Doctoral Dissertation **Academic year 2015**

Bayesian inference on structures using differential evolution adaptive Metropolis algorithm considering measurement uncertainty

Keio University

Graduate School of Science and Technology

School of Science for Open and Environmental Systems

Student ID Number 81047834 Name Zhou, Jin

Thesis Advisor Professor Name Mita, Akira

Keio University
Graduate School of Science and Technology
September 2015

Abstract

Recent years witness the fast advancement of system identification which is widely applied in civil engineering such as health monitoring and non-destructive damage detection. Because of its wide applicability, many identification methods have been studied for various purposes and a wide range of analytical methods exist for linear and nonlinear structural system. Many common among these methods are based on an inverse problem using heuristic algorithms such as genetic algorithms (GAs), particle swarm optimization (PSO), and differential evolution (DE) algorithm, etc.. The structural identification is formulated as an inverse problem which is concerned with the derivation of mathematical models from experimental measured data. Given the measured response, a set of candidate models are built up and choosing the optimal one based on a predefined fitness by which the residual error, measuring the fitness between the measure output of the actual system and the response of simulated model, is minimized. However, these heuristic algorithms based identification studies are all treated as a deterministic issue, which inevitably obtain a biased solution if taking uncertainties such as measurement noise or model error into account.

Bayesian posterior density estimation is a classic method to quantify the uncertainty based on a probabilistic model that is defined by stochastic model classes. The model set is a class of parameterized probability models, each of which predicts the behavior of the actual system with a prior probability density. In Bayesian estimation, the identification problem is to infer the plausibility of each candidate model with a posterior density conditioned by the measured data; it is not a quest for the true structural parameters. The posterior density of structural parameters indicates how

plausible each model is when considering the uncertainty of predictive errors.

However, it is usually difficult for the Bayesian identification method to obtain the posterior probability density of parameters conditioned by the measured response, because its calculation often requires an evaluation of multidimensional integrals that cannot be easily calculated. The Markov chain Monte Carlo (MCMC) method is a widespread medium for Bayesian inference but its convergence is often slow. It is known that because of the noise corrupted system response, the surface of the residual error lies in a hyper-surface of a multi-dimensional parametric space, which will cause the convergence of the Markov chains difficult to be approached. Moreover, most of these existing MCMC based identification methods use a single Markov Chain, which may be inefficient and unreliable when the surface of posterior probability density function is complicated.

The purpose of this study is to surmount this difficulty that solving the convergence of the Markov chains when the Bayesian inference framework is applied in the structural system. In this thesis, the ability of heuristic algorithms to search for the global optimum will have to be merged with the advantage of the Metropolis-Hasting (MH) algorithm for inferring the posterior probability. We present an improved differential evolution adaptive Metropolis-Hasting algorithm (IDREAM) strategy to estimate the posterior density of structural parameters. The main benefit of IDREAM is its efficient MCMC simulation through its use of the adaptive Metropolis (AM) method with a mutation strategy for ensuring quick convergence and robust solutions. Its effectiveness was demonstrated in simulations on identifying the structural parameters with limited output data and noise polluted measurements.

Moreover, the estimator (maximum a posterior estimator, MAP) of the Bayesian inference is inevitably biased, which may be attributed to the surface of the posterior probability density owning to multiple local optima. Another purpose of this thesis is to improve the accuracy of the MAP estimator. To solve this problem, which is defined as the “equifinality” of Bayesian inference, a two-step Bayesian identification method is proposed. In step 1, the formal likelihood measure is used to obtain the MAP

estimator; in step 2, the first-two derivative of the log-likelihood measure is proposed to formulate a new fitness function to improve the accuracy of the estimator. The benefit of the proposal was demonstrated in simulations on identifying the structural parameters with limited output data and considering noise polluted measurements. Finally, the conclusion is given that the proposal could not only improve the accuracy of the MAP estimator but also reduce the standard deviation (uncertain range) of the posterior samples. The identification using the proposed method is applied into the measured data of a shake-table experiment, called the E-Defense. Comparison with the results that using the existing methods show that our proposed methodology is indeed a powerful tool for the Bayesian identification of building structures.

Acknowledgements

The completion of this thesis would not have been possible but for the support and guidance of my well-beloved supervisor, Professor Akira Mita, and because of this I would like to express my most sincere gratitude. Firstly, I'd like to thank him for giving me the opportunity to study in Mita Lab at Keio University. Secondly, I will be eternally grateful for his broad expertise and untold levels of patience, and most importantly for remaining positive and having faith in me when desperate difficulties arise on me when finishing this thesis. Thanks a million to you, my supervisor; whatever I have learned and obtained during my Ph.D.'s studies is a valuable treasure for me. Many thanks more than I can say.

Furthermore, I would like to express my gratitude to Professor Shuichi Adachi, Professor Masayuki Kohiyama and Professor Masaki Takahashi for reviewing my thesis and giving me many valuable comments.

Many thanks to the students at Mita Laboratory and friends at Keio University who gave me countless warm help and friendship.

Finally, the greatest of thanks must go to my beloved wife and parents, who without the emotional and not inconsiderable support, I would never have had the ambition to attempt this work, let alone stubbornness to see it through to the end.

Contents

List of symbols

$\alpha(\boldsymbol{\theta}_{(s)}, \boldsymbol{\theta}_{(s+1)})$	Metropolis acceptance index
$\alpha_i, (i = 1,2)$	Parameter that control the size of the hysteretic loop
$\beta_i, (i = 1,2)$	Parameter that control the shape of the hysteretic loop
$\boldsymbol{C}$	Damping matrix
C_R	Crossover constant
d	Dimension of the optimal vector
δ	Number of chosen pairs for the mutation strategy
$\boldsymbol{e}_d$	Small random vector drawn from a uniform distribution
$\boldsymbol{E}(\hat{\boldsymbol{x}}, t)$	Matrix of residual error
$\boldsymbol{f}(t)$	Vector of restoring force for the nonlinear structural system
$\mathcal{F}(\cdot)$	Objective function
γ	Scaling factor used for deciding the transition scale of posterior sample

$\boldsymbol{G}$	Gaussian vector of the likelihood measure
$\boldsymbol{H}$	Hessian matrix of the likelihood measure
$\boldsymbol{K}$	Stiffness matrix
L	Likelihood measure
$\boldsymbol{M}$	Mass matrix
$\boldsymbol{\mathcal{M}}$	Model set
n_{CR}	Number of samples chosen for the crossover strategy
$n_i \ \ (i = 1,2)$	Parameter that control the post-elastic branch of the hysteretic loop
N_d	Parametric dimension of the structural parameters
N_m	Number of measurements/outputs
N_s	Number of posterior samples
N_t	Number of observed points in time domain
$\boldsymbol{\Omega_\theta}$	Probability space
$p(\cdot)$	Probability density function
$\pi(\cdot)$	Target distribution for the Markov samples

$p(\boldsymbol{\theta}\|\boldsymbol{Y}^M(t))$	Posterior/updated PDF of the structural parameter
$r_1(j), r_2(n)$	Random integers that chosen form the integer set
$\widehat{\boldsymbol{R}}$	Matrix of Gelman-Rubin convergence criteria
$\hat{R}_j$	Gelman-Rubin convergence criteria for j-dimensional parameter
$\mathbb{R}$	Set of real numbers
$Ratio_{in}$	The including ratio reflects the quality of stochastic identification
s	Iteration/state of the posterior samples
σ_j	Variance of the prediction error for the j^{th} measurement
t	Time step of the structural response
$\boldsymbol{\theta}$	Matrix of random variables denoting the stochastic structural parameters
$\boldsymbol{\theta}_{\text{MAP}}$	The Maximum posterior estimator of the parameter
$\boldsymbol{\theta}_0$	The true value of the structural parameter
$\widehat{\boldsymbol{\theta}}_o$	The nominal true value of the structural parameter
$\boldsymbol{\theta}^*$	The optimal point of the proposed likelihood measure function
$\zeta^{(m)}$	Damping ratio of the m^{th} mode

$\boldsymbol{u}(t)$	Vector of ground acceleration for the structural system
$\boldsymbol{u}_d$	N_d-dimensional unit vector
$\widehat{\boldsymbol{x}}$	Parameter vector in deterministic identification method
$\boldsymbol{x_0}$	Actual parameter vector (deterministic issue)
$\widehat{\boldsymbol{x}}^*$	Optimal parameter vector (deterministic issue)
$\boldsymbol{y}(t)$	Vector of relative displacement response
$\dot{\boldsymbol{y}}(t)$	Vector of relative velocity response
$\ddot{\boldsymbol{y}}(t)$	Vector of relative acceleration response
$\boldsymbol{Y}^M(t)$	The measured response of the structural system
$\widehat{\boldsymbol{Y}}(\widehat{\boldsymbol{x}}, t)$	The output of candidate model (deterministic issue)
$\widehat{\boldsymbol{Y}}(\boldsymbol{\theta}, t)$	Stochastic process/response in Bayesian inference method
$\widehat{\boldsymbol{Y}}^{\mathrm{low}}(t)$	lower bound of the 95% probability insurance uncertainty range
$\widehat{\boldsymbol{Y}}^{\mathrm{up}}(t)$	upper bound of the 95% probability insurance uncertainty range

Figures

Tables

CHAPTER 1
Introduction

1.1 Structural health monitoring (SHM) and system identification

The art of structural health monitoring (SHM) can be regarded as the routine checkups of human bodies by the doctors, who gathered the signals that are relevant to diseases to find out the status of body organs or human systems. The findings are translated by the doctors to provide diagnosis and prognosis to the patients. SHM that applied in civil structures can thus be understood as the process of carrying out a health status (damage detection strategy) for buildings, bridges or other civil infrastructures. It thus involves the observation of a structural system over time using dynamic response from an array of sensors, the extraction of damage sensitive features from these measured response, and the statistical analysis of these features to determine the current state of the structural health (Sohn and Farrar, 2001; Sohn et al., 2004). In short, a complete SHM framework includes three parts: Monitoring, Diagnosing and State evaluation, as is shown in Figure 1.1. Seen from the Fig. 1.1, it can be found that system identification plays an important role in the SHM framework, which is the process of identifying the

parameters of a structural model using the recorded seismic output of the actual building system (Alvin et al., 2003 and Ljung, 1999). Because of its extensive application, currently, a wide range of analytical methods exist for linear or nonlinear system (Peeters et al., 2001; Kessler et al., 2002; Kim et al., 2003; Qiao et al., 2007; Ye and Wang, 2007).

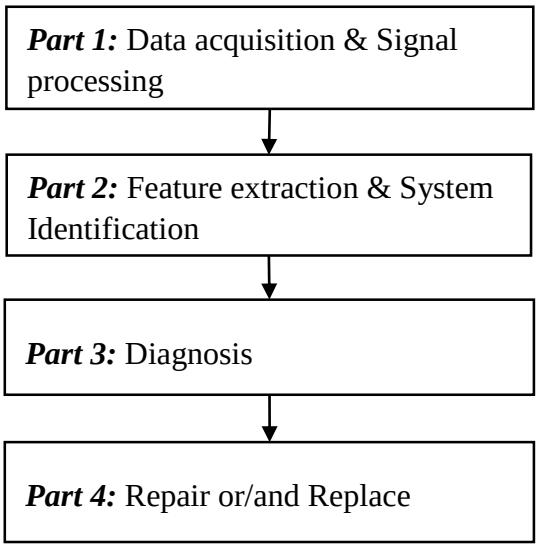


Figure 1.1 SHM framework

The conventional methodologies can be divided into two categories: parametric and nonparametric identification approaches. The parametric identification methods, such as least square method (Yang et al., 2007), the extended Kalman filter (Yang et al., 2006), and the H∞ filter method (Sato and Qi, 1998) etc., achieved some progress but hit the bottleneck because most of these methods require an initial guess so that the process can start. Xie and Mita (2010) used subspace method and component mode synthesis into the restoring force identification of a base-isolated structural system considering nonlinearity. Besides the traditional parametric estimation methods,

numerous nonparametric approaches in literature have been used in applications including wavelet analysis (Hung et al., 2003), neural network (Qian and Mita, 2008), and least squares support vector machine (Tang et al., 2006; Mita, 2003) technique etc. However, these estimated 'nonparametric' coefficients lack, in general, physical meaning, which makes these methods be very difficult to extract the physical characteristics like stiffness or damping unless some of these are assumed known a priori.

1.2 The optimization principle of identification

This study focuses on parametric identification, where a model within a chosen model class is defined by assigning values to a set of model parameters. Identification, is performed by searching the "best" set of structural parameters from the given model class according to a prescribed criterion such as least-square errors, which measures the fitness between the measured response and the output of simulation model. If the resulting model can be deemed to be accurate enough for its intended application, predictions are made using this so-called "best" (optimal) model; otherwise, the process of model updating is repeated.

To detail the optimization principle of identification problem, let $\boldsymbol{Y}^M(t)$ denote the measured response at each time interval ($t = 1, \dots, N_t$) and $\widehat{\boldsymbol{Y}}(\hat{\boldsymbol{x}}, t)$ denotes the output of candidate models, where $\hat{\boldsymbol{x}} = (\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{N_d})^{\mathrm{T}} \in \mathbb{R}^{N_d}$ denotes an N_d dimensional parameterized system to capture the behavior of the physical system. The difference between the measured response and model outputs is defined as the residual error: $\boldsymbol{E}_j(\hat{\boldsymbol{x}}, t) = \boldsymbol{Y}_j^M(t) - \widehat{\boldsymbol{Y}}_j(\hat{\boldsymbol{x}}, t)$, where j=1,…, N_m, and N_m is the number of outputs. Therefore, the identification problem is formulated as minimizing the residual

error norm of the outputs, e.g., the following mean least square error function:

$$\mathcal{F}(\hat{\boldsymbol{x}}) = \frac{1}{N_m N_t} \sum_{j=1}^{N_m} \sum_{t=1}^{N_t} \left\| \boldsymbol{Y}_j^M(t) - \hat{\boldsymbol{Y}}_j(\hat{\boldsymbol{x}}, t) \right\|, \tag{1.1}$$

where $\|\cdot\|$ is a Euclidean norm of vectors. Formally, the identification problem requires finding a set of N_d parameters $\hat{\boldsymbol{x}}^* \in \mathbb{R}^{N_d}$, so that the error norm $\mathcal{F}(\cdot)$ is minimized (Different from the deterministic identification method, where the structural parameters are denoted by a N_d dimensional deterministic vector $\hat{\boldsymbol{x}}$, a random matrix of N_d by N_s, $\boldsymbol{\theta}$, is used for representing the uncertain structural parameters in model set for Bayesian inference method, where N_s is the number of samples in random model set).

Obviously, from (1.1) we can find that a successful implementation of system identification requires to overcome several difficulties, such as dealing with ill-conditioned inverse problems with multiple optimal (local optimal) solutions. In general, the parameters that defining the physical system are identified through noise corrupted measured data. The predictive error surface of this kind of problem inevitably exists as a hyper-surface in the parameter space. The hyper-surface potential is often highly multi-modal in the sense that it is a non-quadratic surface that possesses local minima, in addition to a global minimum that represents the optimal solution. Due to the ill-conditioning and multi-modality, traditional local methods, like Levenberg-Marquardt or Gauss-Newton, may fail to identify the global solution and may converge into a local minimum. Because these local methods are performed by point-to-point search strategy. Such optimization methods work satisfactorily when the error surface contains no local minima. A good initial guess of the parameter and gradient or higher-order derivatives of the objective function is generally required. There is always a possibility to fall into a local minimum rather than the global minimum. In order to

surmount these difficulties, there is a distinct requirement for using heuristic global optimization methods, which do not converge to local minima and thus expected to provide more accurate estimates of the system parameters.

1.3 Heuristic algorithm based identification problem

Heuristic computational intelligence methods belonging to the global optimization category have proven to be promising tools to solve many multi-modal optimization problems. These have been found to be powerful methods in domains where local methods have not been proved to be effective. Among the most important global heuristic optimization methods, such as genetic algorithms (GAs) (Eiben et al., 1994), evolution strategy (ES) (Koziel and Michalewicz, 1999), particle swarm optimization (PSO) (Kennedy and Eberhart, 1995), differential evolution (DE) (Storn and Price, 1997) and big bang-big crunch (BB-BC) algorithm (Jaradat and Ayob, 2010) have been successfully utilized to solve this multi-modal optimization problem.

1.3.1 Current studies of structural identification using heuristic algorithms

The heuristic algorithms are highly adaptive methods originated from the laws of nature or biology. Unlike the traditional local methods, one of the important characteristics of these algorithms is their effectiveness and robustness in coping with ill-conditioned inverse identification problem with insufficient information, and noise. The advantages of using heuristic algorithms for structural parameter estimation have been increasingly recognized in recent years. Koh and Perry (2007) applied GAs to solve the global

problem of systems identification in shear-type building structures. Cunha et al. (1999) used GAs to identify the elastic constants of composite materials. Barbieri et al. (2010) applied GAs to identify the physical parameters of sandwich beams. Chou and Ghaboussi (2001) introduced GAs to identify damage severity of trusses. Wang (2009) used a hybrid GA to identify structural systems. Perry et al. (2006) utilized a modified GAs to identify structural systems. Trinh and Koh (2012) presented a GA-based substructural identification strategy for large structural systems. Franco et al. (2004) used ES to identify multiple degree of freedom (DOF) systems. Jeong and Lee (1996) proposed an adaptive simulated annealing genetic algorithm for system identification. Levin and Lieven (1998) applied SA method to optimize a finite element model for describing the dynamic behavior of structures. Xue et al. (2009) introduced PSO to identify the parameters of linear and nonlinear structural system. Tang et al. (2008, 2010) proposed an improved DE and BB-BC optimization strategy to structural parameter estimation. Li et al. (2013) combined the symbolization method with the DE strategy to improve the accuracy of the identification results.

Despite the success of these heuristic algorithm based identification methods on paper, there are several problems, the reason for which is that they are all on the basis of a deterministic mechanical model to quixotically quest for the output of true system with uncertain prediction errors. As is questioned by Beck (2010) on the deterministic identification methods: No candidate model is expected to exactly represent I/O behavior of actual system, which raises questions about the basis for choosing only one and biased model. If we make a unique estimate for the actual system it is sure to produce biased predictions. Hence, it is necessary to use stochastic model to quantify the uncertainty of prediction errors. An alternative approach, which has gained a substantial following in many research disciplines in recent years, is to evaluate candidate models using the probability of the models given data from the system of interest. The roots of the probability logic approach are in work performed by Bayes

(1763). He presented a method for updating probability distributions for parameters based on available data that would come to be known as Bayesian Theorem, and it forms the foundation of a framework for probabilistic inference. It was Laplace (1951), however, who showed the power of Bayesian Theorem by applying it to problems using real data and demonstrating the ability of probabilistic inference to separate "signal" from "noise." The probability logic approach is applied to system-identification problems using the probabilistic framework developed by Beck (1989, 1996) and Beck and Katafygiotis (1991, 1998).

1.4 Bayesian probabilistic inference for structural identification

Bayesian probabilistic inference (Lee, 1997; Box and Cox, 1964; Box and Tiao, 1973; Blasco et al. 1998) provides a rigorous framework for quantifying the uncertainty based on probability models defined by stochastic model classes, $\boldsymbol{\mathcal{M}}$. The model class is composed as a set of parameterized probability models for predicting the behavior of actual system with prior probability model indicating the initial relative plausibility of each model. In this framework, the identification problem is viewed as inferring the plausibility of system models with posterior probability distribution conditioned by the measurement data, but not a quest for true parameter values.

For the Bayesian probabilistic inference identification method it is usually difficult to obtain the posterior probability density function (PDF) of the structural parameters, $p(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$, conditioned by the measured response, $\boldsymbol{Y}^M(t)$, where $\boldsymbol{\theta}$ is a stochastic parameter vector defining each possible model within the model set ($\boldsymbol{\Omega_\theta}$ denotes the probability space for the random parameters). The posterior PDF describes how plausible each model is if one accounts for the uncertainty of the prediction errors. The

posterior density, $p(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$, is needed to make robust predictions of the performance of the system based on past observation, as was illustrated by Papadimitriou et al. (2001), Beck and Au (2002) and (Simoen et al., 2013).

1.4.1 Current studies of Bayesian inference for structural identification

Many studies have focused on obtaining the posterior PDF because its calculation often requires an evaluation of multidimensional integrals that cannot be easily obtained. In particular, Laplace's method of asymptotic approximation was utilized by Beck and Au (2002) to obtain a posterior PDF with a small-dimensional parameter space. To solve higher dimensional problems, an adaptive Markov chain Monte Carlo (MCMC) simulation method, Metropolis-Hasting (MH) algorithm, was developed to be used in the Bayesian model updating (Muto and Beck 2008). Since the advent of the MH algorithm, MCMC methods have become the primary means to obtain the posterior PDF in structural identification. Gibbs sampling and transitional Markov Chain Monte Carlo (TMCMC) were used by Ching and Chen (2007). Cheung and Beck (2009) used a hybrid Monte Carlo method, known as the Hamiltonian Markov chain method, to solve the higher dimensional Bayesian model updating problems. Vanik et al. (2000) and Yuen and Beck (2003) successfully applied the probabilistic logic framework into the model updating for reliability and structural health monitoring fields. The Bayesian inference framework has also been applied to more challenging problems such as non-linear systems with uncertain input (Beck and Yuen, 2004) and reliability-based control robust to probabilistic model uncertainty (Scruggs et al., 2006).

However, all of these MCMC-based identification methods use a single Markov Chain, which may be inefficient and unreliable when the posterior surface is complicated. It is known that because of the noise corrupted system response, the surface of the prediction

error lies in a hyper-surface of a multi-dimensional parametric space. The complicated surface of the prediction error will definitely cause the surface of the posterior model samples to have multiple regions of attraction and numerous local optima. In short, the biggest challenge of Bayesian inference in structural system, therefore, lies in the efficient-convergence of the Markov chain.

1.5 The generalized likelihood uncertainty estimation

For traditional Bayesian structural identification framework, it is inevitably facing the problem of "equifinality" (Beven, 2006), which originates from the imperfect knowledge of the system under identification, and many sets of models, parameters and variables may therefore be considered equal or almost equal simulators of the unknown system. The generalized likelihood uncertainty estimation (GLUE) methodology (Beven and Freer, 2001), which adopts the concept of equifinality of models, parameters and variables is thus popularly studied recent decades. In GLUE, it deals with the variable degree of membership of the sets. And the degree of membership is determined by assessing the extent to which solutions fit the model, which in turn is determined by subjective likelihood functions. Therefore, the application of GLUE requires the proper definition of a likelihood measure.

The term "likelihood" in the framework of GLUE is used less formally than the likelihood measures of classical statistics, which make specific assumptions about the nature of the error associated with the model simulations. The traditional GLUE likelihood measure is model efficiency function which is given by Beven and Binley (1992). In this definition, the likelihood equals one if all residuals are zero, and zero if the weighted variance of the residuals is larger than the weighted variance of the

observations. Beven (1996) and Smith et al. (2008) listed a few example likelihood measures and discussed the limitations of using the maximum likelihood model. Ratto et al. (2007) considered the subjectivity in the GLUE method, which allows the modeler to interact in the modeling process. Zhang et al. (2011) proposed a modified GLUE framework to address the problem of the subjectivity of likelihood measure selection.

1.6 Review of the previous identification studies at Mita Lab

The previous studies concerning with the structural identification problem at Mita Lab are given by Qian (2008) and Li (2013). Qian explored the PSO algorithm to the acceleration-based damage localization and quantification of identification problem. She compared the PSO-based identification results with the usage of the Simulated Annealing (SA) and Genetic Algorithm (GA). The conclusion that the PSO-based identification method outperforms the SA and GA based estimation approach was drawn in her thesis. Differential evolution (DE) algorithm was utilized by Li to the identification problem combing with the advantages of the symbolic technology. The DE based identification results using the symbolic time series and the raw acceleration are compared in his study, where he got the conclusion that the application of symbolized acceleration into the DE based identification could improve the accuracy of the results obtained by the raw acceleration. Both of the two previous identification studies are formulated as a deterministic optimization problem. The limitations of formulating the parameter estimation as a deterministic issue have already discussed by Beck (2010). This study is to solve the identification problem considering the measurement uncertainties under the Bayesian logic framework.

1.7 The purposes of this study

The purpose of this thesis is to solve the two aspects of the difficulties in Bayesian inference. In first aspect, it would like to solve the problem of slow-convergence of Markov chain of the posterior samples. To surmount this difficulty, the ability of heuristic algorithms to search for the global optimum will have to be merged with the advantage of the MH algorithm for inferring the posterior PDF. The convergence speed of the Markov chains is enhanced by an improved differential evolution adaptive Metropolis-Hasting algorithm (IDREAM) strategy in estimating the posterior density of structural parameters is proposed in Section 3. The results of the numerical simulation and the verification of the proposed method that applied in a shake table experiment shows its potentials in Bayesian identification.

The second aspect of this study is to solve the accuracy problem of the estimator that obtained by Bayesian inference. The definition of the likelihood measure plays an important role in the success of Bayesian identification, because the surface of the likelihood measure often has multiple regions of local optima due to the noise corrupted measurement of the model errors. To solve the “equifinality” problem, in Section 4, the first-two derivative of the log-likelihood measure is used to formulate a new objective function for the sake of improving the accuracy of the estimator.

1.8 Organization of the Thesis

This thesis is divided into five chapters as below.

Chapter 1 gives an introduction system identification based on optimization algorithms

and the background of Bayesian posterior density estimation for structural system.

Chapter 2 presents problem statement of Bayesian inference for structural system, the Metropolis-Hasting algorithm and the Differential Evolution strategy.

Chapter 3 proposes a posterior density estimation method for structural parameters using improved differential evolution adaptive Metropolis-Hasting algorithm. Numerical simulation of the proposed approach in a linear and nonlinear structural system demonstrates its effectiveness.

Chapter 4 proposes a method of reducing "equifinality" of the structural identification. An improved Bayesian structural identification method using the first-two derivative of the log-likelihood measure is presented. The experimental verification of the proposed identification method in the identification of E-Defense structural system shows the advantage of the proposed method.

Chapter 5 summarizes contributions of this thesis, and points out the direction for future works.

CHAPTER 2

Differential Evolution Adaptive Metropolis-Hasting Algorithm

2.1 Introduction

In this section, differential evolution adaptive Metropolis-Hasting algorithm, proposed by Vrugt et al. (2009), is explained along with its components. The difficulty of the Bayesian posterior density estimation lies in a high dimensional integral, which is called Markov integration. The Markov integration, the Markov chain and its property is introduced. Then a classic Markov chain Monte Carlo (MCMC) method, called Metropolis-Hasting (MH) algorithm is presented. However, the convergence of this traditional MCMC algorithm is often observed to be slow, which is frequently caused by an inappropriate selection of the proposal distribution used to generate the trails of the Markov chain. This difficulty especially happens when the target distribution is typically a high dimensional posterior distribution in the Bayesian analysis. The combination of heuristic algorithms (DE algorithm) and MH algorithm solves an

important problem in MCMC in real parameter spaces, which is of choosing an appropriate scale and orientation for the jumping distribution.

2.2 Bayesian probabilistic identification framework

Bayesian probabilistic inference (Box and Tiao, 1973) provides a rigorous way of quantifying this uncertainty based on a probabilistic model that is defined by stochastic model classes. The model set, $\boldsymbol{\mathcal{M}}$, is a class of parameterized probability models, each of which predicts the behavior of the actual system with a prior probability density. The prior probability of each model indicates the initial plausibility of the individual model. In Bayesian parametric posterior density estimation, the identification problem is to infer the plausibility of each candidate model with a posterior density conditioned by the measured data; it is not a quest for the true structural parameters.

For the Bayesian probabilistic inference identification method, it is usually difficult to obtain the posterior probability density function (PDF) of the structural parameters, $p(\boldsymbol{\theta}|\boldsymbol{Y}^M(t), \boldsymbol{\mathcal{M}})$, conditioned by the measured response, $\boldsymbol{Y}^M(t)$, where $\boldsymbol{\theta}$ is a stochastic parameter vector defining each possible model within the model set in probability space $\boldsymbol{\Omega_\theta}$. Let $\boldsymbol{Y}^M(t)$ denote the measured response at each time step ($t = 1, \dots, N_t$). The stochastic model set, $\boldsymbol{\mathcal{M}}$, is defined by stochastic parameter matrix, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \cdots, \boldsymbol{\theta}_{N_s}) \in \boldsymbol{\Omega_\theta} \subset \mathbb{R}^{N_d \times N_s}$, where N_d is the parametric dimension, N_s is the number of posterior samples. The initial plausibility of each model parameterized by $\boldsymbol{\theta}_k$ ($k = 1,2, \dots, N_s$) is defined as a prior density function, $p(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})$. The updated plausibility of each I/O model considering the uncertainty of the measured response is defined as the posterior density, $p(\boldsymbol{\theta}|\boldsymbol{Y}^M(t), \boldsymbol{\mathcal{M}})$, which from Bayes' theorem gives:

$$p(\boldsymbol{\theta}|\boldsymbol{Y}^M(t),\boldsymbol{\mathcal{M}}) = p(\boldsymbol{Y}^M(t)|\boldsymbol{\theta},\boldsymbol{\mathcal{M}}) \cdot p(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}})/p(\boldsymbol{Y}^M(t)|\boldsymbol{\mathcal{M}}), \tag{2.1}$$

where $p(\boldsymbol{Y}^M(t)|\boldsymbol{\theta},\boldsymbol{\mathcal{M}})$ is obtained from a probabilistic model that accounts for the uncertainty of the prediction errors between the measured response and the output of the each probability model in the model set, $\widehat{\boldsymbol{Y}}(\boldsymbol{\theta},t)$. Let $\boldsymbol{E}_j(\boldsymbol{\theta},t)= \boldsymbol{Y}_j^M(t) - \widehat{\boldsymbol{Y}}_j(\boldsymbol{\theta},t)$ denote the residual error of j^{th} measured response at each time interval ($t = 1, \dots, N_t$; and j=1,…, N_m, N_m is the number of measured response). The predictive PDF for the model output (white noise is considered as the measurement error, it thus obeys the normal distribution) at each time interval is

$$p\left(\boldsymbol{Y}_j^M(t)\middle|\boldsymbol{\theta},\boldsymbol{\mathcal{M}}\right) = \frac{1}{\sqrt{2\pi}\sigma_j} e^{\left[-\frac{1}{2{\sigma_j}^2}\sum_{t=1}^{N_t}(\boldsymbol{Y}_j^M(t)-\widehat{\boldsymbol{Y}}_j(\boldsymbol{\theta},t))^2\right]}, \quad j=1,\dots,N_m. \tag{2.2}$$

Hence, the predictive PDF (which is the likelihood function) seen from the whole time history is

$$p(\boldsymbol{Y}^M(t)|\boldsymbol{\theta},\boldsymbol{\mathcal{M}}) = \frac{1}{(\prod_{j=1}^{Nm}\sqrt{2\pi}\sigma_j)^{N_t}} e^{\left[-\sum_{j=1}^{Nm}\frac{1}{2{\sigma_j}^2}\sum_{t=1}^{N_t}(\boldsymbol{Y}_j^M(t)-\widehat{\boldsymbol{Y}}_j(\boldsymbol{\theta},t))^2\right]}. \tag{2.3}$$

The vector of the prediction error variance, σ_j^2, is an independent parameter corresponding to each candidate model in the model class, $\boldsymbol{\mathcal{M}}$. The term, $p(\boldsymbol{Y}^M(t)|\boldsymbol{\mathcal{M}})$, is called the evidence of the model class, and it equals

$$p(\boldsymbol{Y}^M(t)|\boldsymbol{\mathcal{M}}) = \int_{\boldsymbol{\Omega_\theta}} p(\boldsymbol{Y}^M(t)|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}}) \cdot p(\boldsymbol{\theta}|\boldsymbol{\mathcal{M}}) d\boldsymbol{\theta}. \tag{2.4}$$

The difficulty in estimating the Bayesian posterior density is none other than approximating this multi-modal integral, which is called Markov integration.

2.3 Markov integration and Markov chain property

From Eq. (2.4), we can find that in the implementation of Bayesian inference for identification, the challenge lies in obtaining the posterior distribution which requires an evaluation of multidimensional integration. This integration that is computationally difficult to solve, which is often named as Monte Carlo Integration (Geyer, 1992; Evans and Swartz, 1995). It can be rewritten as:

$$I = \int_{\boldsymbol{\Omega_\theta}} f(\boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{2.5}$$

where $p(\boldsymbol{\theta})$ is a PDF of random variable $\boldsymbol{\theta}$; $f(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$ in our interest (for instance $f(\boldsymbol{\theta}) = \boldsymbol{\theta}$ for the mean and $f(\boldsymbol{\theta}) = (\boldsymbol{\theta} - E(\boldsymbol{\theta}))^2$ for the variance); and $\boldsymbol{\Omega_\theta}$ denotes the probability space of random variable $\boldsymbol{\theta}$. An easy-to-realizing method called Monte Carlo simulation can solve this integration, but usually the calculation efficiency is low. Recent decades MCMC methods have found widespread use to estimate this integration in a Bayesian framework. The Markov chain can be defined as a stochastic process, a consecutive set of random quantities that defined on probability space. It can be defined as a sequence of a dependent random variables, $\boldsymbol{\theta}_{(0)}$, $\boldsymbol{\theta}_{(1)}$,

$\boldsymbol{\theta}_{(2)}, \ldots, \boldsymbol{\theta}_{(s)}, \ldots$, such that the probability distribution of $\boldsymbol{\theta}_{(s)}$ given the past variables depends only on $\boldsymbol{\theta}_{(s-1)}$. This means the future states of Markov chain samples are independent of past states given the present state (considering a draw of $\boldsymbol{\theta}_{(s)}$ to be a position of current state s, the next draw $\boldsymbol{\theta}_{(s+1)}$ is dependent only on the current draw $\boldsymbol{\theta}_{(s)}$, and not on any past draws). This principle is called the Markov property of MC Chains:

$$p(\boldsymbol{\theta}_{(s+1)} | \boldsymbol{\theta}_{(0)}, \boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \ldots, \boldsymbol{\theta}_{(s)}) = p(\boldsymbol{\theta}_{(s+1)} | \boldsymbol{\theta}_{(s)}). \tag{2.6}$$

Thus the Markov chain is a bunch of draws of $\boldsymbol{\theta}$ that are each slightly dependent on the previous one. In short, it wanders around the parameter space, remembering only where it has been in the last period. The jumping rules, which determines the probability of moving to some other state based on the current state, are governed by a transition kernel. In classic MH algorithm a simple but popular-used transition method is random walk Metropolis (RWM).

2.4 Metropolis-Hasting algorithm

The MH algorithm is one of the best known of MCMC methods. It was developed by Metropolis and Ulam (1949) and subsequently generalized by Hasting (1970) and its impact on Bayesian statistics has been immense as detailed in many studies (Haario et al., 2001, 2006).The basis of the MH algorithm is a Markov chain that generates a

random walk through the parameter space and successively visits solutions with stable frequency stemming from a fixed probability distribution. The transition of Markov chain using RWM method in MH algorithm satisfies:

$$\boldsymbol{\theta}_{(s+1)} = \boldsymbol{\theta}_{(s)} + \boldsymbol{\epsilon}_{(s)}, \tag{2.7}$$

where $\boldsymbol{\epsilon}_{(s)}$ is independent of the Markov samples $\boldsymbol{\theta}_{(s)}$, which obeys normal distribution, $\mathcal{N}(0,1)$. From Eq.(2.7), the transition kernel of the MC chain corresponds to a $\mathcal{N}(\boldsymbol{\theta}_{(s)}, 1)$ density. Assuming that a RWM has already sampled points $\{\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(s)}\}$, the MH algorithm proceeds in the following three steps. In the first step, a candidate point $\boldsymbol{\theta}_{(s+1)}$ is sampled from a proposal distribution $q(\cdot\,|\cdot)$ that is symmetric $q(\boldsymbol{\theta}_{(s)}|\boldsymbol{\theta}_{(s+1)}) = q(\boldsymbol{\theta}_{(s+1)}|\boldsymbol{\theta}_{(s)})$ and depends on the current location, $\boldsymbol{\theta}_{(s)}$. In the second step, the candidate sample is either accepted or rejected using the Metropolis acceptance probability:

$$\alpha(\boldsymbol{\theta}_{(s)}, \boldsymbol{\theta}_{(s+1)}) = \begin{cases} \min\left[\frac{p(\boldsymbol{\theta}_{(s+1)})}{p(\boldsymbol{\theta}_{(s)})}, 1\right] & \text{if } p(\boldsymbol{\theta}_{(s)}) > 0 \\ 1 & \text{if } p(\boldsymbol{\theta}_{(s)}) = 0 \end{cases}. \tag{2.8}$$

Finally, if the proposal is accepted, the chain moves to the candidate sample, $\boldsymbol{\theta}_{(s+1)}$, otherwise the chain remains at its current location $\boldsymbol{\theta}_{(s)}$. From the Eq.(2.7), it is clear that the transition of the Markov chain highly depends on the current location of sample $\boldsymbol{\theta}_{(s)}$ and the random variable $\boldsymbol{\epsilon}_{(s)}$. The scale of the transition is depend on the variance of the kernel distribution, which is a constant value, 1. The adaptive Metropolis (AM) scheme is proposed, in which a single Markov chain continuously adapt the variance of the transition distribution, $\boldsymbol{V}_{(s)}$, which is a N_d dimensional vector. The adaptive

variance (denoted as Cov signifying the variance of samples of different states) is decided by the samples of $\boldsymbol{V}_{(s)} = s_{N_d}\mathrm{Cov}(\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(s)}) + s_{N_d}\varepsilon \boldsymbol{I}_{N_d}$, where s_{N_d} represents a scaling factor that depends only on the dimensionality of the problem, $\boldsymbol{I}_{N_d}$ is a unit vector of N_d dimensions, and ε is a small constant. The $\mathrm{Cov}(\boldsymbol{\theta}_{(1)}, \dots, \boldsymbol{\theta}_{(s)})$ is the covariance of the samples from the initial state to the current s^{th} state. Although the AM scheme has some positive effect on many relatively simple Bayesian inference problems, but its inefficiency and slow-speed of convergence when confronted with posterior distribution with heavy tails or with complex posterior surfaces that contains multiple regions of attraction and local optima. Since the traditional MCMC method, such as MH (Kuczera and Parent, 1998; Gelman et al. 1997), AM (Haario et al., 2006) and Gribbs sampler (Casella and George, 1992) algorithm, yields only one single MC chain, it unavoidably trap into the local optima and traverse badly in pursuit of sampling the target distribution.

Recognizing the limitations of the existing MCMC algorithm, it is inevitably thinking of exploring the evolution mechanism of the heuristic algorithm, for instance the mutation and crossover strategy of DE algorithm, into the MH algorithm. For sake of integrity, the mutation strategy and the crossover operation of DE algorithm is briefly described below.

2.5 Differential Evolution algorithm

DE algorithm (Price 1999) is a population based evolution algorithm designed for optimization in real parameter spaces. Each population of the DE algorithm in different state (iteration) evolves by the mutation strategy and crossover operation, which can be regarded as a single MC chain. This is to say that there are N_s populations (MC chains) updating in parallel, which can fasten the convergence of the classic MH algorithm. For

short, the mutation and the crossover strategy of the DE algorithm can be described as follows:

2.5.1 The Mutation Strategy of DE

The objective of mutation strategy is to enable search diversity in the parameter space as well as to direct the existing object vectors with suitable amount of parameter variation in a way which will lead to better results at a suitable time. According to the mutation strategy, for each individual, $\boldsymbol{\theta}_{(s)}^{k}$, $k = 1,2,\dots,N_s$ (N_s denotes the number of populations), at the state (generation/iteration) s, a mutation vector $\boldsymbol{\theta}_{(s+1)}^{k}$ is determined:

$$\boldsymbol{\theta}_{(s+1)}^{k} = \boldsymbol{\theta}_{(s)}^{r1} + \gamma(\boldsymbol{\theta}_{(s)}^{r2} - \boldsymbol{\theta}_{(s)}^{r3}), \tag{2.9}$$

where $\boldsymbol{\theta}_{(s+1)}^{k}$ represents the k^{th} individual sample at next state s+1 that evolved from the population of the state s; γ denotes the real parameter, called mutation constants, which control the amplification of difference between two individuals so as to avoid search stagnation; and r_1, r_2, r_3, are mutually different integers, randomly selected from the set $\{k = 1,2,\dots,k-1,k+1,\dots,N_s\}$. The DE mutation strategy of an objective function (two-dimensional example) and its contour lines are shown in Figure 2.1.

In order to combing the DE mutation strategy into a Markov chain for drawing samples from a target distribution, the proposal and acceptance scheme must be such that there is detailed balance with respect to the target distribution. To ensure this, the first mutation strategy called "DE/1/bin" in Price and Storn (1997) is utilized to fit the

updating of Markov chains using DE mutation strategy as:

$$\boldsymbol{\theta}_{(s+1)}^{k} = \boldsymbol{\theta}_{(s)}^{k} + \gamma\left(\boldsymbol{\theta}_{(s)}^{r1} - \boldsymbol{\theta}_{(s)}^{r2}\right) + \boldsymbol{e}, \tag{2.10}$$

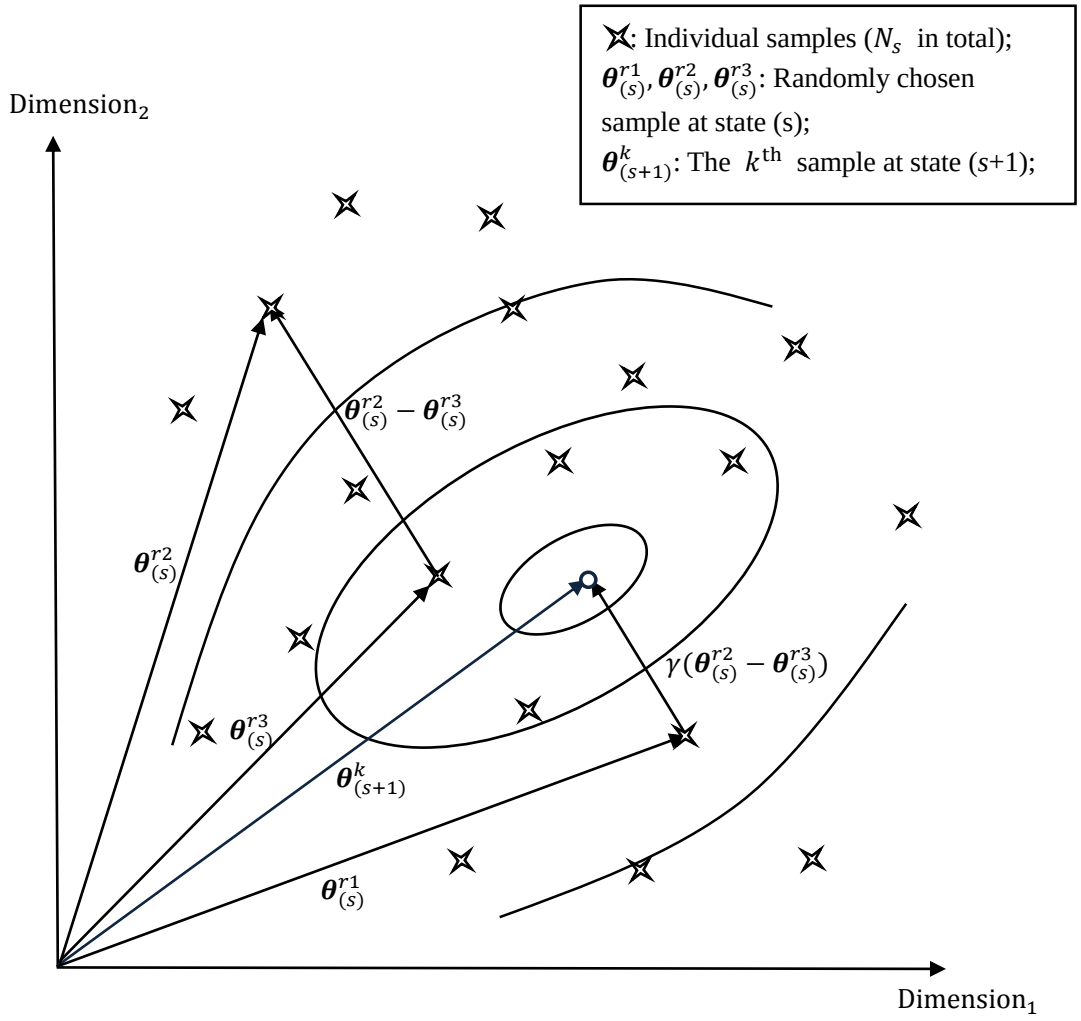


Figure 2.1 The operation DE mutation strategy (2-dimensional example)

where $\boldsymbol{e}$ is a vector drawn from a symmetric distribution with a small variance compared to that of the target, but with unbounded support, e.g. $\boldsymbol{e} \sim N(0, \boldsymbol{\sigma})^{N_d}$ with small vector $\boldsymbol{\sigma}$ of N_d-dimension. Comparing the transition of Markov chain in MH

algorithm as in Eq.(2.7), Eq.(2.10) can be seen as the transition of the k^{th} Markov chain at the state (s+1). From this point, if combing the mutation strategy of DE into MH algorithm, the problem of transition scale is well-solved. More importantly, the transition scale is automatically evolving with the reducing search domain of each population, which can be considered as an adaptive updating Markov progress in parallel (N_s Markov chains).

2.5.2 The Crossover Strategy of DE algorithm

Following the mutation phase, the crossover operator is applied on the population in DE algorithm. For each mutant vector, $\boldsymbol{\theta}_{(s+1)}^{k}$, is generated, with

$$\theta_{(s+1)}^{k,j} = \begin{cases} \theta_{(s+1)}^{k,j} & \text{if } (\text{rand}(j) \le C_R)\ ) \\ \theta_{(s)}^{k,j} & \text{if } (\text{rand}(j) > C_R)\ ) \end{cases}, \tag{2.11}$$

where $k = 1,2,\dots,N_s$; $j = 1,2,\dots,N_d$; $\text{rand}(j)$ is the j^{th} independent random number that uniformly distributed in the range of [0, 1]; C_R is pre-defined crossover constant $\in [0,1]$ that controls the diversity of the population (Storn and Price, 1997).

2.6 Differential Evolution Adaptive Metropolis-Hasting algorithm

The advantage of combining of DE and Metropolis-Hasting algorithm is due to the fact that it can run multiple different Markov chains in parallel and the posterior samples in

each iteration can be evolved using DE strategy respectively in each MC chains (Vrugt et al., 2009). This enhances the efficiency of optimum searching and the balance of ergodicity, which makes it able to solve higher dimensional probability inference problem. The main procedures of the differential evolution adaptive Metropolis-Hasting (DREAM) algorithm is summarized as:

1. Draw the initial samples, the k^{th} sample is denoted as $\boldsymbol{\theta}^k$ $(k = 1,2, \dots N_s)$, using the prior distribution, where the N_s is the number of population.
2. Compute the density of each samples $p(\boldsymbol{\theta}^k)$ for $k = 1,2, \dots N_s$.
3. Generating the candidate sample, $\boldsymbol{\theta}^k_{(s+1)}$, in chain k using the Eq. (2.12).

$$\theta^{k,i}_{(s+1)} = \theta^{k,i}_{(s)} + \gamma(\delta, N_d)[(u_i + e_i)\left(\sum_{p=1}^{n} \theta^{r_1(p),i}_{(s)} - \sum_{n=1}^{\delta} \theta^{r_2(p),i}_{(s)}\right)] + \varepsilon_i, \ i=1,\dots, \ N_d, \quad (2.12)$$

where $\boldsymbol{\varepsilon}$ is a N_d dimensional random vector that is drawn from $\boldsymbol{\varepsilon} \sim N\left(0, \boldsymbol{V}_{(s)}\right)^{N_d}$ (ε_i is the i^{th} element of $\boldsymbol{\varepsilon}$). This variable is the same as the jumping scale vector in the adaptive Metropolis algorithm. $r_1(p)$ and $r_2(p)$ are respectively different and random integers that are chosen from the integer set $\{1, 2,\dots k-1, k+1,\dots, N_s\}$. The term u_i is the i^{th} element of a N_d dimensional unit vector, and e_i signifies the i^{th} element of a small random vector drawn from a uniform distribution to assure the ergodicity of each individual Markov chain. The scaling factor $\gamma(\delta, N_d)$ is a function decided by the values of δ and N_d, where δ is the number of chosen pairs and N_d denotes the parametric dimension.

4. Replace each dimension $(j = 1, \dots, N_d)$ of the candidate sample $\theta^{k,j}_{(s+1)}$ with $\theta^{k,j}_{(s)}$ using a binomial scheme with probability $1-C_R$, where C_R is the predefined

crossover probability. With C_R=1, all dimensions of the new sample are updated jointly, which is same as the crossover strategy of DE algorithm.

5. Compute the $p(\boldsymbol{\theta}^k_{(s+1)})$, and $\alpha\left(\boldsymbol{\theta}^k_{(s)}, \boldsymbol{\theta}^k_{(s+1)}\right)$ of the candidate sample.

6. Accept the new state $\boldsymbol{\theta}^k_{(s+1)}$ with probability min(1, $\alpha\left(\boldsymbol{\theta}^k_{(s)}, \boldsymbol{\theta}^k_{(s+1)}\right)$), and keeps the current state $\boldsymbol{\theta}^k_{(s)}$ with probability 1−min(1, $\alpha\left(\boldsymbol{\theta}^k_{(s)}, \boldsymbol{\theta}^k_{(s+1)}\right)$.

7. Remove outlier chains using the Interquartile Range (IQR) statistic, a measure of statistical dispersion also called middle fifty (Upton and Cook, 1996).

8. Compute the Gelman-Rubin $\hat{R}_j$ convergence diagnostic for each dimension, $j = 1, \ldots, N_d$), using the last 50% of the samples in each chain after considering the burn-in period (the algorithm runs for many iterations until the initial state of the samples is "forgotten", and these iterations of the discarded samples are defined as "burn-in" period).

9. If the Gelman-Rubin's convergence diagnostic, *R*-Hat for each dimension is less than 1.2 ($\hat{R}_j < 1.2$, (Gelman and Rubin, 1992)), stop the algorithm, otherwise updating the samples in new iteration.

Seen from Eq.(2.12), it can be easily found that the DREAM algorithm automatically selects an appropriate scale and orientation of the transition distribution that routed to the target distribution. Because of the strong search ability of DE algorithm, heavy-tailed and multimodal target distributions are more efficiently accommodated when comparing with the transition of Markov chain in the original MH algorithm as in Eq.(2.7).

The desired posterior PDF, $p(\cdot)$, can be obtained from these Markov Chain samples excluding the ones in the burn-in period. Same as the standard MH algorithm, in

DREAM algorithm it was required to design reversibility condition to maintain the balance at each step in the chain:

$$p(\boldsymbol{\theta}_{(s+1)})p(\boldsymbol{\theta}_{(s)}|\boldsymbol{\theta}_{(s+1)})=p(\boldsymbol{\theta}_{(s)})p(\boldsymbol{\theta}_{(s+1)}|\boldsymbol{\theta}_{(s)}). \quad (2.13)$$

This reversibility condition mainly denotes that the transition rates between the samples at two different states are equal, which ensures Markov chain samples are distributed as the target density.

2.7 Operational parameters for the DREAM algorithm

2.7.1 The mutation scale index

Different from Eq.(2.10), to increase the diversity of the samples the DREAM generates the proposals, $\boldsymbol{\theta}^{k}_{(s+1)}$ as in Eq. (2.12), using higher-order pairs, δ was chosen larger than 2. A good choice for $\gamma(\delta, N_d)$ is suggested as be equal to $2.38/\sqrt{2\delta N_d}$ (Vrugt and Ter Braak, 2011). To enable the Markov chain jump between different modes of the posterior samples, the $\gamma(\delta, N_d)$ is suggested to be set as 1.0 every 5 iteration.

2.7.2 Outlier-chains checking and removing

It is required to remove the outlier chains to facilitate the convergence of the Markov samples into a limiting distribution because the outlier chains may deteriorate the performance of MCMC samplers. To detect aberrant trajectories, DREAM stores the mean of logarithm of the posterior densities of the last 50% of the samples in each chain.

From these stored samples, the interquartile range statistic, $IQR = Q_3 - Q_1$ is computed, in which Q_1 and Q_3 denote the first quartile and the third quartile of the Markov samples. Chains with $\Omega < Q_1 - 2IQR$ are considered as outlier chains, of which are removed from the current state of the posterior samples. This operation cannot maintain detailed balance and is therefore only be used during burn-in period. If an outlier chain is being detected another burn-in period is applied before summarizing the posterior moments.

2.8 Conclusions

In this section, a classic MCMC method named as the MH algorithm and a heuristic evolution algorithm, DE algorithm, was explained. The significant improvements to the efficiency of MCMC simulation can be made by running multiple interacting chains simultaneously when combing the DE mutation and crossover strategy with the MH algorithm. This algorithm entitled as DREAM, which automatically tunes the scale and orientation of the transition distribution during the search of MC samples, adapts subspace sampling to maximize the average normalized jumping distance in each chain, and explicitly handles outlier chains to avoid convergence problems on difficult response surfaces with numerous local optima, was presented. The operational parameters for the DREAM algorithm was briefly introduced in this section.

CHAPTER 3

Posterior Density Estimation for Structural Parameters using Improved Differential Evolution Adaptive Metropolis Algorithm

3.1 Introduction

In this section, an improved differential evolution adaptive Metropolis-Hasting algorithm (IDREAM) is proposed for updating the posterior PDF of the structural identification model. The contribution of the proposal in this section lies in the enhancing of convergence for the transition of the Markov chains, which will greatly save the time of Bayesian identification for structural system. The proposed algorithm combines an improved mutation and crossover strategy, enhancing the ability of global search, with the MH algorithm. It runs different parallel Markov chains simultaneously and the posterior samples mutually exchange information along the iteration. Comparing with the original DREAM algorithm, the convergence speed of the Markov

chain is fasten without reducing the accuracy of the maximum a posteriori estimation (MAP). Numerical examples of updating the posterior PDF of a 5-DOFs linear structural system and a 2-DOFs nonlinear hysteretic system are presented, with which the effectiveness and efficiency of IDREAM are investigated. The influence of the incomplete measurements and noise errors on the posterior PDF of the parameters is discussed.

3.2 IDREAM for Bayesian Inference of Parameter Estimation

The IDREAM algorithm starts by choosing stochastic samples represented as a dimensional vector $\boldsymbol{\theta}$, and the sample k on the Markov chains is denoted as $\boldsymbol{\theta}^k$ $(k = 1,2, \dots N_s)$. The initial states of the Markov chain samples are drawn from the search domain by using Latin hypercube sampling (LHS). The density function of each sample in the initial state can be computed as a prior density $p\left(\boldsymbol{\theta}_{(1)}^k\right)$ for $k = 1,2, \dots N_s$.

3.2.1 The improved transition strategy for the Markov samples

In the DREAM algorithm the samples are updated by using the difference between randomly chosen pairs of samples in the current state. Let $\Delta_{\boldsymbol{\theta}}^k = \boldsymbol{\theta}_{(s+1)}^k - \boldsymbol{\theta}_{(s)}^k$ denote the jumping scale between the updating state (s+1) and current state (s) of the sample k in the Markov chain. Here a new update pattern in which the sequence having the largest probability density in the current state (s^{th} state) and the one with the maximum posterior PDF from the initial state (1^{st} state) to the current state (s^{th} state) are used for the updating of the Markov chain samples:

$$\Delta_{\theta}^{k,i} = w_1^k\left(\theta_i^{\text{cbest}} - \theta_{(s)}^{k,i}\right) + w_2^k\left(\theta_i^{\text{gbest}} - \theta_{(s)}^{k,i}\right) + \gamma(\delta, N_d)[(u_i + e_i) \left(\sum_{p=1}^{n} w^{*k}_j \left(\theta_{(s)}^{r_1(p),i} - \theta_{(s)}^{r_2(p),i}\right)\right)] + \varepsilon_i, \quad i=1, \dots, N_d, \tag{3.1}$$

where $\Delta_{\theta}^{k,i}$ denotes the i^{th} element of the jumping scale vector for k^{th} sample, $\Delta_{\boldsymbol{\theta}}^{k}$. $\theta_{(s)}^{k,i}$ is the i^{th} element of the k^{th} posterior sample at the current state (s^{th} state), $\boldsymbol{\theta}_{(s)}^{k}$. $\boldsymbol{\theta}^{\text{cbest}}$ denotes the sample with the maximum PDF in the current state (s^{th} state) and is called the "current-best individual sample" (θ_i^{cbest} is the i^{th} element of the vector, $\boldsymbol{\theta}^{\text{cbest}}$), and $\boldsymbol{\theta}^{\text{gbest}}$ denotes the sample with the maximum PDF during all the previous states (from the initial state (1^{st} state) to the current state (s^{th} state)) and is called the "global-best individual sample" (θ_i^{gbest} is the i^{th} element of the vector, $\boldsymbol{\theta}^{\text{gbest}}$). $\boldsymbol{\theta}_{(s)}^{r_1(p)}$ and $\boldsymbol{\theta}_{(s)}^{r_2(p)}$ are randomly chosen pairs of samples but mutually different individuals in the current state (s^{th} state) (totally *n* pairs). The weighting factors, w_1^k, w_2^k and w^{*k}_j, are obtained from the distance between the individual samples as follows:

$$\begin{cases} w_1^k = d_1^k/d_{sum}^k \,, w_2^k = d_2^k/d_{sum}^k, w^{*k}_j = d_j^k/d_{sum}^k \\ d_{sum}^k = d_1^k + d_2^k + \sum_{j=1}^{\delta} d_j^k \end{cases}, \tag{3.2}$$

where d_1^k and d_2^k are respectively the Euclidean distance from the sample $\boldsymbol{\theta}_{(s)}^{k}$ to the "current-best" sample $\boldsymbol{\theta}^{\text{cbest}}$, and the distance from the sample $\boldsymbol{\theta}_{(s)}^{k}$ to the "global-best" sample $\boldsymbol{\theta}^{\text{gbest}}$, and d_j^k is the Euclidean distance between randomly chosen sample pairs excluding the sample $\boldsymbol{\theta}_{(s)}^{k}$ in the Markov chain. From Eq. (2.12), one can see that the biggest difference of IDREAM from DREAM is that the posterior samples are updated using both the maximum PDF in the current state, $\boldsymbol{\theta}^{\text{cbest}}$, and that of the previous states, $\boldsymbol{\theta}^{\text{gbest}}$, while the updating samples of DREAM are orientated between the chosen pairs of samples only in the current state. IDREAM can enhance

the convergence speed especially in the early stage because the difference between the individual and the "best" sample is definitely large owing to the diversity of samples in the initial state. From Eq. (3.1), it is clear that samples farther away from the "best" sample possess a larger jumping scale because the weighting factor is positively proportional to the distance between the samples. As for the "best" sample, the update method is the same as in DREAM, i.e., Eq. (2.12). The convergence becomes slower in the later stages when the diversity of the samples in the Markov chain decreases. Because of this, a dynamic crossover strategy is used to keep the diversity of the Markov-chain samples high.

3.2.2 The dynamic adaptation of the crossover probability

The improved-DREAM (IDREAM) also explores the DE crossover strategy as in Eq. (2.11). To speed up convergence to the target distribution, the IDREAM estimates a distribution of crossover probabilities, which adapted dynamically during the burn-in period, favoring large jumps over smaller ones in each of the N_s chains. A discrete number of candidate points for the crossover value is generated as $\{m/n_{CR}\,|m = 1,2,\dots,n_{CR}\}$, where n_{CR} denotes the number of samples chosen for the crossover strategy.

To realize the adaptation of the crossover probability for each individual C_R value in Eq. (2.11), the crossover probability is adapted as follows:

Step 1: Set *s*=1, L_m=0, $p_m = 1/n_{CR},\ m = 1,2,\dots,n_{CR}$.

Step 2: Sample m from the numbers $1,2,\dots,n_{CR}$ using the multinomial distribution. Set C_R=m/n_{CR}, and $L_m = L_m + 1$; Create a candidate point $\boldsymbol{\theta}^k_{(s+1)}$ using Eqs. (3.1) and (2.11) with crossover probability C_R.

Step 3: Accept or reject the candidate Markov samples by the Metropolis acceptance.

Step 4: Compute the squared normalized jumping distance as:

$$\Delta_m = \sum_{j=1}^{N_d} (\boldsymbol{\theta}_{(s+1)}^{k,j} - \boldsymbol{\theta}_{(s)}^{k,j})^2 / r_j^2,\ j=1,2,\dots,N_d. \tag{3.3}$$

where r_j denotes the current standard deviation of dimension *j* of all the samples.

Step 5: Update the probability of the different C_R values as:

$$p_m = sN_s \cdot (\Delta_m / L_m) / \sum_{j=1}^{n_{CR}} \Delta_j,\ m=1,2,\dots,n_{CR}. \tag{3.4}$$

Step 6: Set s=s+1. If s is in burn-in period, the crossover probability is adapted dynamically, otherwise is a predefined constant value.

Therefore, to obtain an adaptive crossover probability, the parameter that required to be initially set is n_{CR}. For instance, set $n_{CR} = 3$, the crossover probability C_R will be equal to $\{1/3, 2/3, 1\}$, and the probability distributions p for the crossover probability, $p = \{p_m | p_m = p(C_R = 1/m), m = 1,2,\dots,n_{CR}\}$, will be obtained dynamically by Eqs. (3.3) and (3.4).

3.2.3 The Metropolis acceptance

The density of the new sample, $p(\boldsymbol{\theta}_{(s+1)}^{k})$, and the Metropolis acceptance (Chib and

Greenberg, 1995; Haario et al., 2006) can be calculated with the updated samples in the Markov chain:

$$\alpha\left(\boldsymbol{\theta}_{(s)}, \boldsymbol{\theta}_{(s+1)}\right) = \min\left[\frac{p(\boldsymbol{\theta}_{(s+1)})}{p(\boldsymbol{\theta}_{(s)})}, 1\right]. \tag{3.5}$$

The algorithm accepts the candidate state $\boldsymbol{\theta}_{(s+1)}$ with probability $\alpha\left(\boldsymbol{\theta}_{(s)}, \boldsymbol{\theta}_{(s+1)}\right)$, and keeps the current state $\boldsymbol{\theta}_{(s)}$ with probability $1-\alpha\left(\boldsymbol{\theta}_{(s)}, \boldsymbol{\theta}_{(s+1)}\right)$. This process is repeated in several iterations, and after a burn-in period, the chain of samples approaches a stationary distribution. The desired posterior PDF can be obtained from these Markov-chain samples, excluding the ones in the burn-in period.

3.2.4 Gelman-Rubin convergence condition

The IDREAM algorithm ends by checking the Gelman-Rubin convergence condition (Gelman and Rubin 1992), and calculates the $\hat{R}_j$-statistic by using the last 50% of the samples in each chain. Let *k* be the number of the sequences used to calculate $\hat{R}_j$, and let $\boldsymbol{B}$ denote the variance between the N_s chains, and $\boldsymbol{W}$ signify the average of the *k* within-chain variances ($\boldsymbol{B}$=$N_s \sum_{i=1}^{k}(\boldsymbol{\theta}_i - \overline{\boldsymbol{\theta}})^2 /(k-1)$, and $\boldsymbol{W}$=$\frac{1}{k}\sum_{i=1}^{k}\boldsymbol{s}_i^2$, where $\boldsymbol{s}_i^2$ is the variance of the samples, $\boldsymbol{s}_i^2 = \frac{1}{N_s-1}\sum_{i=1}^{N_s}(\boldsymbol{\theta}_i - \overline{\boldsymbol{\theta}})^2$). Each j^{th} dimension of the Gelman-Rubin convergence diagnostics vector, $\widehat{\boldsymbol{R}}$, can be calculated as:

$$\hat{R}_j = \sqrt{\frac{N_s - 1}{N_s} + \frac{k+1}{kN_s}\frac{B_j}{W_j}},\ j = 1, \ldots, N_d. \tag{3.6}$$

If $\hat{R}_j < 1.2$ for all dimensions (Gelman et al. 2004), it means that the Markov chain has converged to a stationary distribution (with big number of chains, the multivariate matrix, $\widehat{\boldsymbol{R}}$, should converge less than 1.2 for each dimension of the parameter); otherwise, the posterior sample is updated with Eq. (3.1).

3.2.5 Procedures of IDREAM based parametric posterior density estimation

The procedure of IDREAM-based Bayesian probability inference parameter estimation is as:

Procedure 1: The initial N_s sequences for the Markov chain are drawn by Latin hypercube sampling with the predefined maximum and minimum boundary of structural parameters and the number of samples, N_s, while respecting the prescribed limits of the search space. Calculate the prior density and the likelihood function using Eq. (2.3).

Procedure 2: Update the posterior sample using Eq. (3.1) and Eq. (3.2).

Procedure 3: The samples are updated according to the crossover probability calculated with Eqs. (2.11) and (3.3)~(3.4). Calculate the density for the each updated sample in the Markov chain.

Procedure 4: Use the Metropolis acceptance (Eq. (3.5)) to decide whether to accept the updated samples.

Procedure 5: Considering the burn-in period and calculate the Gelman and Rubin

diagnostic using Eq. (3.6) for each dimension of the stochastic parameter. If the stopping criteria are met, ($\hat{R}_j < 1.2$ or the iteration number reached to the predefined maximum iteration) stop the algorithm; otherwise, return to Procedure 2.

The block diagram of the algorithm is shown in Fig. 3.1.

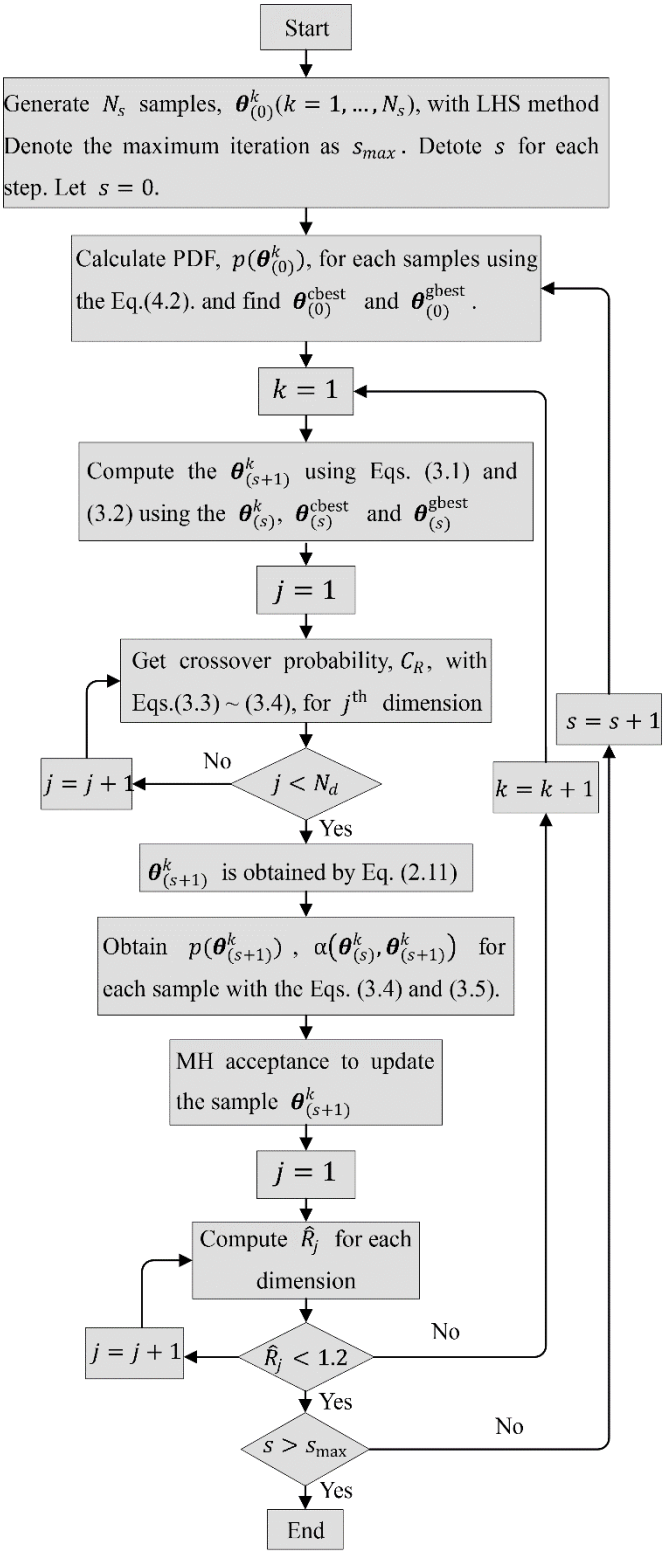


Figure 3.1 Block diagram of the IDREAM algorithm

3.2.6 The assertion of convergence of the IDREAM algorithm

IDREAM yields a Markov chain, which is irreducible and aperiodic with unique stationary distribution with target distribution $p(\cdot)^{N_s}$. *Proof*: the assertion consist of two parts as:

1. Markov chains are updated sequentially and conditionally on the other chains. It is thus an N_s-component Metropolis-within-Gibbs algorithm that defines a single Markov chain on the state space. The conditional distribution of each sample is $p(\cdot)$.

If the sample of k^{th} chain coincides with the position of the global-best and the current-best individual, $p(\cdot)$ is a stationary distribution of the k^{th} chain, because the chain is reversible. The reason for this is that the jumps in each MC sample chain satisfy detailed balance with respect to $\pi(\cdot)$ at each step. For the k^{th} chain, the conditional probability to jump from $\boldsymbol{\theta}_{(s)}^{k}$ to $\boldsymbol{\theta}_{(s+1)}^{k}$, which owns the conditional probability, $p\left(\boldsymbol{\theta}_{(s)}^{k} \rightarrow \boldsymbol{\theta}_{(s+1)}^{k}\right)$, being equal to the reverse jump, $p\left(\boldsymbol{\theta}_{(s+1)}^{k} \rightarrow \boldsymbol{\theta}_{(s)}^{k}\right)$. This is because the distribution of $\boldsymbol{e}_d$ and $\boldsymbol{\varepsilon}$ is symmetric and the pair $(\boldsymbol{\theta}_{(s)}^{r_1}, \boldsymbol{\theta}_{(s)}^{r_2})$ is as likely as $(\boldsymbol{\theta}_{(s)}^{r_2}, \boldsymbol{\theta}_{(s)}^{r_1})$. If $\boldsymbol{\theta}_{(s+1)}^{k} \sim p(\cdot)$, then the detailed balance is achieved point wise by accepting the proposal with probability $\min(p(\boldsymbol{\theta}_{(s+1)}^{k})/p(\boldsymbol{\theta}_{(s)}^{k}), 1)$.

For the individuals who are different from the "best" sample in current state, *s*, the distance and the corresponding weighting factor reduced fast with the convergence of the populations in the later stages (especially following a sufficient burn-in period). The deviations (random walk) of the randomly chosen sample pairs mainly decides the transition of each chain. The $N_s - 1$ samples can been seen as updated conditionally on the other chains obeying the reversibility of the Markov chain, because the Jacobian of the transformation (Chauveau and Vandekerkhove, 2002; Hastie and Green, 2012) implied in Eq. (3.6) is close to 1 in absolute value, (Vrugt et al. 2009), and the first two

items of Eq. (3.1) plays limited role on the transition Markov chains during the later stages (Noted that: the $\boldsymbol{\theta}^{\mathrm{cbest}}$ is equal to the $\boldsymbol{\theta}^{\mathrm{gbest}}$ in the later stage).

2. Detailed balance can be achieved by the accepting rule with probability $\min(p(\boldsymbol{\theta}_{(s+1)})/p(\boldsymbol{\theta}_{(s)}), 1)$. Because the joint stationary probability distribution of the N_s chains factorizes to $p(\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^{N_s}) = p(\boldsymbol{\theta}^1)p(\boldsymbol{\theta}^2) \dots p(\boldsymbol{\theta}^{N_s})$, each of the N_s chains are independent at any generation after the algorithm has become independent of its initial value after the burn-in period. And each chain maintains conditional detailed balance, because the chains are aperiodic and not transient with a random walk that generated by the Eq.(3.1). With the unbounded support of the distribution of $\boldsymbol{\varepsilon}$ in Eq.(3.1), the N_s chains are irreducible. This thus concludes the assertion.

3.3 Numerical Simulation

3.3.1 Linear structural system

A numerical simulation of a 5-DOF linear time invariant (LTI) system was carried out to verify the IDREAM algorithm. The actual system output was simulated from a linear structural system (shear frame structure), as is shown in Fig.3.2. For sake of clearly exhibiting that only the measurement uncertainty is considered, the second-order differential dynamic equation of the structural system is described by the state-space representation as:

$$\begin{bmatrix} \dot{\boldsymbol{v}}_1(t) \\ \dot{\boldsymbol{v}}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_1(t) \\ \boldsymbol{v}_2(t) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ -\mathbf{I} \end{bmatrix} \boldsymbol{\Gamma}^T \boldsymbol{u}(t), \tag{3.7}$$

where $\boldsymbol{M}$, $\boldsymbol{C}$, and $\boldsymbol{K}$ are the mass, damping and the stiffness matrices, $\mathbf{I}$ is a $N_d \times N_d$

identity matrix, $\boldsymbol{\Gamma} = [1,1,\ldots,1]^T$ is a $1 \times N_d$ position vector ($N_d = 5$). $\boldsymbol{v}_1(t)$, and $\boldsymbol{v}_2(t)$, are state space vectors respectively representing the vector of displacement and velocity, and $\boldsymbol{u}(t)$ is the input ground acceleration of the system. Equation (3.7) includes a Rayleigh damping matrix $\boldsymbol{C}$, where modal damping ratio ($\boldsymbol{\zeta}$) is set as 5% for the first two modes (Mita, 2003).

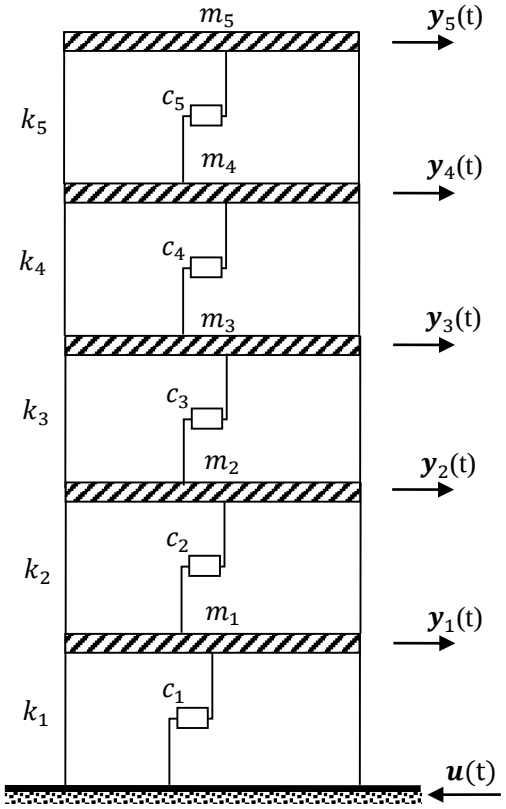


Figure 3.2 5-DOF LTI structural system

$$\boldsymbol{C} = \alpha \boldsymbol{M} + \beta \boldsymbol{K} \tag{3.8}$$

$$\alpha = 2\,\omega_i\omega_j(\zeta_i\omega_i - \zeta_j\omega_j)/(\omega_i^2 - \omega_j^2);\ \beta = 2\,(\zeta_i\omega_i - \zeta_j\omega_j)/(\omega_i^2 - \omega_j^2) \tag{3.9}$$

where i=1 and j=2 respectively denotes the first two mode; ω_i and ω_j thus denotes the natural frequency for the first and second mode of the system. The system output is an relative acceleration value that is assumed to be contaminated by Gaussian white noise $\boldsymbol{w}_j(t) \sim N(0, \sigma_j^2)$, ($j$=1,..$N_m$). The measured output vector is written as

$$\boldsymbol{y}(t) = [-\boldsymbol{M}^{-1}\boldsymbol{K} \quad -\boldsymbol{M}^{-1}\boldsymbol{C}] \begin{bmatrix} \boldsymbol{v}_1(t) \\ \boldsymbol{v}_2(t) \end{bmatrix} - \boldsymbol{\Gamma}^T \cdot \boldsymbol{u}(t) + \boldsymbol{w}(t). \tag{3.10}$$

The ground acceleration was an the 1940 El Centro ground motion (N-S acceleration at the Imperial Valley Irrigation District substation in El Centro, CA, during the 1940 Imperial Valley earthquake) lasting 40 s normalized so that its peak is 10 cm/s^2 and the sampling frequency was 50 Hz (Fig. 3.3). The Newmark-beta method was used to calculate the structural response. The output relative acceleration (*acc.*) with different Gaussian white noise levels (Eq. (3.10)) was assumed. To show the effectiveness of the proposed method for identification problem with large noise level, the noise level (nl) was chosen to have a standard deviation that is 30% or 100% of the corresponding signal; i.e., if $\sigma_{acc.,j}$ is the standard deviation of the j^{th} relative acceleration, then the noise on the measurement of that floor's acceleration has an RMS $\sigma_j = nl \times \sigma_{acc.,j}$ (j=1,..N_m). The measurement of the 5$^{\text{th}}$ DOF without and with the 100% noise is shown in Fig. 3.3.

The influence of the limited availability of measurements on the proposed method is also assessed in this study. In the "full output" scenario, measurements of all DOFs are available, whereas in the "partial output" scenario, only data from floors 3 and 5 are available. The mass is assumed to be known and deterministic; hence, a N_d-DOF system with N_m-available measurements can be described by a model set, of which the stochastic parameterized vector for each model is

$$\boldsymbol{\theta}=\{k_1,\dots,k_{N_d},\zeta_1,\zeta_2,\sigma_1,\dots,\sigma_{N_m}\} \tag{3.11}$$

where k_i denotes the stiffness of the i^{th} floor, N_d is the number of structural DOFs; ζ_1 and ζ_2 denotes the damping ratio of the first two modes; σ_i denotes the standard deviation of i^{th} measurement noise, N_m is the number of available measurements; $\boldsymbol{\theta}$ denotes the random variables in probability space, $\boldsymbol{\Omega_\theta}$.

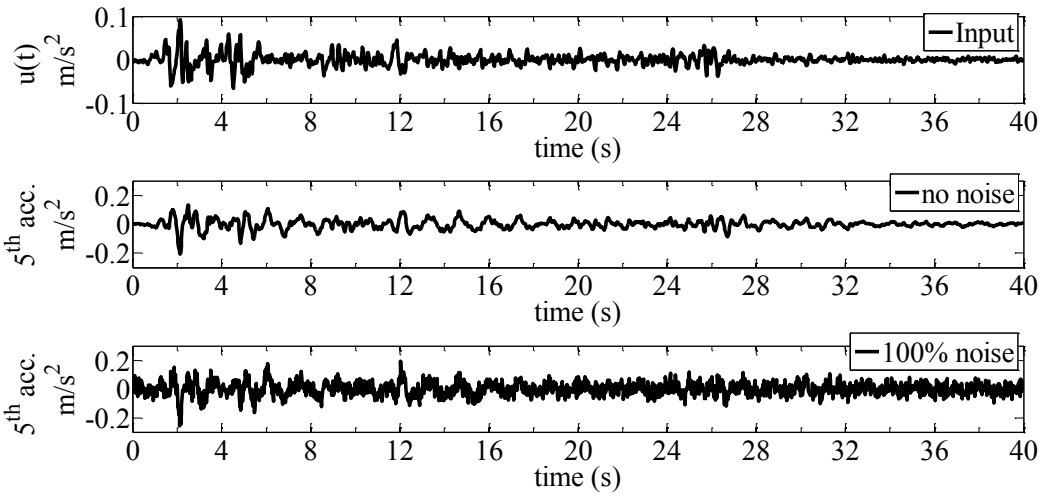


Figure 3.3 Input and output without and with noise

Table 3.1 lists the structural properties together with the simulated variance of the prediction-error for each available measured response. The parametric dimension was 12 for the full output scenario, and 9 corresponding to the partial output scenario.

The parameters of the IDREAM algorithm were set as follows: the number of Markov chain samples (N_s) was 20, the crossover probability (C_R) was 0.85, and the number of

sample pairs (δ) was 5. The maximum iteration was set as 4000. The search domain was taken as $0.5-2.0$ times the exact values. The prior distribution obeys uniform distribution in the search boundary. The results obtained by the original DREAM algorithm are shown for comparison in the full output scenario (Tables 3.2 and 3.3) and partial output scenario (Tables 3.4 and 3.5).

Table 3.1 Structural properties and the prediction-error variances

Stiffness (k)	(10^4N/m)		(σ_j; j=1,..N_m) m/s^2	30% noise	100% noise
Floor 1	2.000		Floor 1	0.0049	0.0162
Floors 2-5	1.000		Floor 2	0.0074	0.0248
Mass (m)	(kg)	Case 1	Floor 3	0.0076	0.0252
Floors 1-4	50		Floor 4	0.0077	0.0256
Floor 5	45		Floor 5	0.0095	0.0317
Damping ratio					
ζ_1	0.05	Case 2	Floor 3	0.0076	0.0252
ζ_2	0.05		Floor 5	0.0095	0.0317

From Tables 3.2 and 3.3, it is clear that both algorithms performed very well in the noise-free scenario. With the increasing magnitude of the measured noise the relative error of the identified solutions are also not large. The maximum relative error for the most plausible value of structural stiffness, the sample with the maximum posterior density ($\boldsymbol{\theta}_{\mathrm{MAP}}$), ranged from zero in the no-noise case to 0.502% in the 30% noise case and up to 1.987% in the 100% noise case when using the DREAM algorithm.

Table 3.2 Identified results of structural parameters (full output scenario)

		no noise		30% noise		100% noise	
		DREAM	IDREAM	DREAM	IDREAM	DREAM	IDREAM
k_1	error	0.000	0.000	0.352	0.164	0.656	0.250
	cov.	0.000	0.000	0.906	0.533	3.263	1.641
k_2	error	0.000	0.000	0.324	0.242	1.558	0.114
	cov.	0.000	0.000	0.683	0.399	2.462	1.286
k_3	error	0.000	0.000	0.298	0.353	1.987	0.198
	cov.	0.000	0.000	0.812	0.459	2.946	1.474
k_4	error	0.000	0.000	0.172	0.045	0.658	0.717
	cov.	0.000	0.000	0.813	0.465	2.789	1.564
k_5	error	0.000	0.000	0.502	0.299	1.155	1.083
	cov.	0.000	0.000	0.937	0.530	3.227	1.835
ζ_1	error	0.000	0.000	1.307	0.813	3.442	2.035
	cov.	0.000	0.000	1.591	0.871	5.545	2.946
ζ_2	error	0.000	0.000	0.887	1.217	1.677	0.499
	cov.	0.000	0.000	1.158	0.635	3.951	2.095

* the error in the table is in %; the cov. in the table is in %.

The accuracy of $\boldsymbol{\theta}_{\mathrm{MAP}}$ for IDREAM is mostly better than that of the DREAM algorithm. The maximum relative error fell to 0.353% in the 30% noise case and 1.083% in the 100% noise case. The parameter with the largest uncertainty obtained by the two algorithms was in the damping ratio. The largest coefficient of variance (*cov.*, calculated by the standard deviation of the posterior samples divided by the mean value of the posterior samples, $100 \times \sigma(\boldsymbol{\theta})/\mu(\boldsymbol{\theta})$) of the damping ratio was 1.591% in the 30% noise case, and it increased to 5.545% in the 100% noise case for the DREAM

algorithm. For IDREAM, the coefficient of variance (cov.) of the damping ratio ranged from 0.871% to 2.946%.

Table 3.3 Identified results of prediction errors (full output scenario)

		no noise		30% noise		100% noise	
		DREAM	IDREAM	DREAM	IDREAM	DREAM	IDREAM
σ_1	error	0.000	0.000	0.967	0.080	2.060	0.204
	cov.	0.000	0.000	0.136	0.082	0.172	0.055
σ_2	error	0.000	0.000	0.966	0.034	0.725	2.562
	cov.	0.000	0.000	0.074	0.042	0.183	0.043
σ_3	error	0.000	0.000	0.445	0.731	2.341	0.678
	cov.	0.000	0.000	0.088	0.051	0.084	0.032
σ_4	error	0.000	0.000	0.185	1.879	1.302	1.128
	cov.	0.000	0.000	0.075	0.049	0.070	0.042
σ_5	error	0.000	0.000	1.108	1.741	2.237	0.128
	cov.	0.000	0.000	0.096	0.022	0.111	0.048

* the error in the table is in %; the cov. in the table is in %.

It is clear that the parametric uncertainty was additive as the measurement error increased. Table 3.3 shows that the two algorithms can identify the exact value of the prediction-error variance in Table 3.1. In the noise free scenario, both solutions were as small as zero. For the 100% noise case, the maximum error emerged in the 3rd floor at 2.341% for DREAM and at 2.562% for IDREAM. The results for the partial output are listed in Tables 3.4 ~3.5.

Table 3.4 Identified results of structural parameters (partial output scenario)

		no noise		30% noise		100% noise	
		DREAM	IDREAM	DREAM	IDREAM	DREAM	IDREAM
k_1	error	0.019	0.010	1.404	1.501	2.227	2.546
	cov.	0.000	0.000	4.267	2.606	16.58	7.761
k_2	error	0.082	0.025	1.025	0.794	4.372	1.450
	cov.	0.000	0.000	2.216	1.351	10.33	4.584
k_3	error	0.072	0.018	1.855	0.196	5.602	0.725
	cov.	0.000	0.000	2.710	1.530	12.69	5.538
k_4	error	0.089	0.014	1.041	0.511	2.722	2.376
	cov.	0.000	0.000	2.145	1.222	8.633	4.224
k_5	error	0.106	0.028	1.233	1.189	0.859	4.485
	cov.	0.000	0.000	2.243	1.358	12.32	4.171
ζ_1	error	0.050	0.030	1.066	0.406	1.667	5.214
	cov.	0.000	0.000	2.110	1.179	6.774	4.024
ζ_2	error	0.016	0.006	1.564	1.067	0.273	0.972
	cov.	0.000	0.000	2.142	1.258	7.930	4.020

* the error in the table is in %; the cov. in the table is in %.

From Table 3.4, we find that when there is a noise error, the maximum relative errors of $\boldsymbol{\theta}_{\mathrm{MAP}}$ are mostly smaller for IDREAM than for DREAM. In the 30% noise case, the maximum relative error for DREAM was 1.855%, but only 1.501% for IDREAM, while in the 100% noise case, the maximum error decreased from 5.602% to 5.214%. Comparing Table 3.4 with Table 3.2, it is clear that the loss of available measurements leads to an increase in parametric uncertainty, because the maximum coefficient of variance for the structural parameters adds from 2.946% to 7.761% at the same noise

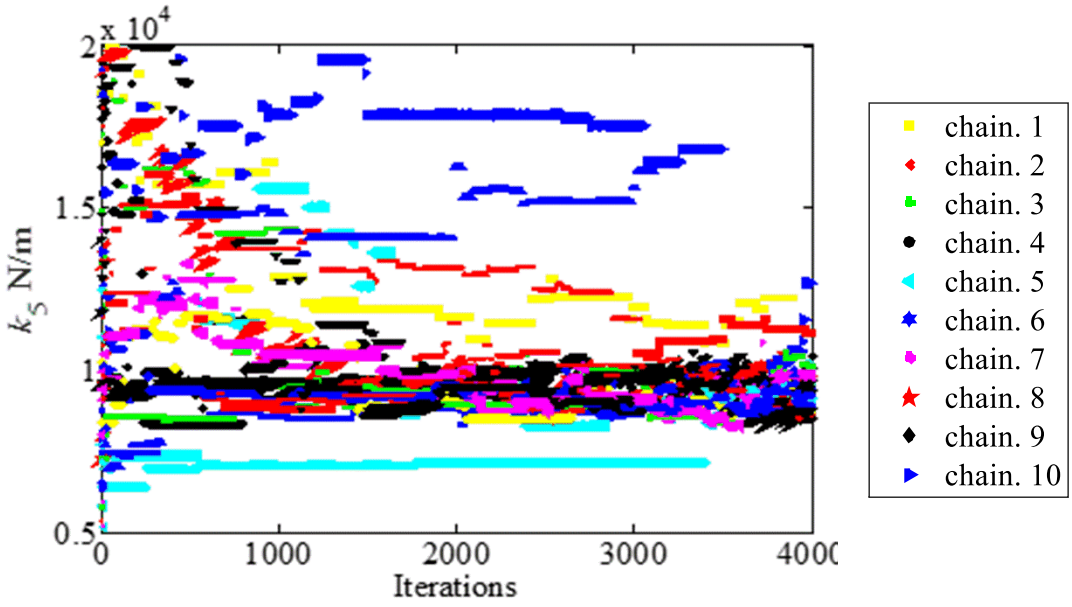
level.

Table 3.5 Identified results of prediction errors (partial output scenario)

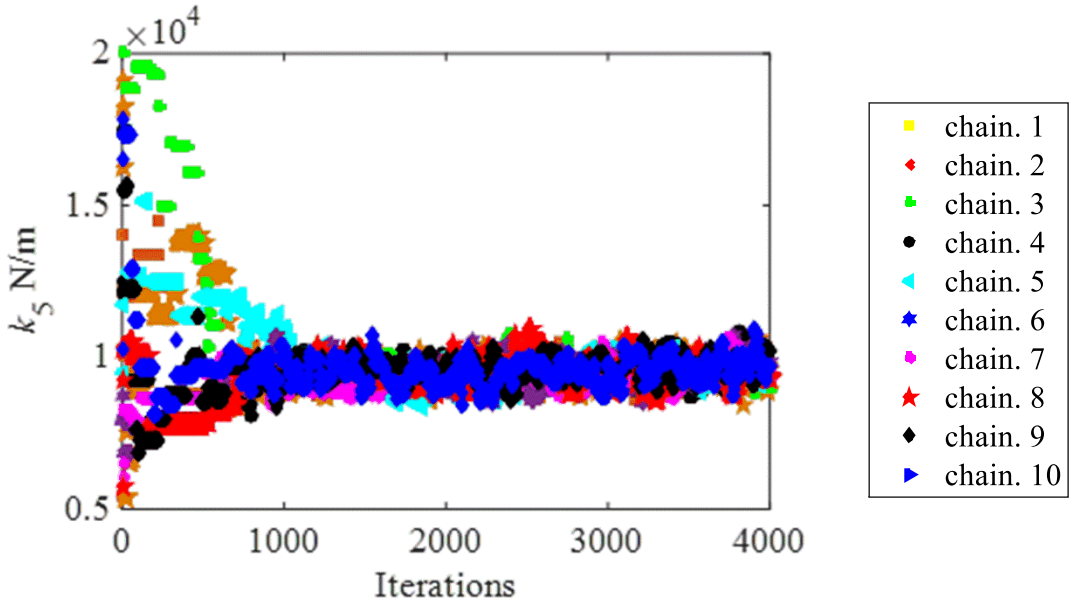
		no noise		30% noise		100% noise	
		DREAM	IDREAM	DREAM	IDREAM	DREAM	IDREAM
σ_3	error	0.000	0.000	0.950	2.406	2.675	0.669
	cov.	0.000	0.000	0.145	0.114	0.207	0.089
σ_5	error	0.000	0.000	1.240	0.982	0.490	0.557
	cov.	0.000	0.000	0.096	0.094	0.116	0.091

* the error in the table is in %; the cov. in the table is in %.

Table 3.5 shows that the actual prediction error standard deviation can be well estimated even if only measurements of the 3rd and 5th floors are available. The maximum error of the estimated prediction errors, $\boldsymbol{\theta}_{\mathrm{MAP}}$, is 2.406% in the 30% noise case and 0.669% in the 100% noise case. For the DREAM algorithm, the corresponding errors are 1.240% in the 30% noise case and 2.675% in the 100% noise case. (Noted that the maximum error of the estimator prediction errors seems larger for 30% noise than for 100% noise scenario. The reason for this phenomenon is the denominator for calculating the relative error in the case of 30% noise is smaller than that in the case of 100% noise.)

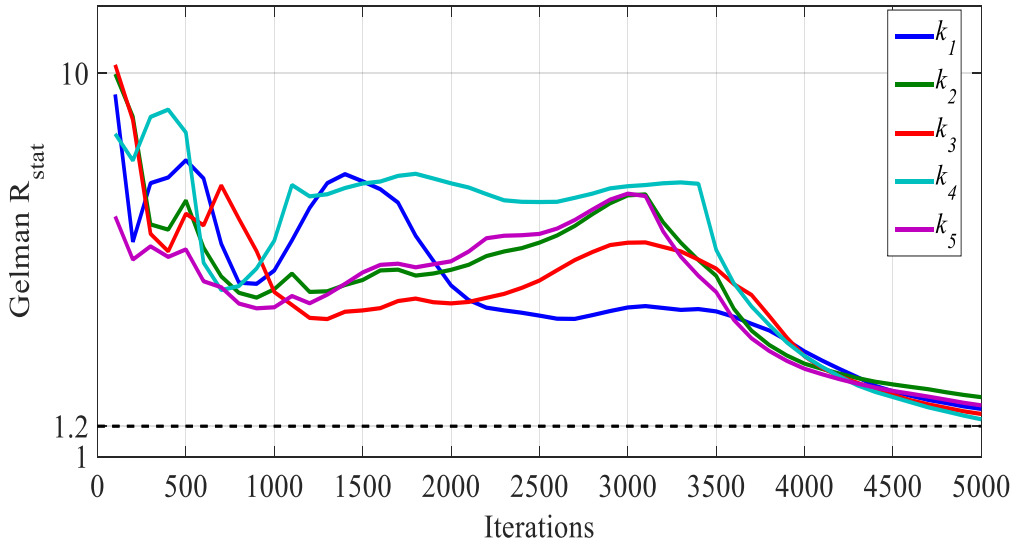


(a) DREAM

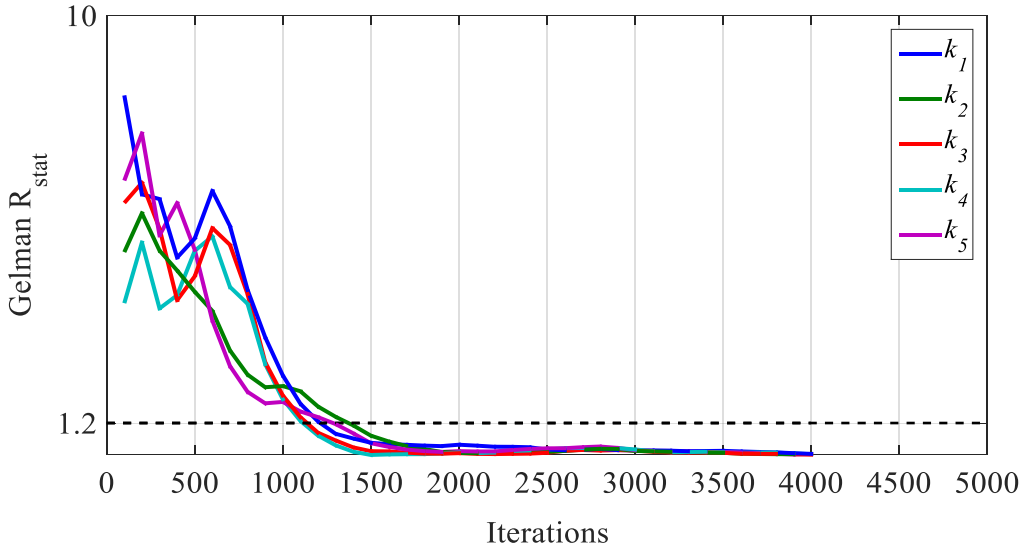


(b) IDREAM

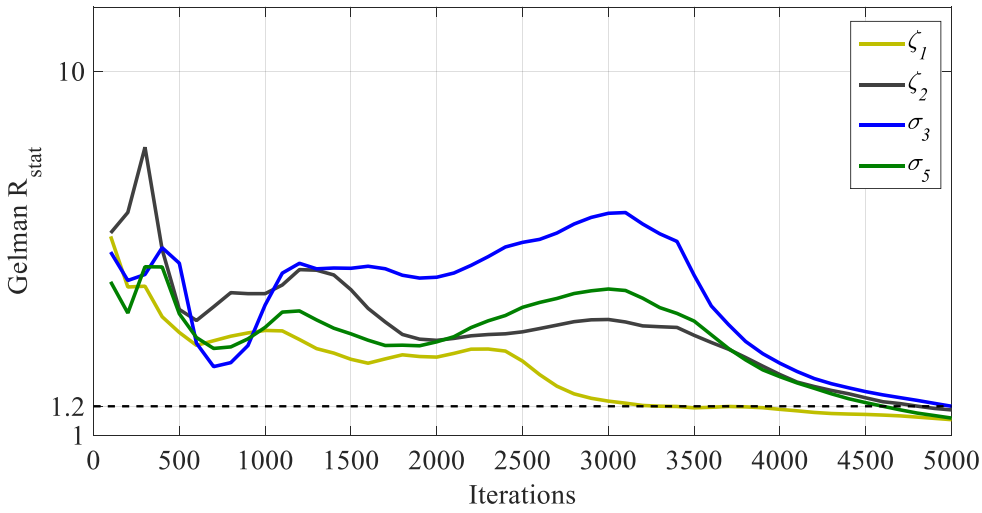
Figure 3.4 Identification progress for stiffness of 5th floor (partial output, 100% noise)



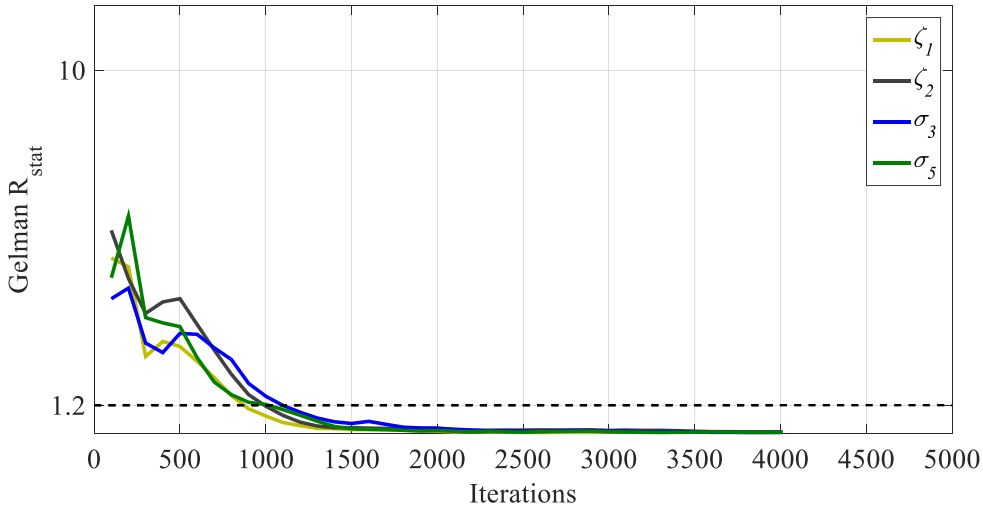
(a) DREAM



(b) IDREAM



(c) DREAM



(d) IDREAM

Figure 3.5 Convergence of Markov Chain (100% noise, partial output)

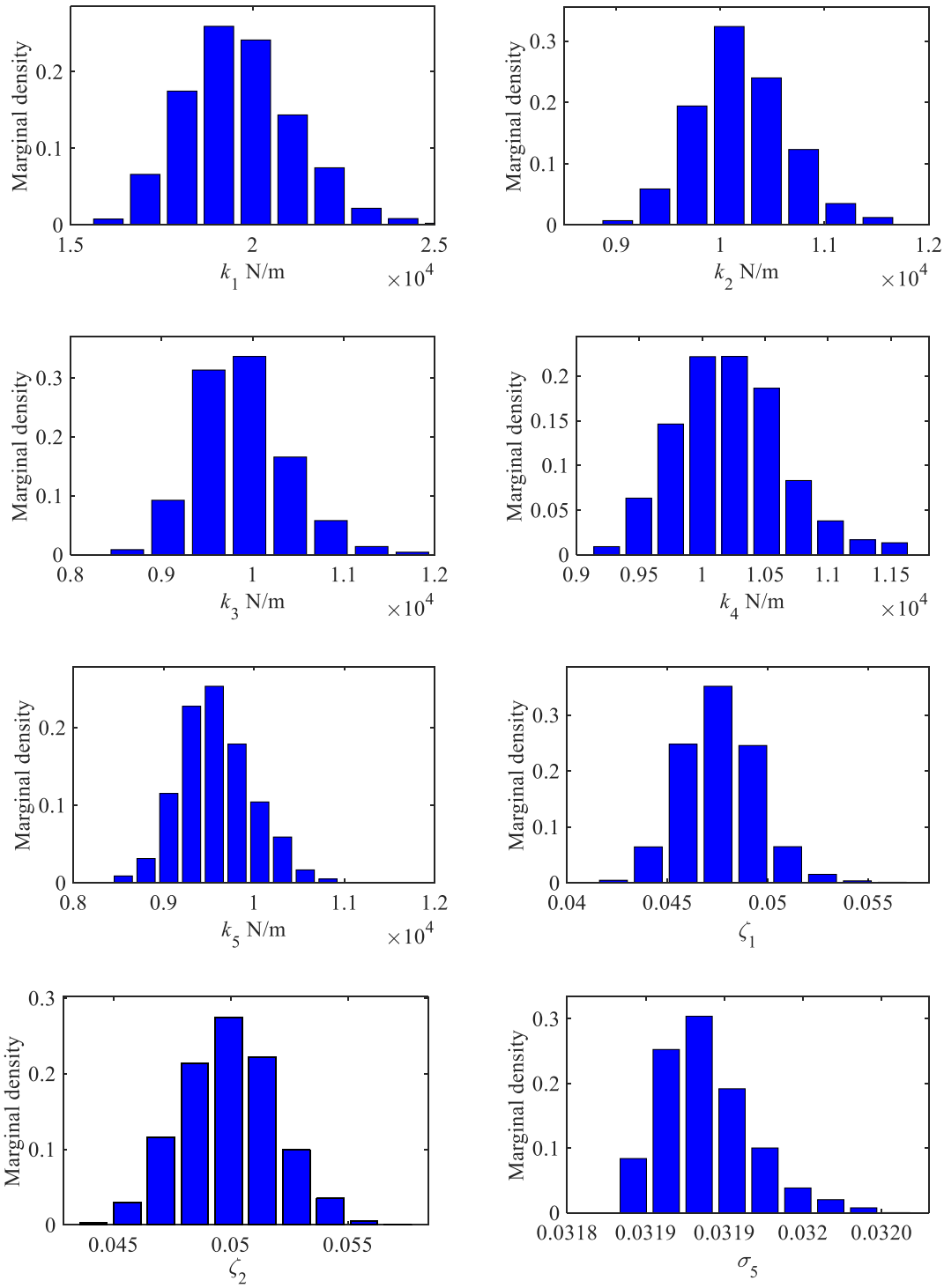
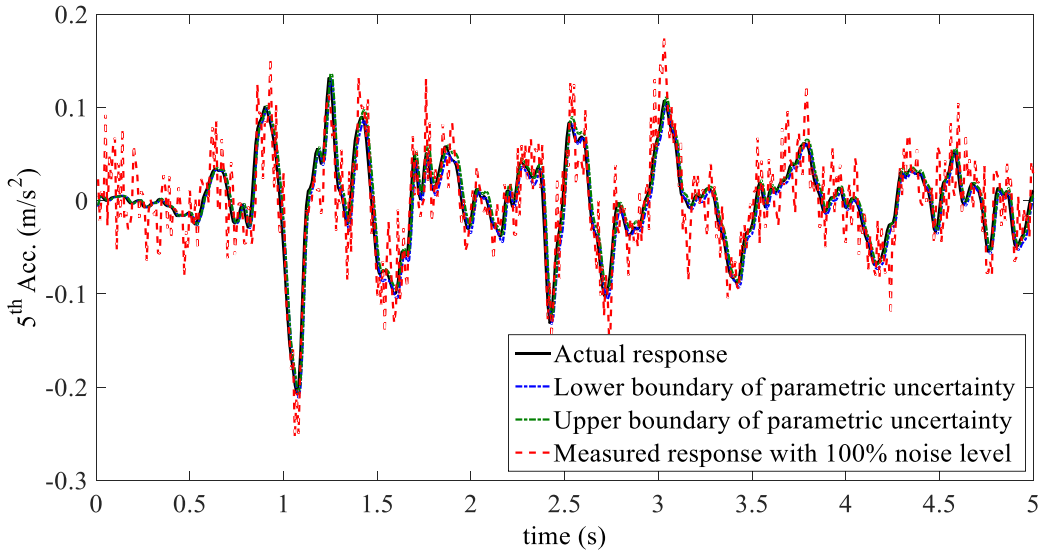
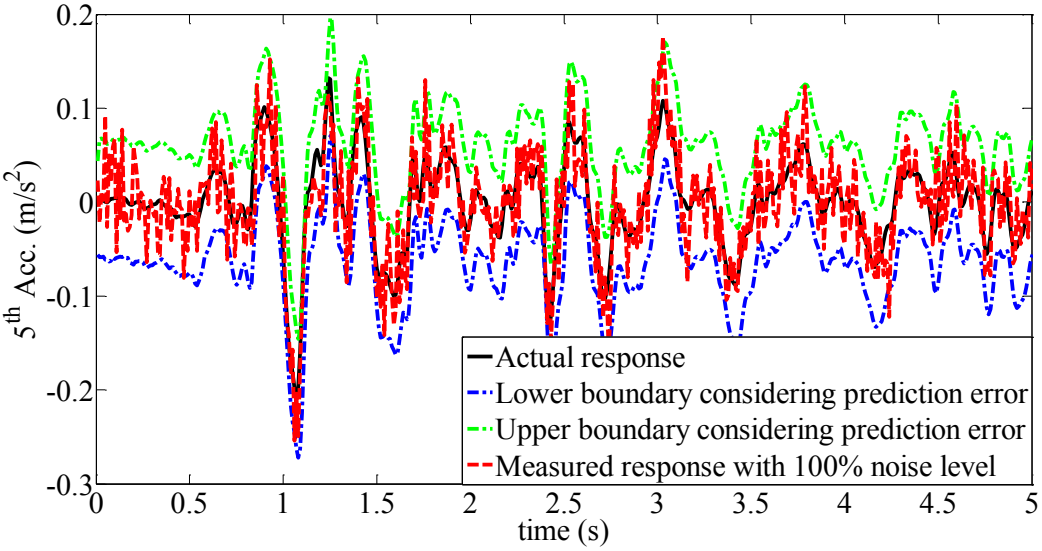


Figure 3.6 Histogram for each parameter (100% noise, partial output)



(a) Response range considering the identified parametric uncertainty



(b) Response range considering the identified prediction error

Figure 3.7 95% uncertainty ranges for acceleration of 5th DOF (100% noise, partial output)

Figure 3.4 shows the progress of identification of the stiffness of the 5th floor at the

100% noise level in the partial output scenario. In Fig. 3.4, each line with different color marks denotes the progress of a Markov chain, which means there are 20 posterior MC samples at each iteration. We can see that the posterior samples of the Markov chain obtained by the proposed IDREAM were more stable than those of DREAM, which leads to a smaller uncertain range.

Figures 3.5 indicates that the Markov chain converged for each identified parameter when using DREAM and IDREAM in the scenario of partial outputs and 100% noise. Comparing Fig. 3.5 (a) with Fig. 3.5 (b) and Fig. 3.5 (c) with Fig. 3.5 (d), however, makes it clear that IDREAM already converged by the time 4000 iterations were reached, but the Markov chains of DREAM were still unstable. Therefore, IDREAM converged faster than DREAM.

Combing the solutions in Tables 3.2~3.5, we can conclude that IDREAM outperformed DREAM because of its earlier convergence and robustness of the posterior samples without decreasing the accuracy of the results. The marginal posterior density of the parameters using IDREAM can be obtained by using kernel density estimation on the stationary Markov chain samples excluding the sequence during the burn-in period, as is shown in Fig. 3.6.The posterior uncertain range that assures a reliability of 95% can be obtained from the posterior samples of the model class which denotes the plausibility of each I/O system.

Figure 3.7 shows the ranges for part of time history (5 seconds). Figure 3.7 (a) is the uncertain response range of a stochastic I/O system parameterized by identified parameters of the posterior candidate model set with 95% assurance at each time interval considering only the parametric uncertainty. On the other hand, Fig. 3.7 (b) shows the uncertain range of the response with 95% assurance at each time interval considering the measurement error by incorporating the identified standard deviation of the prediction error. Figure 3.7(a) illustrates the effectiveness of the Bayesian

identification method because the parametric uncertain response range is close to the actual system response. Moreover, in Fig. 3.7(b), the percentage of the response considering a 100% measurement error within the uncertain range that considers prediction error is 94.75%.

3.3.2 2-DOFs of a Nonlinear Hysteretic System

In this section, a two degrees freedom (DOF) system of hysteretic Bouc-Wen model is considered to verify the effectiveness of the IDREAM based Bayesian inference, as is shown in Fig. 3.8. The dynamic equation for the nonlinear structural system can be written as

$$\boldsymbol{M}\ddot{\boldsymbol{y}}(t) + \boldsymbol{C}\dot{\boldsymbol{y}}(t) + \boldsymbol{f}(t) = \boldsymbol{u}(t), \tag{3.12}$$

where $\ddot{\boldsymbol{y}}(t)$, $\dot{\boldsymbol{y}}(t)$ and $\boldsymbol{y}$(t) are the acceleration, velocity and displacement; $\boldsymbol{u}(t)$ is the external excitation, and $\boldsymbol{f}(t) = (f_1 - f_2, f_2)^T$ is the restoring force; $\boldsymbol{M}$, $\boldsymbol{C}$ and are the mass, damping matrices, respectively. The restoring force of the nonlinear system can be described as

$$\dot{f}_1(t) = k_1\dot{y}_1(t) - \alpha_1|\dot{y}_1(t)||f_1(t)|^{n_1-1}f_1(t) - \beta_1\dot{y}_1(t)|f_1(t)|^{n_1}, \tag{3.13}$$

$$\begin{aligned}\dot{f}_2(t) \ = k_2\big(\dot{y}_2(t) - \dot{y}_1(t)\big) - \alpha_2|\dot{y}_2(t) - \ \dot{y}_1(t)||f_2(t)|^{n_2-1}f_2(t) \\ - \beta_2(\dot{y}_2(t) - \dot{y}_1(t))|f_2(t)|^{n_2},\end{aligned} \tag{3.14}$$

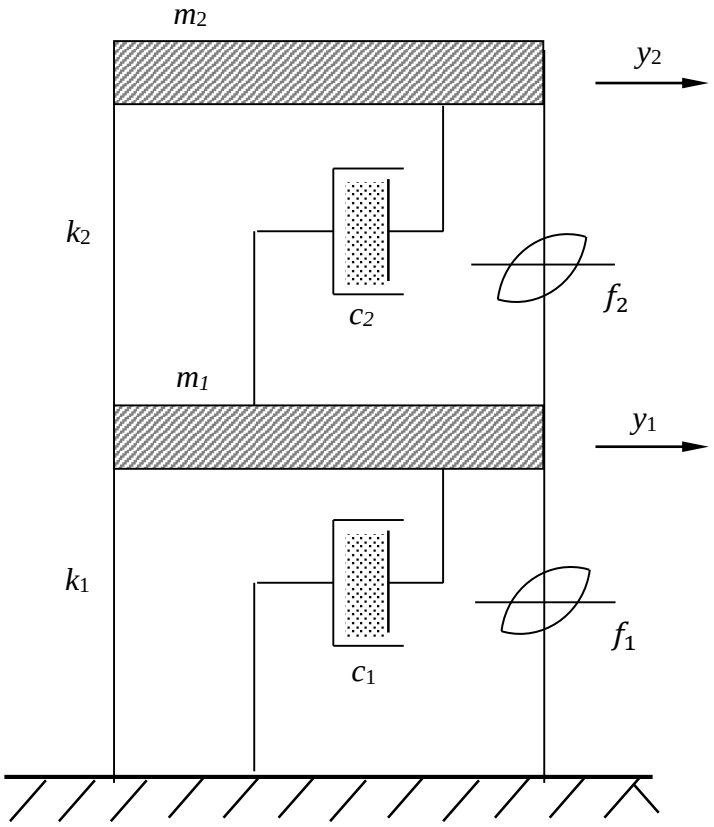


Figure 3.8 Bouc-Wen model

where α_i, β_i and n_i $(i = 1,2)$ are dimensionless quantities controlling the behavior of the hysteretic model. If the positive exponential parameter n_i is small, the transition from the elastic to the post-elastic branch is smooth, whereas for large n_i the transition becomes abrupt, approaching that of a bilinear model. Parameters α_i and β_i control the size and shape of the hysteretic loop. A simple 4th-order Runge-Kutta method is utilized to obtain the simulated response of the nonlinear dynamic equation of (3.12).

The properties of each story unit are: m_1 =1 kg, k_1 = 30 N/m, c_1 = 0.55 Ns/m, α_1 = 1, β_1 = 2, n_1 = 3, m_2 = 0.8 kg, k_2 = 24 N/m, c_2 = 0.5 Ns/m, α_2 = 2, β_2 = 1, n_2 = 2. The excitation is assumed to be a known force (Niigata earthquake, Oct. 23, 2004) as is shown in Fig.3.9 and the sample frequency is 100Hz (Xue et al., 2009). The response of the

structure, in terms of acceleration, is recorded at some given points. The nonlinear prosperity of the simulated system is presented in Fig. 3.10.

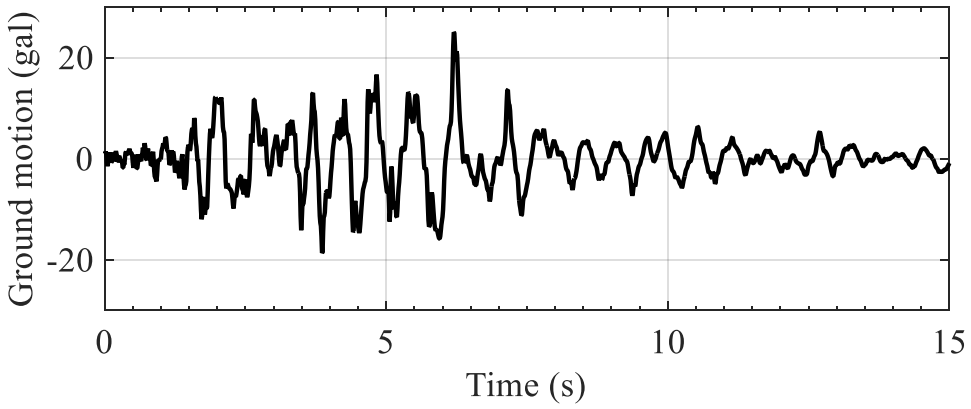


Figure 3.9 The Input excitation (Niigata 2004, NS)

The influence of the limited availability of measurements on the performance of the IDREAM for Bayesian inference is also discussed in this section. The following cases of available measurements will be treated as:

Case 1: A full set of accelerations is available.

$$\boldsymbol{Y}^{M}(t) = (\ddot{\boldsymbol{y}}_1(t), \ddot{\boldsymbol{y}}_2(t)). \tag{3.15}$$

Case 2: A partial set of accelerations is available.

$$\boldsymbol{Y}^{M}(t) = \ddot{\boldsymbol{y}}_2(t). \tag{3.16}$$

The mass of the structure is supposed to be known a priori. Therefore, the system is fully described by the set of stochastic variables:

$$\boldsymbol{\theta} = (k_1, k_2, c_1, c_2, \alpha_1, \alpha_2, \beta_1, \beta_2, n_1, n_2). \tag{3.17}$$

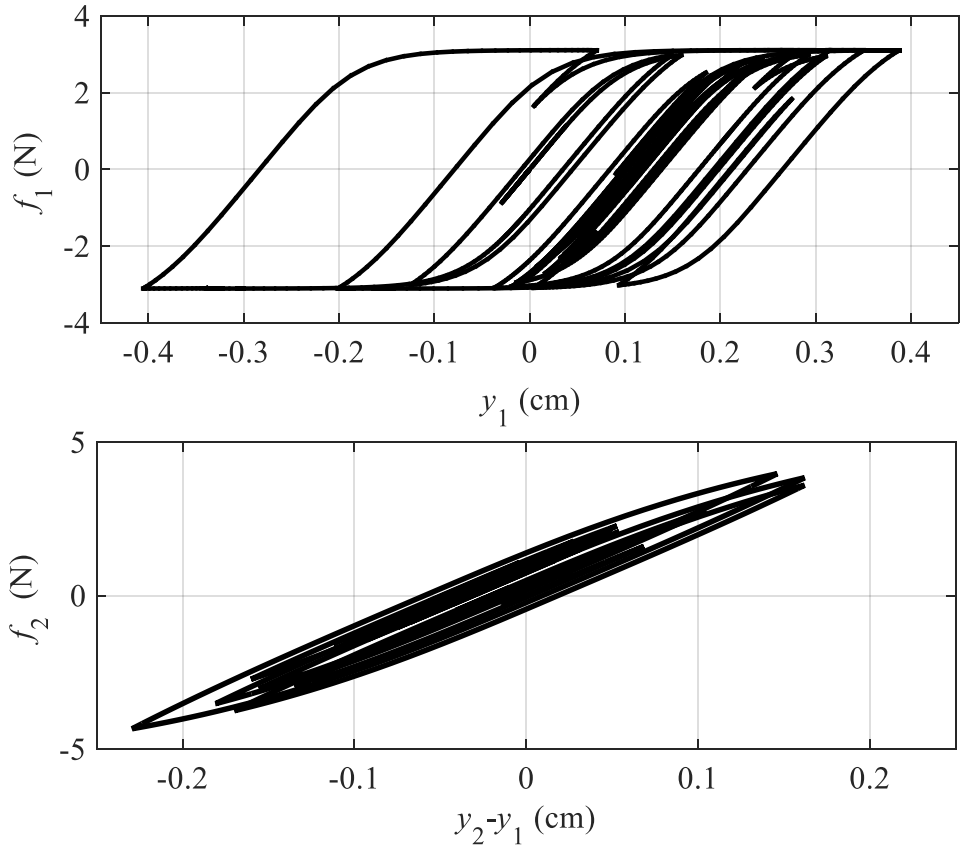


Figure 3.10 Nonlinear prosperity of the actual system

The output response is polluted (in the cases considering noise) with Gaussian, zero mean, white-noise sequences, whose root mean square (RMS) value is adjusted to a certain percentage of the unpolluted time histories.

Table 3.6 Identified results for structural parameters

		full outputs scenario			partial outputs scenario		
		no noise	5% noise	10% noise	no noise	5% noise	10% noise
k_1	error	0.000	0.254	0.371	0.000	0.302	0.785
	cov.	0.000	0.275	0.560	0.000	0.485	0.925
k_2	error	0.000	0.062	0.457	0.000	0.241	0.563
	cov.	0.000	0.180	0.385	0.000	0.265	0.558
c_1	error	0.000	0.909	1.397	0.000	1.789	4.610
	cov.	0.000	0.897	1.772	0.000	1.990	3.792
c_2	error	0.000	0.288	1.410	0.000	0.273	1.488
	cov.	0.000	1.062	2.209	0.000	1.787	3.279
α_1	error	0.000	0.433	1.965	0.000	1.075	2.846
	cov.	0.000	6.573	13.35	0.000	8.499	17.72
α_2	error	0.000	1.126	2.810	0.000	1.369	5.722
	cov.	0.000	7.656	15.06	0.000	11.15	17.89
β_1	error	0.000	0. 260	0.517	0.000	1.112	2.410
	cov.	0.000	6.622	13.49	0.000	9.248	18.97
β_2	error	0.000	3.187	4.192	0.000	3.899	7.568
	cov.	0.000	8.366	18.16	0.000	11.73	23.12
n_1	error	0.000	0.162	2.940	0.000	3.602	4.861
	cov.	0.000	1.883	3.918	0.000	2.541	5.471
n_2	error	0.000	0.295	1.435	0.000	1.050	3.314
	cov.	0.000	3.011	6.302	0.000	4.209	7.825

* the error in the table is in %; the cov. in the table is in %.

The parameters of the IDREAM algorithm were set as follows: the number of Markov chain samples (N_s) was 50, the crossover probability (C_R) was 0.85, and the number of

sample pairs (δ) was 5. The identified results for the clean signals and output signals corrupted with 5% and 10% RMS noise are summarized in Table 3.6 corresponding with full outputs scenario and partial outputs scenario.

Seen from Table 3.6, it can been concluded that the IDREAM algorithm can be successfully applied in the Bayesian inference for structural parameters. The structural parameters, such as stiffness and damping, can be relatively inferred by the MAP estimator. In no noise cases, the maximum relative error of the MAP estimator is close to zero both in full outputs and partial outputs scenarios.

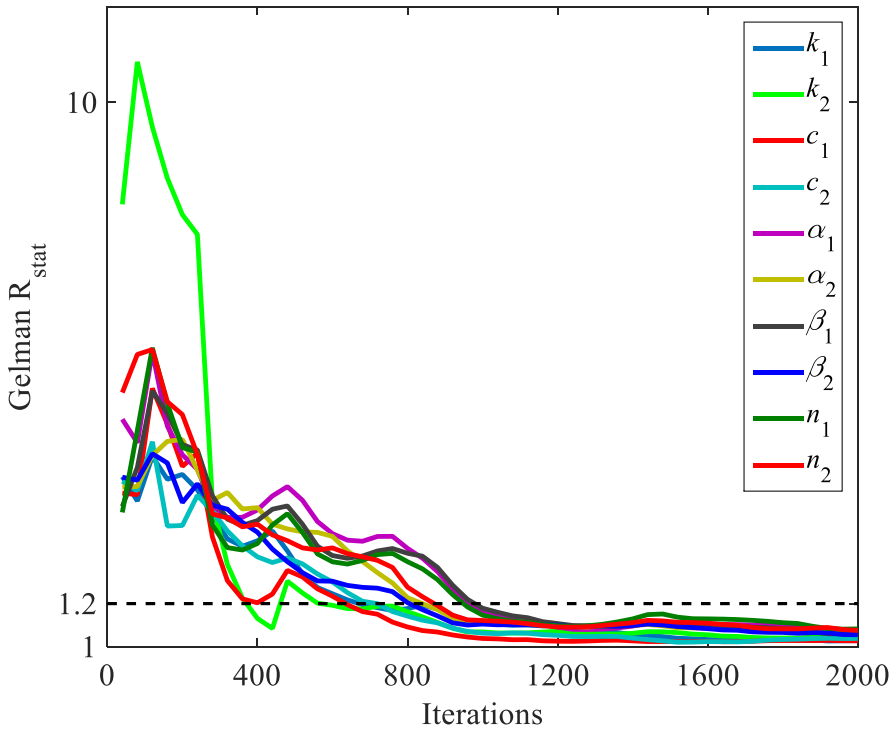


Figure 3.11 The Gelman convergence of the MC chains (10% noise, partial outputs)

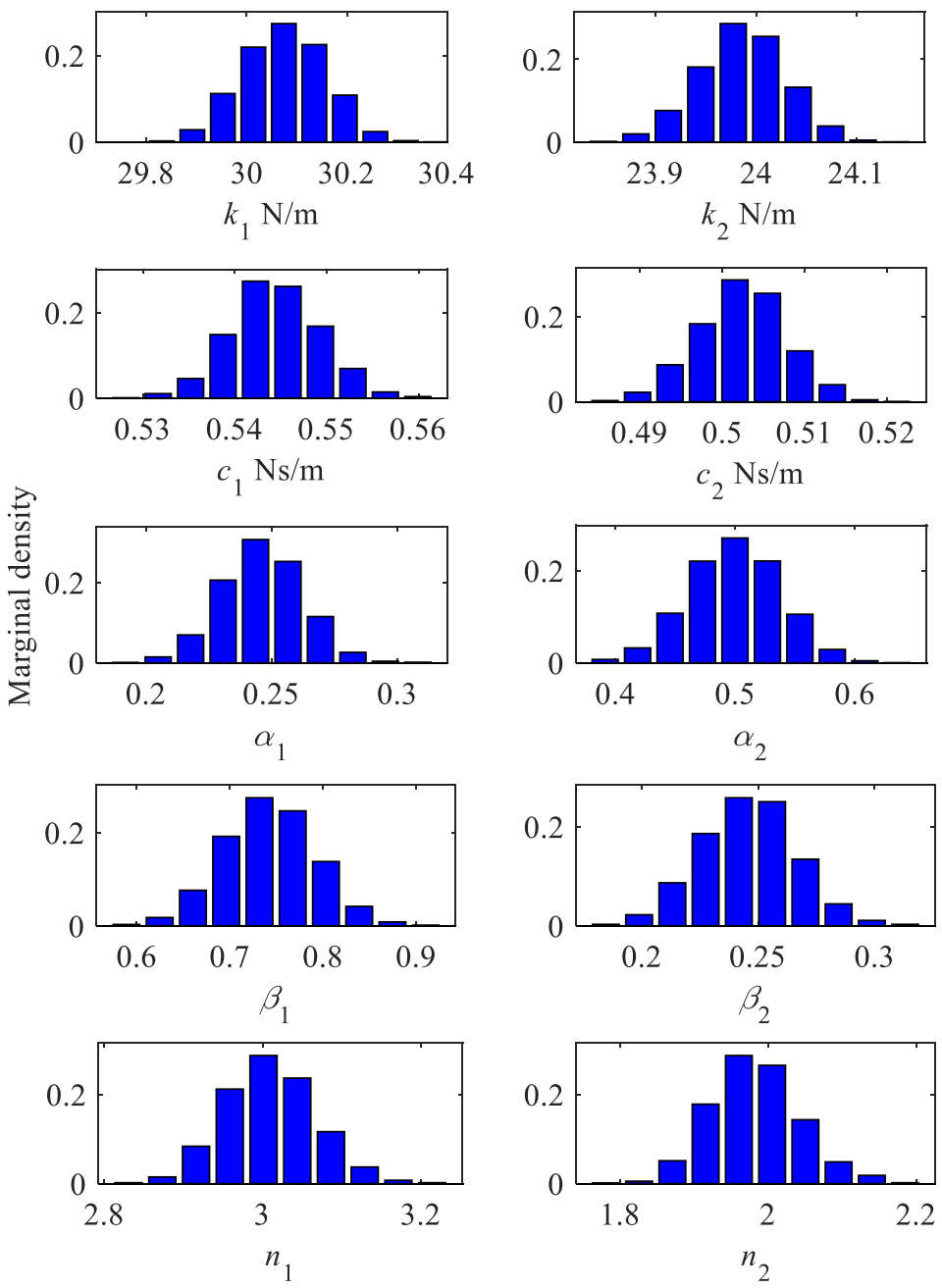


Figure 3.12 Histograms for the identified parameters (5% noise, full outputs)

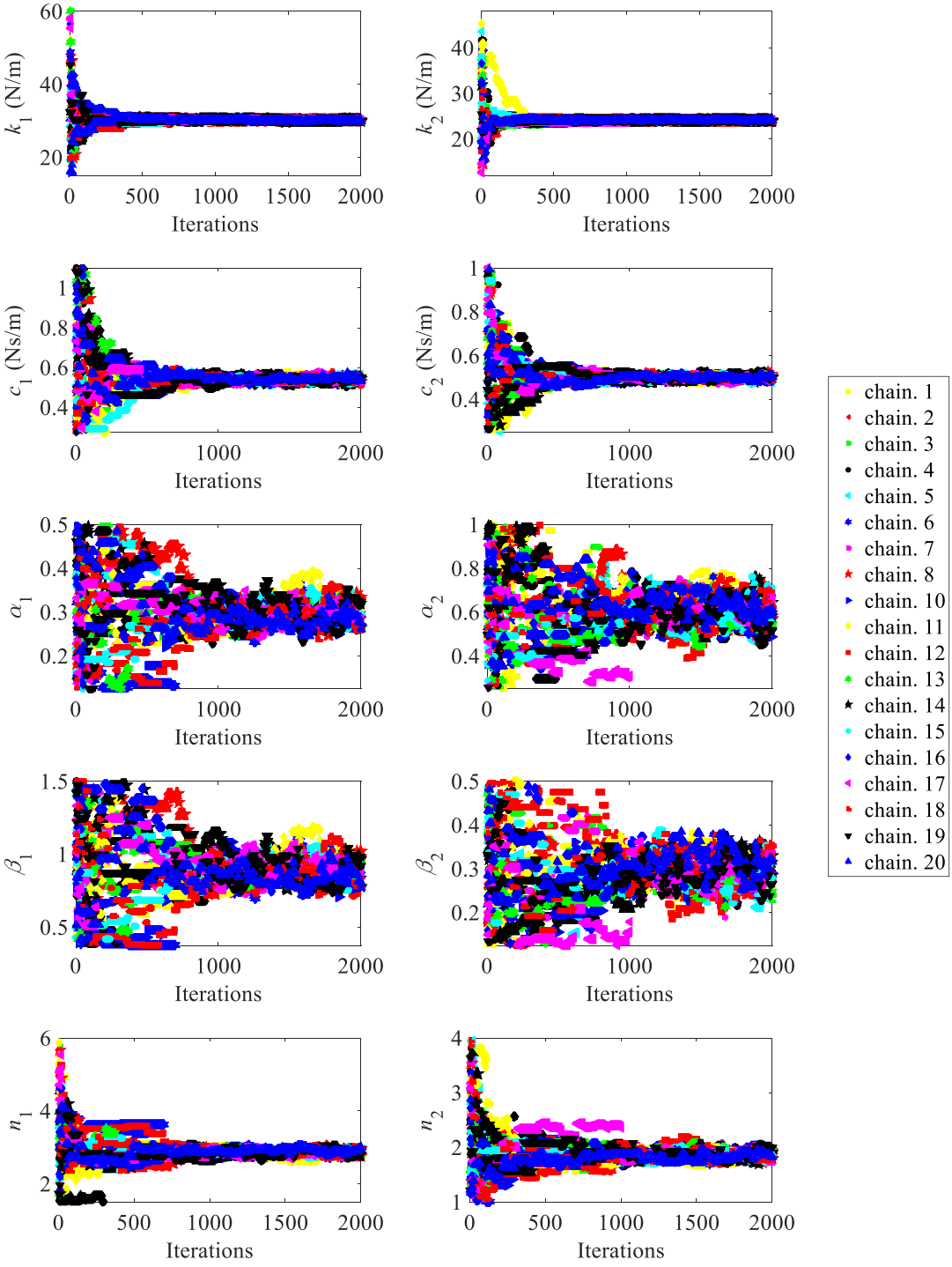


Figure 3.13 The convergence of individual parameters (10% noise, partial outputs)

The maximum relative error of the MAP estimator for structural parameters increased

to 1.062% in 5% noise case and to 2.209% in 10% noise case corresponding with the full outputs scenario; and the maximum relative error of the MAP for structural parameters added to 1.789% in 5% noise case and to 4.610% in the 10% noise case corresponding with the partial outputs scenario.

The relative error for the Bouc-Wen model parameters, such as $\alpha, \beta,$ and n, is relatively larger than the error of the identified MAP estimator for the structural parameters, such as the stiffness, k, and damping, c. The maximum relative error for the identified MAP estimator for model parameters is 3.187% and 3.899% in 5% noise case and adds to 4.192% and 7.568% in 10% noise case corresponding with full outputs and partial outputs scenario respectively. The Gelman convergence condition of the MC chains for each parameters is shown in Fig. 3.11. And the histogram for the identified stochastic parameters in 5% noise and full outputs case is presented in Fig.3.12. The convergence of the posterior samples for each parameters in the 10% noise and partial output scenario is shown in Fig.3.13.

3.4 Conclusions

A framework of Bayesian probability inference for identification based on an improved differential evolution adaptive Metropolis-Hasting (IDREAM) algorithm was proposed. Compared with the DREAM algorithm, its novelty lies in a new sample updating pattern that speeds convergence and improves the stability of the posterior samples. IDREAM runs the MCMC simulations in parallel and keeps the diversity of samples by using a DE crossover strategy. This gives it a strong ability to search for the global optimum and to resolve the problem that the MH algorithm has in choosing an approximate jump scale. A numerical simulation of a 5-DOF linear system and a

simulation of identification for a 2-DOF nonlinear system demonstrated its potential for solving identification problems with a high noise level and with partial outputs data. In conclusion, IDREAM is a new approach to obtaining the posterior density of a model class that cannot be easily found with the classic Monte-Carlo method due to the difficulty in calculating high-dimensional integrals.

CHAPTER 4

An Improved Bayesian Structural Identification Using the First-Two Derivative of Log-Likelihood Measure

4.1 Introduction

In this section, the purpose is to solve the "equifinality" problem of the uncertainty estimation for structural system (Beven, 2006). The main contribution of the study lies in the accuracy-improving of the estimator. A new objective function using the first two derivation of log-likelihood measure is proposed for the Bayesian inference.

The difficulty of Bayesian estimation lies in the efficiency in the convergence of posterior samples in the Markov chain to the acceptable model set. Because of the noise corrupted measurement, the surface of the prediction error lies in a hyper-surface of a multi-dimensional parametric space. It will cause the surface of the probability density for the posterior sequences to have multiple regions of attraction and numerous local optima. It thus inevitably yields a biased estimator (maximum a posteriori estimator, MAP). This problem is defined as the "equifinality" (Schulz et al., 1999). To detail it,

the surfaces of the prediction error using formal likelihood measures are studied.

To solve this problem, the bias between the MAP estimator and the true value are deduced by the Taylor expansion. It's found that the gradient and Hessian matrix of the likelihood measure can bridge the biased estimator and the true value, which is thus proposed to improve the accuracy of the posterior samples. The identification problem is therefore proposed as a two-step strategy. In the first step, the MAP estimator is obtained by the formal Bayesian likelihood measures using the differential evolution adaptive Metropolis-Hasting (DREAM) algorithm. The second step starts with the convergence of the Markov chains for each parametric dimension, where a new fitness function is proposed under the framework of the generalized likelihood uncertainty estimation (GLUE) (Stedinger et al. 2008; Freni et al., 2008). Numerical examples of a linear structural system are presented, with which the effectiveness and efficiency of the proposed method are investigated. Moreover, the proposed method is verified in the identification of a full-scale experiment named as "E-Defense" to show the potential of its application.

4.2 Problem of "Equifinality"

4.2.1 Least squares (LS) estimator (the deterministic inverse problem)

Let $\boldsymbol{Y}^{M}(t)$ denote the measured response at each time interval $(t = 1, \dots, N_t)$ and $\widehat{\boldsymbol{Y}}(\widehat{\boldsymbol{x}}, \mathrm{t})$ denotes the output of candidate models. The difference between the measured response and model outputs is defined as the residual error: $\boldsymbol{E}_j(\widehat{\boldsymbol{x}}, t) = \boldsymbol{Y}_j^{M}(t) - \widehat{\boldsymbol{Y}}_j(\widehat{\boldsymbol{x}}, t)$, where j=1,…, N_m, and N_m is the number of measured outputs. The common measure for the inverse problem is to attempt to force the residual vector as close to zero as possible by tuning the model parameter vector, $\widehat{\boldsymbol{x}}$. Thus the fitness measure can be defined:

$$LS(\hat{\boldsymbol{x}}) = -\frac{1}{N_m N_t}\sum_{j=1}^{N_m}\sum_{t=1}^{N_t}\left\|\boldsymbol{E}_j(\hat{\boldsymbol{x}},t)\right\|^2, \quad \hat{\boldsymbol{x}}^* = \text{argmax}(LS(\hat{\boldsymbol{x}})). \tag{4.1}$$

This is an N_d-dimensional optimization issue which maximizes the likelihood measure of SSR (equivalent to minimize the measure of Least-Squares (*LS*) formulation), N_d is the parametric dimension of the identification problem. But such measure can only provide an estimate of optimal value of $\hat{\boldsymbol{x}}^*$, which usually is a biased estimator when considering the measurement or model error.

If we need to quantify the uncertainty of the measurement error or the model error, it would be a desire to estimate the posterior PDF of parameter, $p(\boldsymbol{\theta}|\boldsymbol{Y}^M(t), \boldsymbol{\mathcal{M}})$, which is under the framework of Bayesian probabilistic estimation.

4.2.2 Bayesian estimate using formal log-likelihood (LL) measures

From Eqs. (2.1)~(2.3) in Section 2, it can be found that the Bayesian estimate mostly relies on the definition of the likelihood measure, $p(\boldsymbol{Y}^M(t)|\boldsymbol{\theta}, \boldsymbol{\mathcal{M}})$. It is more convenient to use the logarithm of the likelihood measures ($L(\boldsymbol{Y}^M(t)|\boldsymbol{\theta})$) as

$$L(\boldsymbol{Y}^M(t)|\boldsymbol{\theta}) = -\frac{N_t}{2}ln(2\pi) - \frac{N_t}{2}\sum_{j=1}^{N_m} ln\left(\sigma_j^2\right) - \frac{1}{2}\sum_{j=1}^{N_m}\sigma_j^{-2}\sum_{t=1}^{N_t}(\boldsymbol{Y}_j^M(t) - \widehat{\boldsymbol{Y}}_j(\boldsymbol{\theta},t))^2, \tag{4.2}$$

where $t = 1, \ldots, N_t$; and j=1,…, N_m, N_m is the number of measured response and N_t is the length of the time history; $\widehat{\boldsymbol{Y}}(\boldsymbol{\theta},t)$ denotes the stochastic response of the each stochastic model and $\boldsymbol{\theta}$ denotes a $N_s \times N_d$ matrix of the stochastic structural parameters that required to be identified. N_s denotes the stochastic samples or the

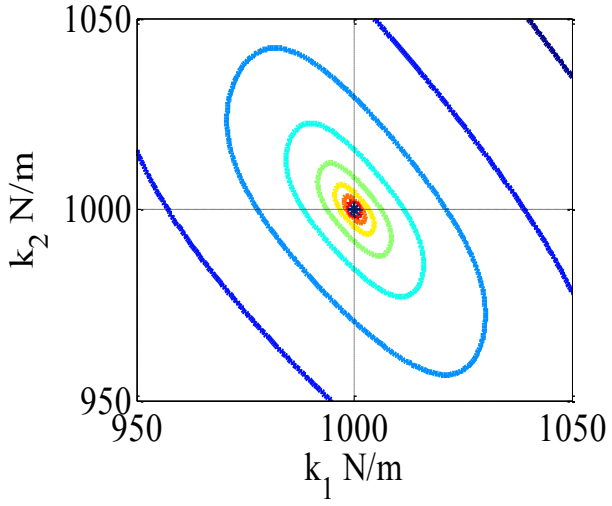
number parameterized probability models in the random model set, $\boldsymbol{\mathcal{M}}$.

In the framework of Bayesian identification, either the log-likelihood measure, as is in Eq.(4.2), or the least square measure, as is in Eq.(4.1), obeys the rule of "goodness-of-fit". This is because only the model with high probabilistic value of likelihood in the MH method will be accepted.

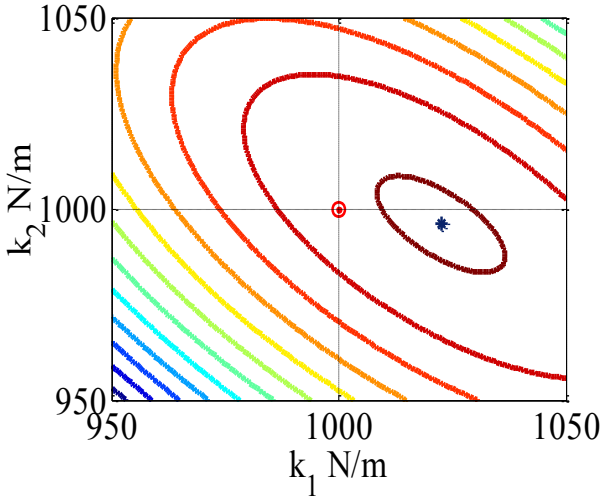
4.2.3 The surface of the likelihood measures

To illustrate the problem of "equifinality" in Bayesian inference, the surfaces of the common-used log-likelihood measure, as is in Eq. (4.2), are studied using an example of a 2-DOF linear dynamic system. The state space of the system is written as the Eq. (3.7). The system output is an acceleration which assumed to be contaminated by Gaussian white noise. The measured output vector is simulated as the Eq.(3.9). The mass, stiffness of each DOFs is defined as 100kg, and 1000N/m. Rayleigh damping *C* (Mita, 2003), where first two modal damping ratio (ζ_r) is set as 5% is considered for the simulation of linear example. The parametric domain is meshed by 5% deviation of the true value. The output acceleration (*acc.*) with different noise levels (Eq. (3.9)) was used, in which the noise level was denoted as nl. The standard deviation of the j^{th} measurement noise is equal to $\sigma_j = nl \times \sigma_{acc.,j}$ (where $\sigma_{acc.,j}$ denotes the standard deviation of the j^{th} relative acceleration, $j = 1,2$). The Contour-plot of the likelihood measure, as is in Eq. (4.2), in the scenarios of noise-free and different noise-level scenarios are exhibited in Fig 4.1.

a. (no noise):



b. (30% noise):



c. (100% noise):

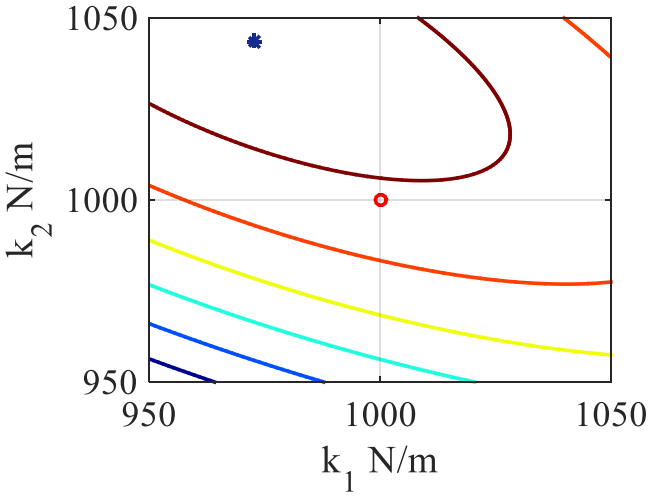


Figure 4.1 Contour plot of the likelihood measures
("∘" denotes the true value; "∗" denotes the MAP estimator)

From Fig.4.1, it can be found that only under the ideal scenario that the measurement error is ignored the MAP estimator will be unbiased (the blue marked "$*$" coincides with the red mark "$\circ$"); however, if taking the noise into account, the sample with maximum posterior PDF deviates from the true value. And the deviation enlarges with the increasing of the noise level. The reason to illustrate this phenomenon is that around the neighborhoods of the optimal solution (the MAP estimator), there are many local optimums, which means there are several possible models that can also give high values of likelihood (high probability) around the neighborhood of the MAP estimator. This will cause the "equifinality" problem (Zak and Beven, 1999) of the Bayesian inference method using the formal likelihood measure, in which it may underestimate or overestimate the uncertain intervals of the posterior samples.

4.3 The proposed accuracy-improving method

4.3.1 The first-two deviation of the likelihood measure

It can conclude that when considering the measurement error the common likelihood measures as in Eqs. (4.1) and (4.2) are weak to solve the problem of "equifinality". The bias of the estimator will increase with the adding number the parametric dimensions and the noise level. It is thus necessary to improve the identified MAP estimator. With Taylor's expansion, the formal likelihood measure can be deduced as:

$$L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t)) = L(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) + \boldsymbol{L'}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t))\Delta\boldsymbol{\theta^T} + \frac{\mathbf{1}}{\mathbf{2}}\Delta\boldsymbol{\theta}\boldsymbol{L''}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t))\Delta\boldsymbol{\theta^T} + \boldsymbol{o}((\Delta\boldsymbol{\theta})^{\mathbf{3}}), \quad (4.3)$$

where the $\Delta\boldsymbol{\theta}$ denotes $\boldsymbol{\theta}-\boldsymbol{\theta}_o$, and $\boldsymbol{\theta}_o$ denotes the true value of the structural parameter. The derivative of Eq.(4.3) with respect to the $\boldsymbol{\theta}$ obtains:

$$\boldsymbol{L}'(\boldsymbol{\theta}|\boldsymbol{Y}^M(t)) = \boldsymbol{L}'(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) + \Delta\boldsymbol{\theta}\boldsymbol{L}''(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) + \boldsymbol{o}((\Delta\boldsymbol{\theta})^{\boldsymbol{2}}). \tag{4.4}$$

Let $\boldsymbol{G}(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$ denote the first order derivative of likelihood measure, $\boldsymbol{L}'(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$, which is the gradient vector of the $L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$ and $\boldsymbol{H}(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$ denote the second order derivative of likelihood measure, $\boldsymbol{L}''(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$, which is the Hessian matrix of the $L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$. The $\boldsymbol{G}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t))$ and the $\boldsymbol{H}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t))$ is the gradient and Hessian matrix at the point of true value, $\boldsymbol{G}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) = \frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta_o}}$ and $\boldsymbol{H}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) = \frac{\partial}{\partial\boldsymbol{\theta}}(\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\boldsymbol{\theta}})\Big|_{\boldsymbol{\theta_o}}$.

Multiplying $\boldsymbol{H}^{-1}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t))$ on the right side of the Eq.(4.4) and ignoring the high order derivative series, then it can be written as:

$$\boldsymbol{G}(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))\boldsymbol{H}^{-1}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) = \boldsymbol{G}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t))\boldsymbol{H}^{-1}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) + \boldsymbol{\theta}-\boldsymbol{\theta}_o + \boldsymbol{o}((\Delta\boldsymbol{\theta})^{\boldsymbol{2}}). \tag{4.5}$$

After the convergence of the Markov chain, the posterior sample with maximum PDF is obtained and $\boldsymbol{G}(\boldsymbol{\theta}_{\mathrm{MAP}}|\boldsymbol{Y}^M(t)) = 0$, then yields:

$$\boldsymbol{\theta}_o - \boldsymbol{\theta}_{\mathrm{MAP}} - \boldsymbol{G}(\boldsymbol{\theta}_o|\boldsymbol{Y}^M(t))\boldsymbol{H}^{-1}(\boldsymbol{\theta_o}|\boldsymbol{Y}^M(t)) = \boldsymbol{o}((\Delta\boldsymbol{\theta})^{\boldsymbol{2}}). \tag{4.6}$$

From Eq.(4.6), it's clear that the bias between the MAP estimator, $\boldsymbol{\theta}_{\mathrm{MAP}}$, and the true

value, $\boldsymbol{\theta}_o$, is equal to the negative product of the gradient and the inverse Hessian matrix at the true value. In this study, the Eq.(4.6) is used to formulate a new fitness when the MAP estimator is obtained. The identification procedure is therefore divided into two steps, where the $\boldsymbol{\theta}_{\mathrm{MAP}}$ will be obtained as the posterior sample with maximum posterior PDF in the first step and the second step is to search for the $\widehat{\boldsymbol{\theta}}_o$ in the posterior samples. The proposed objective function can be written as:

$$L^*(\boldsymbol{\theta}|\boldsymbol{Y}^M(t)) = -\|\boldsymbol{\theta} - \boldsymbol{\theta}_{\mathrm{MAP}} - \boldsymbol{G}(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))\boldsymbol{H}^{-1}(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))\| \qquad (4.7)$$

The proposed method is to search for one/two peaks corresponding with different cases: Case 1: $L^*(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$ owns one peak, which corresponds with the no noise scenario. In such case, the MAP estimator, $\boldsymbol{\theta}_{\mathrm{MAP}}$, coincides with the estimator of true value, $\widehat{\boldsymbol{\theta}}_o$. Case 2: $L^*(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$ owns two peaks corresponding with the scenario that considering the white noise, then $\widehat{\boldsymbol{\theta}}_o$ will be the second peaks of the fitness function, which is more accurate thank the MAP estimator (where the optimal point of the proposed function, $\widehat{\boldsymbol{\theta}}_o$, meets $\widehat{\boldsymbol{\theta}}_o - \boldsymbol{\theta}_{\mathrm{MAP}} - \boldsymbol{G}\big(\widehat{\boldsymbol{\theta}}_o|\boldsymbol{Y}^M(t)\big)\boldsymbol{H}^{-1}\big(\widehat{\boldsymbol{\theta}}_o|\boldsymbol{Y}^M(t)\big) \cong \mathbf{0}$).

4.3.2 Illustration of the proposed fitness function

Since the Taylor equation, as in Eq. (4.3) can also be expanded at other parametric value, $\boldsymbol{\theta}_*$, which is different from the true value, $\boldsymbol{\theta}_o$. Similarly, when $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_*$, with Taylor expansion we have:

$$L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t)) = L(\boldsymbol{\theta}_*|\boldsymbol{Y}^M(t)) + \boldsymbol{L}'(\boldsymbol{\theta}_*|\boldsymbol{Y}^M(t))\Delta\boldsymbol{\theta}^{\boldsymbol{T}} + \frac{\mathbf{1}}{\mathbf{2}}\Delta\boldsymbol{\theta}\boldsymbol{L}''(\boldsymbol{\theta}_*|\boldsymbol{Y}^M(t))\Delta\boldsymbol{\theta}^{\boldsymbol{T}} + \boldsymbol{o}((\Delta\boldsymbol{\theta})^{\mathbf{3}}), \qquad (4.8)$$

Then with the derivation of this equation with respect to the $\boldsymbol{\theta}$, it can draw the equation as: $\boldsymbol{\theta}_{\mathrm{MAP}} = \boldsymbol{\theta}_* - \boldsymbol{G}(\boldsymbol{\theta}_*|\boldsymbol{Y}^M(t))\boldsymbol{H}^{-1}(\boldsymbol{\theta}_*|\boldsymbol{Y}^M(t)) + \boldsymbol{o}((\Delta\boldsymbol{\theta})^{\boldsymbol{2}})$. It is thus necessary to discuss whether the optimal point of the proposed function, $\widehat{\boldsymbol{\theta}}_o$, is near the true value or not when the proposed fitness, $L^*(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$, approaches to its maximum value. Seen from Eqs.(4.4) and (4.7), making the function $L^*(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$ approach to the maximum value can be able to regarded as yielding the item $-\|(\boldsymbol{\theta}_{\mathrm{MAP}} - \boldsymbol{\theta})^2\boldsymbol{L}'''(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))\boldsymbol{H}^{-1}(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))\|$ reach to the extreme value, in which the $\boldsymbol{L}'''(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$ is a N_d −dimensional third order tensor, and the inverse hessian matrix, $\boldsymbol{H}^{-1}(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$, can be viewed as a N_d −dimensional second order tensor.

To illustrate the problem simply, let us consider 1-dimensional problem. And let $f(\theta)$ denote the item $\theta - \theta_{\mathrm{MAP}} - G(\theta|Y^M(t))H^{-1}(\theta|Y^M(t))$ in Eq.(4.7), where the θ, $G(\theta|Y^M(t))$ and $H(\theta|Y^M(t))$ respectively is a scalar. Then the derivative of $f(\theta)$ with respect to the θ is:

$$f'(\theta) = \frac{G(\theta|Y^M(t))L'''(\theta|Y^M(t))}{H(\theta|Y^M(t))^2}. \tag{4.9}$$

From the Eq.(4.9), it's clear that the optimal solution of the fitness function in Eq.(4.7) is the sample who meets the equation $G(\theta|Y^M(t)) = 0$ and $L'''(\theta|Y^M(t)) = 0$. The first equation is satisfied by the MAP estimator, where $G(\theta_{\mathrm{MAP}}|Y^M(t)) = 0$. The proposed method is able to find the true value as the optimal point if the condition $L'''\left(\hat{\theta}_o|Y^M(t)\right) = 0$ is satisfied. Under the assumption of the white noise is added as the measurement error, the proposition that the third derivative of the likelihood measure equal to zero at the true value, $\hat{\theta}_o$, as much as to say, the extreme of the second derivative of a normal distribution lies at its mean value (taking the standard normal distribution, $f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mu = 0, \sigma = 1$, as an example, the second and third derivative of the normal distribution are shown in Fig. 4.2).

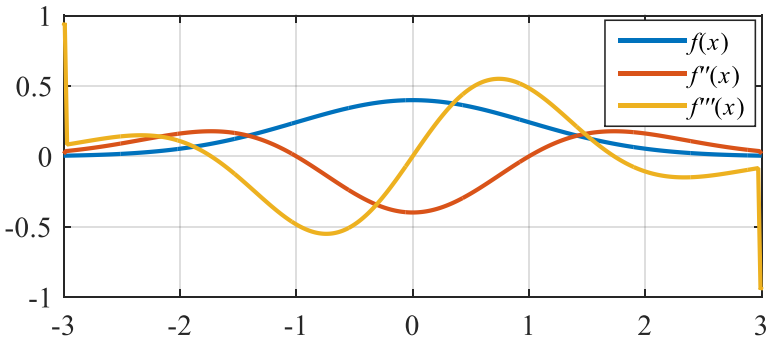
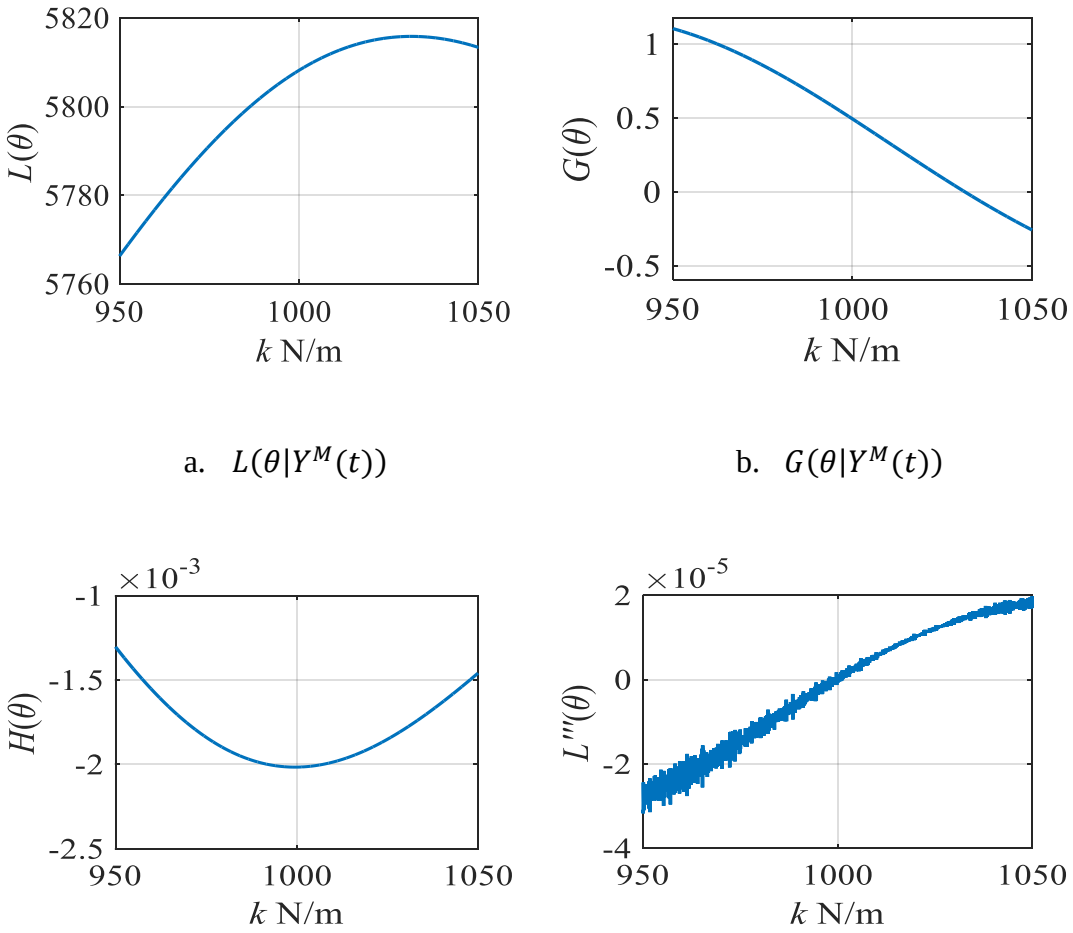


Figure 4.2 Second and third derivative of a standard normal distribution

A 1-DOF structural system with the stiffness equal to 1000 N/m is considered. The input and measurement noise is simulated same as the example in Section 4.2.3. The likelihood measure, $L(\theta|Y^M(t))$, the gradient of the likelihood measure,$G(\theta|Y^M(t))$, the hessian of the likelihood measure, $H(\theta|Y^M(t))$ and the third derivative of the likelihood measure, $L'''(\theta|Y^M(t))$ are respectively shown in Fig. 4.3.

Seen from the Figs. 4.3 (c) and (d), it can be easily found that the second derivative of the likelihood measure reaches its minimum value near the true parametric value, which yields the condition, $L'''\left(\hat{\theta}_o|Y^M(t)\right) = 0$ in the simulation of the 1-DOF structural system. To clear show the performance of the likelihood measure under the scenario that ignoring the measurement error the simulation of 1-DOF structural system in no noise scenario is compared and shown in Fig.4.4. In the simulation of no noise scenario, the proposed fitness function, $L^*(\theta|Y^M(t))$, owns only one extreme at the true value, as is shown in Fig. 4.5 (a); and in the scenario considering 100% noise level, it owns two peaks, as is shown in Fig. 4.5 (b), by excluding the ever-obtained MAP estimator, the second extreme value of the proposed function will be the more accurate estimator of the true value.



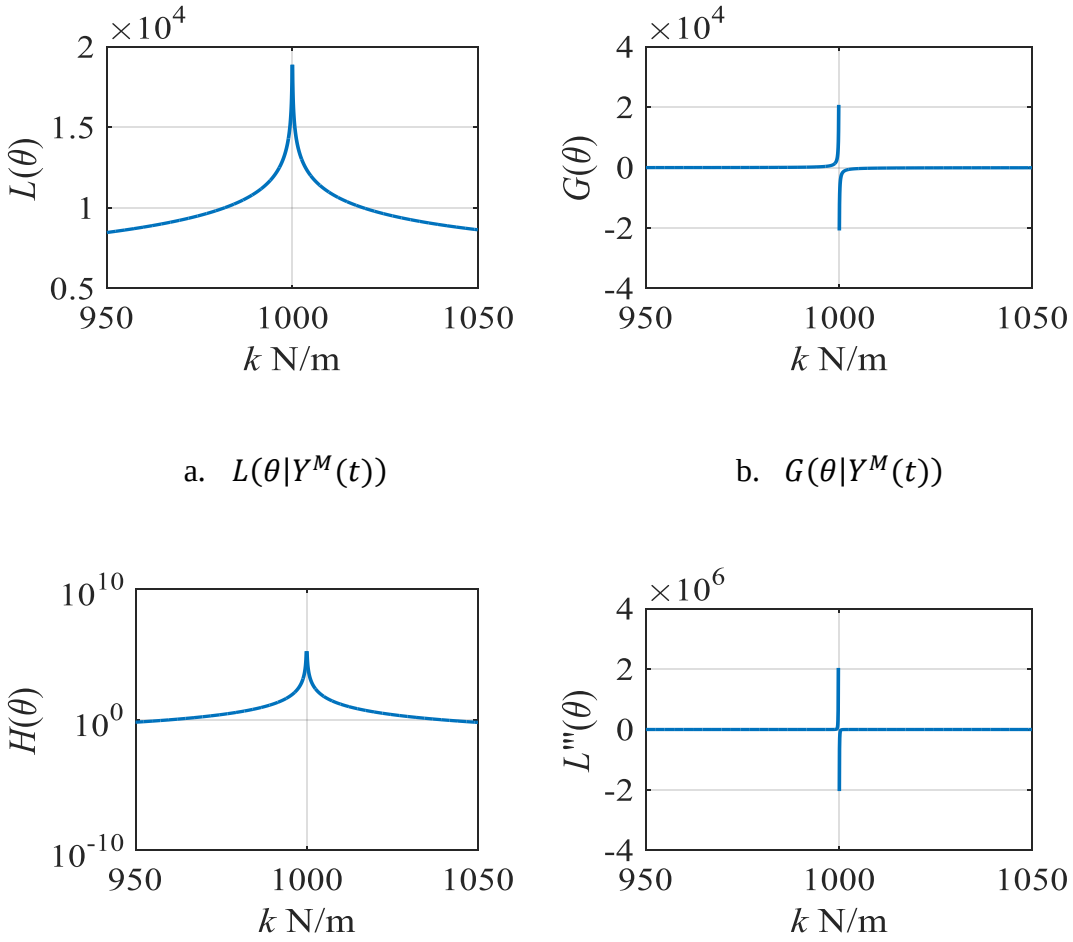
a. $L(\theta|Y^M(t))$

b. $G(\theta|Y^M(t))$

c. $H(\theta|Y^M(t))$

d. $L'''(\theta|Y^M(t))$

Figure 4.3 1-DOF simulation (100% noise scenario)



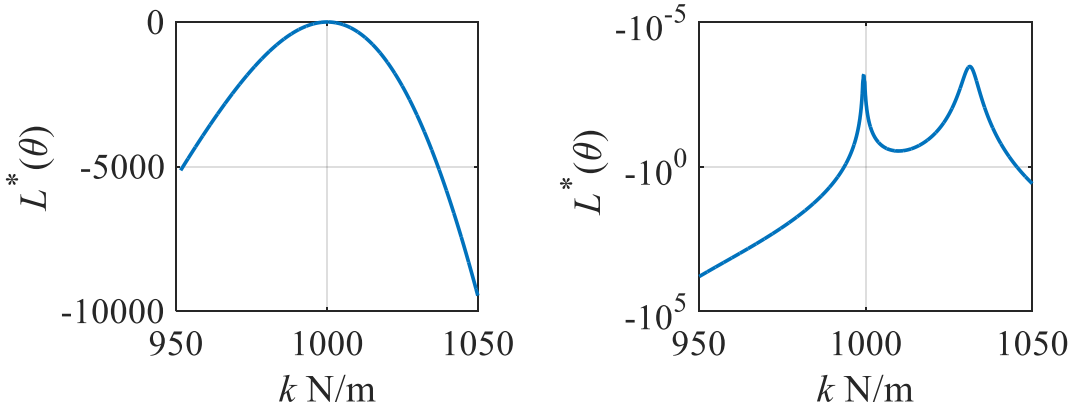
a. $L(\theta|Y^M(t))$

b. $G(\theta|Y^M(t))$

c. $H(\theta|Y^M(t))$

d. $L'''(\theta|Y^M(t))$

Figure 4.4 1-DOF simulation (no noise scenario)



No noise scenario 100% noise scenario

Figure 4.5 Proposed fitness function (1-dimensional problem)

To verify the proposed idea, the simulation of a 2-DOF linear system, of which the simulation in 100% noise scenario same as that in Section 4.2.3, is also considered. The surface and the contour plot of the proposed fitness function is shown in Fig. 4.6. Comparing with the Fig. 4.1(c), we can find that the sample with maximum PDF in the Fig. 4.6 is closer to the true value than the MAP estimator in the Fig. 4.1(c), which means that the accuracy of the estimator can be improved. Moreover, in Fig. 4.6 it can be found that most of the posterior samples concentrate mainly on the neighborhood of the true value, which will increase the stability of the posterior samples on Markov chains.

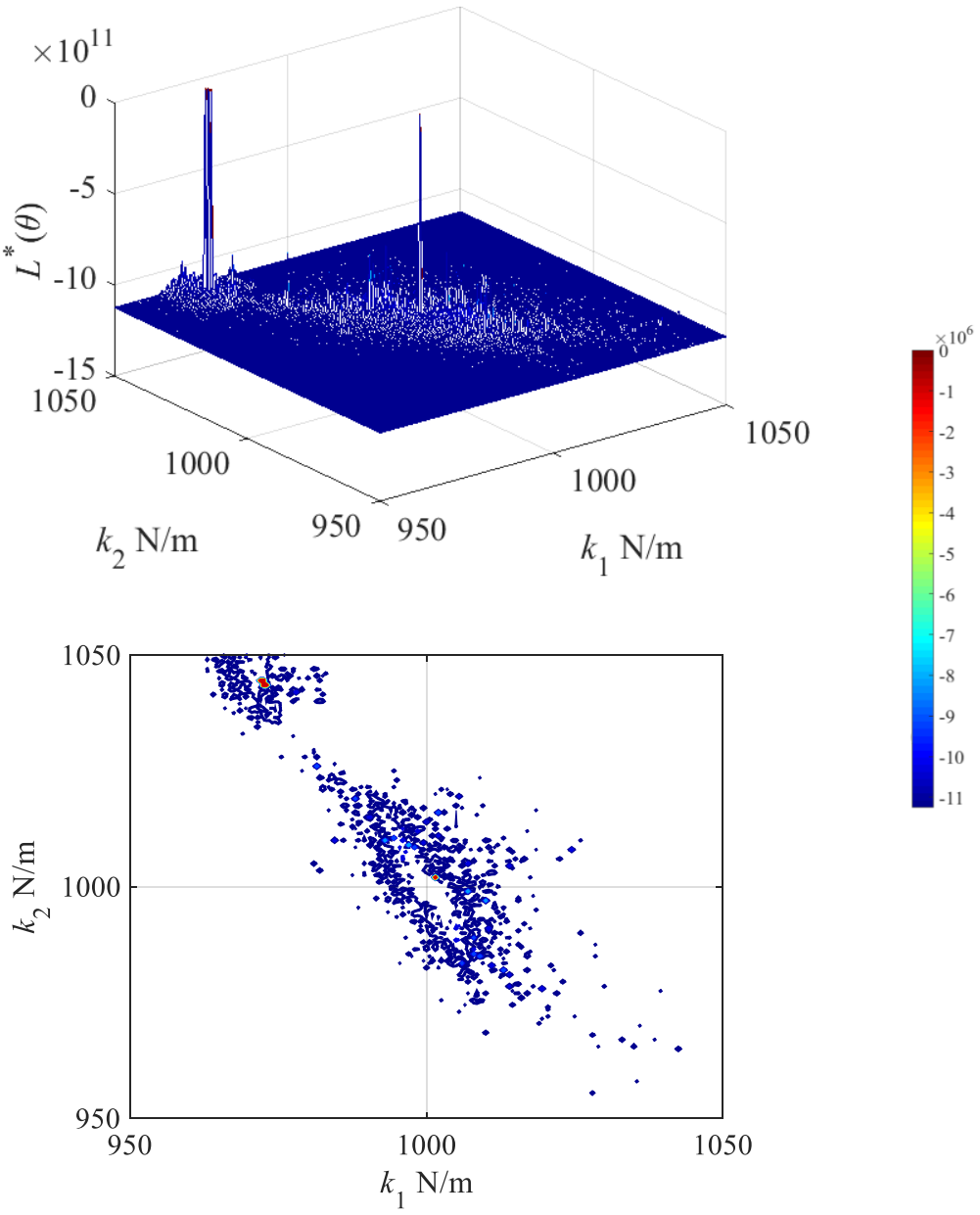


Figure 4.6 Optimal-solutions set for 100% noise scenario

(the red point denotes the estimator of $\widehat{\boldsymbol{\theta}}_o$)

4.3.3 Two-step of the IDREAM based Bayesian estimation

Step 1: MAP estimator using IDREAM algorithm

The MAP estimator, $\boldsymbol{\theta}_{\mathrm{MAP}}$, is obtained with the maximum posterior PDF. Also the standard deviation of the posterior samples, $\boldsymbol{\sigma}(\boldsymbol{\theta})$, can be obtained. The first step of Bayesian inference using IDREAM algorithm is same as that mentioned in Section 3. Then the algorithm forwards to the second step.

Step 2: Density updating of the samples that satisfy the proposed criteria

Calculating the gradient and Hessian matrices of the posterior samples which are located in the boundary of $[\boldsymbol{\theta}_{\mathrm{MAP}} - 3\boldsymbol{\sigma}(\boldsymbol{\theta}), \boldsymbol{\theta}_{\mathrm{MAP}} + 3\boldsymbol{\sigma}(\boldsymbol{\theta})]$ at each iteration of Step 2. The proposed fitness function as in Eq.(4.7) will be used for the updating of the posterior samples. The estimation procedure will be stopped till the prescribed maximum iteration is satisfied. The gradient and the Hessian matrices for the k^{th} posterior sample can be obtained as:

$$\boldsymbol{G}(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t)) = \left[\frac{\partial L}{\theta_1}, \frac{\partial L}{\theta_2}, \ldots, \frac{\partial L}{\theta_{N_d}}\right],\ \boldsymbol{H}(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t)) = \begin{bmatrix} \frac{\partial^2 L}{\partial \theta_1^2} & \frac{\partial^2 L}{\partial \theta_1 \theta_2} & \cdots & \frac{\partial^2 L}{\partial \theta_1 \theta_{N_d}} \\ \frac{\partial^2 L}{\partial \theta_2 \theta_1} & \frac{\partial^2 L}{\partial \theta_2^2} & \cdots & \frac{\partial^2 L}{\partial \theta_2 \theta_{N_d}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 L}{\partial \theta_{N_d} \theta_1} & \frac{\partial^2 L}{\partial \theta_{N_d} \theta_2} & \cdots & \frac{\partial^2 L}{\partial \theta_{N_d}^2} \end{bmatrix}. \tag{4.10}$$

The diagonal and off-diagonal elements of the $\boldsymbol{H}(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t))$ are obtained as:

$$
\begin{aligned}
H^{l,l}(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t)) &= \left[\frac{\partial}{\partial\theta_l}\left(\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\theta_l}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_k} \\
&\approx \frac{1}{\Delta\boldsymbol{\theta}_l}\left[\left.\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_l/2} - \left.\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_l/2}\right] \\
&\approx \frac{1}{\Delta\boldsymbol{\theta}_l}\left[\frac{L(\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t)) - L(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t))}{\Delta\boldsymbol{\theta}_l}\right. \\
&\qquad \left. - \frac{L(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t)) - L(\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_l|\boldsymbol{Y}^M(t))}{\Delta\boldsymbol{\theta}_l}\right] \\
&= \frac{L(\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_l|\boldsymbol{Y}^M(t)) - 2\times L(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t)) + L(\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_l|\boldsymbol{Y}^M(t))}{(\Delta\boldsymbol{\theta}_l)^2}
\end{aligned}
\tag{4.11}
$$

$$
\begin{aligned}
H^{l,l'}(\boldsymbol{\theta}_k|\boldsymbol{Y}^M(t)) &= \left[\frac{\partial}{\partial\theta_{l'}}\left(\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\theta_l}\right)\right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_k} \\
&\approx \frac{1}{2\times\Delta\boldsymbol{\theta}_{l'}}\left[\left.\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_{l'}} - \left.\frac{\partial L(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))}{\partial\theta_l}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_{l'}}\right] \\
&\approx \frac{1}{2\times\Delta\boldsymbol{\theta}_{l'}}\left[\frac{L(\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t)) - L(\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t))}{2\times\Delta\boldsymbol{\theta}_l}\right. \\
&\qquad \left. - \frac{L(\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t)) - L(\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t))}{2\times\Delta\boldsymbol{\theta}_l}\right] \\
&= \frac{1}{4\Delta\boldsymbol{\theta}_l\Delta\boldsymbol{\theta}_{l'}}[L(\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t)) \\
&\qquad - L(\boldsymbol{\theta}_k+\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t)) - L(\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_l+\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t)) \\
&\qquad + L(\boldsymbol{\theta}_k-\Delta\boldsymbol{\theta}_l-\Delta\boldsymbol{\theta}_{l'}|\boldsymbol{Y}^M(t))]
\end{aligned}
\tag{4.12}
$$

where $\Delta\boldsymbol{\theta}_l$ and $\Delta\boldsymbol{\theta}_{l'}$ are vectors with all elements being zero except the l^{th} and l'^{th} elements equal to $\Delta\theta_l$ and $\Delta\theta_{l'}$, respectively. For example, $\Delta\boldsymbol{\theta}_l = [0,\cdots,0,\Delta\theta_l,0,\cdots,0]^T$ is a $1\times N_d$ vector with l^{th} element as a proper step $\Delta\theta_l$.

4.3.4 Identification procedures and the flowchart

The procedure of the proposed posterior density estimation is as follows:

Procedure 1: Use the Latin hypercube sampling (LHS) method to generate N_s sequences for the initial state of MC chains, respecting the prescribed limits of the search space. The likelihood measure of each samples is obtained with the Eq. (4.2).

Procedure 2: Compute the PDF, $p(\boldsymbol{\theta}_{(0)}^{k})$,for each samples, $k = 1, \dots, N_s$. Find the $\boldsymbol{\theta}_{(0)}^{\text{cbest}}$ and $\boldsymbol{\theta}_{(0)}^{\text{gbest}}$; Update the posterior sample of the Markov chain by mutation strategy using Eq. (3.1) and by the crossover probability using Eq. (3.3) to Eq. (3.4). Calculate the density for the updated samples, $p(\boldsymbol{\theta}_{(s+1)}^{k})$, of the iteration, $(s+1)$.

Procedure 3: The Metropolis acceptance (Eq. (3.5)) is used for chosen of accepted posterior samples.

Procedure 4: Repeat Procedure 2 to Procedure 3, after the burn-in period and calculating the convergence criteria using Eq. (3.6) for each dimension of the structural parameter. If the convergence criteria of the MC chain are met, ($\hat{R}_j < 1.2$), the MAP estimator, $\boldsymbol{\theta}_{\text{MAP}}$, and the standard deviation of the posterior samples, $\boldsymbol{\sigma}(\boldsymbol{\theta})$, are obtained.

Procedure 5: Calculates the gradient and Hessian matrix at the point of each samples within the interval of $[\boldsymbol{\theta}_{\text{MAP}} - 3\boldsymbol{\sigma}(\boldsymbol{\theta}), \boldsymbol{\theta}_{\text{MAP}} + 3\boldsymbol{\sigma}(\boldsymbol{\theta})]$ by the Eq. (4.10). The fitness function of these samples, $L^*(\boldsymbol{\theta}|\boldsymbol{Y}^M(t))$, using the Eq.(4.7) are obtained at the iteration when the $\boldsymbol{\theta}_{\text{MAP}}$ is obtained, denoted as the iteration, (s_2). Find the $\boldsymbol{\theta}_{(s_2)}^{*}$ with maximum value in $\left\{L^*\left(\boldsymbol{\theta}_{(s_2)}^{1}|\boldsymbol{Y}^M(t)\right), \dots, L^*\left(\boldsymbol{\theta}_{(s_2)}^{N_s}|\boldsymbol{Y}^M(t)\right)\right\}$.

Procedure 6: Update the samples using Eq.(3.1), and calculate the PDF for each samples at the new iteration, (s_2+1). And find the sample, $\boldsymbol{\theta}_{(s_2+1)}^{*}$, that with maximum value

in $\left\{L^*\left(\boldsymbol{\theta}^1_{(s_2+1)}|\boldsymbol{Y}^M(t)\right), \ldots, L^*\left(\boldsymbol{\theta}^{N_s}_{(s_2+1)}|\boldsymbol{Y}^M(t)\right)\right\}$.

Procedure 7: Repeat Procedure 5 to Procedure 6 till the optimal point, $\widehat{\boldsymbol{\theta}}_0 = \text{argmax}(L^*(\widehat{\boldsymbol{\theta}}_0|\boldsymbol{Y}^M(t)))$, is searched. And compare the PDF of the $\widehat{\boldsymbol{\theta}}_0$ with the PDF of the $\boldsymbol{\theta}_{\text{MAP}}$; if $p(\widehat{\boldsymbol{\theta}}_0) < p(\boldsymbol{\theta}_{\text{MAP}})$ end the algorithm; otherwise, reset the iteration state and let $\boldsymbol{\theta}^{\text{gbest}} = \widehat{\boldsymbol{\theta}}_0$, then the algorithm returns to the Procedure 2. The flowchart of the above-mentioned identification procedures is presented in Fig.4.7.

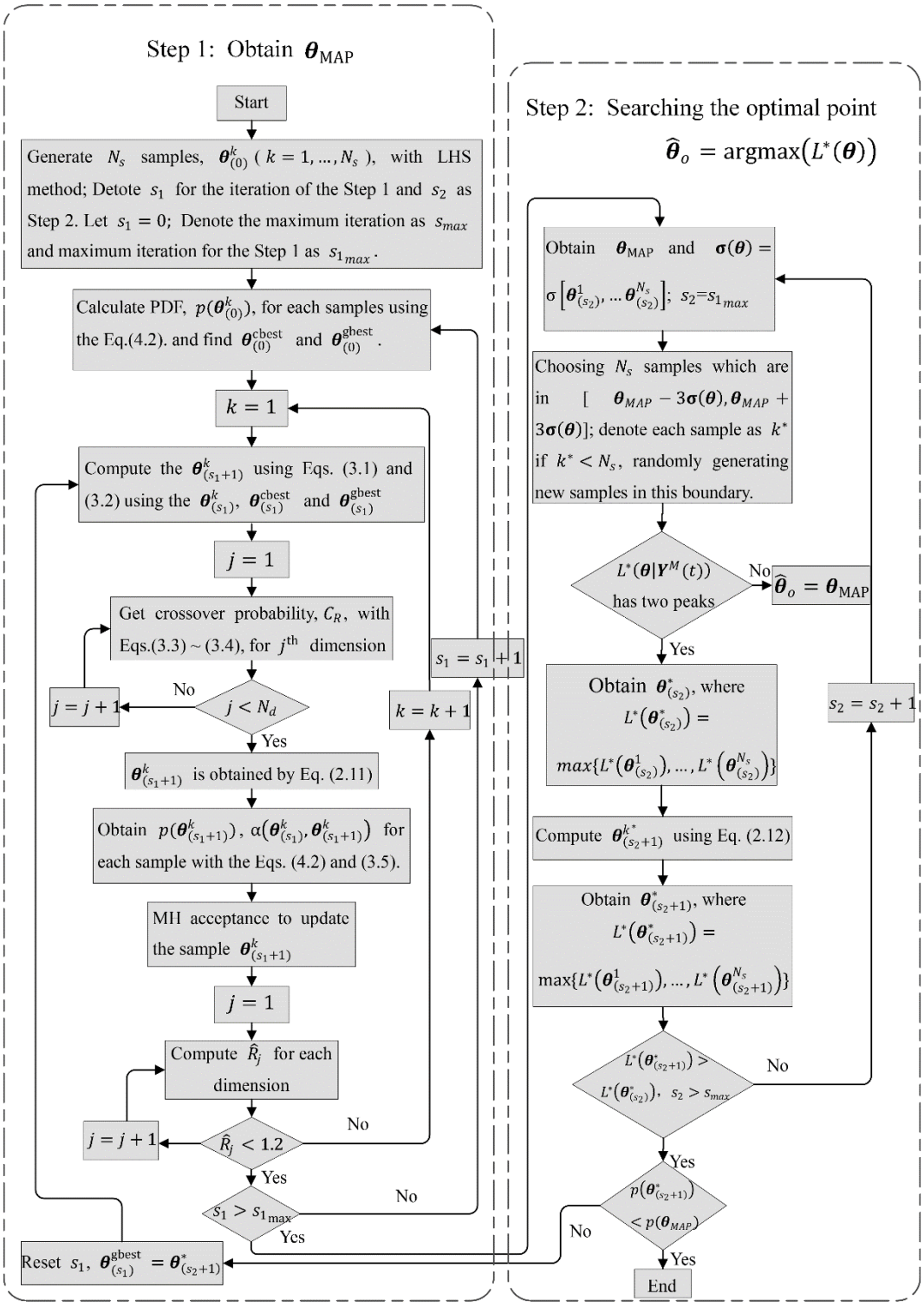


Figure 4.7 Flow-chart of the proposed two-step identification method

4.4 Numerical Simulation

4.4.1 Identification of a 10-DOF LTI system

Numerical simulation of a 10-DOF LTI system was carried out to verify the proposed method. The structural system is simulated as Eq.(3.7) and the measured signal is as Eq. (3.10). The input was an El-Centro wave (Fig. 4.8) lasting 40s and the sampling frequency was 100 Hz. Table 4.1 shows the structural properties of the dynamic system. The influence of the limited availability of measurements on the proposed method is also assessed in this study. In the "full output" scenario, measurements of all floors are available, whereas in the "partial output" case, only the even floors (2^{nd}, 4^{th}, 6^{th}, 8^{th}, and 10^{th}) are assumed to be available.

Table 4.1 Structural properties

Stiffness (k)	(N/m)
Floors 1-10	5.0×10^3
Mass (m)	(kg)
Floors 1-10	50
Damping ratio	
$\zeta_{1,2}$	0.05

The mass is assumed to be known; hence, the 10-DOF system is described by a stochastic model set, of which the parameterized vector is:

$$\boldsymbol{\theta}=\{k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, k_9, k_{10}, \zeta_1, \zeta_2\}. \quad (4.13)$$

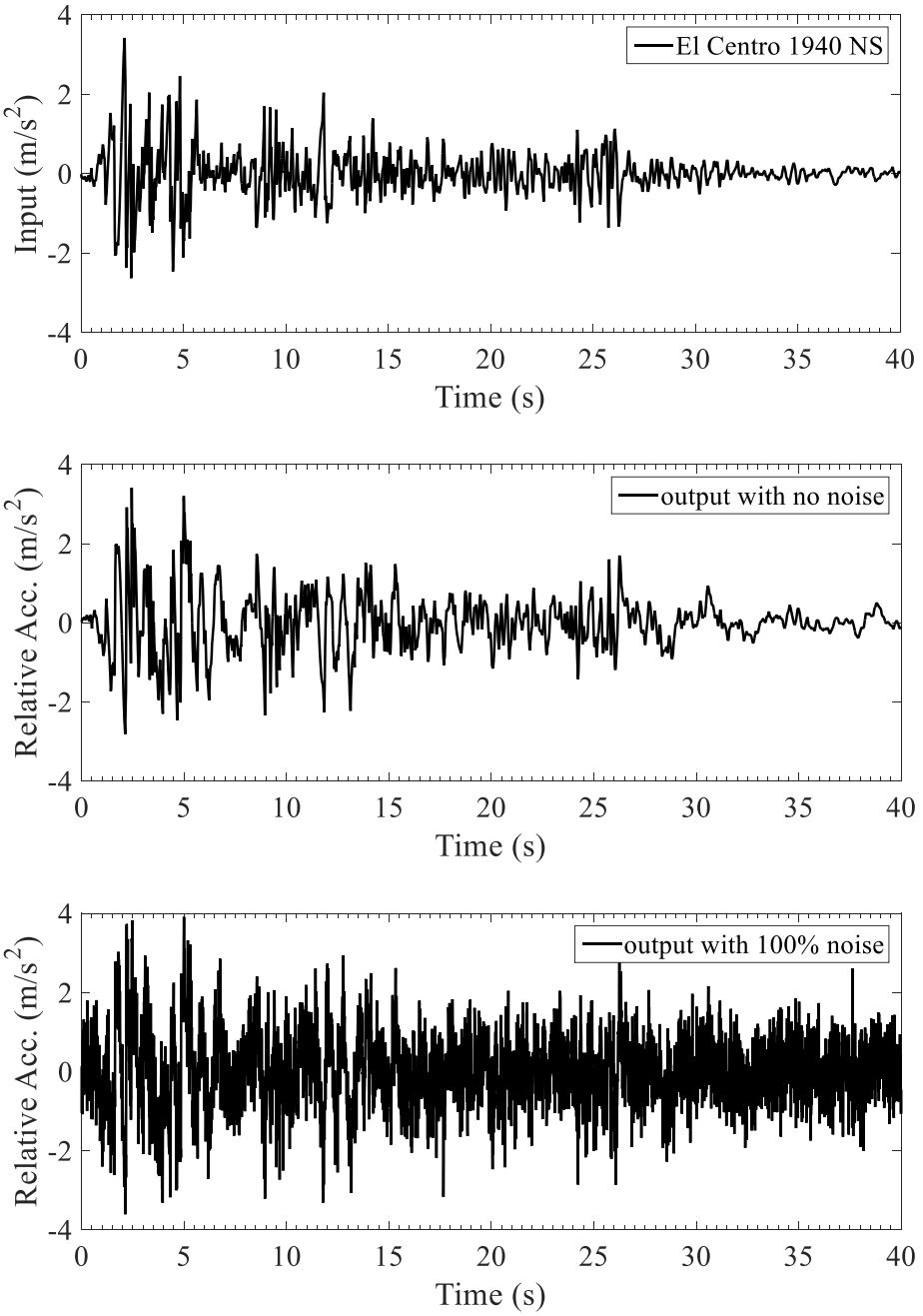


Figure 4.8 Input and output of 10th floor with and without noise

The white noise is added to the measured response, that simulated by Eq.(3.9), where the noise level is assumed to be 30% and 100%. The search domain was taken to be 0.5-2.0 times the true value. The initial parameters for the IDREAM algorithm is set same as those in Section 3.1.1. The results are shown in Table 4.2 and 4.3.

Table 4.2 Identification results in the scenario of "full outputs"

		No noise		30% noise		100% noise	
		$L(\boldsymbol{\theta})$	$L^*(\boldsymbol{\theta})$	$L(\boldsymbol{\theta})$	$L^*(\boldsymbol{\theta})$	$L(\boldsymbol{\theta})$	$L^*(\boldsymbol{\theta})$
k_1	error	0.000	0.000	0.279	0.091	0.517	0.441
	Cov.	0.000	0.000	0.530	0.537	1.828	1.096
k_2	error	0.000	0.000	0.259	0.012	3.484	1.329
	Cov.	0.000	0.000	0.631	0.521	1.951	1.732
k_3	error	0.000	0.000	0.218	0.122	1.401	0.986
	Cov.	0.000	0.000	0.661	0.617	2.281	2.868
k_4	error	0.000	0.000	0.819	0.463	1.277	1.094
	Cov.	0.000	0.000	0.659	0.548	2.273	2.014
k_5	error	0.000	0.000	0.224	0.399	1.432	0.912
	Cov.	0.000	0.000	0.797	0.626	2.484	0.786
k_6	error	0.000	0.000	1.042	0.693	2.382	1.479
	Cov.	0.000	0.000	0.967	0.718	3.215	0.902
k_7	error	0.000	0.000	0.303	0.713	1.017	1.008
	Cov.	0.000	0.000	0.878	0.924	2.854	1.662
k_8	error	0.000	0.000	0.019	0.334	1.904	0.140
	Cov.	0.000	0.000	0.939	0.622	3.261	1.809
k_9	error	0.000	0.000	0.518	0.010	1.370	1.103
	Cov.	0.000	0.000	0.999	0.734	3.173	2.562
k_{10}	error	0.000	0.000	0.338	0.754	6.491	1.575
	Cov.	0.000	0.000	1.238	1.124	3.807	2.052
ζ_1	error	0.000	0.000	1.999	1.175	1.057	1.766
	Cov.	0.000	0.000	1.232	1.426	4.261	3.312
ζ_2	error	0.000	0.000	0.354	0.343	1.286	1.192
	Cov.	0.000	0.000	0.949	1.668	2.492	2.124

* the error in the table is in %; the cov. in the table is in %.

Table 4.3 Identification results in the scenario of "partial outputs"

		No noise		30% noise		100% noise	
		$L(\boldsymbol{\theta})$	$L^*(\boldsymbol{\theta})$	$L(\boldsymbol{\theta})$	$L^*(\boldsymbol{\theta})$	$L(\boldsymbol{\theta})$	$L^*(\boldsymbol{\theta})$
k_1	error	0.000	0.000	0.595	0.342	1.183	0.812
	Cov.	0.000	0.000	2.624	2.096	2.332	2.556
k_2	error	0.000	0.000	3.018	1.431	4.195	1.390
	Cov.	0.000	0.000	2.715	2.032	7.561	3.852
k_3	error	0.000	0.000	2.369	0.985	4.258	2.081
	Cov.	0.000	0.000	2.146	1.869	5.363	3.218
k_4	error	0.000	0.000	1.035	0.173	5.195	3.514
	Cov.	0.000	0.000	2.417	0.543	6.007	4.774
k_5	error	0.000	0.000	2.753	1.497	5.908	2.138
	Cov.	0.000	0.000	1.872	1.931	7.477	3.728
k_6	error	0.000	0.000	2.478	1.058	5.945	2.122
	Cov.	0.000	0.000	2.217	2.137	8.178	3.647
k_7	error	0.000	0.000	3.174	2.065	7.342	1.609
	Cov.	0.000	0.000	1.652	1.662	9.378	6.198
k_8	error	0.000	0.000	1.743	0.809	7.018	3.025
	Cov.	0.000	0.000	2.133	1.058	10.85	4.056
k_9	error	0.000	0.000	5.974	2.032	6.049	1.897
	Cov.	0.000	0.000	2.565	2.561	11.09	6.307
k_{10}	error	0.000	0.000	5.723	2.175	8.451	3.543
	Cov.	0.000	0.000	2.779	1.595	9.292	6.091
ζ_1	error	0.000	0.000	1.397	1.175	5.819	1.130
	Cov.	0.000	0.000	1.700	2.237	6.819	4.902
ζ_2	error	0.000	0.000	1.957	2.062	6.258	2.572
	Cov.	0.000	0.000	0.629	0.574	10.32	5.067

* the error in the table is in %; the cov. in the table is in %.

Seen from Tables 4.2 and 4.3, in the case of "full outputs", the maximum relative error of $\widehat{\boldsymbol{\theta}}_0$ ranges from zero to 1.175% in 30% noise level and 1.766% in 100% noise level. Correspondingly, the maximum relative error of $\boldsymbol{\theta}_{\mathrm{MAP}}$ using the traditional likelihood measure increases from zero in the case of ignoring measurement noise to 1.999% in 30% noise level, and 6.491% in 100% noise level. The improvement is also clear in the scenario of partial outputs. The maximum relative error of $\widehat{\boldsymbol{\theta}}_0$ using the proposed method increases from zero in noise-free case to 2.175% in 30% noise level, and 3.543% in 100% noise level. While the maximum relative error of $\boldsymbol{\theta}_{\mathrm{MAP}}$ using the traditional method raises from zero in no-noise case, to 5.974% in 30% noise level, and 8.451% in 100% noise level. It can be found that the minimum and maximum relative error of mean posterior samples in model set are all reduced using the proposed method.

Moreover, it can be found from Tables 4.2 and 4.3 that using the Eq.(4.6) for the parametric uncertainty (the cov.) becomes smaller than those obtained by the formal log-likelihood measure as the Eq.(4.2). For instance, when considering the case of 100% noise level and partial outputs are available, the maximum coefficient variance of the MC samples using the proposed method is 6.198% comparing with that obtained by traditional method is as 11.09%. It is therefore can be concluded that the accuracy of the estimator using the proposed likelihood measure is improved.

4.5 Parameter estimation of E-Defense experiment

4.5.1 Description of the E-Defense experiment

In order to investigate the performance of hospital buildings during mega-earthquakes and evaluate the condition of the buildings after earthquakes, a full-scale 4-story reinforced concrete hospital building has been tested by E-Defense Shaking Table in the Miki, Japan (NIED, 2009).

The building is 8 meters in width, 10 meters in depth and 16.25 meters in height. The

RC structure is composed of four reinforced concrete columns and four-pieces shear wall at each floor. Each RC column is at the size of 0.6 by 0.6 meters and each shear-wall is 0.3 meters in thickness and 2 meters in length. The picture of the E-defense experiment is shown in Fig. 4.9. To measure the response of the structure under the input earthquake waves, the acceleration sensors were deployed at two directions (X-direction and Y-direction) on each floor and table-board which is shown in Fig. 4.10 and the sampling rate is 1000Hz.

Figure 4.9 Picture of E-Defense (NIED, 2009)

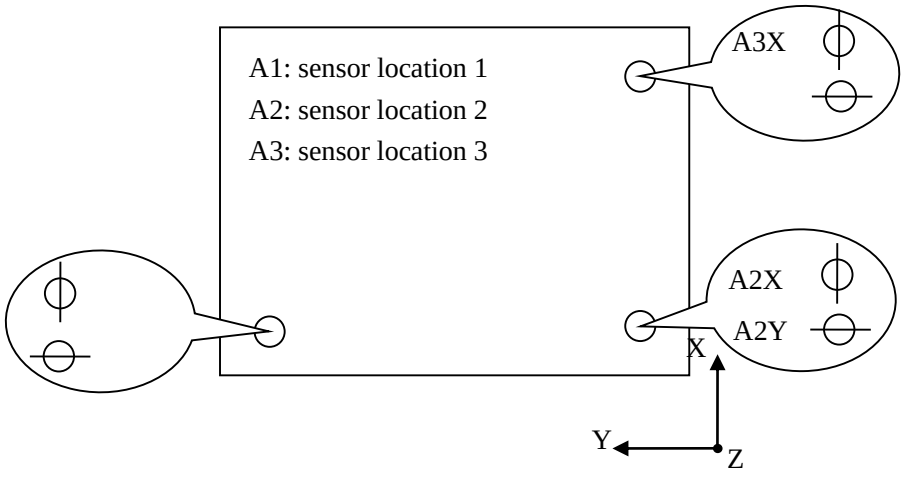


Figure 4.10 Deployment of acceleration sensors

The experiment was completed at three days from 15th to 22nd, January 2009. The whole excitation of the experiment process can be divided as 13 sets, and 3 random waves ahead of real earthquake waves were inputted into the structure at each day. Due to the low amplitude, the input states of random waves at each day (Jan. 15th, Jan. 19th and Jan. 22nd) and the measured responses at X-direction were studied to identify the parameters at such three input sets. An example of the excitation, the Random waves that loaded on the structure in Jan. 15th, is shown in Fig. 4.11.

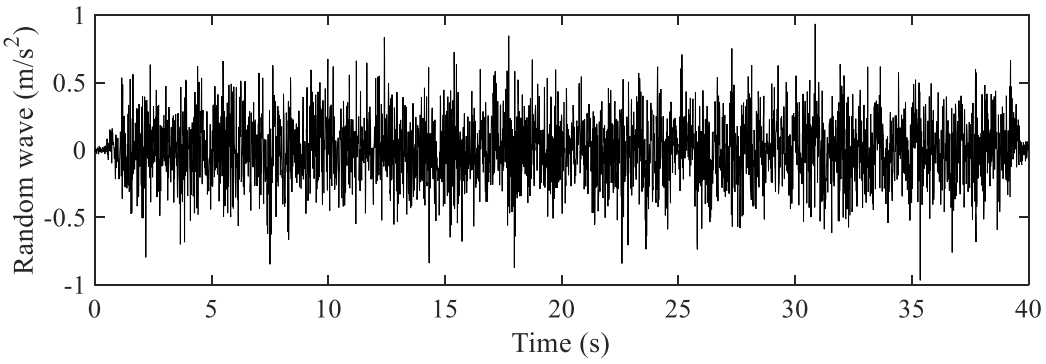


Figure 4.11 An example of input excitation (Random waves, 15th January, 2010)

4.5.2 Identification model

The structural system can be simulated as a linear structure, and the mass of the structure is lumped at each floor level. The dynamic equation is:

$$\ddot{\boldsymbol{y}}(t) + \boldsymbol{M}^{-1}\boldsymbol{C}\dot{\boldsymbol{y}}(t) + \boldsymbol{M}^{-1}\boldsymbol{K}\boldsymbol{y}(t) = -\begin{Bmatrix}1\\1\\1\\1\end{Bmatrix}\ddot{\boldsymbol{u}}(t). \tag{4.14}$$

Let $\boldsymbol{C}'$ denote $\boldsymbol{M}^{-1}\boldsymbol{C}$ and $\boldsymbol{K}'$ denote $\boldsymbol{M}^{-1}\boldsymbol{K}$, they can be written as:

$$\boldsymbol{C}' = \begin{pmatrix} \frac{c_1+c_2}{m_1} & -\frac{c_2}{m_1} & 0 & 0 \\ -\frac{c_2}{m_2} & \frac{c_2+c_3}{m_2} & -\frac{c_3}{m_2} & 0 \\ 0 & -\frac{c_3}{m_3} & \frac{c_3+c_4}{m_3} & -\frac{c_4}{m_3} \\ 0 & 0 & -\frac{c_4}{m_4} & \frac{c_4}{m_4} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{c_1}{m_1}+\frac{c_2}{m_2}\frac{m_2}{m_1} & -\frac{c_2}{m_2}\frac{m_2}{m_1} & 0 & 0 \\ -\frac{c_2}{m_2} & \frac{c_2}{m_2}+\frac{c_3}{m_3}\frac{m_3}{m_2} & -\frac{c_3}{m_3}\frac{m_3}{m_2} & 0 \\ 0 & -\frac{c_3}{m_3} & \frac{c_3}{m_3}+\frac{c_4}{m_4}\frac{m_4}{m_3} & -\frac{c_4}{m_4}\frac{m_4}{m_3} \\ 0 & 0 & -\frac{c_4}{m_4} & \frac{c_4}{m_4} \end{pmatrix}, \tag{4.15}$$

$$\boldsymbol{K}' = \begin{pmatrix} \frac{k_1+k_2}{m_1} & -\frac{k_2}{m_1} & 0 & 0 \\ -\frac{k_2}{m_2} & \frac{k_2+k_3}{m_2} & -\frac{k_3}{m_2} & 0 \\ 0 & -\frac{k_3}{m_3} & \frac{k_3+k_4}{m_3} & -\frac{k_4}{m_3} \\ 0 & 0 & -\frac{k_4}{m_4} & \frac{k_4}{m_4} \end{pmatrix}$$

$$= \begin{pmatrix} \frac{k_1}{m_1}+\frac{k_2}{m_2}\frac{m_2}{m_1} & -\frac{k_2}{m_2}\frac{m_2}{m_1} & 0 & 0 \\ -\frac{k_2}{m_2} & \frac{k_2}{m_2}+\frac{k_3}{m_3}\frac{m_3}{m_2} & -\frac{k_3}{m_3}\frac{m_3}{m_2} & 0 \\ 0 & -\frac{k_3}{m_3} & \frac{k_3}{m_3}+\frac{k_4}{m_4}\frac{m_4}{m_3} & -\frac{k_4}{m_4}\frac{m_4}{m_3} \\ 0 & 0 & -\frac{k_4}{m_4} & \frac{k_4}{m_4} \end{pmatrix}, \quad (4.16)$$

where m_i, k_i and c_i are the mass, stiffness and damping coefficient of each floor, respectively. According to the design drawing and the weight of equipment of the experiment, the ratio of the mass for each floor is obtained as: $\{\frac{m_2}{m_1}, \frac{m_3}{m_2}, \frac{m_4}{m_3}\}$={0.9673, 1, 0.8794}. It is assumed that the mass of each floor doesn't change even when the damage of structure happens, therefore, the ratio of the mass for each floor is considered as known and constant values. Considering the stiffness and damping coefficients, as random variables, the identification parameter is thus fully described by the following stochastic vector:

$$\boldsymbol{\theta} = \left\{\frac{k_1}{m_1}, \frac{k_2}{m_2}, \frac{k_3}{m_3}, \frac{k_4}{m_4}, \frac{c_1}{m_1}, \frac{c_2}{m_2}, \frac{c_3}{m_3}, \frac{c_4}{m_4}\right\}. \quad (4.17)$$

The simulated response of the stochastic system is thus described by:

$$\ddot{\boldsymbol{y}}(\boldsymbol{\theta},t) = -\begin{Bmatrix}1\\1\\1\\1\end{Bmatrix}\ddot{\boldsymbol{u}}(t) - \boldsymbol{C}'\dot{\boldsymbol{y}}(\boldsymbol{\theta},t) - \boldsymbol{K}'\boldsymbol{y}(\boldsymbol{\theta},t), \tag{4.18}$$

where $\ddot{\boldsymbol{u}}(t)$ denotes the excitation input of the structure, which is obtained from the measured data at the base floor.

4.5.3 Identification results

The parameters of IDREAM algorithm are set as follows: maximum iteration=2500 and number of Markov sequences=25, the number of DE pairs =3 and the jumping rate=0.2, other initial parameters for the IDREAM algorithm is set same as in the simulation in Section 3. According to the design materials, the initial guess of the identification parameters is: {5864, 4590, 2940, 2777; 5.643, 7.186, 3.889, 5.100}; the search space is taken as 0.1~10 times the initial guess of the parameters. The estimator of E-Defense parameter via the proposed method are carried out and summarized in Table 4.4.

The convergence for each parameters are shown in Fig. 4.12. From Table 4.4, it can easily draw the conclusion that the ratio of stiffness and mass of each floor is generally in declining trend at the three different days. The uncertain range of the Markov posterior samples considering the identified parametric uncertainty and the measurement uncertainty (total uncertain range) is shown in Fig. 4.13. With the identified estimator of the structural parameters as is shown in Table 4.4, the residual error that is obtained by the deviation between the measured response and the output of simulated system parameterized by the estimator, is analysis. The QQ plot of the residual errors are shown in Fig. 4.14.

Table 4.4 Identification Results

Random waves		k_1/m_1	k_2/m_2	k_3/m_3	k_4/m_4	c_1/m_1	c_2/m_2	c_3/m_3	c_4/m_4
January 15th	1st	5663	4880	3036	2951	21.56	25.20	28.10	6.610
	2nd	5552	4723	3162	3019	21.67	24.92	27.79	6.891
	3rd	5663	4813	3049	2894	20.90	26.32	26.48	6.914
January 19th	1st	3340	3213	2006	1841	27.83	1.801	23.11	6.323
	2nd	3357	3279	1778	1747	29.08	1.870	20.92	6.251
	3rd	3446	3213	2104	1906	28.22	2.093	22.07	6.833
January 22nd	1st	2948	2720	1885	1677	27.67	1.284	20.39	6.341
	2nd	2818	2526	1741	1681	28.22	1.443	19.45	6.193
	3rd	2879	2508	1848	1635	28.22	1.441	19.41	6.494

With the predictive uncertain range, shown in Fig. 4.13, that is formulated by the posterior samples with 95% probability assurance, the percentage of measured response, the red line in Fig 4.13, at each time interval that in beyond the predictive range can be calculated, which can be defined as the including ratio:

$$Ratio_{in} = 1 - \frac{\sum_t \text{find}\left\{t \middle| \boldsymbol{Y}^M(t) < \widehat{\boldsymbol{Y}}^{\text{low}}(t) \parallel \boldsymbol{y}^M(t) > \widehat{\boldsymbol{Y}}^{\text{up}}(t)\right\}}{N_t}, \qquad (4.19)$$

where N_t is the number of measurement data; $\boldsymbol{Y}^M(t)$ denotes the measured response at each time interval $(t = 1,2,\dots,N_t)$, $\widehat{\boldsymbol{Y}}^{\text{low}}(t)$ and $\widehat{\boldsymbol{Y}}^{\text{up}}(t)$ denotes the lower and upper bound of the 95% posterior density insurance uncertainty range. $Ratio_{in}$ reflects the quality of stochastic identification. For this case, the including ratio equals to 94.95%, which means the identification is satisfied because the response distribution

range contains the most part of the measured information.

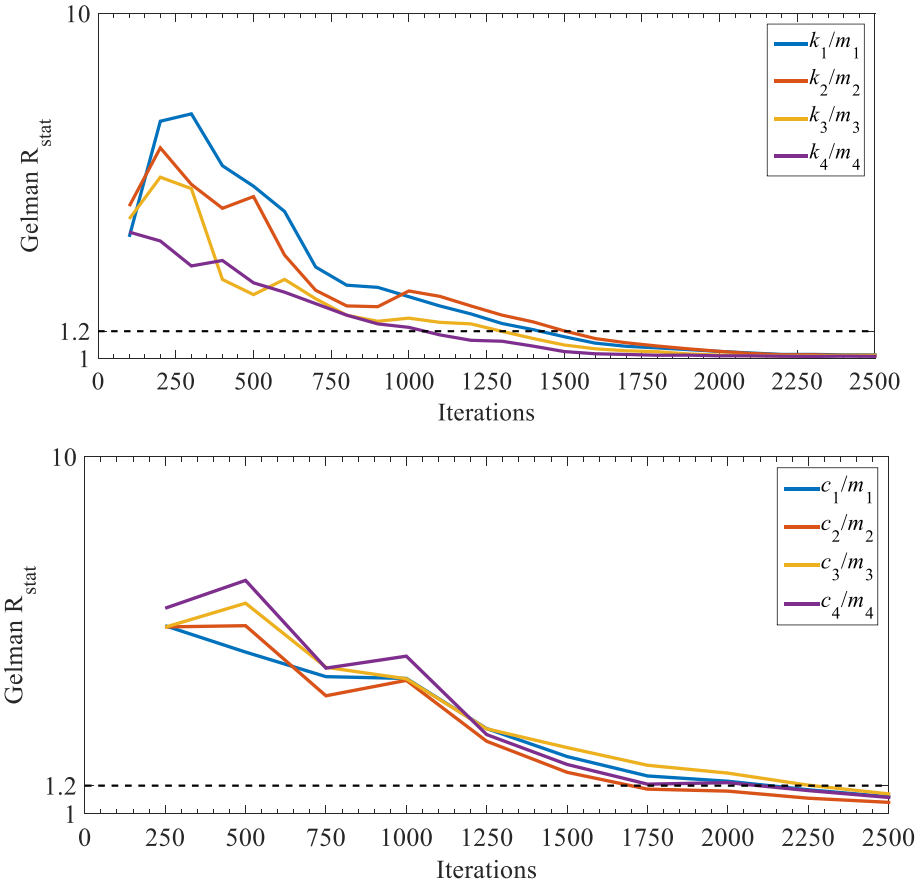


Figure 4.12 The convergence diagnosis for each identified parameters (Jan. 19th)

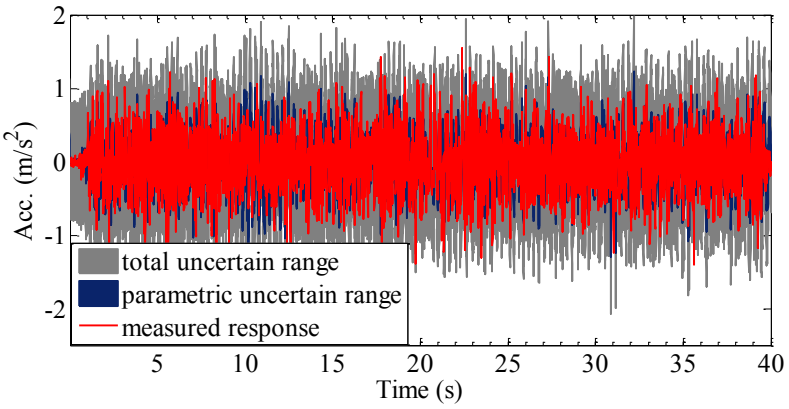


Figure 4.13 95% posterior simulation uncertain ranges (top floor, Jan. 15th)

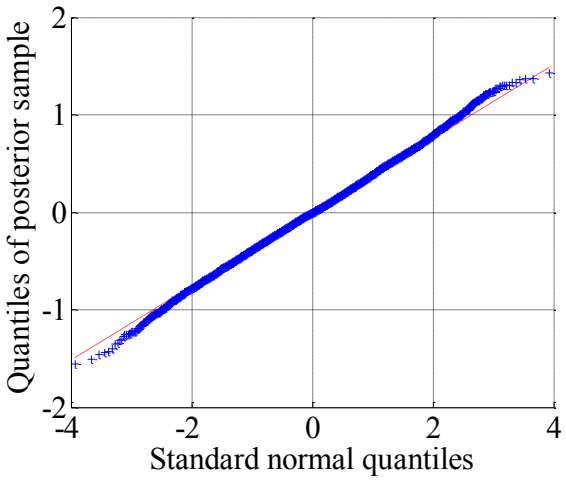


Figure 4.14 QQ plot of posterior sample versus standard normal distribution (top floor, Jan. 15th)

4.6 Conclusions

To improve the accuracy of the MAP estimator that obtained by the traditional likelihood measure using the IDREAM based structural identification, the gradient and Hessian of the log-likelihood measure is proposed to formulate the generalized likelihood measure for the density transition of Markov chains. Comparing with the formal likelihood function, the relative error of the estimator and the uncertain range of the posterior samples using the proposed method becomes smaller. Numerical simulations of a 10-DOF LTI system and experimental verification demonstrated its effectiveness in solving identification problems with a high noise level and loss of measurement data. In conclusion, IDREAM based Bayesian estimation using the proposed improvement has ability to solve the problem of "equifinality" especially when considering large level of measurement error. From the identified results of the experimental verification, it can be found that the IDREAM algorithm based

identification method using the proposed objective function have great potential in the Bayesian inference in the experiment of real structural system.

CHAPTER 5
Conclusions

Bayesian identification of structures using improved Differential evolution adaptive Metropolis algorithm, with the purpose to solve the problem of slow-convergence of the Markov chains in Bayesian inference, and generalized likelihood uncertainty estimation (GLUE) framework of identification using the proposed objective function, for sake of improving the accuracy of the Maximum a posterior estimator (MAP), were addressed in this thesis.

The major difficulty of using Bayesian inference for system identification is to obtain the posterior probability density of parameters conditioned by the measured response. The Markov chain Monte Carlo (MCMC) method is a widespread medium for posterior inference but its convergence is often slow. The first aspect of this study is to solve the convergence problem in the framework of Bayesian structural identification. This research presented a new methodology for posterior density estimation using improved differential evolution adaptive metropolis algorithm (IDREAM), where the transition of Markov chain is promoted by the weighting factor of the sample pairs. The main benefit of IDREAM is its efficient MCMC simulation through its use of the adaptive Metropolis (AM) method with a mutation strategy for ensuring quick convergence and robust solutions. Its effectiveness was demonstrated in simulations on identifying the

structural parameters with limited output data and noise polluted measurements.

Moreover, for traditional Bayesian identification, the MAP estimator will be inevitably biased, which is called the “equifinality” problem. The other aspect of this thesis is to solve this problem, in which the first-two derivative of the log-likelihood measure is proposed to formulate a new fitness function for sake of improving the accuracy of the estimator. The Identification procedure is proposed as two-step strategy. In the first step, the MAP estimator is obtained by the formal Bayesian likelihood measures using the IDREAM algorithm. In the second step, a new fitness measure is proposed under the framework of the generalized likelihood uncertainty estimation (GLUE), to obtain the estimator of the true value in the posterior samples set. Comparing with the formal likelihood function, the relative error of the MAP estimator and the uncertain range of the posterior samples using the proposed method becomes smaller. Numerical simulations of a 10-DOF LTI system demonstrated its effectiveness in solving identification problems with a high noise level and loss of measurement data. Seen from the simulated results, it can be concluded that the IDREAM based Bayesian inference using the proposed strategy has potential to solve the problem of “equifinality”, especially when considering large level of measurement error.

Moreover, in order to prove that the method is indeed applicable to realistic problems, the computing strategy for Bayesian identification was experimentally verified. Data of the experiments using the E-Defense model (the world’s largest full-scale Earthquake (EQ) Shaking Table facility carried out by National Research Institute for Earth Science and Disaster Prevention (NIED) in the Miki, Japan) was used to further verify the proposed methodology.

Finally, the conclusion was given. The Bayesian identification using the improved differential evolution metropolis-Hasting algorithm can enhance the convergence speed of the Markov chain sequences. And the proposed fitness function can be useful for

solving the problem of “equifinality”. These are the two contributions of this study made for the Bayesian inference in civil structures. Comparisons with existing methods showed that our proposed methodology was indeed a powerful tool for Bayesian identification of building structures. For future study, the achievements that obtained in this study will be used for reliability analysis of the structure using the posterior density of the parameters conditioned by the measurement.

References

Alvin K.F., Robertson A.N., Reich G.W. and Park K.C. (2003) Structural system identification: from reality to models. *Journal of Advanced Computational Models and Techniques in Dynamics*, Vol. 81, No. 12, pp. 1149–1176.

Barbieri N., Barbieri R. and Winikes L.C. (2010) Parameters estimation of sandwich beam model with rigid polyurethane foam core. *Journal of Mechanical System and Signal Processing*, Vol. 24, No.2, pp. 406–415.

Bayes T. (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, Vol. 53, pp. 370–418.

Beck, J.L. (1989) Statistical system identification of structures. *In Fifth International Conference on Structural Safety and Reliability*, San Francisco, California, pp. 7–11.

Beck, J.L. (1996) System identification methods applied to measured seismic response. *In Eleventh World Conference on Earthquake Engineering*, New York.

Beck J.L (2010) Bayesian system identification based on probability logic. *Journal of Structural Control and Health Monitoring*, Vol. 17, No. 7, pp. 825–847.

Beck J.L. and Au S.K. (2002) Bayesian Updating of Structural Models and Reliability

using Markov-Chain Monte-Carlo Simulation. *Journal of Engineering Mechanics*, Vol. 128, No. 4, pp. 380–391.

Beck, J.L. and Katafygiotis, L.S. (1991) Updating of a model and its uncertainties utilizing dynamic test data. *In First International Conference on Computational Stochastic Mechanics*, Corfu, Greece, pp. 125–136.

Beck J.L. and Katafygiotis, L.S. (1998) Updating models and their uncertainties. Part I: Bayesian statistical framework. *Journal of Engineering Mechanics*, Vol. 124, No. 4, pp. 455–461.

Beck J.L. and Yuen K.V. (2004) Model selection using response measurements: A Bayesian probabilistic approach. *Journal of Engineering Mechanics*, Vol. 130, No. 2, pp. 192–203.

Beven K.J. (1996) Equifinality and Uncertainty in Geomorphological Modeling. *The Scientific Nature of Geomorphology: Proceeding of* the 27*th Binghamton Symposium in Geomorphology*, pp. 289–313.

Beven K.J. (2006) A manifesto for the equifinality thesis. *Journal of Hydrology*, Vol. 320, No. 1–2, pp. 18–36.

Beven K.J. and Binley A. (1992) The future of distributed models: Model calibration and uncertainty prediction. *Journal of Hydrological Processes*, Vol.6, No. 3, pp. 279–298.

Beven K.J. and Freer J. (2001) Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methology. *Journal of Hydrology*, Vol. 249, No. 1–4, pp. 11–29.

Blasco A., Sorensen D. and Bidanel J.P. (1998) Bayesian inference of genertic parameters and selection response for litter size components in pigs. *Journal of Genetics*, Vol. 149, pp. 301–306.

Box G.E.P. and Cox D.R., (1964) An analysis of transformations. *Journal of the Royal Statistical Society. Series B*, Vol. 26, No. 2, pp. 211–252.

Box G.E.P. and Tiao G.C. (1973) Bayesian Inference in Statistical Analysis. *John Wiley and Sons, INC*, ISBN: 0471574287.

Casella G. and George E. I. (1992) Explaining the Gibbs sampler. *Journal of American Statistician*, Vol. 46, No. 3, pp. 167–174.

Chandrashekhar M. and Ganguli R. (2009) Uncertainty handling in structural damage detection using fuzzy logic and probabilistic simulation. *Journal of Mechanical Systems and Signal Processing*, Vol. 23, No. 2, pp. 384–404.

Chauveau D and Vandekerkhove P. (2002) Improving convergence of the Hasting Metropolis algorithm with a learning proposal. *Journal of Statistics*, Vol. 29, No. 1, pp. 13–29.

Cheung S.H. and Beck J.L. (2009) Bayesian model updating using Hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters. *Journal of Engineering Mechanics*, Vol. 135, No. 4, pp. 243–255.

Chib S. and Greenberg E. (1995) Understanding the Metropolis-Hastings algorithm. *Journal of American Statistician*, Vol. 49, No.4, pp. 327–335.

Ching J. and Chen Y.C. (2007) Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. *Journal of Engineering Mechanics*, Vol. 133, No. 7, pp. 816–832.

Chib S. and Greenberg E. (1995) Understanding the Metropolis-Hasting algorithm. *Journal of American statistical association*, Vol. 49, No. 4, pp. 327–335.

Chou J.H. and Ghaboussi J. (2001) Genetic algorithm in structural damage detection. *Journal of Computers and Structures*, Vol. 79, No. 14, pp. 1335–1353.

Cunha J., Cogan S. and Berthod C. (1999) Application of genetic algorithms for the identification of elastic constants of composite materials from dynamic tests. *International Journal for Numerical Methods in Engineering*, Vol. 45, No. 7, pp. 891–900.

Eiben A.E., Raue P. E. and Ruttkay Z. (1994) Genetic algorithms with multi-parent recombination. *Lecture Notes in Computer Science*, Vol. 866, pp.78-87.

Evans M. and Swartz T. (1995) Methods for approximating integrals in statistics with special emphasis on Bayesian integration problems. *Journal of Statistical Science*, Vol. 10, No. 3, pp. 254–272.

Franco G., Betti R. and Lus H. (2004) Identification of Structural Systems Using an Evolutionary Strategy, *Journal of Engineering Mechanics*, Vol.130, No.10, pp. 1125–1139.

Freni G., Mannina G. and Viviani G. (2008) Uncertainty in urban stormwater quality modelling: The effect of acceptability threshold in the GLUE methodology. *Journal of Water Research*, Vol.42, No. 8–9, pp. 2061–2072.

Gelman A., Carlin J.B., Stern H.S., Dunson D.B., Vehtari A. and Rubin D.B. (2004) Bayesian data analysis, Second edition. London, Chapman & Hall.

Gelman A., Gilks W.R. and Roberts G.O. (1997) Weak convergence and optimal scaling of random walk Metropolis algorithm. *Journal of the Annals of Applied Probability*, Vol. 7, No. 1, pp. 110–120.

Gelman A. and Rubin D.B. (1992) Inferences from iterative simulation using multiple sequences. *Journal of Statistical Science*, Vol. 7, No. 4, pp.457–511.

Geyer C.J. (1992) Practical Markov chain Monte Carlo. *Journal of Statistical Science*, Vol. 7, No. 4, pp. 473–511.

Haario H., Laine M., Mira A. and Saksman E. (2006) DRAM: Efficient adaptive MCMC. *Journal of Statistics and Computing*. Vol. 16, No. 4, pp. 339–354.

Haario H., Saksman E. and Tamminen J. (2001) An adaptive Metropolis algorithm, *Journal of Bernouli*, Vol. 7, No.2, pp. 223–242.

Hastings W.K. (1970) Monte Carlo sampling methods using Markov Chains and their applications. *Journal of Biometrika*, Vol. 57, No. 1 pp. 97–109.

Hastie D.I. and Green P.J. (2012) Model choice using reversible jump Markov chain Monte Carlo. *Journal of Statistica Neerlandica*, Vol. 66, No. 3, pp. 309–338.

Hung S.L., Huang C.S., Wen C.M. and Hsu Y.C. (2003) Nonparametric identification of a building structure from experimental data using wavelet neural network. *Journal of Computer-Aided Civil and Infrastructure Engineering*, Vol. 18, No. 5, pp.358–368.

Jaradat G.M. and Ayob M. (2010) Big bang-big crunch optimization algorithm to solve the course timetabling problem. *IEEE*, 10th *International Conference on Intelligent Systems Design and Applications*, pp. 1448-1452.

Jeong I.K. and Lee J.J. (1996) Adaptive simulated annealing genetic algorithm for system identification. *Journal of Engineering Applications of Artificial Intelligence*, Vol. 9, No. 5, pp. 523–532.

Kennedy J. and Eberhart R. (1995) Particle swarm optimization. Proceedings. *IEEE International conference on neural networks*, pp. 1942–1948.

Kessler S., Spearing S.M., Atalla, J.M., Cesnik, E.C. and Soutis C. (2002) Damage detection in composite materials using frequency response methods. *Journal of Composites Part B: Engineering*, Vol. 33, No. 1, pp. 87–95.

Kim J.T., Ryu Y.S., Cho H.m. and Stubbs N. (2003) Damage identification in beam-

type structures: frequency-based method vs mode-shape-based method. *Journal of Engineering Structures*, Vol. 25, No. 1, pp. 57–67.

Koh C.G. and Perry M.J. (2007) Structural damage quantification by system identification. *Journal of Earthquake and Tsunami*, Vol. 01, No. 3, pp. 211–231.

Koziel S. and Michalewicz Z. (1999) Evolutionary algorithms, homomorphous mappings and constrained parameter optimization. *Journal of Evolutionary Computation*, Vol. 7, No. 1, pp. 19–44.

Kuczera G. and Parent E. (1998) Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the metropolis algorithm. *Journal of Hydrol*ogy, Vol. 211, No. 1–4, pp. 69–85.

Laplace, P.S. (1951) A Philosophical Essay on Probabilities. *Translated from the 6th French ed. by Frederick Wilson Truscott and Frederick Lincoln Emory*. Dover.

Lee P.M. (1997) Bayesian Statistics: An introduction. *John Wiley*, New York.

Levin R.I. and Lieven N.A.J. (1998) Dynamic finite element model updating using simulated annealing and genetic algorithm. *Journal of Mechanical Systems and Signal Processing*. Vol. 12, No. 1, pp. 91–120.

Li R.S. (2013) Identification of structural parameters using symbolic time series analysis and intelligent algorithms. *PhD Thesis*, Keio University, Japan.

Li R.S., Mita A. and Zhou J. (2013) Symbolization-based differential evolution strategy for identification of structural parameters. *Journal of Structural Control and Health Monitoring*, Vol. 20, No. 10, pp. 1255–1270.

Ljung L. (1999) *System identification: theory for the user*. Prentice Hall PTR: Upper Saddle River, New Jersey.

Makowski P. and Todini E. (2006) Hydrological forecasting uncertainty assessment:

Incoherence of the GLUE methodology. *Journal of Hydrology*, Vol. 330, No. 1–2, pp. 368–381.

Makowski D., Wallach D. and Tremblay M. (2002) Using a Bayesian approach to parameter estimation: comparison of the GLUE and MCMC methods. *Journal of Agronomie*, Vol. 22, No. 2, pp. 191–203.

Metropolis N. and Ulam S. (1949) The Monte Carlo method. *Journal of the American Statistical Association.* Vol. 44, No. 247, pp. 335–341.

Mita A., (2003) Structural Dynamics for Health Monitoring, *Senkeisha Co., Ltd*, Nagoya.

Muto M. and Beck J.L. (2008) Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *Journal of Vibration and Control*, Vol.14, No. 1–2, pp. 7–34.

NIED (2009) Report of Functional maintenance of medical facility, December 2008 & Jan. 2009, http://www.bosai.go.jp/hyogo/ehyogo/research/movie/movie-detail.

Papadimitriou C., Beck J.L. and Katafygiotis L.S. (2001) Updating robust reliability using structural test data. *Journal of Probabilistic Engineering Mechanics*, Vol. 16, No. 2, pp. 103–113.

Peeters B., Maeck J. and Roeck G.D. (2001) Vibration-based damage detection in civil engineering: excitation sources and temperature effects. *Journal of Smart Materials and Structures*, Vol. 10, pp. 518–527.

Perry M.J., Koh C.G., and Choo Y.S. (2006), Modified genetic algorithm strategy for structural identification. *Journal of Computers and Structures*, Vol. 84, No. 8–9, pp. 529–540.

Price K.V. (1999). An introduction to differential evolution. *New ideas in optimization*,

McGraw-Hill, London, England, pp. 79–108.

Price K.V. and Storn R. (1997) Differential evolution: numerical optimization made easy. *Dr. Dobb's Journal*, Vol. 78, pp. 18–24.

Qian Y.Y. (2008) A time domain damage identification technique for building structures under arbitrary excitation. *PhD Thesis*. Keio University, Japan.

Qian Y.Y. and Mita A. (2008) Acceleration-based damage indicators for building structures using neural network emulators. *Journal of Structural Control and Health Monitoring*, Vol. 15, No. 6, pp. 901–920.

Qiao P., Lu K., Lestari W. and Wang J.L. (2007) Curvature mode shape-based damage detection in composite laminated plates. *Journal of Composite Structures*, Vol. 80, No.3, pp. 409–428.

Ratto M., Young P.C., Romanowicz R., Pappenberger F., Saltelli A. and Pagano A. (2007) Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. *Journal of Hydrology and Earth System Sciences*, Vol. 11, pp. 1249–1266.

Romanowicz R. and Beven K. (2006) Comments on generalised likelihood uncertainty estimation. *Journal of Reliability Engineering and System Safety*, Vol. 91, No. 10–11, pp. 1315–1321.

Sato T. and Qi K. (1998) Adaptive H∞ filter: its application to structural identification. *Journal of Engineering Mechanics*, Vol. 124, No. 11, pp.1233–1240.

Schulz K., Beven K. and Huwe B. (1999) Equifinality and the problem of robust calibration in nitrogen budget simulations. *Journal of Soil Science Society of America*, Vol. 63, No. 6, pp. 1934–1941.

Scruggs J.T., Taflanidis A.A. and Beck J.L. (2006). Reliability-based control

optimization for active base-isolation systems. *Journal of Structural Control and Monitoring*, Vol. 13, No. 2–3, pp. 705–723.

Simoen E., Papadimitriou C. and Lombaert G. (2013) On prediction error correlation in Bayesian model updating. *Journal of Sound and Vibration*, Vol. 332, No. 18, pp. 4136–4152.

Smith P., Beven K.J. and Tawn J.A. (2008) Informal likelihood measures in model assessment: theoretic development and investigation. *Journal of Advances in Water Resources*, Vol. 31, No. 8, pp. 1087–1100.

Sohn H. and Farrar C.R. (2001). Damage diagnosis using time-series analysis of vibrating signals. *Journal of Smart Materials and Structures*, Vol. 10, No. 3, pp. 446–451.

Sohn H., Farrar C.R., Hemez F.W., Shunk D.D., Stinemates D.W., Nadler B.R. and Czarnecki J.J. (2004) A Review of Structural Health Monitoring Literature: 1996-2001. *Los Alamos National Laboratory Report*, LA–13976–MS.

Spear R.C., Grieb T.M. and Shang N. (1994) Parameter uncertainty and interaction in complex environmental models. *Journal of Water Resources Research*, Vol. 30, No. 11, pp. 3159–3169.

Stedinger J.R., Vogel R.M. and Lee S.U. and Batchelder R. (2008) Apprisal of the generalized likelihood uncertainty estimation (GLUE) method. *Journal of Water Resources Research*, Vol. 44, No. 12, pp. 1–17.

Storn R. and Price K. (1997) Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, Vol. 11, No. 4, pp. 341–359.

Tang H.S., Xue S.T., Chen R. and Sato T. (2006) Online weighted LS-SVM for hysteretic structural system identification. *Journal of Engineering Structures*, Vol.

28, No. 12, pp. 1728–1735.

Tang H.S., Xue S.T. and Fan C.X. (2008) Differential evolution strategy for structural system identification. *Journal of Computers and Structures*, Vol. 86, No. 21–22, pp. 2004–2012.

Tang H.S., Zhou J., Xue S.T. and Xie L.Y. (2010) Big bang-big crunch optimization for parameter estimation in structural systems. *Journal of Mechanical Systems and Signal Processing*, Vol. 24, No. 8, pp. 2888–2897.

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, *Journal of Statistics and Computing*, Vol. 16, No. 3, pp. 239–249.

Trinh T.N. and Koh C.G. (2012). An improved substructural identification strategy for large structural systems. *Journal of Structural Control and Health Monitoring*, Vol. 19, No. 8, pp. 686–700.

Upton G. and Cook I. (1996) Understanding Statistics. Oxford University Press. ISBN 0-19-914391-9, pp. 55.

Vanik M.W., Beck J.L. and Au S.K. (2000) Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics*, Vol. 126, No. 7, pp. 738–745.

Vrugt J.A., Ter Braak C.J.F., Diks C.G.H., Robinson B.A., Hyman J.M. and Higdon D. (2009) Acceleration Markov Chain Monte Carlo simulation by differential evolution with self-adaptive randomized subspace sampling. *International Journal of Nonlinear Science and Numerical Simulation*, Vol. 10, No. 3, pp. 273–290.

Vrugt J.A. and Ter Braak C.J.F. (2011) $\text{DREAM}_{(D)}$: an adaptive Markov Chain Monte Carlo simulation algorithm to solve discrete, non-continuous, and combinatorial

posterior parameter estimation problems. *Journal of Hydrology and Earth System Sciences*, Vol. 15, pp. 3701–3713.

Wang G.S. (2009) Application of hybrid genetic algorithm to system identification. *Journal of Structural Control Health Monitoring*. Vol.16, No. 2, pp. 125–153.

Xie L.Y. and Mita A. (2010) Using component mode synthesis to estimate the restoring force of an isolation layer subjected to earthquakes. *Journal of Structural Control and Health Monitoring*, Vol. 17, No. 2, pp. 152–177.

Xue S.T., Tang H.S. and Zhou J. (2009) Identification of structural systems using particle swarm optimization. *Journal of Asian Architecture and Building Engineering*, Vol. 8, No. 2, pp. 101–112.

Xu Y.L., Zhang J., Li J. and Wang X.M. (2011) Stochastic damage detection method for building structures with parametric uncertainties. *Journal of Sound and Vibration*, Vol. 330, No. 20, pp. 4725–4737.

Yang J.N., Lin S., Huang H. and Zhou L. (2006) An adaptive extended Kalman filter for structural damage identification. *Journal of Structural Control and Health Monitoring*, Vol. 13, No. 4, pp. 849–867.

Yang J.N., Pan S. and Lin S. (2007) Least-Squares estimation with unknown excitations for damage identification of structures. *Journal of Engineering Mechanics*, Vol. 133, No. 1, pp. 12–21.

Ye M. and Wang X. (2007) Parameter estimation of the Bouc-Wen hysteresis model using particle swarm optimization. *Journal of Smart Materials and Structures*, Vol. 16, No. 16, pp. 2341–2349.

Yuen K.V. and Beck J.L. (2003) Reliability-based robust optimal control for uncertain dynamical systems using feedback of noisy measurements. *Journal of Earthquake Engineering and Structural Dynamics*, Vol. 32, No. 5, pp.751–770.

Zak S.K. and Beven K.J. (1999) Equifinality, senstivity and uncertainty in the estimation of critical loads. *Journal of Science of the Total Environ*ment, Vol. 236, No. 1–3, pp. 191–214.

Zak S.K., Beven K.J. and Reynolds B. (1997) Uncertainty in the estimation of critical loads: a practical methodology. *Journal of Water, Air and Soil Pollution*, Vol. 98, No. 3–4, pp. 297–316.

Zhang Y., Liu H. and Houseworth J. (2011) Modified generalized likelihood uncertainty estimation (GLUE) Methodology for considering the subjectivity of likelihood measure selection. *Journal of Hydrologic Engineering*, Vol. 16, No. 6, pp. 558–561.

Author’s biography

Jin Zhou was born in JiangSu Province of China in 1983. In 2006, Jin Zhou received the B.S. degree from Suzhou University of Science and Technology in Suzhou and majored in building design and construction of civil engineering. He went to the Tongji University, Shanghai, China, for the study on wind engineering and disaster reduction. In 2010, he obtained the M.S. degree from Tongji University with the thesis “Big Bang-Big Crunch algorithm for the structural identification and its application”. From 2010, he studies in Mita Lab, Keio University, Japan. His research is about Bayesian inference for the structural identification considering measurement uncertainty.

List of publications

Articles on periodicals (related to thesis)*:*

1. Jin Zhou, Akira Mita and Liu Mei, "Posterior density estimation for structural parameters using improved differential evolution adaptive Metropolis algorithm", *Journal of Smart Structures and Systems*, Vol. 15 (3), pp. 735-749, 2015, http://dx.doi.org/10.12989/sss.2015.15.3-4.000.

2. Jin Zhou, Akira Mita and Liu Mei, "An improved differential evolution adaptive Metropolis algorithm based Bayesian structural identification using the gradient of likelihood measures", *Journal of Structures*, Vol. 2015, Article ID 236475, DOI: http: dx.doi.org/101155/2015/236475.

Articles on international conference proceedings*:*

1. Jin Zhou, Akira Mita, Li Rongshuai, "Multi-objective differential evolution algorithm for stochastic system identification", *Proc. SPIE* 8692, San Diego, USA, 86923G, April, 2013, DOI: 10.1117/12.2006578.

2. Jin Zhou, Akira Mita, Li Rongshuai, "Multi-objective optimization strategies for damage detection using cloud model theory", *Proc. SPIE* 8348, San Diego, USA, 83482R, April, 2012, DOI: 10.1117/12.914231.

3. Rongshuai Li, Akira Mita and Jin Zhou, "Identification of structural parameters based on symbolic time series analysis and differential evolution strategy", *Proc. SPIE* 8692, San Diego, USA, 86923F, April, 2013, DOI: 10.1117/12.1000146.

Other articles on periodicals*:*

1. Jin Zhou, Akira Mita and Liu Mei, "Multi-objective differential evolution strategy for stochastic system identification considering uncertainties", *Journal of Mechanical Systems and Signal Processing*, (under review).

2. Liu Mei, Akira Mita and Jin Zhou, "An improved substructural damage detection approach of shear structure based on ARMAX model residual", *Journal of Structural Control and Health Monitoring*, DOI: 10.1002/stc.1766.

3. Liu Mei, Akira Mita and Jin Zhou, "A substructural damage identification approach for shear structure based on changes in the first AR model coefficient matrix", *Journal of Structures*, Vol. 2015, Article ID 976349, DOI: http://dx.doi.org/10.1155/2015/976349.

4. Rongshuai Li, Akira Mita and Jin Zhou, "Abnormal State Detection of Building Structures Based on Symbolic Time Series Analysis and Negative Selection", *Journal of Structural Control and Health Monitoring*, Vol.21 (1), pp. 80-97, 2014, DOI: 10.1002/stc.1555.

5. Rongshuai Li, Akira Mita and Jin Zhou, "Symbolization-Based Differential Evolution Strategy for Identification of Structural Parameters", *Journal of Structural Control and Health Monitoring*, Vol. 20 (10), pp. 1255-1270, 2013, DOI: 10.1002/stc.1530.