior can be obtained by combining equations (30) in (31), which is called *truncated Gaussian distribution*, that is,

$$\pi(P_j) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{1}{2} \frac{(P_j - \mu_j)^2}{\sigma_j^2}\right] & , \quad a < P_j < b \\ 0 & , \quad elsewhere \end{cases} \quad (32)$$

- where $a < \mu_j < b$.

Practical Issues for MCMC Methods

Priors

Other distributions that satisfy positive constraints are available. For example, the *Rayleigh distribution* $P_j \sim R(\gamma_0)$ is given by

$$\pi(P_j) = \frac{P_j}{\gamma_0^2} \exp\left[-\frac{1}{2}\left(\frac{P_j}{\gamma_0}\right)^2\right] \text{ for } P_j > 0 \quad (33)$$

- and depends only on the scale parameter (centerpoint) γ_0 . The mean and the variance of Rayleigh's distribution are given by $\gamma_0 \sqrt{\frac{\pi}{2}}$ and $\frac{4-\pi}{2} \gamma_0^2$, respectively.

Practical Issues for MCMC Methods

The *Gamma distribution* with parameters α and β , denoted as $P_j \sim G(\alpha, \beta)$, has the following density

$$\pi(P_j) = \frac{1}{\beta^\alpha \Gamma(\alpha)} P_j^{\alpha-1} \exp\left(-\frac{P_j}{\beta}\right) \quad \text{for } P_j > 0 \quad (34)$$

with mean $\alpha\beta$ and variance $\alpha\beta^2$, where $\Gamma(\alpha)$ is the gamma function. For $\beta=1$, the so-called one-parameter gamma distribution is obtained. The density that results by making $\alpha=1$ is called exponential distribution.

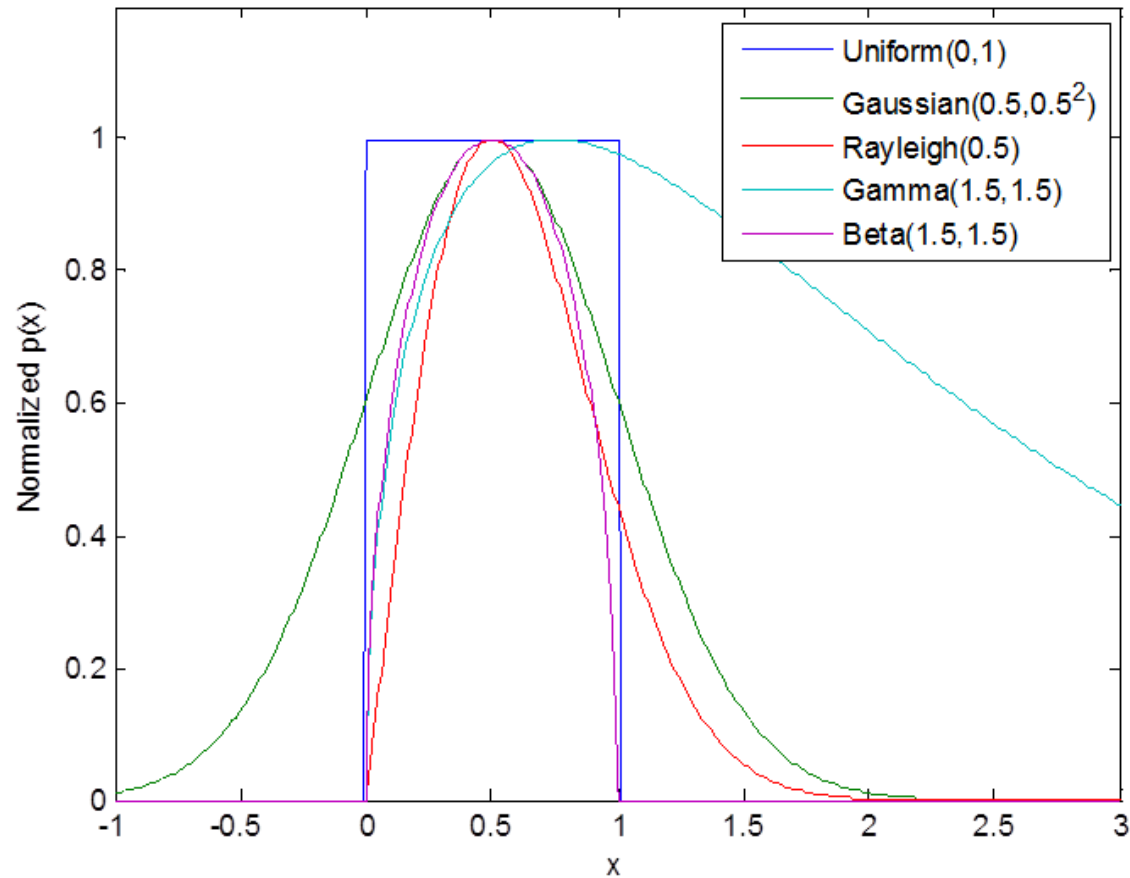
The *Beta distribution* $P_j \sim Be(\alpha, \beta)$ has support in $0 < P_j < 1$. The density of this distribution is given by

$$\pi(P_j) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} P_j^{\alpha-1} (1 - P_j)^{\beta-1} \quad \text{in } 0 < P_j < 1 \quad (35)$$

- with mean $\frac{\alpha}{\alpha + \beta}$ and variance $\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$.

Practical Issues for MCMC Methods

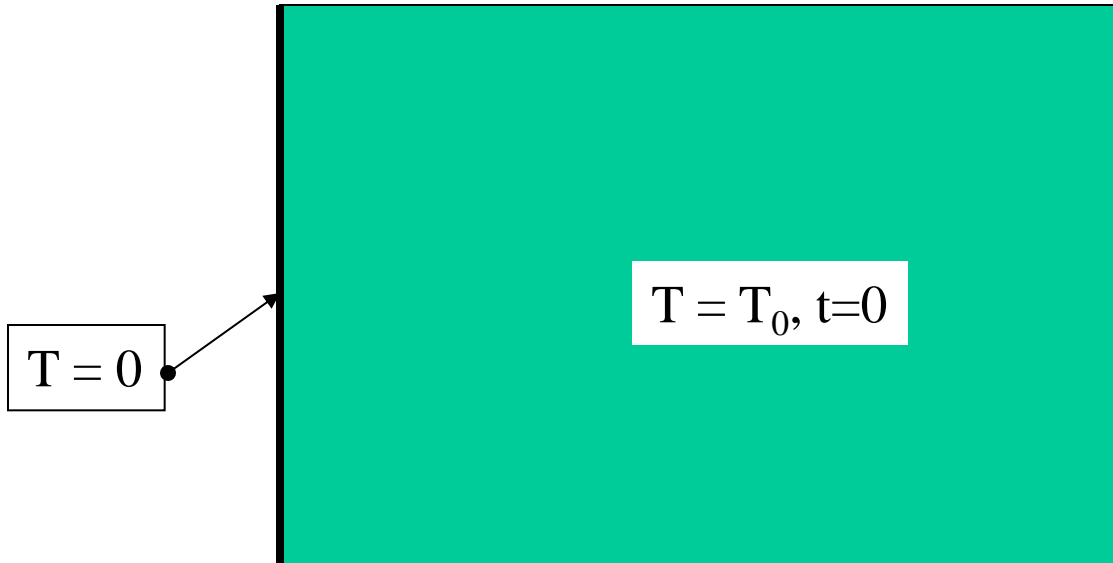
Priors





Markov Chain Monte Carlo (MCMC) methods

EXAMPLE



$$T(x,t) = T_0 \operatorname{erf} \left(\frac{x}{\sqrt{4\alpha t}} \right)$$

$$T_0 = 50 \text{ }^\circ\text{C}$$

$$\text{Concrete: } \alpha = 4.9 \times 10^{-7} \text{ m}^2/\text{s}$$



Markov Chain Monte Carlo (MCMC) methods

EXAMPLE

- Simultaneous estimation of T_0 and α .
- Prior for T_0 : Uniform distribution (40, 65) °C
- Prior for α : Uniform distribution ($10^{-7}, 10^{-5}$) m²/s
- Start the chain in the middle of the intervals

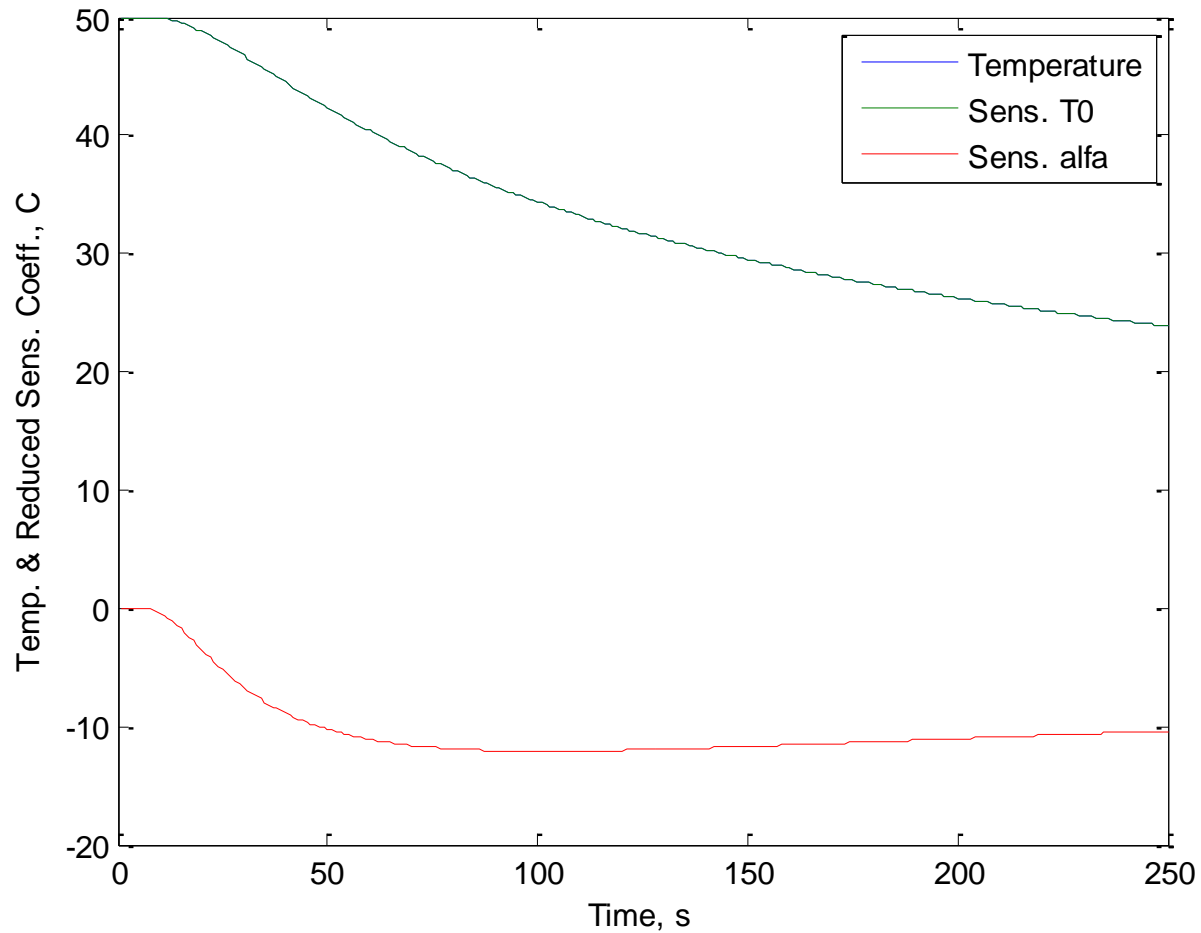
$$\frac{\partial T(x,t)}{\partial T_0} = \operatorname{erf}\left(\frac{x}{\sqrt{4\alpha t}}\right)$$

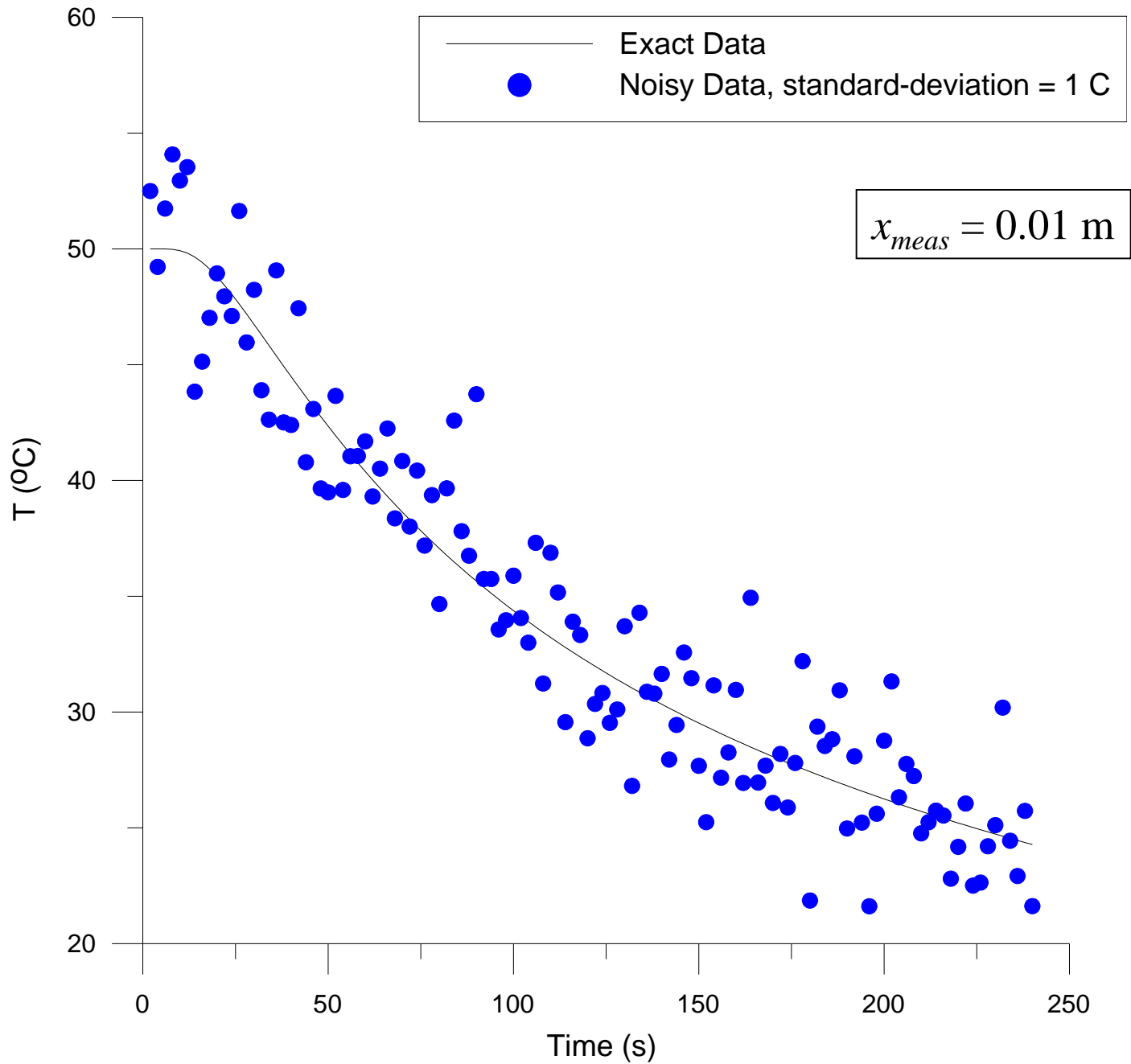
$$\frac{\partial T(x,t)}{\partial \alpha} = -\frac{x t T_0 \exp\left(-\frac{x^2}{4\alpha t}\right)}{2\sqrt{\pi}(\alpha t)^{3/2}}$$

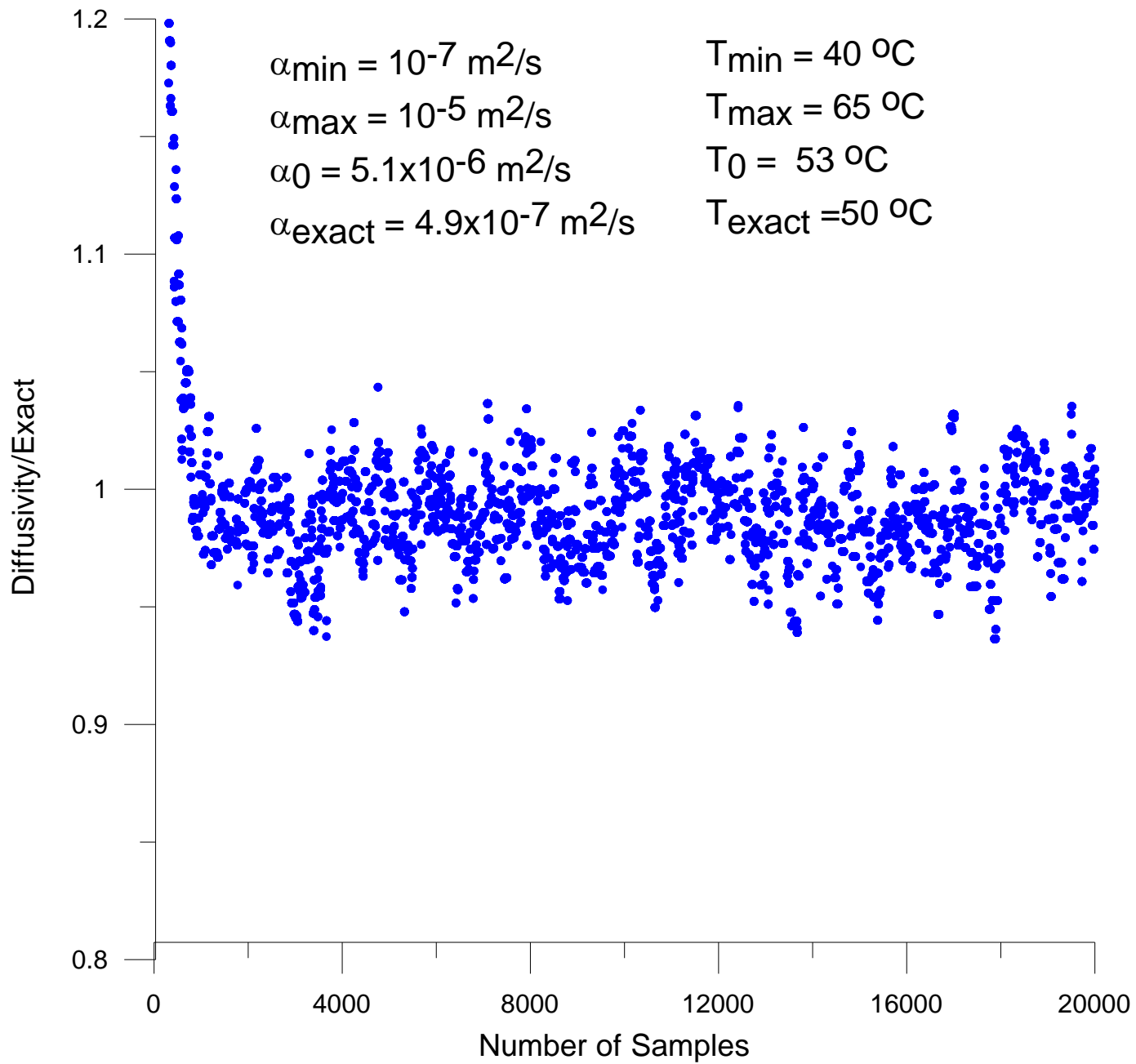


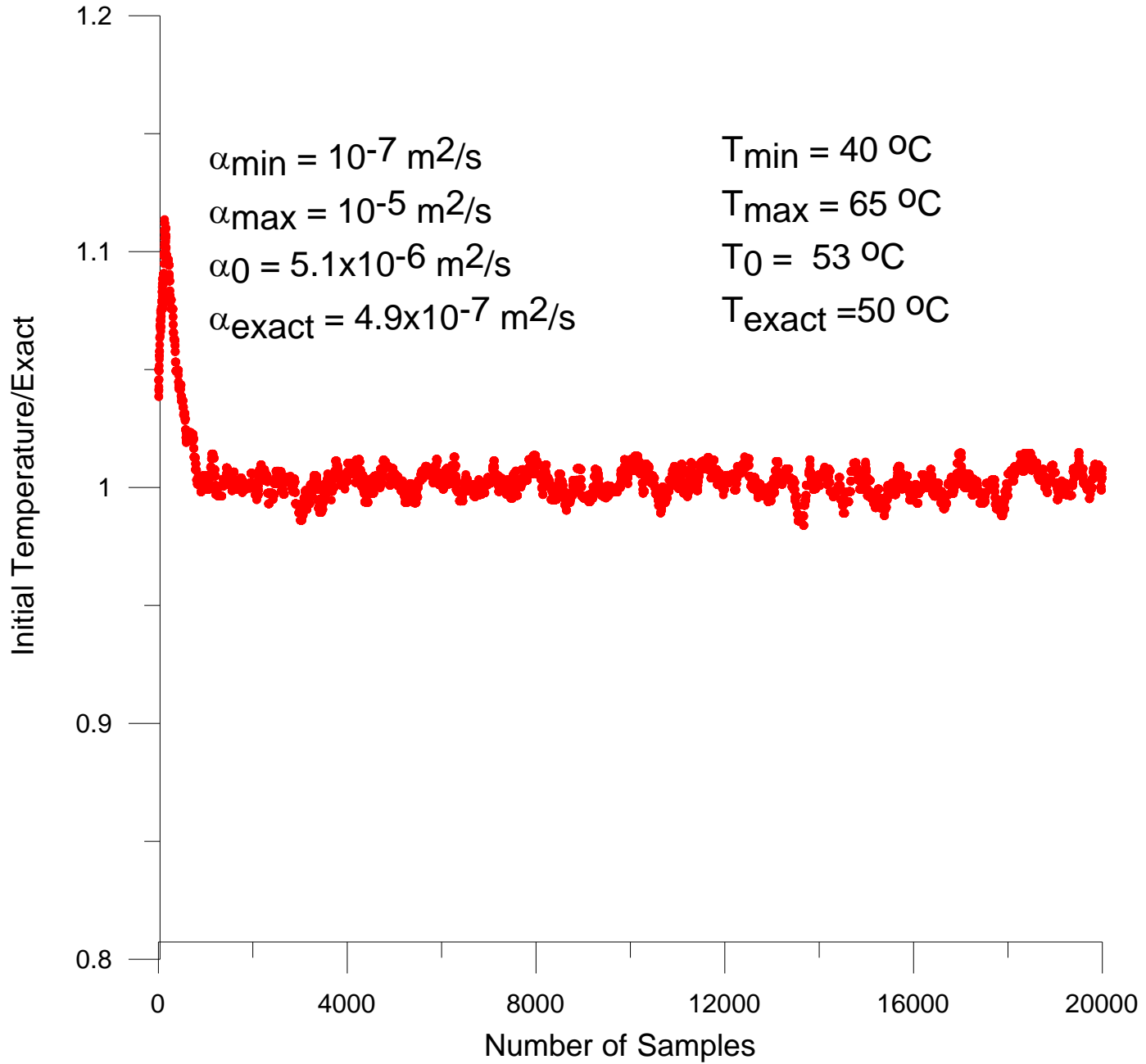
Markov Chain Monte Carlo (MCMC) methods

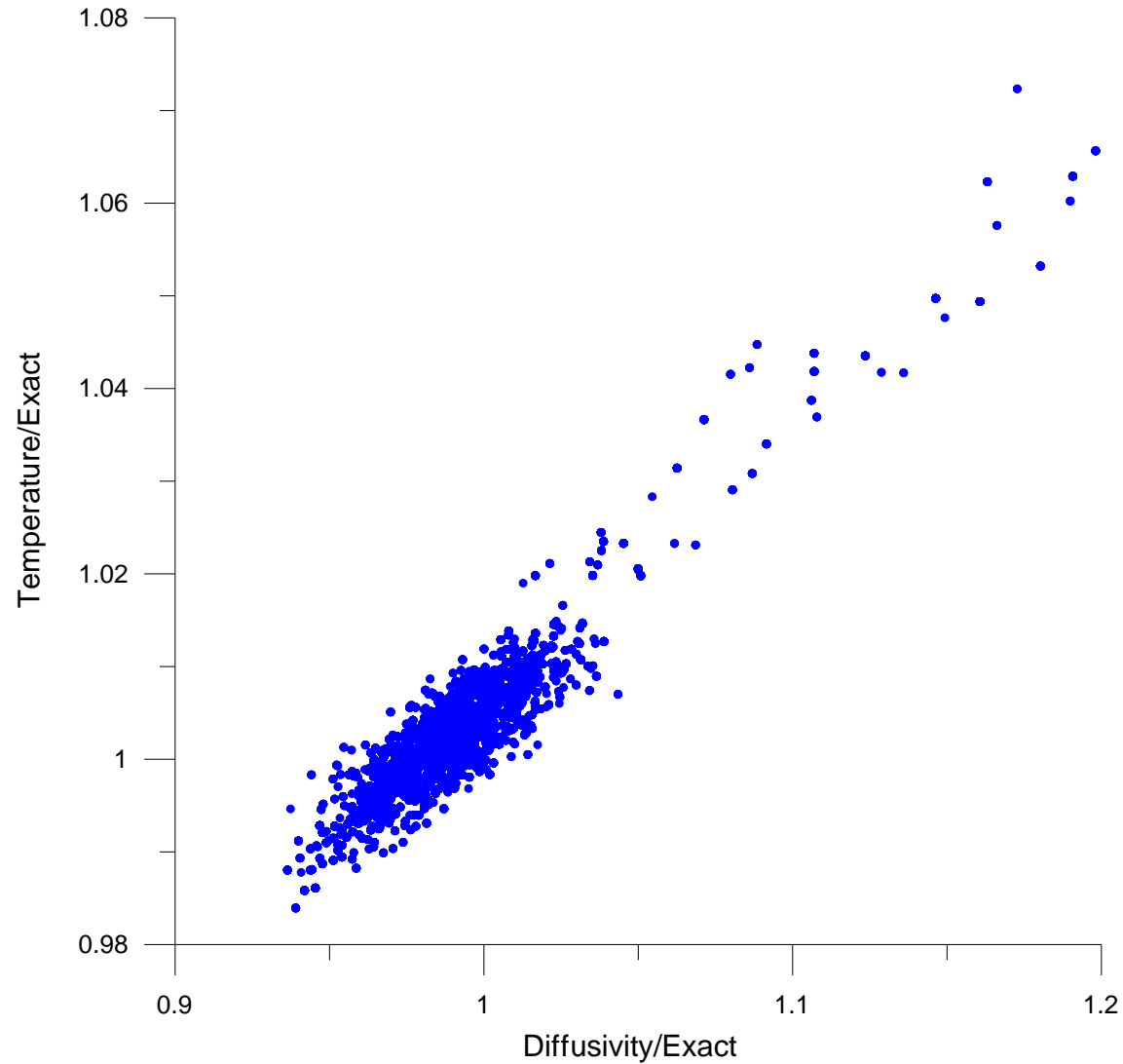
EXAMPLE



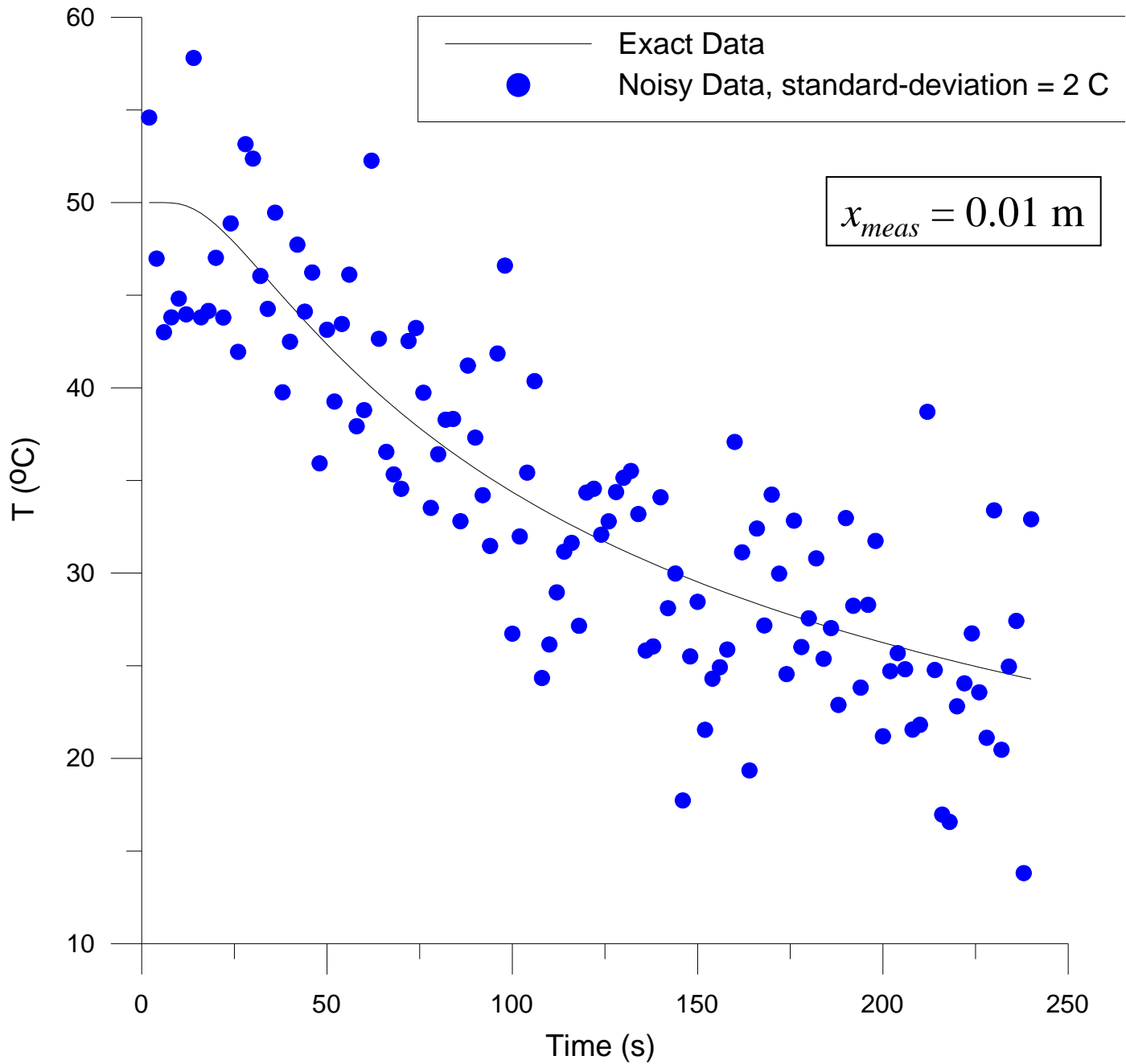


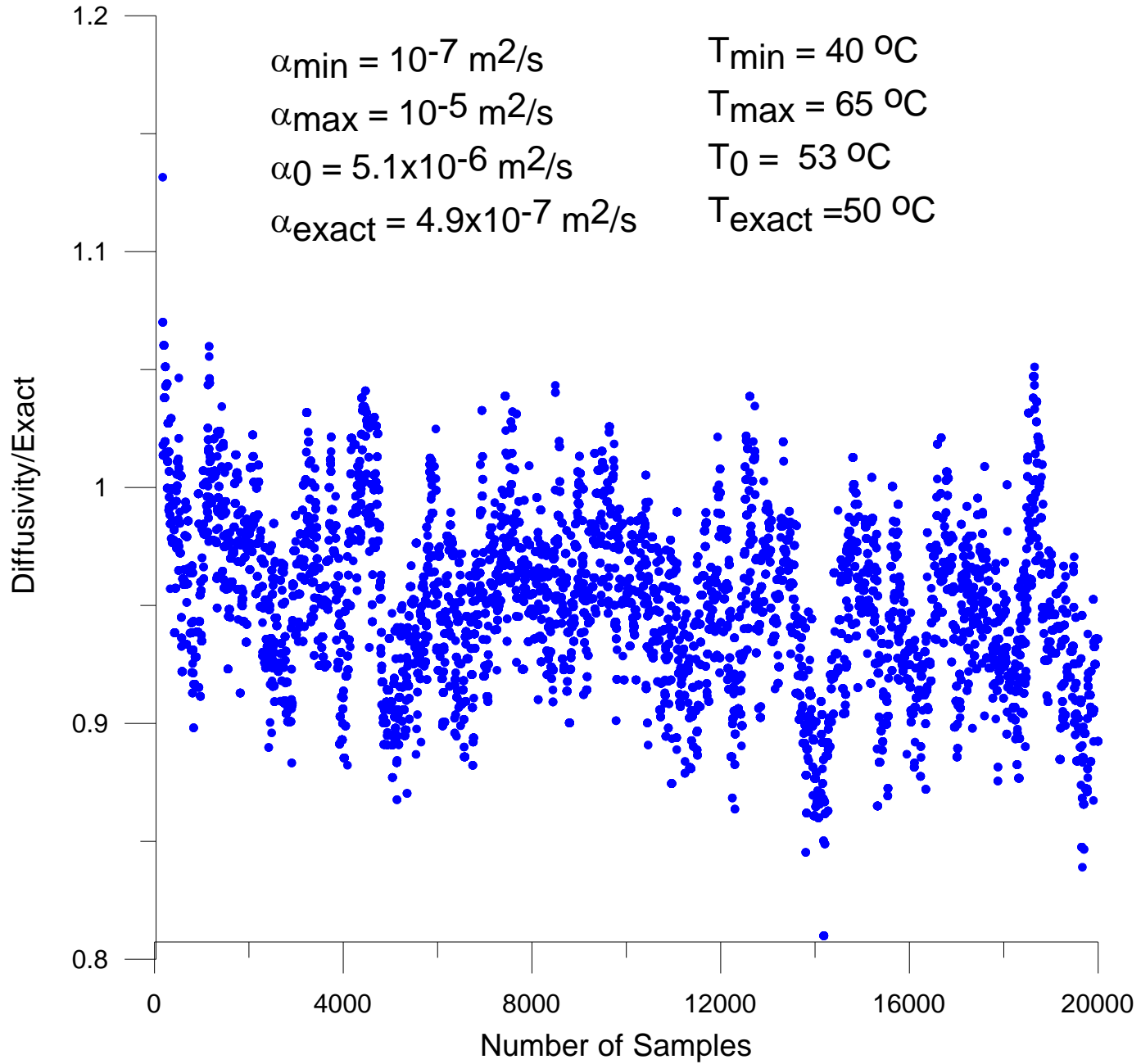


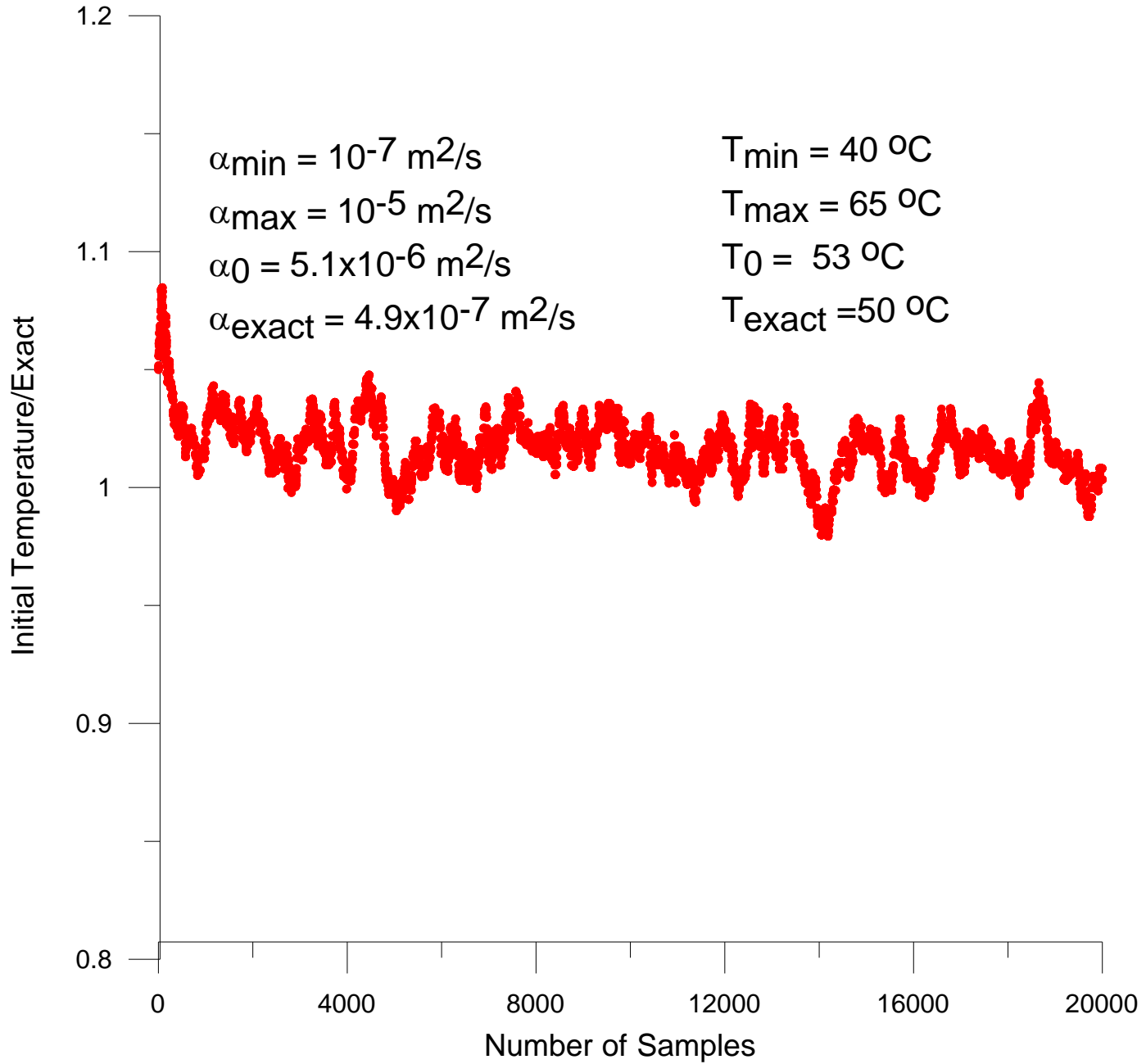


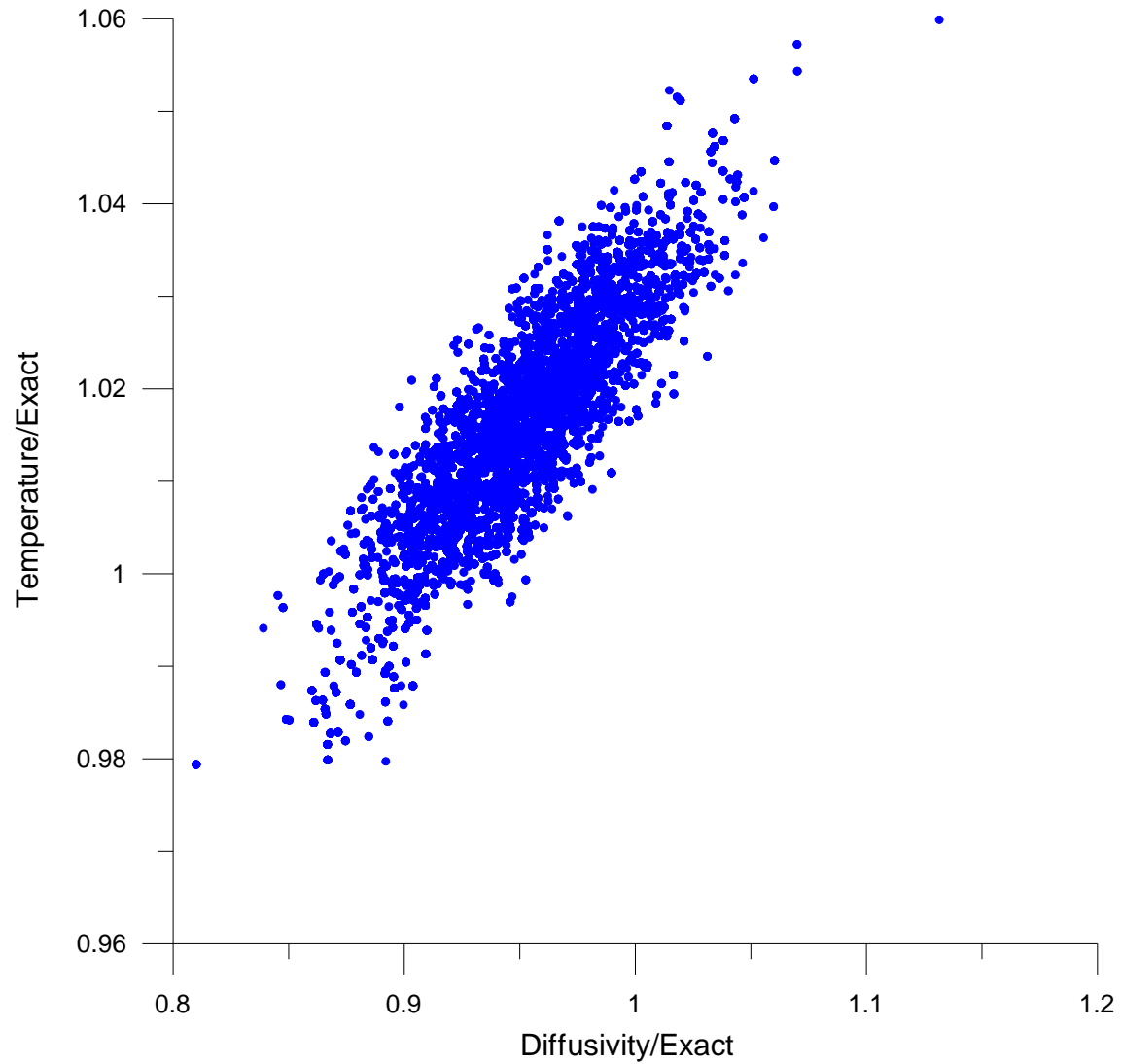


	Diffusivity/Exact	Temperature/Exact
Mean	0.991879086	1.003147487
Standard-Deviation	0.030681073	0.009687663









	Diffusivity/Exact	Temperature/Exact
Mean	0.950617158	1.016337295
Standard-Deviation	0.034386526	0.011192818

Priors – Markov Random Field

A collection $\{P_1, P_2, \dots, P_N\}$ is a *Markov Random Field* if the full conditional distribution of P_j depends only on its set of neighbors [28].

A common use of a Markov Random Field is for priors that resemble Tikhonov's regularization [5], written in the following general form

$$\pi(\mathbf{P}) \propto \exp\left[-\frac{1}{2}\gamma\|\mathbf{D}(\mathbf{P} - \tilde{\mathbf{P}})\|^2\right] \quad (36)$$

where $\|\cdot\|$ denotes the L_2 norm. The constant γ is a parameter associated with uncertainties in the prior and $\tilde{\mathbf{P}}$ is a reference value for \mathbf{P} . The matrix \mathbf{D} is such that each line of $\mathbf{D}(\mathbf{P} - \tilde{\mathbf{P}})$ involves the parameter P_j corresponding to that line and its neighbors

Practical Issues for MCMC Methods

Priors – Markov Random Field

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & -1 & 1 \end{bmatrix} \quad \text{with size } (N-1) \times N$$

$$\mathbf{D} = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \\ & & & & & & \end{bmatrix} \quad \text{with size } (N-2) \times N$$

Practical Issues for MCMC Methods

Equation (36) can be rewritten as

$$\pi(\mathbf{P}) \propto \exp \left[-\frac{1}{2} \gamma (\mathbf{P} - \tilde{\mathbf{P}})^T \mathbf{Z} (\mathbf{P} - \tilde{\mathbf{P}}) \right] \quad (38.a)$$

where

$$\mathbf{Z} = \mathbf{D}^T \mathbf{D} \quad (38.b)$$

Equation (38.a) is in a form similar to that of a Gaussian distribution. For this reason, it is also called a Gaussian Markov Random Field [28] or a Gaussian Smoothness Prior [5]. By comparing equation (38.a) with the canonical Gaussian multivariate distribution, one can notice that the mean and the covariance matrix of this prior are given by $\tilde{\mathbf{P}}$ and $\gamma^{-1} \mathbf{Z}^{-1}$, respectively. Therefore, we can write the Gaussian Smoothness Prior as

$$\pi(\mathbf{P}) = (2\pi)^{-N/2} \gamma^{N/2} |\mathbf{Z}^{-1}|^{-1/2} \exp \left[-\frac{1}{2} \gamma (\mathbf{P} - \tilde{\mathbf{P}})^T \mathbf{Z} (\mathbf{P} - \tilde{\mathbf{P}}) \right] \quad (39)$$

An important remark about this prior is that, with \mathbf{D} given by equations (37.a,b), its variance is unbounded, since the matrix \mathbf{Z} is singular and \mathbf{Z}^{-1} does not exist. Densities with unbounded variances are denoted as *improper* [5,28].

Priors – Markov Random Field

We now discuss another Markov Random Field prior, which gives high probabilities for piecewise regular solutions with sparse gradients. The *Total Variation (TV) prior* satisfies these characteristics, being quite appropriate for spatially varying functions that contain large variations at few boundaries within the domain and with small variations within the regions limited by such boundaries [5]. The TV prior is given by [5]:

$$\pi(\mathbf{P}) \propto \exp[-\gamma TV(\mathbf{P})] \quad (40)$$

where

$$TV(\mathbf{P}) = \sum_{j=1}^N V_j(\mathbf{P}) \quad V_j(\mathbf{P}) = \frac{1}{2} \sum_{i \in N_j} l_{ij} |P_i - P_j| \quad (41.a,b)$$

being N_j the set of neighbors to P_j and l_{ij} the length of the edge between neighbors.

Priors – Markov Random Field

The TV prior is improper, such as the Gaussian smoothness prior. The representation of equation (40) in terms of a canonical probability density would require the derivation of an expression for the normalizing constant $\int_{R^N} \pi(\mathbf{P}) d\mathbf{P}$, or, at least, practical means for its computation. Although improper priors need to be used with caution, they do not pose difficulties for the application of the Metropolis-Hastings algorithm, since the normalizing constants of such densities are cancelled when $\alpha(\mathbf{P}^* | \mathbf{P}^{(t)})$ is computed with equation (26). On the other hand, both the Gaussian smoothness prior and the TV prior involve an additional parameter γ that needs to be specified for the application of MCMC methods. The specification of a value for such parameter can be made by numerical experiments, by using simulated experimental data that serve as a reference for the inverse problem under analysis. On the other hand, within the Bayesian framework, if a parameter is not known it shall be regarded as part of the inference problem, leading to the use of hierarchical (*hyperprior*) models, as described below.

Hierarchical Models

The parameter γ appearing in the Gaussian smoothness prior given by equation (39) can be treated as a *hyperparameter*, that is, be estimated as part of the inference problem [5]. Consider, for example, the *hyperprior density* for γ in the form of a Rayleigh distribution (see equation 33), where the scale parameter γ_0 can be chosen as sufficiently large in order to avoid any restriction on possible values for γ . Therefore, the posterior distribution, with the Gaussian likelihood given by equation (6.b), can be written as:

$$\pi(\gamma, \mathbf{P} | \mathbf{Y}) \propto \gamma^{(N+2)/2} \exp \left\{ -\frac{1}{2} [\mathbf{Y} - \mathbf{T}(\mathbf{P})]^T \mathbf{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P})] - \frac{1}{2} \gamma (\mathbf{P} - \tilde{\mathbf{P}})^T \mathbf{Z} (\mathbf{P} - \tilde{\mathbf{P}}) - \frac{1}{2} \left(\frac{\gamma}{\gamma_0} \right)^2 \right\} \quad (42)$$

On the other hand, the parameter γ appearing in the TV prior given by equation (40) cannot be treated as a hyperparameter. Such is the case because the normalizing constant of such prior is of difficult calculation and also depends on γ .

Output Analysis

We basically follow references [22,28] for the material presented in this section and consider the analysis on a single component P_j of the vector of parameters \mathbf{P} . Let $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\}$ be a homogeneous and reversible Markov chain for P_j . A function $f(P_j^{(n)})$ from the sample $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\}$ is called a *statistic* if it does not depend on any other unknown parameters. Some useful statistics are:

$$\text{Minimum Value: } f(P_j^{(n)}) = P_{j,\min}^{(n)} = \min\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\} \quad (43.a)$$

$$\text{Maximum Value: } f(P_j^{(n)}) = P_{j,\max}^{(n)} = \max\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\} \quad (43.b)$$

$$\text{Median: } f(P_j^{(n)}) = \tilde{P}_j^{(n)} = \text{med}\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\} \quad (43.c)$$

$$\text{Mean: } f(P_j^{(n)}) = \bar{P}_j^{(n)} = \frac{1}{n} \sum_{t=1}^n P_j^{(t)} \quad (43.d)$$

$$\text{Variance: } f(P_j^{(n)}) = \text{var}(P_j^{(n)}) = \frac{1}{n-1} \sum_{t=1}^n (P_j^{(t)} - \bar{P}_j^{(n)})^2 \quad (43.e)$$

Output Analysis

Since $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\}$ are realizations of a random variable, a statistic is itself a random variable as well. A statistic of the sample will be a good representation of a statistic of the population if the sample is a good representation of the population. This certainly depends on the size n and on the independence of the individuals of the sample. Furthermore, since the sample $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\}$ is obtained from a Markov chain, the chain should already have reached equilibrium before statistics can be computed for the solution of the inverse problem. For this reason, states of the Markov chain are discarded before the chain reaches equilibrium, which is called the burn-in period. If m states are needed for the chain to reach equilibrium, the sample used for the computation of the statistics is $\{P_j^{(m+1)}, P_j^{(m+2)}, \dots, P_j^{(n)}\}$. The index of this sample is changed from $t = m+1, \dots, n$ to $r = 1, \dots, s$ for simplicity in the notation, where $s = n - m$ is the number of samples used for the computation of the statistics.

Output Analysis

The *mean* of the sequence $f(P_j^{(r)}) \equiv \{f(P_j^{(1)}), f(P_j^{(2)}), \dots, f(P_j^{(s)})\}$ is

$$\bar{f}_s(P_j^{(r)}) = \frac{1}{s} \sum_{r=1}^s f(P_j^{(r)}) \quad (44)$$

If the chain is ergodic, this mean based on the chain values $f(P_j^{(r)})$ provides a strongly consistent estimate of the mean of the limiting distribution, that is,

$$\bar{f}_s(P_j^{(r)}) \rightarrow E[f(P_j)] \quad \text{as } s \rightarrow \infty \quad (45)$$

This result is the equivalent of the law of large numbers for a Markov chain.

Output Analysis

If $\{f(P_j^{(1)}), f(P_j^{(2)}), \dots, f(P_j^{(s)})\}$ are independent samples, then the *variance of the mean* $\bar{f}_s(P_j^{(r)})$ is

$$\text{var}[\bar{f}_s(P_j^{(r)})] = \frac{\text{var}[f(P_j^{(r)})]}{s} \quad (46.a)$$

where $\text{var}[f(P_j^{(r)})]$ is the variance of $\{f(P_j^{(1)}), f(P_j^{(2)}), \dots, f(P_j^{(s)})\}$. On the other hand, since the samples are in general correlated, equation (46.a) is rewritten as

$$\text{var}[\bar{f}_s(P_j^{(r)})] = \frac{\tau \text{var}[f(P_j^{(r)})]}{s} \quad (46.b)$$

where τ is the *integrated autocorrelation time* (IACT), which represents the number of correlated samples between independent samples in the chain

$\{f(P_j^{(1)}), f(P_j^{(2)}), \dots, f(P_j^{(s)})\}$. Therefore, the effective chain size, which gives the number of independent samples in the chain, is $s_{\text{eff}} = s / \tau$.

Output Analysis

The *autocovariance function of lag k* of the chain for $\{f(P_j^{(1)}), f(P_j^{(2)}), \dots, f(P_j^{(s)})\}$ is defined by:

$$C_{ff}(k) = \text{cov}[f(P_j^{(r)}), f(P_j^{(r+k)})] \quad (47)$$

Clearly, the variance of $f(P_j^{(r)})$ is $\text{var}[f(P_j^{(r)})] = C_{ff}(0)$.

The *normalized autocovariance function of lag k* is given by

$$\rho_{ff}(k) = \frac{C_{ff}(k)}{C_{ff}(0)} \quad (48)$$

so that $\rho_{ff}(0) = 1$, which means that $f(P_j^{(r)})$ is perfectly correlated with itself.

Output Analysis

The integrated autocorrelation time is related to the *normalized autocovariance function* by

$$\tau = 1 + 2 \sum_{k=1}^{\infty} \rho_{ff}(k) \quad (49)$$

For the calculation of τ , the summation in equation (49) needs to be truncated at a finite number of terms $s^* \leq s$. In fact, $\rho_{ff}(k)$ is expected to tend to zero as k increases, but it will be dominated by noise for large k . Therefore, s^* can be selected by increasing k until $\rho_{ff}(k)$ is first approximately zero, thus avoiding the terms that are dominated by noise in $f(P_j^{(r)})$.

Output Analysis

For s sufficiently large and for an uniformly ergodic chain, the distribution of $\frac{\bar{f}_s(P_j^{(r)}) - E[f(P_j)]}{\sqrt{\text{var}[\bar{f}_s(P_j^{(r)})]}}$, where $\text{var}[\bar{f}_s(P_j^{(r)})]$ is given by equation (46.b), tends to a standard

Gaussian distribution, with zero mean and unitary standard deviation. One can write

$$\frac{\bar{f}_s(P_j^{(r)}) - E[f(P_j)]}{\sqrt{\text{var}[\bar{f}_s(P_j^{(r)})]}} \xrightarrow{d} N(0,1) \quad \text{as } s \rightarrow \infty \quad (50)$$

where \xrightarrow{d} indicates that the distribution of the random variable on the left tends to the distribution on the right. Equation (50) is an statement of the central limit theorem of the distribution of $\bar{f}_s(P_j^{(r)})$.

Output Analysis

A statistic of great interest is the mean of $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(s)}\}$. Therefore, by assuming that the appropriate assumptions are satisfied, equation (50) shows that

$$\frac{\bar{P}_j^{(r)} - E[P_j]}{\sqrt{\text{var}[\bar{P}_j^{(r)}]}} \xrightarrow{d} N(0,1) \quad \text{as } s \rightarrow \infty \quad (51)$$

where

$$\bar{P}_j^{(r)} = \frac{1}{s} \sum_{r=1}^s P_j^{(r)} \quad ; \quad \text{var}[\bar{P}_j^{(r)}] = \frac{\tau \text{var}[P_j^{(r)}]}{s} \quad ; \quad \text{var}[P_j^{(r)}] = \frac{1}{s-1} \sum_{r=1}^s (P_j^{(r)} - \bar{P}_j^{(r)})^2 \quad (52.a,b,c)$$

are the mean of $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(s)}\}$, the variance of this mean and the variance of $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(s)}\}$, respectively.

Practical Issues for MCMC Methods

Output Analysis

The main result of equation (51) is that it provides a reasonable manner of presenting the solution of the inverse problem of estimating the parameter P_j , from inference over the Markov chain $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(s)}\}$, as $\bar{P}_j^{(r)} \pm C \sqrt{\text{var}[\bar{P}_j^{(r)}]}$, where C is a constant that defines the approximate confidence interval of $\bar{P}_j^{(r)}$. For a 99% confidence interval, $C = 2.576$.

Practical Issues for MCMC Methods

The convergence of the Markov chain to an equilibrium distribution can be verified by plotting the chains of each parameter $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(n)}\}$, $j = 1, \dots, N$, and the posterior distribution $\pi_{posterior}(\mathbf{P}^{(t)})$, $t = 1, \dots, n$. Geweke [34] proposed a method for convergence diagnosis based on means computed with different ranges of the Markov chain. Let

$$\bar{P}_j^a = \frac{1}{s_a} \sum_{r=1}^{s_a} P_j^{(r)} \quad \text{and} \quad \bar{P}_j^b = \frac{1}{s_b} \sum_{r=s^*}^s P_j^{(r)} \quad (53.a,b)$$

where

$$s^* = s - s_b + 1 \quad ; \quad s_a = 0.1s \quad ; \quad s_b = 0.5s \quad (54.a-c)$$

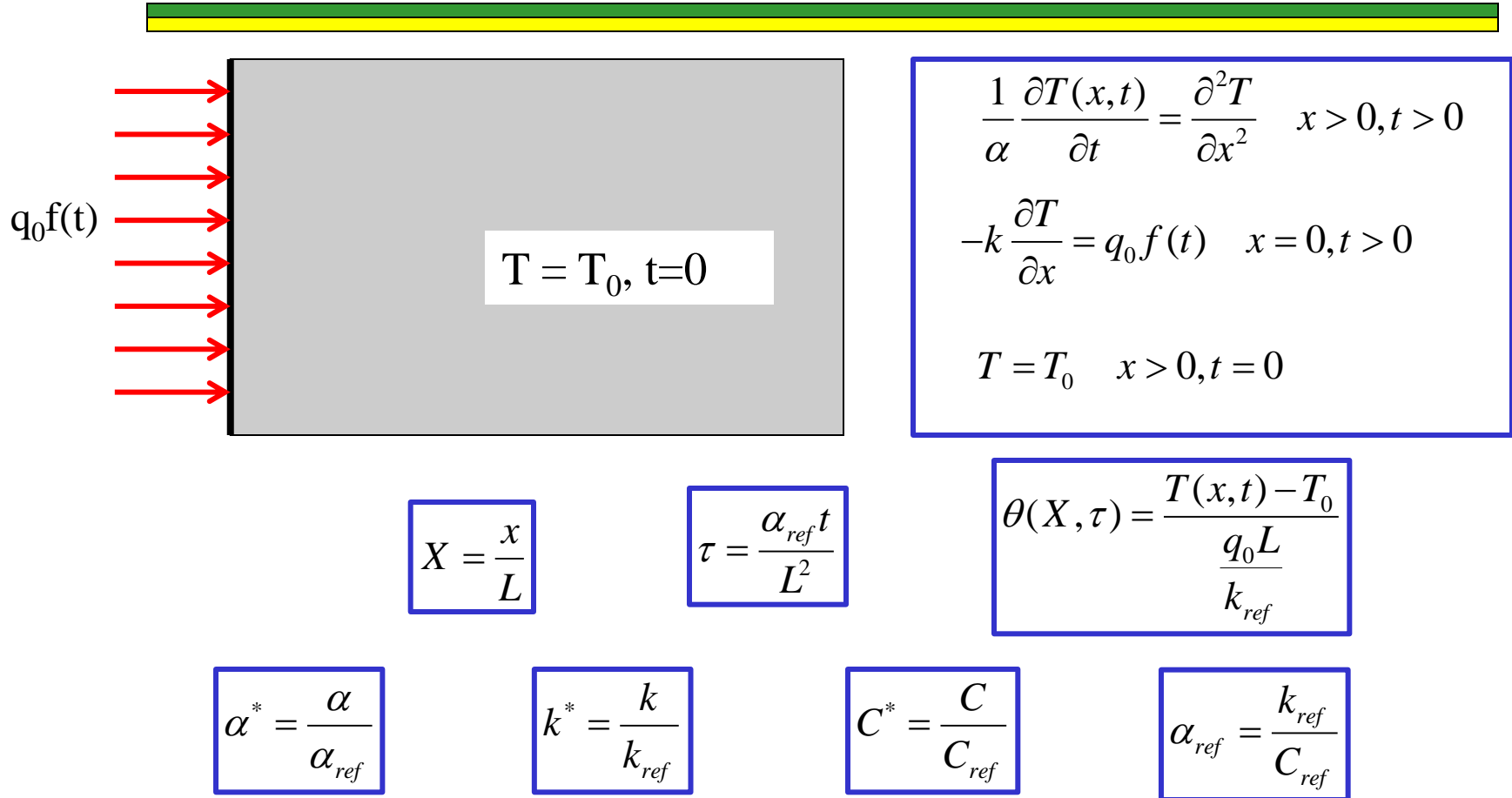
Geweke [34] has demonstrated that $(\bar{P}_j^a - \bar{P}_j^b) \rightarrow 0$ as the chain $\{P_j^{(1)}, P_j^{(2)}, \dots, P_j^{(s)}\}$ approaches equilibrium.

It is also a good practice to repeat such procedures for convergence analysis by generating Markov chains from different initial states. A method for inference from multiple chains was developed by Gelman and Rubin [35].



Markov Chain Monte Carlo (MCMC) methods

EXAMPLE





Markov Chain Monte Carlo (MCMC) methods

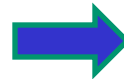
EXAMPLE

$$\frac{1}{\alpha^*} \frac{\partial \theta(X, \tau)}{\partial \tau} = \frac{\partial^2 \theta}{\partial X^2} \quad X > 0, \tau > 0$$

$$-k^* \frac{\partial \theta}{\partial X} = f(\tau) \quad X = 0, \tau > 0$$

$$\theta = 0 \quad X > 0, \tau = 0$$

$$f(\tau) = \phi = \text{constant}$$

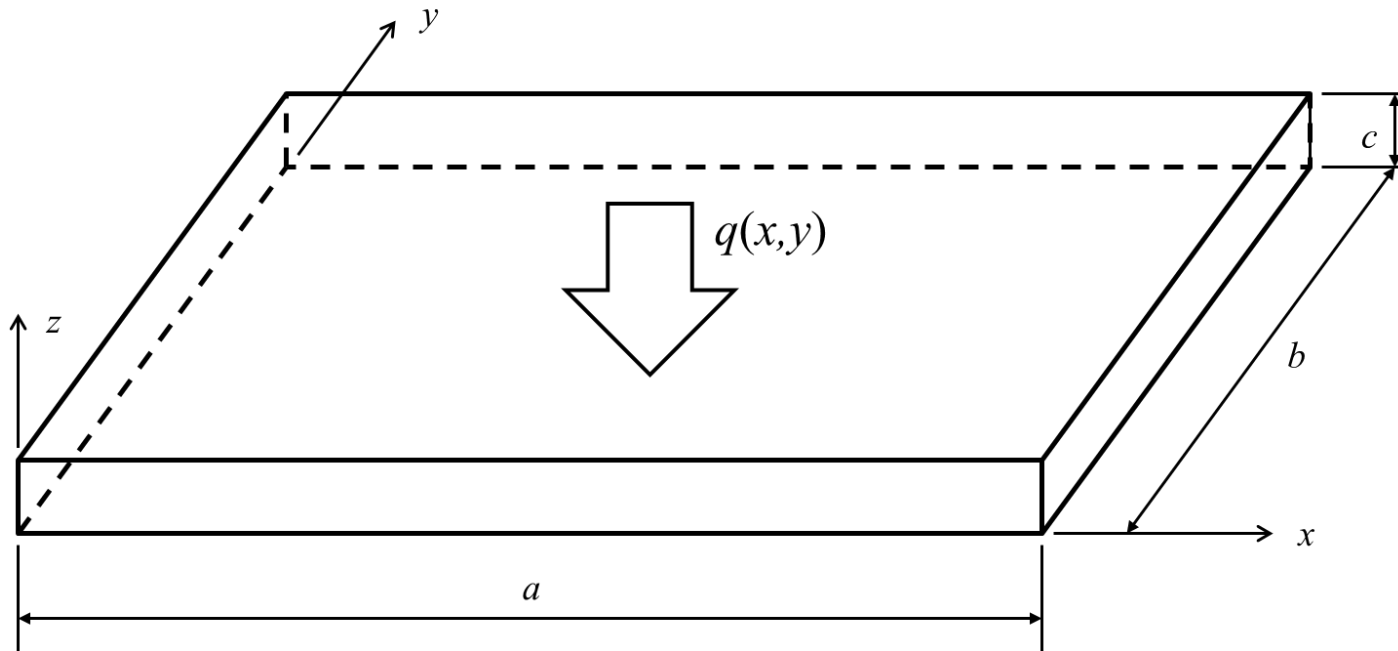


$$\theta(X = 0, \tau) = \frac{2\phi\sqrt{\tau}}{\sqrt{\pi}\sqrt{k^*C^*}}$$

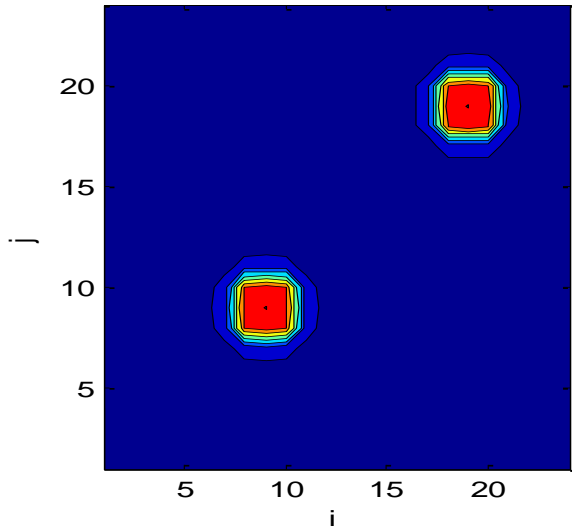
$$\text{Thermal Effusivity} = e^* = \sqrt{k^*C^*}$$

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

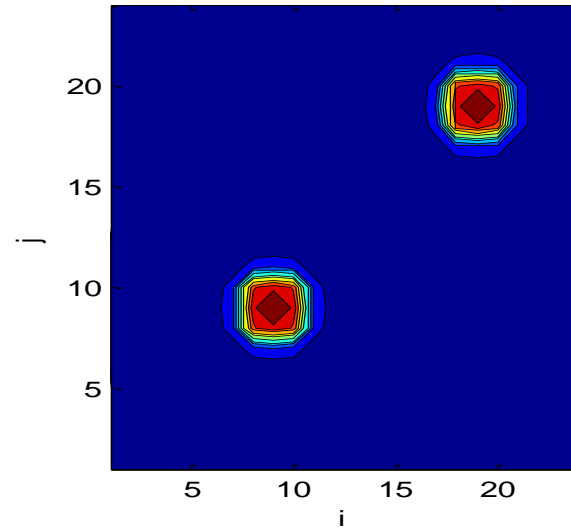
Example: Non linear 3D heat conduction
Estimation of $q(x,y)$ with measurements of $T(x,y,0,t)$



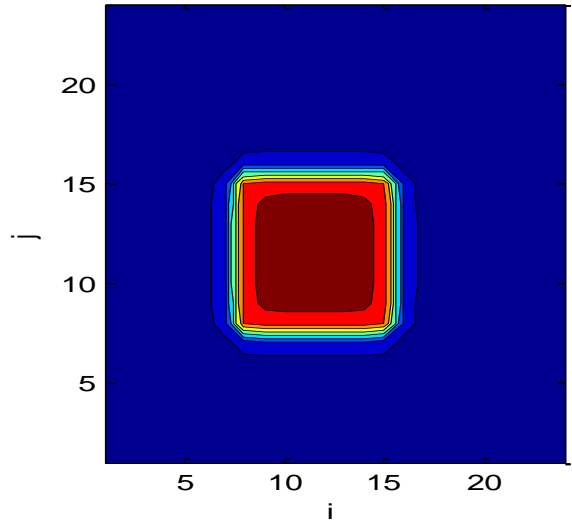
Temperature Top Surface, K



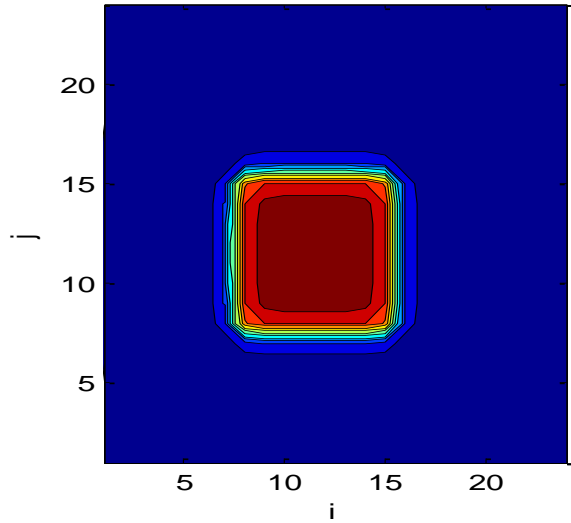
Temperature Bottom surface, K



Temperature Top Surface, K



Temperature Bottom surface, K



500 1000 1500

400 600 800 1000 1200 1400

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

Complete model

$$C(T_c) \frac{\partial T_c(x, y, z, t)}{\partial t} = \frac{\partial}{\partial x} \left[k(T_c) \frac{\partial T_c}{\partial x} \right] + \frac{\partial}{\partial y} \left[k(T_c) \frac{\partial T_c}{\partial y} \right] + \frac{\partial}{\partial z} \left[k(T_c) \frac{\partial T_c}{\partial z} \right]$$

in $0 < x < a, 0 < y < b, 0 < z < c, \text{ for } t > 0$

$$\frac{\partial T_c}{\partial x} = 0 \quad \text{at } x = 0 \text{ and } x = a, 0 < y < b, 0 < z < c, \text{ for } t > 0$$

$$\frac{\partial T_c}{\partial y} = 0 \quad \text{at } y = 0 \text{ and } y = b, 0 < x < a, 0 < z < c, \text{ for } t > 0$$

$$\frac{\partial T_c}{\partial z} = 0 \quad \text{at } z = 0, 0 < x < a, 0 < y < b, \text{ for } t > 0$$

$$k(T_c) \frac{\partial T_c}{\partial z} = q(x, y) \quad \text{at } z = c, 0 < x < a, 0 < y < b, \text{ for } t > 0$$

$$T_c = T_0 \quad \text{for } t = 0, \text{ in } 0 < x < a, 0 < y < b, 0 < z < c$$

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

Reduced models: Linear problem with properties at T^*

$$C^* \frac{\partial \bar{T}(x, y, t)}{\partial t} = \frac{\partial}{\partial x} \left[k^* \frac{\partial \bar{T}}{\partial x} \right] + \frac{\partial}{\partial y} \left[k^* \frac{\partial \bar{T}}{\partial y} \right] + \frac{q(x, y)}{c}$$

in $0 < x < a$, $0 < y < b$, for $t > 0$

$$\frac{\partial \bar{T}}{\partial x} = 0 \quad \text{at } x = 0 \text{ and } x = a , 0 < y < b , \text{ for } t > 0$$

$$\frac{\partial \bar{T}}{\partial y} = 0 \quad \text{at } y = 0 \text{ and } y = b , 0 < x < a , \text{ for } t > 0$$

$$\bar{T} = T_0 \quad \text{for } t = 0 , \text{ in } 0 < x < a , 0 < y < b$$

where

$$\bar{T}(x, y, t) = \frac{1}{c} \int_{z=0}^c T(x, y, z, t) dz$$

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

Reduced models: Linear problem with properties at T^*

Classical Lumped Formulation:

Temperature gradients across the thickness of the plate are fully neglected.

$$T(x, y, 0, t) = T(x, y, c, t) = \bar{T}(x, y, t)$$

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

Improved Lumped Formulation:

Temperature gradients across the thickness of the plate are not neglected, but taken into account in an approximate form (Cotta, R.M., Mikhailov, M.D., *Heat Conduction: Lumped Analysis, Integral Transforms, Symbolic Computation*, Wiley-Interscience, New York, USA, 1997.).

H_{1,1} formula (correct trapezoidal rule):

$$\bar{T}(x, y, t) \approx \frac{1}{2} [T(x, y, 0, t) + T(x, y, c, t)] + \frac{c}{12} \left[\left. \frac{\partial T}{\partial z} \right|_{z=0} - \left. \frac{\partial T}{\partial z} \right|_{z=c} \right]$$

H_{0,0} formula (trapezoidal rule):

$$\int_{z=0}^c \frac{\partial T(x, y, z, t)}{\partial z} dz = T(x, y, c, t) - T(x, y, 0, t) \approx \frac{c}{2} \left[\left. \frac{\partial T}{\partial z} \right|_{z=0} + \left. \frac{\partial T}{\partial z} \right|_{z=c} \right]$$

$$T(x, y, 0, t) = \bar{T}(x, y, t) - \frac{c}{6k^*} q(x, y)$$

$$T(x, y, c, t) = \bar{T}(x, y, t) + \frac{c}{3k^*} q(x, y)$$

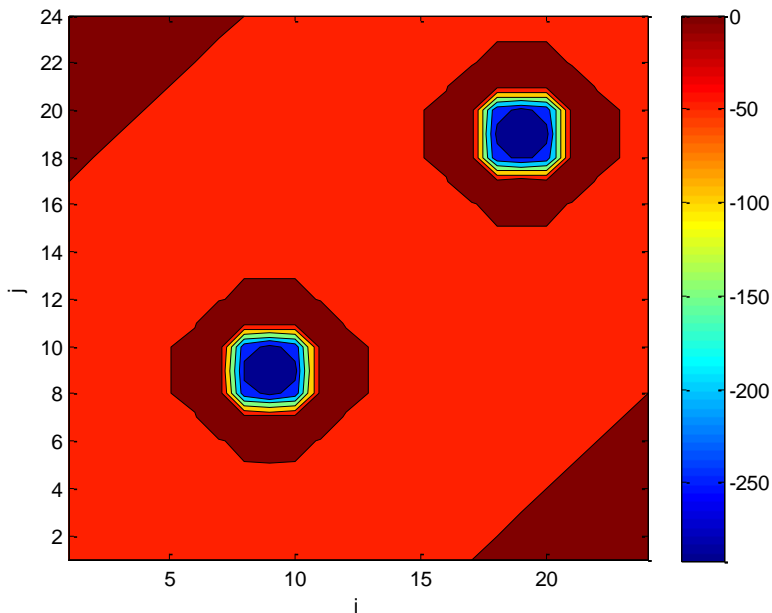
In general, the direct problem solution with the **complete model took around 7.2 s**, while the solution with the **reduced model took around 0.09 s** of CPU time.

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

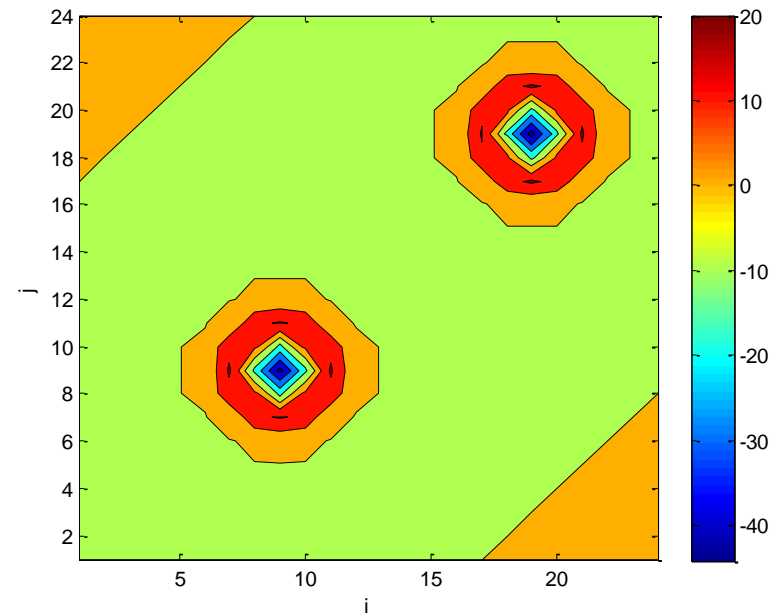
Error of the Direct Problem Solution at the final time:

$$q(x_i, y_j) = \begin{cases} 10^7 \text{ Wm}^{-2} & , \text{ for } 8 \leq i \leq 10 \text{ and } 8 \leq j \leq 10 \\ 10^7 \text{ Wm}^{-2} & , \text{ for } 18 \leq i \leq 20 \text{ and } 18 \leq j \leq 20 \\ 0 & , \text{ elsewhere} \end{cases}$$

Classical Lumped Model



Improved Lumped Model



Orlande, H.R.B., Dulikravich, G., Inverse Heat Transfer Problems and their Solutions within the Bayesian Framework, ECCOMAS Special Interest Conference, Numerical Heat Transfer 2012, 4-6 September 2012, Gliwice-Wrocław, Poland

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

DELAYED ACCEPTANCE METROPOLIS-HASTINGS ALGORITHM

(Christen, J. and Fox, C., Markov chain Monte Carlo Using an Approximation, *Journal of Computational and Graphical Statistics*, vol. 14, no. 4, pp. 795–810, 2005)

The regular Metropolis-Hastings algorithm is applied with the surrogate model for the computation of the likelihood function. If a proposal state is accepted, then another test of Hastings is performed with the complete model, to finally decide if such proposal should be accepted or not.

Helcio R. B. Orlande, George S. Dulikravich, Markus Neumayer, Daniel Watzenig, Marcelo J. Colaço, Accelerated Bayesian Inference For The Estimation Of Spatially Varying Heat Flux In A Heat Conduction Problem, *Numerical Heat Transfer – Part A*, 65:1, 1-25, 2014

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

DELAYED ACCEPTANCE METROPOLIS-HASTINGS ALGORITHM

(Christen, J. and Fox, C., Markov chain Monte Carlo Using an Approximation, *Journal of Computational and Graphical Statistics*, vol. 14, no. 4, pp. 795–810, 2005)

1. Sample a *Candidate Point* \mathbf{P}^* from a proposal distribution $p(\mathbf{P}^*, \mathbf{P}^{(t-1)})$.
2. Calculate the acceptance factor with the surrogate model:

$$\alpha = \min \left[1, \frac{\pi(\mathbf{P}^* | \mathbf{Y}) p(\mathbf{P}^{(t-1)}, \mathbf{P}^*)}{\pi(\mathbf{P}^{(t-1)} | \mathbf{Y}) p(\mathbf{P}^*, \mathbf{P}^{(t-1)})} \right]$$

3. Generate a random value U that is uniformly distributed on $(0,1)$.
4. If $U \leq \alpha$, proceed to step 5. Otherwise, return to step 1.
5. Calculate a new acceptance factor with the complete model:

$$\alpha_c = \min \left[1, \frac{\pi_c(\mathbf{P}^* | \mathbf{Y}) p(\mathbf{P}^{(t-1)}, \mathbf{P}^*)}{\pi_c(\mathbf{P}^{(t-1)} | \mathbf{Y}) p(\mathbf{P}^*, \mathbf{P}^{(t-1)})} \right]$$

6. Generate a new random value U_c which is uniformly distributed on $(0,1)$.
7. If $U_c \leq \alpha_c$, set $\mathbf{P}^{(t)} = \mathbf{P}^*$. Otherwise, set $\mathbf{P}^{(t)} = \mathbf{P}^{(t-1)}$.
8. Return to step 1.

where $\pi(\mathbf{P} | \mathbf{Y})$ and $\pi_c(\mathbf{P} | \mathbf{Y})$ are the posterior distributions with the likelihoods computed with the surrogate model and with the complete model, respectively.

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

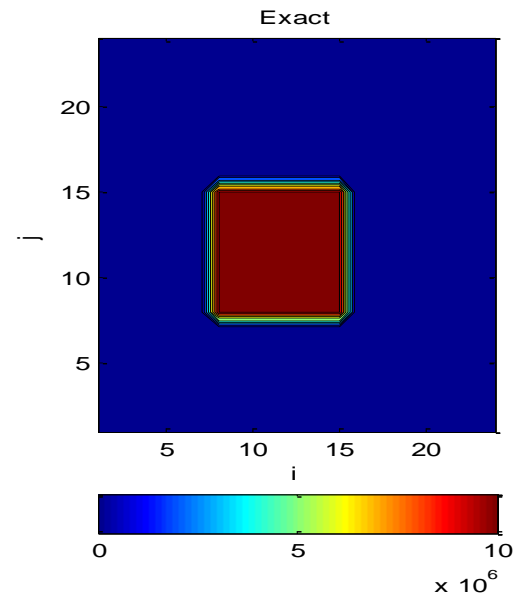
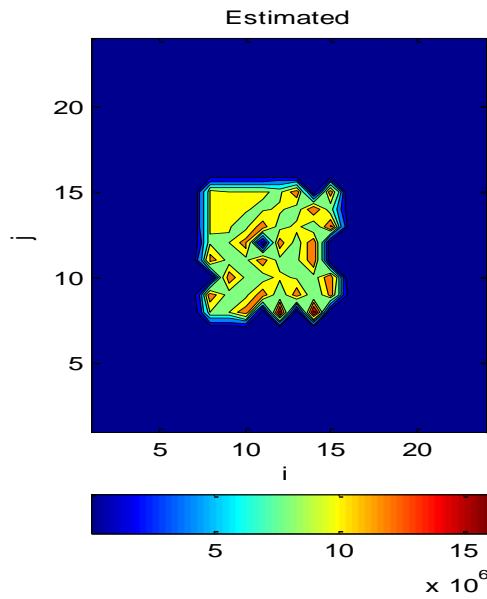
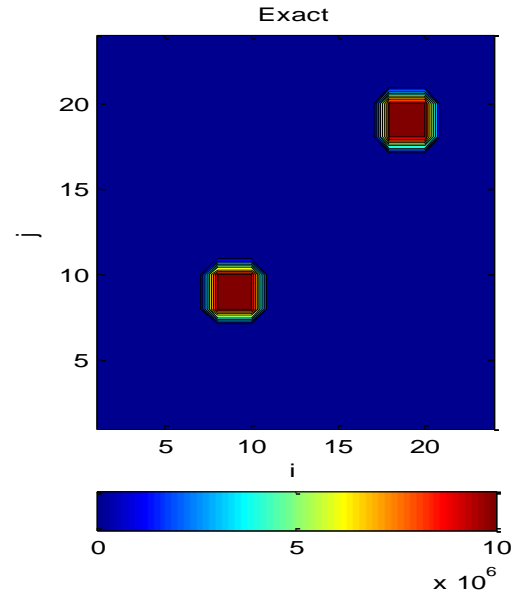
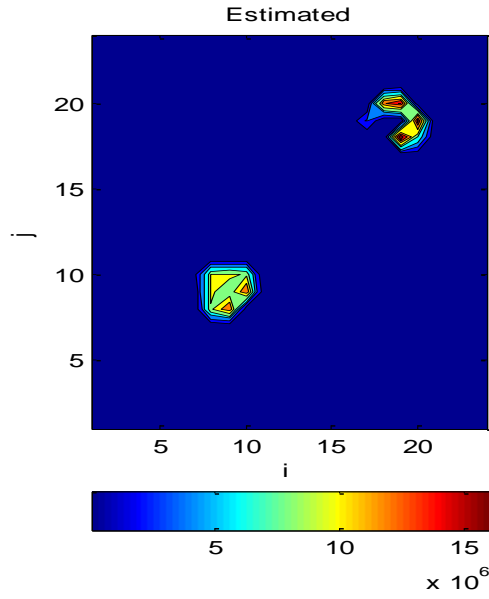
PRIOR DISTRIBUTIONS

Total variation non-informative prior

$$\pi(\mathbf{P}) \propto \exp[-\alpha TV(\mathbf{P})]$$

$$TV(\mathbf{P}) = \sum_{i=2}^{I-1} \sum_{j=2}^{J-1} \Delta y \left[\left| q(x_i, y_j) - q(x_{i+1}, y_j) \right| + \left| q(x_i, y_j) - q(x_{i-1}, y_j) \right| \right] + \\ + \Delta x \left[\left| q(x_i, y_j) - q(x_i, y_{j+1}) \right| + \left| q(x_i, y_j) - q(x_i, y_{j-1}) \right| \right]$$

TV prior + DAMH - $\sigma = 1.25$ K



REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

APPROXIMATION ERROR MODEL

- Kaipio, J. and Somersalo, E., *Statistical and Computational Inverse Problems*, Applied Mathematical Sciences 160, Springer-Verlag, 2004
- Kaipio, J., and Somersalo, E., Statistical Inverse Problems: Discretization, Model Reduction and Inverse Crimes, *Journal of Computational and Applied Mathematics*, vol. 198, pp. 493–504, 2007.

In the approximation error model (AEM) approach, the statistical model of the approximation error is constructed and then represented as additional noise in the measurement model [1,19-23]. With the hypotheses that the measurement errors are additive and independent of the parameters \mathbf{P} , one can write

$$\mathbf{Y} = \mathbf{T}_c(\mathbf{P}) + \mathbf{e} \quad (16)$$

where $\mathbf{T}_c(\mathbf{P})$ is the sufficiently accurate solution of the complete model given by equations (1.a-h). The vector of measurement errors, \mathbf{e} are assumed here to be Gaussian, with zero mean and known covariance matrix \mathbf{W} .

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

APPROXIMATION ERROR MODEL

$$\mathbf{Y} = \mathbf{T}(\mathbf{P}) + [\mathbf{T}_c(\mathbf{P}) - \mathbf{T}(\mathbf{P})] + \mathbf{e}$$

By defining the error between the complete and the surrogate model solutions as

$$\boldsymbol{\varepsilon} = [\mathbf{T}_c(\mathbf{P}) - \mathbf{T}(\mathbf{P})]$$

equation (17) can be written as

$$\mathbf{Y} = \mathbf{T}(\mathbf{P}) + \boldsymbol{\eta}$$

where

$$\boldsymbol{\eta} = \boldsymbol{\varepsilon} + \mathbf{e}$$

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

APPROXIMATION ERROR MODEL

η is modeled as a Gaussian variable

Enhanced error model:

$$\bar{\eta} \approx \bar{\varepsilon}$$

$$\tilde{W} \approx W_{\varepsilon} + W$$

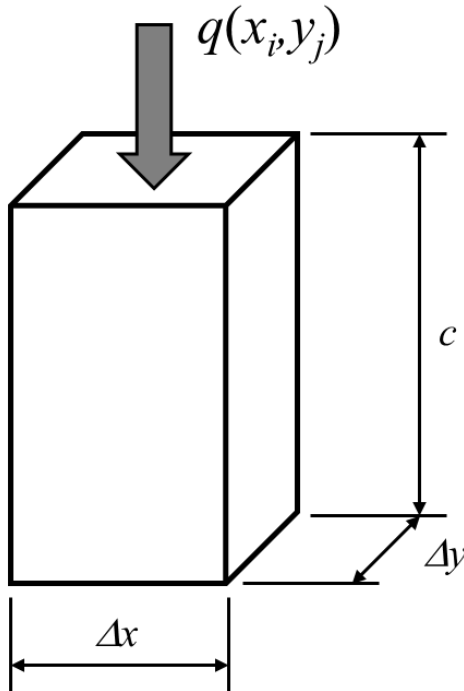
$$\tilde{\pi}(\gamma, \mathbf{P} | \mathbf{Y}) \propto \gamma^{(IJ+2)/2} \exp \left\{ -\frac{1}{2} [\mathbf{Y} - \mathbf{T}(\mathbf{P}) - \bar{\eta}]^T \tilde{W}^{-1} [\mathbf{Y} - \mathbf{T}(\mathbf{P}) - \bar{\eta}] - \frac{1}{2} \gamma (\mathbf{P} - \boldsymbol{\mu})^T \Gamma^{-1} (\mathbf{P} - \boldsymbol{\mu}) - \frac{1}{2} \left(\frac{\gamma}{\gamma_0} \right)^2 \right\}$$

Modified Likelihood

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

Gaussian prior

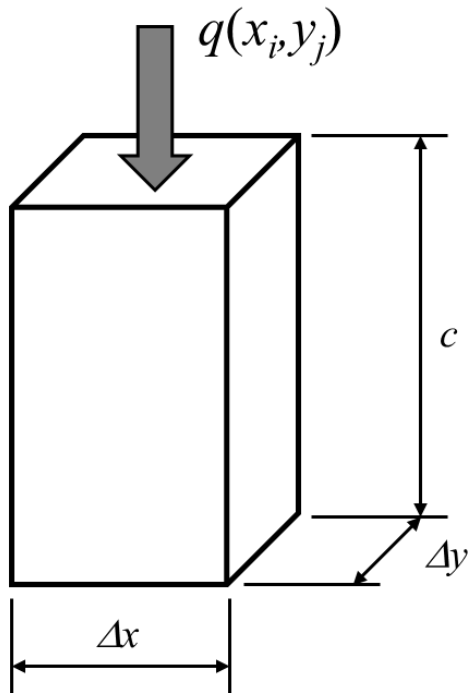
Energy Balance:
$$q(x_i, y_j) = C^* c \frac{dT(x_i, y_j)}{dt}$$



In order to generate this physically motivated Gaussian prior, and at the same time not violate the Bayesian principle that the prior is the information for the unknowns (coded in the form of probability distribution functions) that is available before the measurements are taken, we assume here that another kind of measurements is also available. Such other kind of measurements is only used to generate the prior, and is considered independent of the temperature measurements used in the inverse analysis, that is, for the computation of the likelihood.

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

APPROXIMATION ERROR MODEL



As for the prior, the local heat flux was calculated with the local *temperature* increase rate computed with the transient temperature measurements. From the means and variances of the local heat fluxes at each time step, 100 samples from a Gaussian distribution were generated for the spatially varying heat flux. The solutions of the complete and surrogate models were then computed, in order to calculate the modeling error, for each of these samples. Hence, the mean and the covariance matrix were computed, to be used in the enhanced error model.

$$\sigma = 0.02 \text{ K}$$



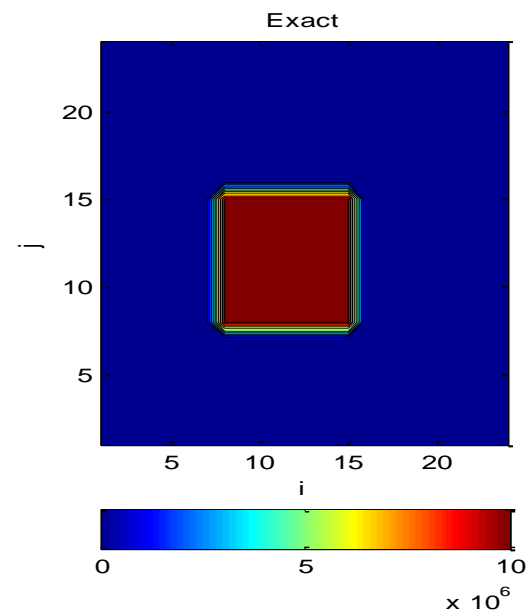
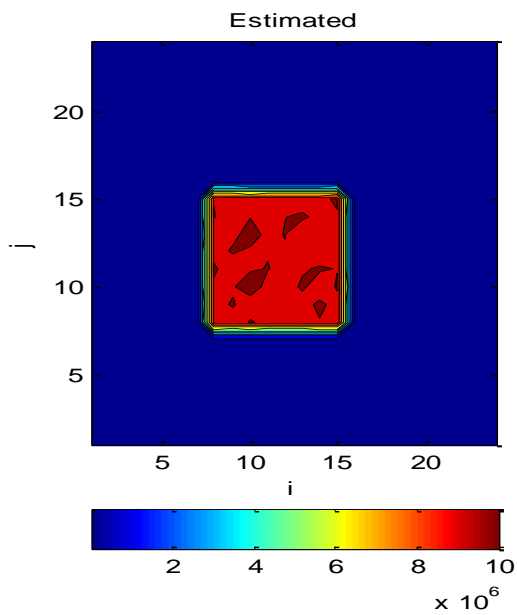
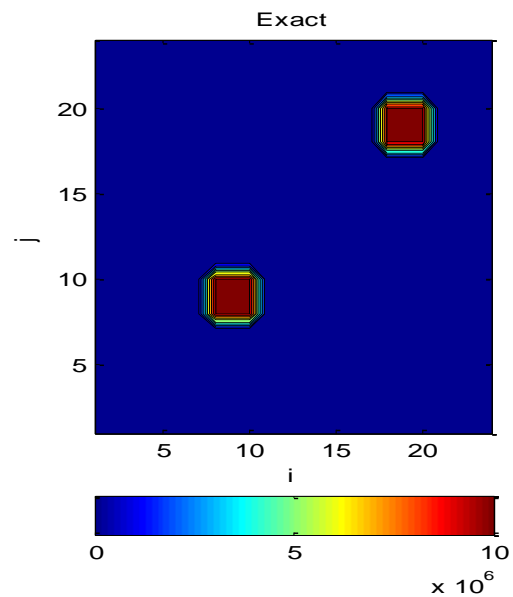
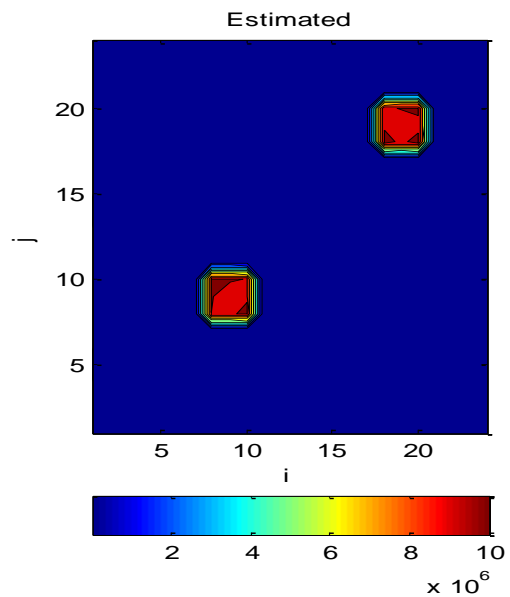
Test case	Flux	Prior	Approach
1	A	TV	-
2	B	TV	-
3	C	TV	-
4	A	TV	DAMH
5	B	TV	DAMH
6	C	TV	DAMH
7	A	Gaussian	-
8	B	Gaussian	-
9	C	Gaussian	-
10	A	Gaussian	AEM
11	B	Gaussian	AEM
12	C	Gaussian	AEM

$$\sigma = 1.25 \text{ K}$$

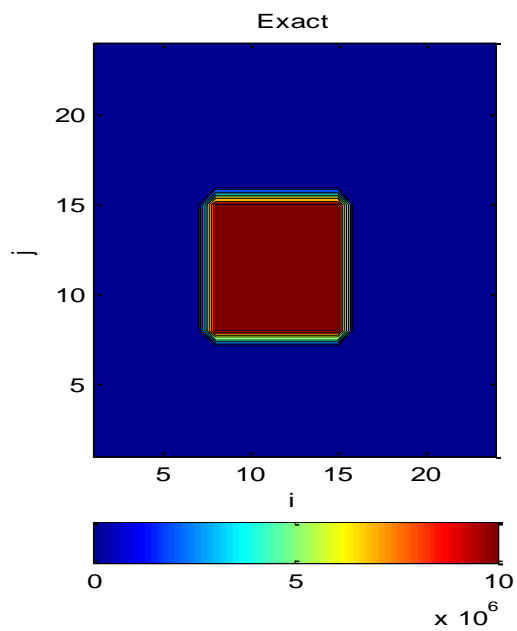
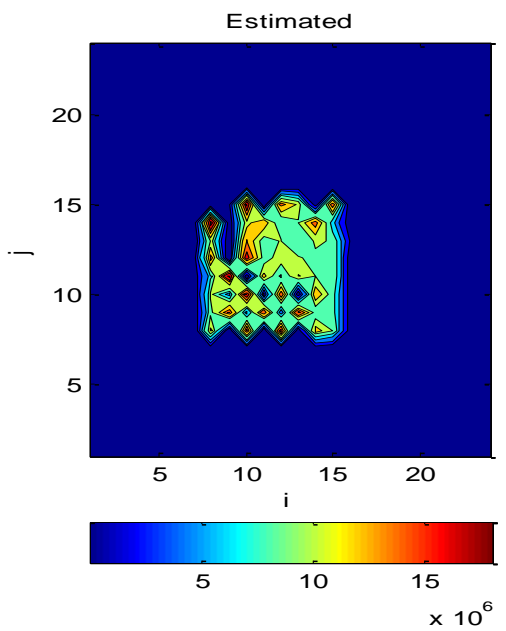
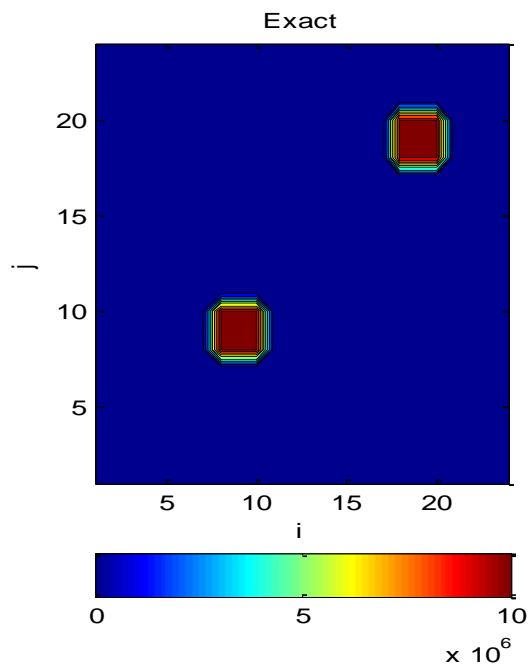
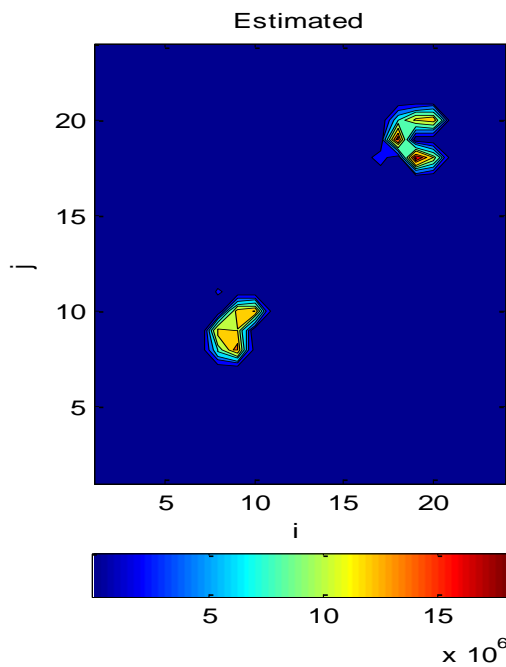


Test case	CPU Time (h)	Acceptance ratio (%)	RMS Error (W/m ²)
1	2.7	10.9	9.3x10 ⁴
2	2.8	9.0	6.6x10 ⁴
3	2.6	9.9	1.1x10 ⁵
4	114.2	46.7 – 5.3	9.8x10 ⁴
5	113.0	47.9 – 4.2	5.9x10 ⁴
6	98.3	40.8 – 5.9	1.4x10 ⁵
7	2.6	11.3	9.3x10 ⁴
8	2.8	9.3	6.6x10 ⁴
9	2.7	10.2	1.1x10 ⁵
10	44.5	12.8	4.1x10 ⁴
11	44.2	11.0	2.6x10 ⁴
12	42.5	11.2	8.5x10 ⁴
1	2.6	9.1	1.1x10 ⁶
2	2.6	7.5	1.0x10 ⁶
3	2.6	9.1	1.8x10 ⁶
4	98.7	41.9 – 6.8	1.1x10 ⁶
5	93.5	44.6 – 5.7	6.9x10 ⁵
6	64.5	34.6 – 5.4	1.4x10 ⁶
7	2.7	9.1	1.1x10 ⁶
8	2.6	8.3	9.8x10 ⁵
9	2.6	9.4	1.3x10 ⁶
10	42.9	9.8	1.2x10 ⁶
11	43.3	11.9	1.2x10 ⁶
12	43.1	8.7	2.0x10 ⁶

Gaussian prior + AEM - $\sigma = 0.02$ K

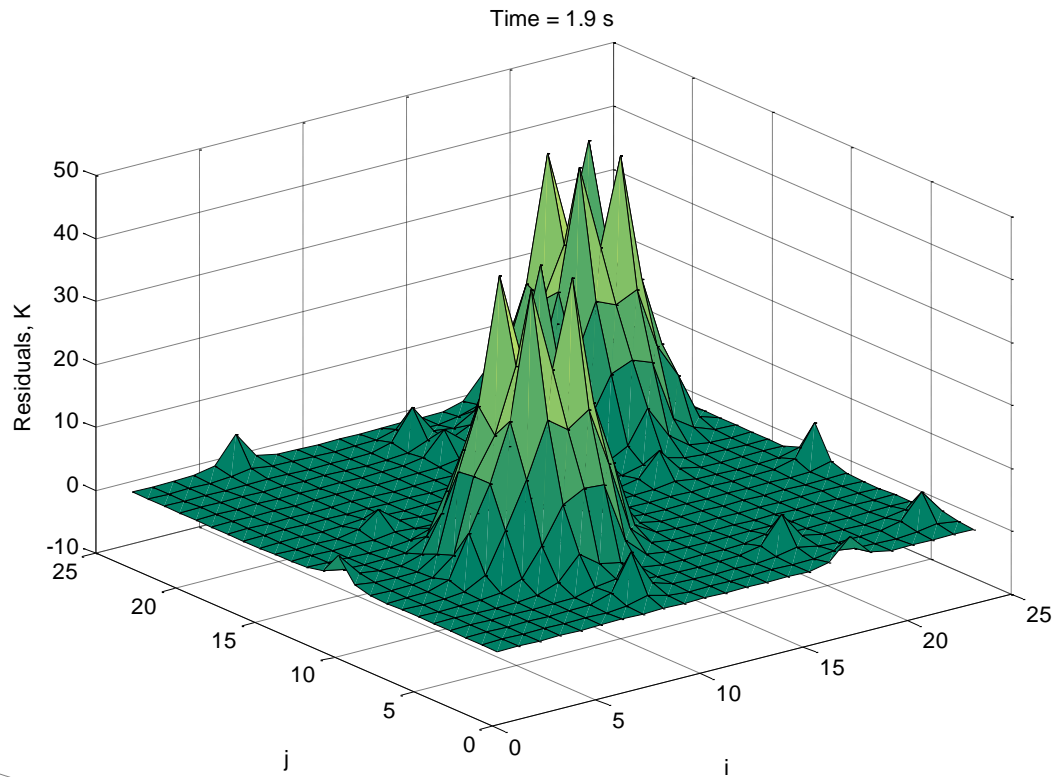


Gaussian prior + AEM - $\sigma = 1.25$ K

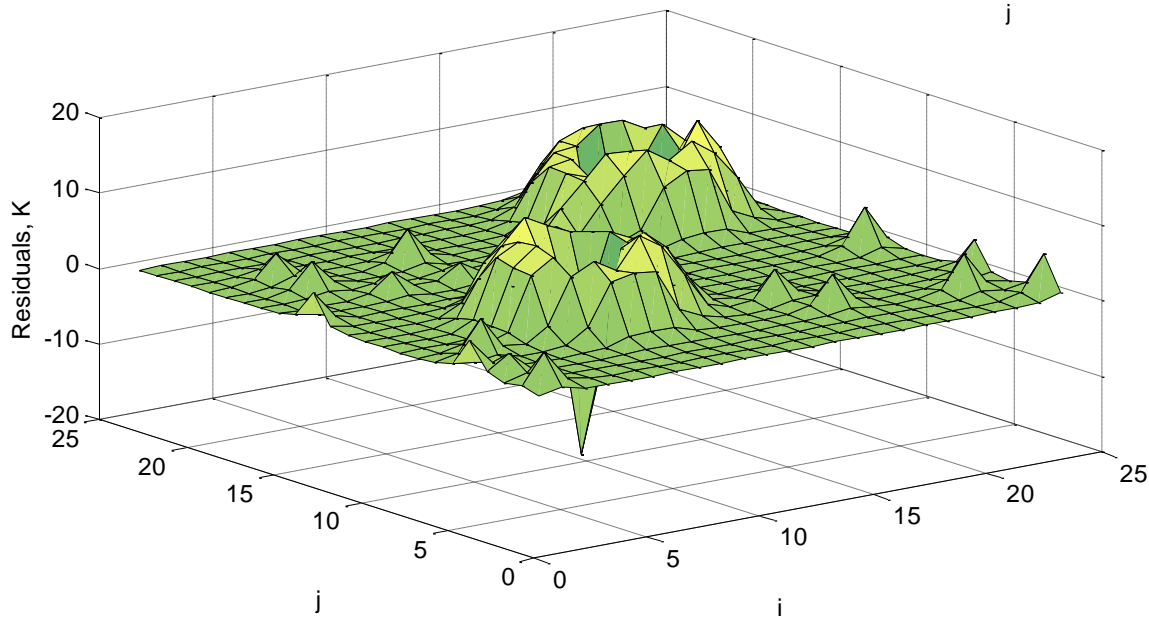


$\sigma = 1.25 \text{ K}$

Gaussian prior



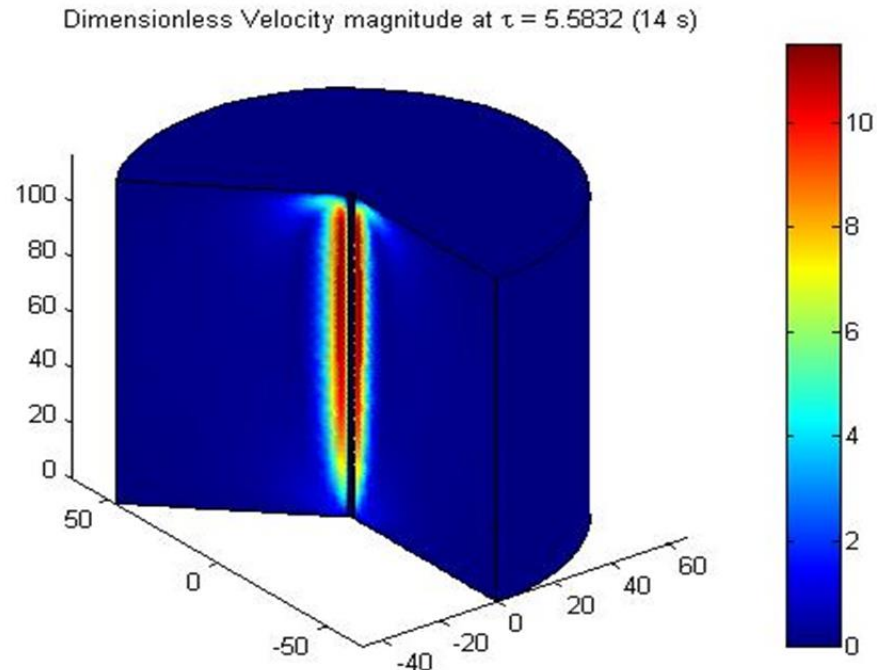
Time = 1.9 s



Gaussian prior + AEM

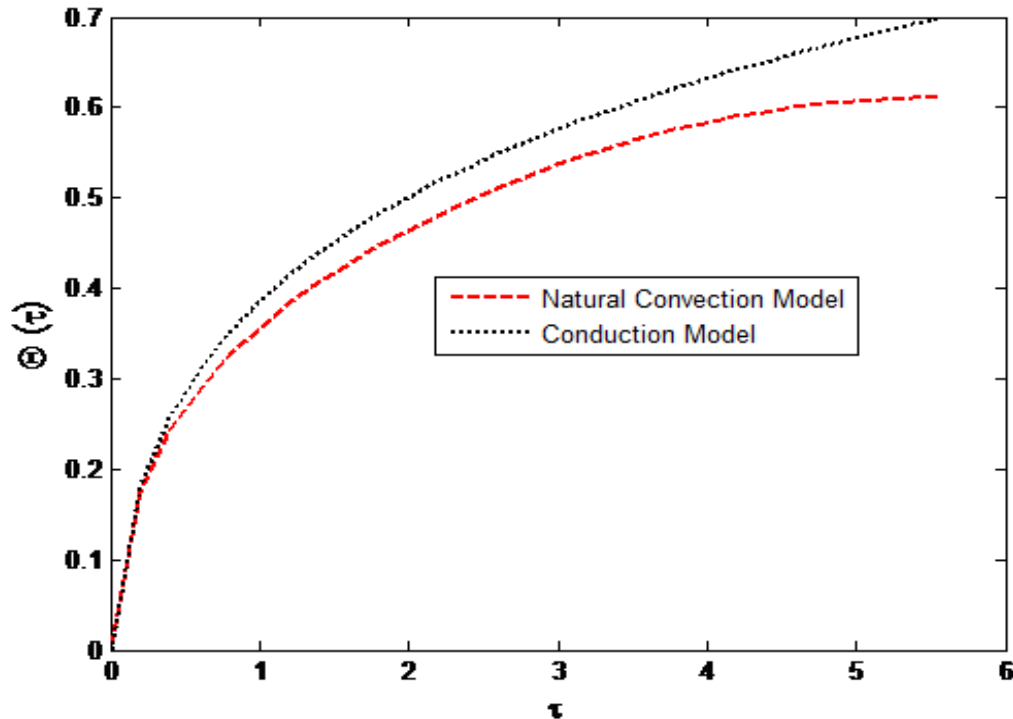
REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

CONVECTIVE EFFECTS IN LIQUIDS CHARACTERIZED BY THE LINE HEAT SOURCE PROBE – AEM MODEL



REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

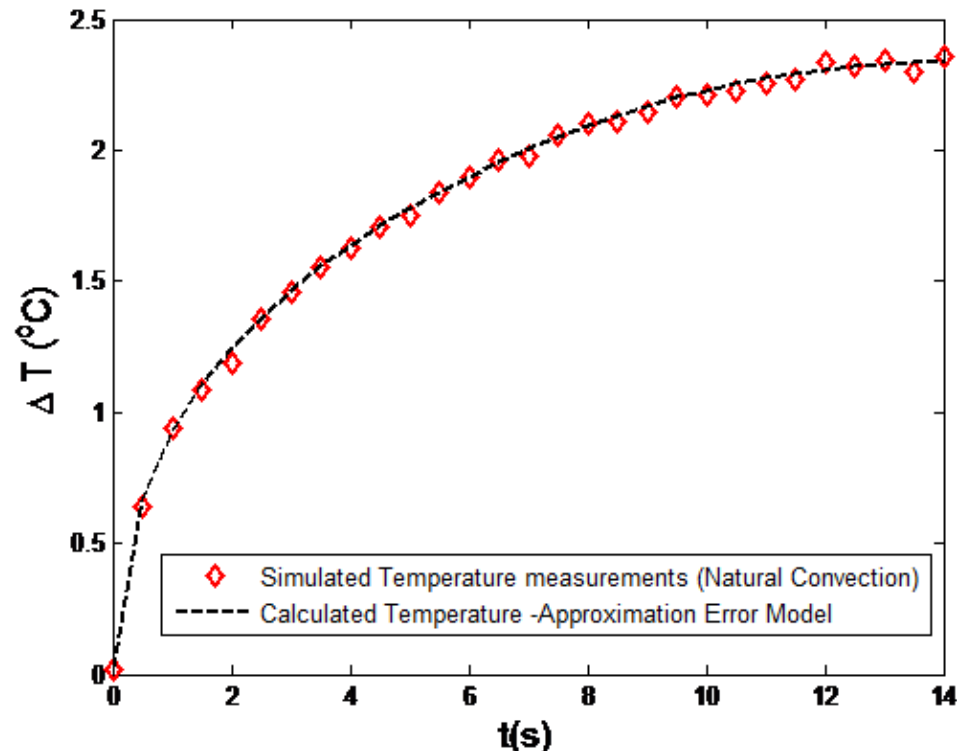
CONVECTIVE EFFECTS IN LIQUIDS CHARACTERIZED BY THE LINE HEAT SOURCE PROBE – AEM MODEL



Bernard Lamien, Helcio R. B. Orlande, Approximation Error Model To Account For Convective Effects In Liquids Characterized By The Line Heat Source Probe, *4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)*, Albi, France, June 26-28, 2013 99

REDUCTION OF THE COMPUTATIONAL TIME FOR MCMC METHODS

CONVECTIVE EFFECTS IN LIQUIDS CHARACTERIZED BY THE LINE HEAT SOURCE PROBE – AEM MODEL



Bernard Lamien, Helcio R. B. Orlande, Approximation Error Model To Account For Convective Effects In Liquids Characterized By The Line Heat Source Probe, *4th Inverse Problems, Design and Optimization Symposium (IPDO-2013)*, Albi, France, June 26-28, 2013 100