STOCHASTIC ANALYSIS, MODEL AND RELIABILITY UPDATING OF COMPLEX SYSTEMS WITH APPLICATIONS TO STRUCTURAL DYNAMICS

Thesis by

Sai Hung Cheung

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2009

(Defended January 28, 2009)

Acknowledgements

I would like to express my sincere gratitude to my advisor, James Beck, for his invaluable mentoring throughout my years in CALTECH on my research and my work as a teaching assistant. He shows me how to be an excellent scientist and instructor. I would also like to thank him for giving me the warm support, confidence, complete freedom and full support to creative and independent researches. I enjoy very much our numerous conversations and discussions about many aspects of our life.

I would like to thank my advisor for my Bachelor and Master Thesis, Lambros Katafygiotis, for his guidance and enthusiastic encouragement during those years. I would also like to thank him and Costas Papadimitriou for their unlimited support throughout these years during my pursuance of a Ph.D. in CALTECH.

I would also like to thank all my committee members, Professor Swaminathan Krishnan, Professor Thomas Heaton, Professor Joel Burdick and Professor Richard Murray for their insightful discussion and comments.

I would like to thank my best friend in CALTECH, Alexandros Taflanidis, for his friendship and support during my Ph.D. studies. I would like to thank my friends in Asia for their support: Ka Veng Yuen (Kelvin), Siu Kui Au (Ivan), Jianye Ching, Heung Fai Lam (Paul). Also thanks to other friends in CALTECH: Masumi Yamada, Chang Kook Oh, Judith Mitrani-Reiser, Matt Muto, Daniel Sutoyo and Jing Yang.

Special thanks to the love of my life, my wife Yunu He (Yuki), my mom, my brother Sai Keung Cheung (Patrick) and my parents-in-law for their unconditional love and support throughout these years.

Abstract

In many engineering applications, it is a formidable task to construct mathematical models that are expected to produce accurate predictions of the behavior of a system of interest. During the construction of such predictive models, errors due to imperfect modeling and uncertainties due to incomplete information about the system and its environment (e.g., input or excitation) always exist and can be accounted for appropriately by using probability logic. To assess the system performance subjected to dynamic excitations, a stochastic system analysis considering all the uncertainties involved has to be performed. In engineering, evaluating the robust failure probability (or its complement, robust reliability) of the system is a very important part of such stochastic system analysis. The word 'robust' is used because all uncertainties, including those due to modeling of the system, are taken into account during the system analysis, while the word 'failure' is used to refer to unacceptable behavior or unsatisfactory performance of the system output(s). Whenever possible, the system (or subsystem) output (or maybe input as well) should be measured to update models for the system so that a more robust evaluation of the system performance can be obtained. In this thesis, the focus is on stochastic system analysis, model and reliability updating of complex systems, with special attention to complex dynamic systems which can have high-dimensional uncertainties, which are known to be a very challenging problem. Here, full Bayesian model updating approach is adopted to provide a robust and rigorous framework for these applications due to its ability to characterize modeling uncertainties associated with the underlying system and to its exclusive foundation on the probability axioms.

First, model updating of a complex system which can have high-dimensional uncertainties within a *stochastic system model class* is considered. To solve the challenging computational problems, stochastic simulation methods, which are reliable and robust to problem complexity, are proposed. The Hybrid Monte Carlo method is investigated and it is shown how this method can be used to solve Bayesian model updating problems of complex dynamic systems involving high-dimensional uncertainties. New formulae for Markov Chain convergence assessment are derived. Advanced hybrid Markov Chain Monte Carlo simulation algorithms are also presented in the end.

Next, the problem of how to select the most plausible model class from a set of competing candidate model classes for the system and how to obtain robust predictions from these model classes rigorously, based on data, is considered. To tackle this problem, Bayesian model class selection and averaging may be used, which is based on the posterior probability of different candidate classes for a system. However, these require calculation of the evidence of the model class based on the system data, which requires the computation of a multi-dimensional integral involving the product of the likelihood and prior defined by the model class. Methods for solving the computationally challenging problem of evidence calculation are reviewed and new methods using posterior samples are presented.

Multiple stochastic model classes can be created even there is only one embedded deterministic model. These model classes can be viewed as a generalization of the stochastic models considered in Kalman filtering to include uncertainties in the parameters characterizing the stochastic models. State-of-the-art algorithms are used to solve the challenging computational problems resulting from these extended model classes. Bayesian model class selection is used to evaluate the posterior probability of an extended model classe and the original one to allow a data-based comparison. The problem of calculating robust system reliability is also addressed. The importance and effectiveness of the proposed method is illustrated with examples for robust reliability updating of structural

systems. Another significance of this work is to show the sensitivity of the results of stochastic analysis, especially the robust system reliability, to how the uncertainties are handled, which is often ignored in past studies.

A model validation problem is then considered where a series of experiments are conducted that involve collecting data from successively more complex subsystems and these data are to be used to predict the response of a related more complex system. A novel methodology based on Bayesian updating of hierarchical stochastic system model classes using such experimental data is proposed for uncertainty quantification and propagation, model validation, and robust prediction of the response of the target system. Recently-developed stochastic simulation methods are used to solve the computational problems involved.

Finally, a novel approach based on stochastic simulation methods is developed using current system data, to update the robust failure probability of a dynamic system which will be subjected to future uncertain dynamic excitations. Another problem of interest is to calculate the robust failure probability of a dynamic system during the time when the system is subjected to dynamic excitation, based on real-time measurements of some output from the system (with or without corresponding input data) and allowing for modeling uncertainties; this generalizes Kalman filtering to uncertain nonlinear dynamic systems. For this purpose, a novel approach is introduced based on stochastic simulation methods to update the reliability of a nonlinear dynamic system, potentially in real time if the calculations can be performed fast enough.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

In many engineering applications, it is a formidable task to construct mathematical models that are expected to produce accurate predictions of the behavior of a system of interest. During the construction of such predictive models, errors due to imperfect modeling and uncertainties due to incomplete information about the system and its environment (e.g., input or excitation) always exist and can be accounted for appropriately by using probability logic. In probability logic, probability is viewed as a multi-valued logic for plausible reasoning that extends Boolean propositional logic to the case of incomplete information (Cox 1946, 1961; Jaynes 2003; Beck 2008; Beck and Cheung 2009). Often one has to decide which proposed candidate models are acceptable for prediction of the system behavior. Behind the above also lies a great engineering interest to assess during the design and operation of a system whether it is expected to satisfy specified engineering performance objectives. To assess the system performance subjected to dynamic excitations, a stochastic system analysis considering all the uncertainties involved should be performed. In engineering, evaluating the robust failure probability (or its complement, robust reliability) of the system is a very important part of such stochastic system analyses. The word 'robust' is used because all uncertainties are taken into account during the system analysis, including those due to modeling of the system while the word 'failure' is used to refer to unacceptable behavior or unsatisfactory performance of the system output(s). Whenever possible, the system (or subsystem) output(s) (or maybe input(s) that include

quantities related to the environment) should be measured to update models for the system so that a more robust evaluation of the system performance can be obtained.

There are several characteristics of complex dynamic systems making the corresponding stochastic analysis, model and reliability updating computationally very challenging: (1). the system outputs or performance measures cannot be analytically expressed in terms of the uncertain modeling parameters (e.g., when dynamic systems are nonlinear); and (2). the number of uncertain modeling parameters can be quite large; for example, a large number of uncertain parameters are typical in modeling structures which have a large number of degrees of freedom subjected to dynamic excitations such as uncertain future earthquakes (requiring uncertain parameters of the order of hundreds or thousands to specify their discretized ground-motion time histories).

Another problem of much recent interest is model validation for a system which has attracted the attention of many researchers (e.g. Babuška and Oden, 2004; Oberkampf et al. 2004; Babuška et al. 2006; Chleboun 2008; Babuška et al. 2008; Grigoriu and Field 2008; Pradlwarter and Schuëller 2008; Rebba and Cafeo 2008) from many different fields of engineering and applied science because of the desire to provide a measure of confidence in the predictions of system models. In particular, in May 2006, the Sandia Model Validation Challenge Workshop brought together a group of researchers to present various approaches to model validation (Hills et al. 2008). The participants could choose to work on any of three problems; one in heat transfer (Dowding et al. 2008), one in structural dynamics (Red-Horse and Paez 2008) and one in structural statics (Babuška et al. 2008). The difficult issue of how to validate models is, however, still not settled; indeed, it is clear that a model that has given good predictions in tests so far might perform poorly under different circumstances, such as an excitation with different characteristics.

In this work, a full Bayesian model updating approach is adopted to provide a robust and rigorous framework for the above problems due to its ability to characterize modeling uncertainties associated with the underlying system and to its exclusive foundation on the

probability axioms. A probability logic approach is used (Beck and Cheung 2009) that is consistent with the Bayesian point of view that probability represents a degree of belief in a proposition but it puts more emphasis on its connection with missing information and information-theoretic ideas stemming from Shannon (1948).

1.1 Stochastic analysis, model and reliability updating of complex systems

Model updating using measured system response, with or without measured excitation, has a wide range of applications in response prediction, reliability and risk assessment, and control of dynamic systems and structural health monitoring (e.g., Vanik et al. 2001; Beck et al. 2001; Papadimitriou et al. 2001; Beck and Au 2002; Katafygiotis et al. 2003; Lam et al. 2004; Yuen and Lam 2006; Ching et al. 2006). There always exist modeling errors and uncertainties associated with the process of constructing a mathematical model of a system and its future excitation, whether it is based on physics or on a black-box 'nonparametric' model. Being able to quantify the uncertainties accurately and appropriately is essential for a robust prediction of future response and reliability of structures (Beck and Katafygiotis 1991, 1998; Papadimitriou et al. 2001; Beck and Au 2002; Cheung and Beck 2007a, 2008a, 2008b). Here in this thesis, a fully probabilistic Bayesian model updating approach is adopted, which provides a robust and rigorous framework due to its ability to characterize modeling uncertainties associated with the system and to its exclusive foundation on the probability axioms.

1.1.1 Stochastic system model classes

In this thesis, for the applications of the Bayesian approach, the Cox-Jaynes interpretation of probability as an extension of binary Boolean logic to a multi-valued logic of plausible inference is adopted where the relative plausibility of each model within a class of models is quantified by its probability (Cox 1961; Jaynes 2003). A key concept in the proposed approach here is a *stochastic system model class* $\mathcal{M}$ which consists of a *set* of probabilistic predictive input-output models for a system together with a probability distribution, the

prior, over this set that quantifies the initial relative plausibility of each predictive model. For simpler presentation, we will usually abbreviate the term "stochastic system model class" to "model class". Based on $\mathcal{M}$, one can use data $\mathcal{D}$ to compute the updated relative plausibility of each predictive model in the set defined by $\mathcal{M}$. This is quantified by the *posterior* PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ for the uncertain model parameters $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^D$ which specify a particular model within $\mathcal{M}$. By Bayes' theorem, this posterior PDF is given by:

$$p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}) = c^{-1} p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M})\, p(\boldsymbol{\theta} \mid \mathcal{M}) \tag{1.1}$$

where $c = p(\mathcal{D}|\mathcal{M}) = \int p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}$ is the normalizing constant which makes the probability volume under the posterior PDF equal to unity; $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})$ is the *likelihood function* which expresses the probability of getting data $\mathcal{D}$ based on the predictive PDF for the response given by model $\boldsymbol{\theta}$ within $\mathcal{M}$; and $p(\boldsymbol{\theta}|\mathcal{M})$ is the prior PDF for $\mathcal{M}$ which one can freely choose to quantify the initial plausibility of each model defined by the value of the parameters $\boldsymbol{\theta}$. For example, through the use of prior information that is not readily built into the predictive PDF that produces the likelihood function, the prior can be chosen to provide regularization of ill-conditioned inverse problems (Bishop 2006). As emphasized by Jaynes (2003), probability models represent a quantification of the state of knowledge about real phenomena conditional on the available information and should not be imagined to be a property inherent in these phenomena, as often believed by those who ascribe to the common interpretation that probability is the relative frequency of "inherently random" events in the "long run".

Based on the topology of $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})$ in the parameter space, and, in particular, the set $\{\boldsymbol{\theta} \in \Theta : \boldsymbol{\theta} = \arg\max p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})\}$ of MLEs (*maximum likelihood estimates*), a model class $\mathcal{M}$ can be classified into 3 different categories (Beck and Katafygiotis 1991, 1998; Katafygiotis and Beck 1998): *globally identifiable* (unique MLE), *locally identifiable* (discrete set of MLEs) and *unidentifiable* (a continuum of MLEs) based on the available data $\mathcal{D}$. Full Bayesian updating can treat all these cases (Yuen et al. 2004).

1.1.2 Stochastic system model class comparison

In many engineering applications, we are often faced with the problem of model class selection, that is, based on system data, choosing the most plausible model class from a set of competing candidate model classes to represent the behavior of the system of interest. A model class is a set of parameterized probability models for predicting the behavior of interest together with a prior probability model over this set indicating the relative plausibility of each predictive probability model. The main goal is to handle the tradeoff between the data-fit of a model and the simplicity of the model so as to avoid "overfitting" or "underfitting" the data. Bayesian methods of model selection and hypothesis testing have the advantage that they only use the axioms of probability. In contrast, analysis of multiple models or hypotheses is very difficult in a non-Bayesian framework without introducing ad-hoc measures (Berger and Pericchi 1996). The common selection criteria using *p*-values (significance tests) are difficult to interpret and can often be highly misleading (Jeffreys 1939, 1961; Lindley 1957, 1980; Berger and Delampady 1987). A common principle enunciated is that, if data is explained equally well by two models, then the simpler model should be preferred (often referred to as Ockham's razor) (Jeffreys 1961). Bayesian methods perform this automatically and systematically (Gull 1988; Mackay 1992; Beck and Yuen 2004) while non-Bayesian methods require introduction of ad-hoc measures to penalize model complexity to prevent overfitting.

There are several simplified data-based model selection methods, the most common of which are the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). AIC was proposed by Akaike (1974) based on providing an estimate to the Kullback-Leibler information (Kullback and Leibler 1951) with the goal of extending Fisher's maximum likelihood theory. Hurvich and Tsai (1989) proposed AICc, a variant of AIC, which provides an empirical but ad-hoc correction to AIC for the case where the sample size is small or the dimension of the uncertain parameters are large relative to the samples size. AICc converges to AIC as the sample size gets sufficiently large.

BIC was derived by Schwarz (1978) using Bayesian updating and an asymptotic approach assuming a sufficiently large sample size and that the candidate models all have unique maximum likelihood estimates. Deviance information criterion (DIC) (Spiegelhalter et al. 2002) is a generalization of AIC and BIC. DIC has an advantage that it can be readily calculated from the posterior samples generated by MCMC (Markov chain Monte Carlo) simulation. BIC and DIC are asymptotic approximations to full Bayesian updating at the model class level as the sample size becomes large and they may be misleading when two model classes give similar fits to the data. It was shown empirically by Kass and Raftery (1993) that BIC biases towards simpler models and AIC towards more complicated models as compared with a full Bayesian updating at the model class level, discussed next. The potential of BIC to produce misleading results was pointed out, for example, in Muto and Beck (2008).

Model class comparison is a rigorous Bayesian updating procedure that judges the plausibility of different candidate model classes, based on their posterior probability (that is, their probability conditional on the data from the system). Its application to system identification of dynamic systems that are globally identifiable or unidentifiable was studied in Beck and Yuen (2004) and Muto and Beck (2008), respectively. In these publications, a model class is referred to as a *Bayesian model class*.

Given a set of candidate model classes $M=\{\mathcal{M}_j: j=1,2,\ldots N_M\}$, we calculate the posterior probability $P(\mathcal{M}_j \mid \mathcal{D}, M)$ of each model class based on system data $\mathcal{D}$ by using Bayes' Theorem:

$$P(\mathcal{M}_j \mid \mathcal{D}, M) = \frac{p(\mathcal{D} \mid \mathcal{M}_j) P(\mathcal{M}_j \mid M)}{p(\mathcal{D} \mid M)} \tag{1.2}$$

where $P(\mathcal{M}_j \mid M)$ is the prior probability of each $\mathcal{M}_j$ and can be taken to be $1/N_M$ if one considers all N_M model classes as being equally plausible a priori; $p(\mathcal{D} \mid \mathcal{M}_j)$ expresses the

probability of getting the data $\mathcal{D}$ based on $\mathcal{M}_j$ and is called the *evidence* (or sometimes marginal likelihood) for $\mathcal{M}_j$ provided by the data $\mathcal{D}$ and it is given by the Theorem of Total Probability:

$$p(\mathcal{D} \mid \mathcal{M}_j) = \int p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_j)\, p(\boldsymbol{\theta} \mid \mathcal{M}_j) d\boldsymbol{\theta} \tag{1.3}$$

Although $\boldsymbol{\theta}$ corresponds to different sets of parameters and can be of different dimension for different $\mathcal{M}_j$, for simpler presentation a subscript j on $\boldsymbol{\theta}$ is not used since explicit conditioning on $\mathcal{M}_j$ indicates which parameter vector $\boldsymbol{\theta}$ is involved.

Notice that (1.3) can be interpreted as follows: the evidence gives the probability of the data according to $\mathcal{M}_j$ (if (1.3) is multiplied by an elemental volume in the data space) and it is equal to a weighted average of the probability of the data according to each model specified by $\mathcal{M}_j$, where the weights are given by the prior probability $p(\boldsymbol{\theta}|\mathcal{M}_j)d\boldsymbol{\theta}$ of the parameter values corresponding to each model. The evidence therefore corresponds to a type of integrated global sensitivity analysis where the prediction $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}_j)$ of each model specified by $\boldsymbol{\theta} \in \Theta$ is considered but it is weighted by the relative plausibility of the corresponding model.

The computation of the multi-dimensional evidence integral in (1.3) is highly nontrivial. The problem involving complex dynamic systems with high-dimensional uncertainties makes this computationally even more challenging. This will be discussed in more detail in a later chapter.

It is worth noting that from (1.3), the log evidence can be expressed as the difference of two terms (Ching et al. 2005; Muto and Beck 2008):

$$\ln[p(\mathcal{D} \mid \mathcal{M}_j)] = E[\ln(p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_j)] - E[\ln \frac{p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j)}{p(\boldsymbol{\theta} \mid \mathcal{M}_j)}] \tag{1.4}$$

where the expectation is with respect to the posterior $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_j)$. The first term is the posterior mean of the log likelihood function, which gives a measure of the goodness of the fit of the model class $\mathcal{M}_j$ to the data, and the second term is the Kullback-Leibler divergence, or relative entropy (Cover and Thomas 2006), which is a measure of the information gain about $\mathcal{M}_j$ from the data $\mathcal{D}$ and is always non-negative.

Comparing the posterior probability of each model class provides a quantitative *Principle of Model Parsimony* or *Ockham's razor* (Gull 1989; Mackay 1992), which have long been advocated qualitatively, that is, simpler models that are reasonably consistent with the data should be preferred over more complex models that only lead to slightly improved data fit. The importance of (1.3) is that it shows rigorously, without introducing ad-hoc concepts, that the log evidence for $\mathcal{M}_j$, which controls the posterior probability of this model class according to (1.2), explicitly builds in a trade-off between the data-fit of the model class and its "complexity" (how much information it takes from the data).

The evidence, and so Bayesian model class selection, may be sensitive to the choice of priors $p(\boldsymbol{\theta}|\mathcal{M}_j)$ for the uncertain model parameters (Berger and Pericchi 1996). The effect of priors on Bayesian hypothesis comparison was first noted in Lindley's paradox (Lindley 1957). The use of excessively diffuse priors for the parameters should be avoided since it will enforce a strong preference towards simpler models. In fact, since the model class includes the prior, for a given likelihood, Bayesian model class selection will give low posterior probability to a model class with a very diffuse prior, which can be deduced from (1.2) and (1.4); more generally, it provides a mechanism to judge priors based on data, as is done, for example, by parameterizing the priors in automatic relevance determination (Mackay 1993; Bishop 2006; Oh et al. 2008).

1.1.3 Robust predictive analysis and failure probability updating using stochastic system model classes

One of the most useful applications of Bayesian model updating is to make robust predictions about future events based on past observations. Let $\mathcal{D}$ denote data from available measurements on a system. Based on a candidate model class $\mathcal{M}_j$, all the probabilistic information for the prediction of a vector of future responses $\mathbf{X}$ is contained in the *posterior robust* predictive PDF for $\mathcal{M}_j$ given by the Theorem of Total Probability (Papadimitriou et al. 2001):

$$p(\mathbf{X} \mid \mathcal{D}, \mathcal{M}_j) = \int p(\mathbf{X} \mid \boldsymbol{\theta}, \mathcal{D}, \mathcal{M}_j) p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) d\boldsymbol{\theta} \tag{1.5}$$

The interpretation of (1.5) is similar to that given for (1.3) except now the prediction $p(\mathbf{X}|\boldsymbol{\theta},\mathcal{D},\mathcal{M}_j)$ of each model specified by $\boldsymbol{\theta} \in \Theta$ is weighted by its posterior probability $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_j)d\boldsymbol{\theta}$ because of the conditioning on the data $\mathcal{D}$. If this conditioning on $\mathcal{D}$ in (1.5) is dropped so, for example, the prior $p(\boldsymbol{\theta}|\mathcal{M}_j)$ is used in place of the posterior $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_j)$, the result $p(\mathbf{X}|\mathcal{M}_j)$ of the integration is the *prior robust* predictive PDF.

Many system performance measures can be expressed as the expectation of some function $\mathbf{g}(\mathbf{X})$ with respect to the posterior robust predictive PDF in (1.5) as follows:

$$E[\mathbf{g}(\mathbf{X}) \mid \mathcal{D}, \mathcal{M}_j] = \int \mathbf{g}(\mathbf{X}) p(\mathbf{X} \mid \mathcal{D}, \mathcal{M}_j) d\mathbf{X} \tag{1.6}$$

Some examples of important special cases are:

1) $\mathbf{g}(\mathbf{X})=I_F(\mathbf{X})$, which is equal to 1 if $\mathbf{X} \in F$ and 0 otherwise, where F is a region in the response space that corresponds to unsatisfactory system performance, then the integral in (1.6) is equal to the robust "failure" probability $P(F|\mathcal{D}, \mathcal{M}_j)$;

2) $\mathbf{g}(\mathbf{X})=\mathbf{X}$, then the integral in (1.6) becomes the robust mean response;

3) $\mathbf{g}(\mathbf{X})=(\mathbf{X}-E[\mathbf{X}|\mathcal{D}, \mathcal{M}_j])(\mathbf{X}-E[\mathbf{X}|\mathcal{D}, \mathcal{M}_j])^T$, then the integral in (1.6) is equal to the robust covariance matrix of $\mathbf{X}$.

The Bayesian approach to robust predictive analysis requires the evaluation of multi-dimensional integrals, such as in (1.5), and this usually cannot be done analytically. For problems involving complex dynamic systems with high-dimensional uncertainties, this can be computationally challenging. This will be discussed in more detail in a later chapter.

If a set of candidate model classes $M=\{\mathcal{M}_j: j=1,2,\ldots N_M\}$ is being considered for a system, all the probabilistic information for the prediction of future responses $\mathbf{X}$ is contained in the *hyper-robust* predictive PDF for M given by the Theorem of Total Probability (Muto and Beck 2008):

$$p(\mathbf{X} \mid \mathcal{D}, M) = \sum_{j=1}^{N_M} p(\mathbf{X} \mid \mathcal{D}, \mathcal{M}_j) P(\mathcal{M}_j \mid \mathcal{D}, M) \tag{1.7}$$

where the robust predictive PDF for each model class $\mathcal{M}_j$ is weighted by its posterior probability $P(\mathcal{M}_j|\mathcal{D}, M)$ from (1.2). Equation (1.7) is also called *posterior model averaging* in the Bayesian statistics literature (Raftery et al. 1997, Hoeting et al. 1999).

Let F denote the events or conditions leading to system failure (unsatisfactory system performance). The *hyper-robust* failure probability $P(F|\mathcal{D},M)$ based on M is then given by (Cheung and Beck 2008g, 2009a, 2009b):

$$P(F \mid \mathcal{D}, M) = \sum_{j=1}^{N_M} P(F \mid \mathcal{D}, \mathcal{M}_j) P(\mathcal{M}_j \mid \mathcal{D}, M) \tag{1.8}$$

The importance of the above is investigated in Chapters 4 and 5.

1.2 Outline of the Thesis

In this thesis, the focus is on stochastic system analysis, model and reliability updating of complex systems, with special attention to complex dynamic systems which can have high-dimensional uncertainties, which are very challenging. New methods are developed to solve these problems. Most of the methods developed in this thesis are intended to be very general without requiring special assumptions regarding the system. A new methodology is also developed to tackle the challenging model validation problem. Novel methods for updating robust failure probability are also developed.

In Chapter 2, model updating problems for complex systems which have high-dimensional parameter uncertainties within a *stochastic system model class* are considered. To solve the challenging computational problems, stochastic simulation methods, which are reliable and robust to problem complexity, are proposed. Markov Chain Monte Carlo simulation methods are presented and reviewed. An advanced Markov Chain Monte Carlo simulation method namely Hybrid Monte Carlo simulation method is investigated. Practical issues for the feasibility of this method to solve Bayesian model updating problems of complex dynamic systems involving high-dimensional uncertainties are addressed. Improvements are proposed to make it more effective and efficient for solving such model updating problems. New formulae for Markov Chain convergence assessment are derived. The effectiveness of the proposed approach is illustrated with an example for Bayesian model updating of a structural dynamic model with many uncertain parameters. New stochastic simulation algorithms created by combining state-of-the-art stochastic simulation algorithms are also presented.

In Chapter 3, the problem of comparison of model classes involving complex dynamic systems with high-dimensional uncertainties is considered. The problem of interest is how to select the most plausible model class from a set of competing candidate model classes for the system, based on data. To tackle this problem, Bayesian model class selection may be used, which is based on the posterior probability of different candidate classes for a

system. Another problem of interest is to tackle cases where more than one model class has significant posterior probability and each of these give different predictions. Bayesian model class averaging then provides a coherent mechanism to incorporate all the considered model classes in the probabilistic predictions for the system. However, both Bayesian model class selection and averaging require calculation of the evidence of the model class based on the system data, which requires the computation of a multi-dimensional integral involving the product of the likelihood and prior defined by the model class. Methods for solving the computationally challenging problem of evidence calculation are reviewed and new methods using posterior samples are presented.

In the past, most applications of Bayesian model updating of dynamic systems have focused on model classes which consider an uncertain prediction error as the difference between the real system output and the model output and model it probabilistically using Jaynes' Principle of Maximum Information Entropy. In Chapter 4, an extension of such model classes is considered to allow more flexibility in treating modeling uncertainties when updating state space models and making robust predictions; this is done by introducing prediction errors in the state vector equation, in addition to those in system output vector equation. These model classes can be viewed as a generalization of the stochastic models considered in Kalman filtering to include uncertainties in the parameters characterizing the stochastic models. State-of-the-art algorithms are used to solve the challenging computational problems resulting from these extended model classes. Bayesian model class selection is used to evaluate the posterior probability of an extended model class and the original one to allow a data-based comparison. To make predictions robust to model uncertainties, Bayesian model averaging is used to combine the predictions of these model classes. The problem of calculating robust system reliability is also addressed. The importance and effectiveness of the proposed method is illustrated with examples for robust reliability updating of structural systems.

In Chapter 5, the problem of model validation of a system is considered. Here, we consider the problem where a series of experiments are conducted that involve collecting data from successively more complex subsystems and these data are to be used to predict the response of a related more complex system. A novel methodology based on Bayesian updating of hierarchical stochastic system model classes using such experimental data is proposed for uncertainty quantification and propagation, model validation, and robust prediction of the response of the target system. The proposed methodology is applied to the 2006 Sandia static-frame validation challenge problem to illustrate our approach for model validation and robust prediction of the system response. Recently-developed stochastic simulation methods are used to solve the computational problems involved.

In Chapter 6, a newly-developed approach based on stochastic simulation methods is presented, to update the robust reliability of a dynamic system. The efficiency of the proposed approach is illustrated by a numerical example involving a hysteretic model of a building.

In Chapter 7, a novel approach is introduced based on stochastic simulation methods, which updates in real time the robust reliability of a nonlinear dynamic system. The performance of the proposed approach is illustrated by an example involving a nonlinear dynamic model using incomplete dynamic data obtained during the 1994 Northridge earthquake from a hotel which is a seven-story reinforced-concrete moment-frame building.

CHAPTER 2

Bayesian updating of stochastic system model classes with a large number of uncertain parameters

In this chapter, model updating problems of a complex system which can have high-dimensional parameter uncertainties within a *stochastic system model class* $\mathcal{M}$ is considered. Since the analysis is conditioned on a single model class, the subscript for $\mathcal{M}$, which denotes different model classes, is dropped in the rest of this chapter. The Bayesian approach to robust predictive analysis requires the evaluation of multi-dimensional integrals, such as in (1.5), and this usually cannot be done analytically. Laplace's method of asymptotic approximation (Beck and Katafygiotis 1991, 1998; Papadimitriou et al. 2001) has been used in the past, which utilizes a Gaussian approximation to the posterior PDF, as mentioned before for (1.3). However, application of this approximation faces difficulties when (i) the amount of data is small so its accuracy is questionable, or (iii) the chosen class of models is unidentifiable based on the available data. Also, such an approximation requires a non-convex optimization in a high-dimensional parameter space, which is computationally challenging, especially when the model class is not globally identifiable and so there may be multiple global maximizing points. It is shown in Cheung and Beck (2008b, g) that the robust failure probability can require information of the posterior PDF in the region of the uncertain parameter space that is not in the high probability region of the posterior PDF. Even if the Laplace analytical approximation gives a good approximation in the region of the uncertain parameter space that contains the high probability content of the posterior PDF, there is no guarantee that it gives sufficient

accuracy in approximating this probability distribution in other regions of the uncertain parameter space. It may therefore lead to a poor estimate of robust failure probability. Other analytical approximations to the posterior PDF such as the variational approximation (Beal 2003) suffer similar problems as Laplace's method of asymptotic approximation.

Thus, in recent years, focus has shifted from analytical approximations to using stochastic simulation methods in which samples consistent with the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ are generated. In these methods, all the probabilistic information encapsulated in $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ is characterized by posterior samples $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K:

$$p(\boldsymbol{\theta} \mid \mathcal{D},\mathcal{M}) \approx \frac{1}{K}\sum_{k=1}^{K}\delta(\boldsymbol{\theta}-\boldsymbol{\theta}^{(k)}) \tag{2.1}$$

With these samples, the integral in (1.5) can be approximated by:

$$p(\mathbf{X} \mid \mathcal{D},\mathcal{M}) \approx \frac{1}{K}\sum_{k=1}^{K} p(\mathbf{X} \mid \boldsymbol{\theta}^{(k)},\mathcal{D},\mathcal{M}) \tag{2.2}$$

Samples of $\mathbf{X}$ can then be generated from each of the $p(\mathbf{X} \mid \boldsymbol{\theta}^{(k)},\mathcal{D},\mathcal{M})$ with equal probability. The probabilistic information encapsulated in $p(\mathbf{X} \mid \mathcal{D},\mathcal{M})$ is characterized by these samples of $\mathbf{X}$.

There are several difficulties related to the sampling of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$: (i) the normalizing constant c in Bayes' Theorem in (1.1), which is actually the evidence in (1.3), is usually unknown a priori and its evaluation requires a high-dimensional integration over the uncertain parameter space; and (ii) the high probability content of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ occupies a much smaller volume than that of the prior PDF, so samples in the high probability region of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ cannot be generated efficiently by sampling from the prior PDF using direct Monte Carlo simulation. To tackle the aforementioned difficulties, Markov Chain Monte Carlo (MCMC) simulation methods (e.g. Robert and Casella 1999, Beck and Au 2002,

Ching et al. 2006, Ching and Cheng 2007, Muto and Beck 2008) were proposed to solve the Bayesian model updating problem more efficiently.

Probably the most well-known MCMC method is the Metropolis-Hastings (MH) algorithm (Metropolis et al. 1953, Hastings 1970) which creates samples from a Markov Chain whose stationary state is a specified target PDF. In principle, this algorithm can be used to generate samples from the posterior PDF but, in practice, its direct use is highly inefficient because the high probability content is often concentrated in a very small volume of the parameter space. Beck and Au (2000, 2002) proposed an approach which combines the idea from simulated annealing with the MH algorithm to simulate from a sequence of target PDFs, where each such PDF is the posterior PDF based on an increasing amount of data. The sequence starts with the spread-out prior PDF and ends with the much more concentrated posterior PDF. The samples from a target PDF in the sequence are used to construct a kernel sampling density which acts as a global proposal PDF for the MH procedure for the next target PDF in the sequence. The success of this approach relies on the ability of the proposal PDF to simulate samples efficiently for each intermediate PDF. However, in practice, this approach is only applicable in lower dimensions since in higher dimensions, a prohibitively large number of samples are required to construct a good global proposal PDF which can generate samples with reasonably high acceptance probability. In other words, if the sample size for the particular level is not large enough, most of the candidate samples generated by the proposal PDF will be rejected by the MH algorithm, leading to many repeated samples, slowing down greatly the exploration of the high probability region of the posterior PDF.

Ching et al. (2006) adopted Gibbs sampling (Geman and Geman 1984) to solve high-dimensional model updating problems that use linear structural models and modal data. Ching and Cheng (2007) proposed the Transitional Markov Chain Monte Carlo (TMCMC) algorithm and Muto and Beck (2008) applied it to the updating of hysteretic structural models. TMCMC adopts the idea as in Beck and Au (2002) of using a sequence of

intermediate PDFs such that the last PDF in the sequence is $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$. The main difference is in the way samples are simulated: TMCMC uses re-weighting and re-sampling techniques on the samples from a target PDF $\pi_i(\boldsymbol{\theta})$ in the sequence to generate initial samples for the next target PDF $\pi_{i+1}(\boldsymbol{\theta})$ in the sequence. A Markov chain of samples is initiated from each of these initial samples using the MH algorithm with stationary distribution $\pi_{i+1}(\boldsymbol{\theta})$: each sample is generated from a local random walk using a Gaussian proposal PDF centered at the current sample of the chain that has a covariance matrix estimated by importance sampling using samples from $\pi_i(\boldsymbol{\theta})$. TMCMC has several advantages over the previous approaches: 1) it is more efficient; 2) it allows the estimation of the normalizing constant c of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$, which is important for Bayesian model class selection (Beck and Yuen 2004). However, TMCMC has potential problems in higher dimensions, which need further attention: 1) the initial samples from re-weighting and re-sampling of samples in $\pi_i(\boldsymbol{\theta})$, in general, do not exactly follow $\pi_{i+1}(\boldsymbol{\theta})$, so the Markov chains must "burn-in" before samples follow $\pi_{i+1}(\boldsymbol{\theta})$, requiring a large amount of samples to be generated for each intermediate level; 2) in higher dimensions, convergence to $\pi_{i+1}(\boldsymbol{\theta})$ can be very slow when using the MH algorithm based on local random walks, as in TMCMC. This adverse effect becomes more pronounced as the dimension increases and it introduces more inaccuracy into the statistical estimates based on the samples.

In this chapter, we show how the Hybrid Monte Carlo method, also known as Hamiltonian Markov Chain method, can be used to solve higher-dimensional Bayesian model updating problems. Additional proof of the validity of the Hybrid Monte Carlo method using the Fokker-Planck equation is also provided. Features and parameters which affect the effectiveness of the Hybrid Monte Carlo method for higher-dimensional updating problems are discussed. Practical issues for feasibility of the method are addressed, and improvements are proposed to make it more effective and efficient for solving higher-dimensional model updating problems for complex dynamic systems. New formulae for Markov Chain convergence assessment are derived. The effectiveness of the proposed approach for Bayesian model updating of complex dynamic systems with many uncertain

parameters is illustrated with a simulate data example involving a 10-story building. Hybrid algorithms based on Markov Chain Monte Carlo simulation algorithms are presented at the end of the chapter. Part of the materials presented in this chapter are presented in Cheung and Beck (2007c;2008a).

2.1 Basic Markov Chain Monte Carlo simulation algorithms

2.1.1 Metropolis-Hastings algorithm and its features

The complete Metropolis-Hastings Algorithm for simulating samples from a target distribution $\pi(\boldsymbol{\theta})$ (where $\pi(\boldsymbol{\theta})$ need not be normalized) can be summarized as follows:

1. Initialize $\boldsymbol{\theta}^{(0)}$ by choosing it deterministically or randomly (see discussion in Section 4.3);
2. Repeat step 3 below for $i = 1,\ldots, N$.
3. In iteration i, let the most recent sample be $\boldsymbol{\theta}^{(i-1)}$, then do the following to simulate a new sample $\boldsymbol{\theta}^{(i)}$.
   i.) Randomly draw a candidate sample $\boldsymbol{\theta}_c$ from some proposal distribution $q(\boldsymbol{\theta}_c \,|\boldsymbol{\theta}^{(i-1)})$;
   ii.) Accept $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}_c$ with probability P_{acc} given as follows:

$$P_{acc} = \min\{1, \frac{\pi(\boldsymbol{\theta}_c) q(\boldsymbol{\theta}^{(i-1)} \,|\, \boldsymbol{\theta}_c)}{\pi(\boldsymbol{\theta}^{(i-1)}) q(\boldsymbol{\theta}_c \,|\, \boldsymbol{\theta}^{(i-1)})}\} \tag{2.3}$$

   If rejected, then $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$, i.e. the $(i\text{-}1)^{\text{th}}$ sample is repeated.

The proposal PDF $q(\boldsymbol{\theta}_c \,|\boldsymbol{\theta}^{(i)})$ should be of a form that allows an easy and direct drawing of $\boldsymbol{\theta}_c$ given $\boldsymbol{\theta}^{(i)}$. The choice of $\boldsymbol{\theta}^{(0)}$ and $q(\boldsymbol{\theta}_c \,|\boldsymbol{\theta}^{(i)})$ affects the convergence rate of the algorithm. The average acceptance probability of the candidate sample cannot be too low, or otherwise a significant number of repeated samples will be obtained, which slows down the convergence significantly and so may lead to biased results. Here the discussion is focused on the effect of the proposal PDF while the effect of $\boldsymbol{\theta}^{(0)}$ will be discussed in a later section.

The most common choice of $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(i)})$ is a symmetric proposal PDF in which $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(i)}) = q(\boldsymbol{\theta}^{(i)}|\boldsymbol{\theta}_c)$; for example, the local random walk Gaussian proposal PDF is popular, which is centered at the current sample $\boldsymbol{\theta}^{(i)}$ with some predetermined covariance matrix C. This proposal PDF allows a local exploration of the neighborhood of the current sample. Its main drawback is that in higher dimensions, it becomes infeasible to construct a proposal PDF which can explore the region of high probability content efficiently and effectively while at the same time maintaining a reasonable acceptance probability of the candidate sample. Another possible choice is the non-adaptive proposal PDF in which the simulation of the candidate sample is independent of the current sample, i.e., $q(\boldsymbol{\theta}_c|\boldsymbol{\theta}^{(i)}) = q(\boldsymbol{\theta}_c)$. For this type of proposal PDF to work, it has to be very similar to the target PDF. However, in general, the construction of such PDFs is infeasible in higher dimensions, even when some samples of the target PDF are available.

2.1.2 Gibbs Sampling algorithm and its features

Consider $\boldsymbol{\theta}$ as a composition of n vector components which do not need to be of the same dimension, i.e., $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_n]$, such that the conditional probability distribution $\pi(\boldsymbol{\theta}_j|\{\boldsymbol{\theta}_{-j}\})$ of $\boldsymbol{\theta}_i$ given all the other components is known. The complete algorithm of Gibbs sampling for simulating samples of a target distribution $\pi(\boldsymbol{\theta})$ (where $\pi(\boldsymbol{\theta})$ need not be normalized) can be summarized as follows:

1. Initialize $\boldsymbol{\theta}^{(0)}$ either deterministically or randomly;
2. Repeat step 3 below for $i = 1, \ldots, N$.
3. In iteration i, let the most recent sample be $\boldsymbol{\theta}^{(i-1)} = [\boldsymbol{\theta}_1^{(i-1)}, \boldsymbol{\theta}_2^{(i-1)}, \ldots, \boldsymbol{\theta}_n^{(i-1)}]$, then do the following to simulate a new sample $\boldsymbol{\theta}^{(i)} = [\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)}, \ldots, \boldsymbol{\theta}_n^{(i)}]$: for each j=1,2,…, n, randomly draw $\boldsymbol{\theta}_j^{(i)}$ from $\pi(\boldsymbol{\theta}_j^{(i)} | \boldsymbol{\theta}_1^{(i)}, \ldots, \boldsymbol{\theta}_{j-1}^{(i)}, \boldsymbol{\theta}_{j+1}^{(i-1)}, \ldots, \boldsymbol{\theta}_n^{(i-1)})$.

The Gibbs sampling algorithm generates a component of $\boldsymbol{\theta}$ from its conditional distribution given the current values of the other components. Gelman et al. (1995) show that the sequence of samples generated by the Gibbs sampling form a Markov Chain with the stationary distribution being the target distribution $\pi(\boldsymbol{\theta})$. Step 3 can be viewed as a special

case of the Metropolis-Hastings algorithm where the acceptance probability is 1 if $\pi(\boldsymbol{\theta}_j^{(i)} | \boldsymbol{\theta}_1^{(i)},\ldots, \boldsymbol{\theta}_{j-1}^{(i)}, \boldsymbol{\theta}_{j+1}^{(i-1)},\ldots,\boldsymbol{\theta}_n^{(i-1)})$ is in a form which allows direct and easy drawing of $\boldsymbol{\theta}_j^{(i)}$; if this is not the case, one can use, for example, the Metropolis-Hastings algorithm: draw a candidate $\boldsymbol{\theta}_j^c$ from some chosen proposal $q(\boldsymbol{\theta}_j^c | \boldsymbol{\theta}_1^{(i)},\ldots,\boldsymbol{\theta}_{j-1}^{(i)}, \boldsymbol{\theta}_j^{(i-1)}, \boldsymbol{\theta}_{j+1}^{(i-1)},\ldots, \boldsymbol{\theta}_n^{(i-1)})$ which allows easy and direct random drawing, and accept $\boldsymbol{\theta}_j^{(i)} = \boldsymbol{\theta}_j^c$ with probability P_{acc} where:

$$P_{acc} = \min\{1, \frac{\dfrac{\pi(\boldsymbol{\theta}_j^c | \boldsymbol{\theta}_1^{(i)},\ldots,\boldsymbol{\theta}_{j-1}^{(i)}, \boldsymbol{\theta}_{j+1}^{(i-1)},\ldots,\boldsymbol{\theta}_n^{(i-1)})}{\pi(\boldsymbol{\theta}_j^{(i-1)} | \boldsymbol{\theta}_1^{(i)},\ldots,\boldsymbol{\theta}_{j-1}^{(i)}, \boldsymbol{\theta}_{j+1}^{(i-1)},\ldots,\boldsymbol{\theta}_n^{(i-1)})}}{\dfrac{q(\boldsymbol{\theta}_j^c | \boldsymbol{\theta}_1^{(i)},\ldots,\boldsymbol{\theta}_{j-1}^{(i)}, \boldsymbol{\theta}_j^{(i-1)}, \boldsymbol{\theta}_{j+1}^{(i-1)},\ldots,\boldsymbol{\theta}_n^{(i-1)})}{q(\boldsymbol{\theta}_j^{(i-1)} | \boldsymbol{\theta}_1^{(i)},\ldots,\boldsymbol{\theta}_{j-1}^{(i)}, \boldsymbol{\theta}^c, \boldsymbol{\theta}_{j+1}^{(i-1)},\ldots,\boldsymbol{\theta}_n^{(i-1)})}}\} \tag{2.4}$$

If rejected, then $\boldsymbol{\theta}_j^{(i)} = \boldsymbol{\theta}_j^{(i-1)}$. It should be noted that the convergence of the Gibbs sampling algorithm can be slowed down if there is a strong correlation between components.

2.2 Hybrid Monte Carlo Method

Hybrid Monte Carlo Method (HMCM) was first introduced by Duane et al. (1987) as a MCMC technique for sampling from complex distributions by combining Gibbs sampling, MH algorithm acceptance rule and deterministic dynamical methods. By avoiding the local random walk behavior exhibited by the MH algorithm through the use of dynamical methods, HMCM can be much more efficient. The advantage of HMCM is even more pronounced when sampling the highly-correlated parameters from posterior distributions that are often encountered in Bayesian structural model updating. However, the potential of HMCM has not yet been explored in Bayesian structural model updating.

In HMCM, a fictitious dynamical system is considered in which auxiliary 'momentum' variables $\mathbf{p} \in \mathbb{R}^D$ are introduced and the uncertain parameters $\boldsymbol{\theta} \in \mathbb{R}^D$ in the target distribution $\pi(\boldsymbol{\theta})$ are treated as the variables for the displacement. The total energy

(Hamiltonian function) of the fictitious dynamical system is defined by: $H(\boldsymbol{\theta},\mathbf{p}) = V(\boldsymbol{\theta}) + W(\mathbf{p})$, where its potential energy $V(\boldsymbol{\theta}) = -\ln\pi(\boldsymbol{\theta})$ and its kinetic energy $W(\mathbf{p})$ depends only on $\mathbf{p}$ and some chosen positive definite 'mass' matrix $\mathbf{M} \in \mathbb{R}^{D\times D}$:

$$W(\mathbf{p}) = \mathbf{p}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{p}/2 \tag{2.5}$$

Since $\mathbf{M}$ can be chosen at our convenience, it is taken as a diagonal matrix with entries M_i, i.e., $\mathbf{M} = \mathrm{diag}(\underline{M_i})$. A joint distribution $f(\boldsymbol{\theta}, \mathbf{p})$ over the phase space $(\boldsymbol{\theta}, \mathbf{p})$ is considered:

$$f(\boldsymbol{\theta},\mathbf{p}) \;= K\exp(-H(\boldsymbol{\theta},\mathbf{p})) \tag{2.6}$$

where K is the normalizing constant. Clearly,

$$f(\boldsymbol{\theta},\mathbf{p}) \;= K\pi(\boldsymbol{\theta})\exp(-\mathbf{p}^{\mathrm{T}}\mathbf{M}^{-1}\mathbf{p}/2) \tag{2.7}$$

Note that $\pi(\boldsymbol{\theta})$ can be unnormalized (the usual situation that arises when constructing a posterior PDF) since its normalizing constant can be absorbed into K. Samples of $\boldsymbol{\theta}$ from $\pi(\boldsymbol{\theta})$ can be obtained if we can sample $(\boldsymbol{\theta}, \mathbf{p})$ from the joint distribution $f(\boldsymbol{\theta}, \mathbf{p})$ in (2.7). Note that (2.7) shows that $\mathbf{p}$ and $\boldsymbol{\theta}$ are independent and the marginal distributions of $\boldsymbol{\theta}$ and $\mathbf{p}$ are respectively $\pi(\boldsymbol{\theta})$ and $\mathcal{N}(\mathbf{0}, \mathbf{M})$, a Gaussian distribution with zero mean and covariance matrix $\mathbf{M}$.

Using Hamilton's equations, the evolution of $(\boldsymbol{\theta}, \mathbf{p})$ through fictitious time t is given by:

$$\frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \boldsymbol{\theta}} = -\nabla V(\boldsymbol{\theta}) \tag{2.8}$$

$$\frac{d\boldsymbol{\theta}}{dt} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{M}^{-1}\mathbf{p} \tag{2.9}$$

There are 4 features worth noting regarding the above evolution:

1. The total energy H remains constant throughout the evolution;
2. The dynamics are time reversible, i.e., if a trajectory initiates at ($\boldsymbol{\theta}$', $\mathbf{p}$') at time 0 and ends at ($\boldsymbol{\theta}$'', $\mathbf{p}$'') at time t, then a trajectory starting at ($\boldsymbol{\theta}$''**,** $\mathbf{p}$'') at time 0 will end at ($\boldsymbol{\theta}$', $\mathbf{p}$') at time $-t$ (or, equivalently, a trajectory starting at ($\boldsymbol{\theta}$'', **-**$\mathbf{p}$'') at time 0 will end at ($\boldsymbol{\theta}$', **-**$\mathbf{p}$') at time t).
3. The volume of a region of phase space remains constant (by Liouville's theorem).
4. The above evolution of ($\boldsymbol{\theta}$, $\mathbf{p}$) leaves $f(\boldsymbol{\theta}, \mathbf{p})$ in (2.7) as the stationary distribution (Duane et al. 1987); in particular, if $\boldsymbol{\theta}(0)$ follows the distribution $\pi(\boldsymbol{\theta})$, then after time t, $\boldsymbol{\theta}(t)$ also follows $\pi(\boldsymbol{\theta})$. Duane et al. (1987) proved this by showing the detailed balance condition for the stationarity of a Markov Chain is satisfied. In Appendix 2A, we provide an alternative proof to show that $f(\boldsymbol{\theta}, \mathbf{p})$ is actually the stationary distribution using the diffusionless Fokker-Planck equation.

If we start with $\boldsymbol{\theta}(0)$ and draw a sample $\mathbf{p}(0)$ from $\mathcal{N}(\mathbf{0}, \mathbf{M})$, then solve the Hamiltonian dynamics (2.8) and (2.9) for some time t, the final values ($\boldsymbol{\theta}(t)$, $\mathbf{p}(t)$) will provide an independent sample $\boldsymbol{\theta}(t)$ from $\pi(\boldsymbol{\theta})$. In practice, (2.8) and (2.9) have to be solved numerically using some time-stepping algorithm such as the commonly-used leapfrog algorithm (Duane et al. 1987). In this latter case, for time step δt, we have:

$$\mathbf{p}(t+\frac{\delta t}{2}) = \mathbf{p}(t) - \frac{\delta t}{2}\nabla V(\boldsymbol{\theta}(t)) \tag{2.10}$$

$$\boldsymbol{\theta}(t+\delta t) = \boldsymbol{\theta}(t) + \delta t \mathbf{M}^{-1}\mathbf{p}(t+\frac{\delta t}{2}) \tag{2.11}$$

$$\mathbf{p}(t+\delta t) = \mathbf{p}(t+\frac{\delta t}{2}) - \frac{\delta t}{2}\nabla V(\boldsymbol{\theta}(t+\delta t)) \tag{2.12}$$

Equations (2.10)-(2.12) can be reduced to:

$$\boldsymbol{\theta}(t+\delta t) = \boldsymbol{\theta}(t) + \delta t \mathbf{M}^{-1}[\mathbf{p}(t) - \frac{\delta t}{2}\nabla V(\boldsymbol{\theta}(t))] \tag{2.13}$$

$$\mathbf{p}(t+\delta t) = \mathbf{p}(t) - \frac{\delta t}{2}[\nabla V(\boldsymbol{\theta}(t)) + \nabla V(\boldsymbol{\theta}(t+\delta t))] \tag{2.14}$$

The gradient of V with respect to $\boldsymbol{\theta}$ needs to be calculated once only for each time instant since its value in the last step in the above algorithm at time t is the same as the first step at time $t+\delta t$.

2.2.1 HMCM algorithm

The complete algorithm of HMCM can be summarized as follows (for some chosen M, δt and L):

1. Initialize $\boldsymbol{\theta}_0$ (discussion of the choice of this is presented in a later section) and simulate $\mathbf{p}_0$ such that $\mathbf{p}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{M})$;
2. Repeat step 3 below for $i = 1, \ldots, N$.
3. In iteration i, let the most recent sample be $(\boldsymbol{\theta}_{i-1}, \mathbf{p}_{i-1})$, then do the following to simulate a new sample $(\boldsymbol{\theta}_i, \mathbf{p}_i)$:
   - i) Randomly draw a new momentum vector $\mathbf{p}'$ from $\mathcal{N}(\mathbf{0}, \mathbf{M})$;
   - ii) Initiate the leapfrog algorithm with $(\boldsymbol{\theta}(0), \mathbf{p}(0)) = (\boldsymbol{\theta}_{i-1}, \mathbf{p}')$ and run the algorithm for L time steps to obtain a new candidate sample $(\boldsymbol{\theta}'', \mathbf{p}'') = (\boldsymbol{\theta}(t+L\delta t), \mathbf{p}(t+L\delta t))$
   - iii) Accept $(\boldsymbol{\theta}_i, \mathbf{p}_i) = (\boldsymbol{\theta}'', \mathbf{p}'')$ with probability $P_{acc} = \min\{1, \exp(-\Delta H)\}$ where $\Delta H = H(\boldsymbol{\theta}'', \mathbf{p}'') - H(\boldsymbol{\theta}_{i-1}, \mathbf{p}')$. If rejected, then $(\boldsymbol{\theta}_i, \mathbf{p}_i) = (\boldsymbol{\theta}_{i-1}, \mathbf{p}')$, so $V(\boldsymbol{\theta}_i) = V(\boldsymbol{\theta}_{i-1})$ and $\nabla V(\boldsymbol{\theta}_i) = \nabla V(\boldsymbol{\theta}_{i-1})$.

2.2.2 Discussion of algorithm

Step 2(i) allows simulation of samples in regions with different H, thereby allowing the Markov chain to move to any point in the phase space of $(\boldsymbol{\theta}, \mathbf{p})$ via the deterministic step in 2(ii). This is an important step since it allows a global exploration of the $\boldsymbol{\theta}$ space in contrast to the local random walk behavior of the MH algorithm with a local proposal PDF. We can represent most integration algorithms used to solve Hamilton's equations by the following general iterative formulae:

$$(\boldsymbol{\theta}(n\delta t), \mathbf{p}(n\delta t)) = \mathbf{h}(\boldsymbol{\theta}((n-1)\delta t), \mathbf{p}((n-1)\delta t)) \tag{2.15}$$

where **h** corresponds to the mapping produced by the time-stepping algorithm, e.g., leap frog. The candidate sample ($\boldsymbol{\theta}_c$, $\mathbf{p}_c$) is then the output of the following:

$$(\boldsymbol{\theta}_c, \mathbf{p}_c) = \underbrace{\mathbf{h}(\mathbf{h}(...\mathbf{h}}_{L}(\boldsymbol{\theta}(0), \mathbf{p}(0))) = \underbrace{\mathbf{h}(\mathbf{h}(...\mathbf{h}}_{L}(\boldsymbol{\theta}(0), \mathbf{M}^{1/2}\mathbf{z})) \tag{2.16}$$

where **z** is a standard Gaussian vector with independent components $\mathcal{N}(0,1)$. Thus Steps 2(i) and (ii) together can be viewed as drawing a candidate sample from a global transition PDF which is non-Gaussian if the mapping **h** is nonlinear (the usual case). Applying mapping **h** multiple times leads to the exploration of the phase space further away from the current point, towards the higher probability region, avoiding the local random walk behavior of most MCMC methods. Therefore, HMCM can be viewed as a combination of Gibbs sampling (Step 2(i)) followed by a Metropolis algorithm step (Step 2(iii)) in an enlarged space with an implied complicated proposal PDF that enhances a more global exploration of the phase space than using a simple Gaussian PDF centered at the current sample, as adopted for the proposal PDF in the random walk Metropolis algorithm.

Although the leapfrog algorithm is volume preserving (sympletic) and time reversible, H does not remain exactly constant due to the systematic error introduced by the discretization of (2.8) and (2.9) with the leapfrog algorithm. To keep $f(\boldsymbol{\theta}, \mathbf{p})$ as the invariant PDF of the Markov chain, and thus keep $\pi(\boldsymbol{\theta})$ invariant, this systematic error needs to be corrected through the Metropolis acceptance/rejection step in Step 2(iii). The probability of acceptance, P_{acc}, in Step 2(iii) depends only on the difference in energy ΔH between H for the candidate sample ($\boldsymbol{\theta}$", $\mathbf{p}$") and H for ($\boldsymbol{\theta}_{i\text{-}1}$, $\mathbf{p}$'), which initiates the current leapfrog steps. The candidate sample ($\boldsymbol{\theta}$", $\mathbf{p}$") with lower H is always accepted while that with higher H is accepted with a probability of $\min\{1, \exp(-\Delta H)\}$.

It is worth noting that when L=1, HMCM is similar to an algorithm in which the evolution of $\boldsymbol{\theta}$ follows the following Itô stochastic differential equation:

$$d\boldsymbol{\theta}(t) = -\frac{1}{2}\mathbf{M}^{-1}\nabla V(\boldsymbol{\theta}(t))dt + \mathbf{M}^{-1/2}d\widetilde{\mathbf{W}}(t) \tag{2.17}$$

where $\widetilde{\mathbf{W}}(t) \in \mathbb{R}^D$ is a standard Wiener process. The discretized version corresponding to (2.17) is:

$$\boldsymbol{\theta}_{\mathrm{c}} = \boldsymbol{\theta}(t) - \frac{1}{2}\mathbf{M}^{-1}\nabla V(\boldsymbol{\theta}(t))\delta t + \mathbf{M}^{-1/2}\sqrt{\delta t}\mathbf{z} \tag{2.18}$$

where $\boldsymbol{\theta}_{\mathrm{c}}$ is the candidate sample and $\mathbf{z}$ is a standard Gaussian vector with independent components that are $\mathcal{N}(0,1)$. Thus, it is interesting to see that when L=1, the candidate sample of HMCM is drawn from the Gaussian proposal PDF:

$$q(\boldsymbol{\theta}_{\mathrm{c}} \mid \boldsymbol{\theta}(t)) = \frac{1}{(2\pi \mid \mathbf{C} \mid)^{D/2}} \exp(-\frac{1}{2}(\boldsymbol{\theta}_{\mathrm{c}} - \widetilde{\mu}(\boldsymbol{\theta}(t)))^T \mathbf{C}^{-1}(\boldsymbol{\theta}_{\mathrm{c}} - \widetilde{\mu}(\boldsymbol{\theta}(t)))) \tag{2.19}$$

where the mean $\widetilde{\mu}(\boldsymbol{\theta}(t))$ and the covariance matrix $\mathbf{C}$ are given by the following:

$$\widetilde{\mu}(\boldsymbol{\theta}(t)) = \boldsymbol{\theta}(t) + \frac{1}{2}\mathbf{M}^{-1}\delta t \nabla \ln \pi(\boldsymbol{\theta}(t)) \tag{2.20}$$

$$\mathbf{C} = \mathrm{E}[(\mathbf{M}^{-1/2}\sqrt{\delta t}\mathbf{z}(t))(\mathbf{M}^{-1/2}\sqrt{\delta t}\mathbf{z}(t))^T] = \delta t \mathbf{M}^{-1} \tag{2.21}$$

It can be seen from (2.20) that the above algorithm can reduce the tendency to do a local random walk by having a drift term that tends to force the Markov Chain samples towards the higher probability region of $\pi(\boldsymbol{\theta})$.

There are 3 parameters, namely **M**, δt and L, that need to be chosen before performing HMCM. If δt is chosen to be too large, the energy H at the end of the trajectory will deviate too much from the energy at the start of the trajectory which may lead to frequent rejections due to the Metropolis step in Step 2(iii). Thus, δt should be chosen small enough so that the average rejection rate due to the Metropolis step is not too large, but not too small that effective exploration of the high probability region is inhibited; a procedure for optimally choosing δt is presented later. For each dynamic evolution in the deterministic Step 2(ii), L can be randomly chosen from a discrete uniform distribution from 1 to some preselected L_{max} to avoid getting into a resonance condition (Mackenzie, 1989) (although it occurs rarely in practice) in which the trajectories from Step 2(ii) go around the same closed trajectory for a number of cycles. Matrix **M** can be chosen to be a diagonal matrix diag(M_1 ,…, M_D) where M_i is 1 for each i if the components of $\boldsymbol{\theta}$ are of comparable scale. This can be ensured by initially normalizing the uncertain parameters $\boldsymbol{\theta}$.

2.3 Proposed improvements to Hybrid Monte Carlo Method

2.3.1 Computation of gradient of $V(\boldsymbol{\theta})$ in implementation of HMCM

In general, $\nabla V(\boldsymbol{\theta}) = -\nabla \ln \pi(\boldsymbol{\theta})$ cannot be found analytically, so numerical methods must be used to find its value. The most common method uses finite differences. The computation of the gradient vector $\nabla V(\boldsymbol{\theta})$ using finite differences requires either D or $2D$ evaluations of V where D is the dimension of the uncertain parameters.

Here, we propose to use "algorithmic differentiation" (Rall, 1981; Kagiwada et al., 1986), in which a program code for sensitivity analysis (gradient calculation) can be created alongside the original program for an output analysis to form a combined code for both output analysis and sensitivity analysis. The program code for the output analysis can always be viewed as a composite of basic arithmetic operations and some elementary intrinsic functions. The main idea of "algorithmic differentiation" is to apply the chain rule

for differentiation judiciously to the elementary functions, the building blocks forming the program for output analysis, and to calculate the output and its sensitivity with respect to the input parameters simultaneously in one code. Unlike the classical finite difference methods which have truncation errors, one can obtain the derivatives within the working accuracy of the computer using algorithmic differentiation.

There are two ways in which the differentiation can be performed: forward differentiation or reverse differentiation. In forward differentiation, the differentiation is carried out following the flow of the program for the output analysis and performing the chain rule in the usual forward manner. To illustrate the idea behind the forward code differentiation, consider the following simple example for the program for computing the output function $y = h(\boldsymbol{\theta}) \in \mathbb{R}$:

$$w_j = \theta_j,\ j = 1, 2, ..., D$$

Repeat for $j=D+1,\ldots, p$

$$w_j = h_j(\{w_k\}_{k \in \{1,2,\ldots,j-1\}})$$

$$y = w_p$$

where h_j's can be elementary arithmetic operations or standard scalar functions on modern computer or mathematical softwares. The computation of the corresponding derivatives is practically free once the function itself has been computed. The corresponding code for computing the sensitivity $\mathbf{S}_y$ of y with respect to $\boldsymbol{\theta}$ is as follows:

$$w_j = \theta_j,\ j = 1, 2, ..., D$$

$$\nabla w_j = \mathbf{e}_j,\ j = 1, 2, ..., D$$

Repeat for $j=D+1,\ldots,p$

$$w_j = h_j(\{w_k\}_{k\in B_j\subseteq\{1,2,\ldots,j-1\}})$$

$$\nabla w_j = \sum_{k\in\{1,2,\ldots,j-1\}} \frac{\partial h_j}{\partial w_k}\nabla w_k$$

$$y = w_p$$

$$\mathbf{S}_y = \nabla w_p$$

where the forward derivative $\nabla w_j = [\partial w_j/\partial\theta_1, \partial w_j/\partial\theta_2, \ldots, \partial w_j/\partial\theta_D]^{\mathrm{T}}$ is the sensitivity of w_j with respect to $\boldsymbol{\theta}$ and $\mathbf{e}_j$ is a D-dimensional unit vector with the j-th component being 1 and all the other components being 0. Assuming the dimension of B_j is N_j and the calculation of each w_j requires at most KN_j arithmetic operations for some fixed constant K, here we can find the amount of computations required to calculate $\mathbf{S}_y$: KN_j+DN_j arithmetic operations are required to calculate each intermediate gradient vector ∇w_j. The total number of arithmetic operations for the calculation of $\mathbf{S}_y$ are $\sum_{j=D+1}^{p}(KN_j + DN_j)$ and that for the calculation of y are $\sum_{j=D+1}^{p} KN_j$. Thus the computational effort required by forward differentiation increases linearly with D. However, as mentioned earlier, forward differentiation does not incur errors as classical finite difference methods do and is accurate to the computer accuracy.

Wolfe (1982) asserted that if care was taken in handling quantities which are common to the function and the derivatives, the ratio of the cost of evaluating the gradient of a scalar function of n input variables and the scalar function itself is on average around 1.5, not n+1. Speelpenning's thesis (1980) proved that this assertion is actually true. Griewank (1989)

later showed that Wolfe's assertion is actually a theorem if the ratio, being on average 1.5, is replaced by an upper bound of 5. Rather than calculating the sensitivity of every intermediate variable with respect to the parameters **θ** as in forward differentiation, reverse differentiation is a form of algorithmic differentiation which starts with the output variables and computes the sensitivity of the output with respect to each of the intermediate variables. The biggest advantage of reverse differentiation is seen when the output variable is a scalar and the corresponding gradient with respect to high-dimensional input parameters is of interest. Under this circumstance, it has been shown (Griewank 1989) that the computational effort required by reverse differentiation to calculate the gradient accurately is only between 1 to 4 times of that required to calculate the output function, regardless of the dimension of the input parameters. This situation applies to our problem since the output variable of interest is the scalar function V.

To illustrate the idea behind the reverse differentiation, consider the same example as for forward differentiation. The code for computing the sensitivity $\mathbf{s}_y$ of y with respect to **θ** using reverse differentiation is as follows:

$w_j = \theta_j,\ j = 1, 2, ..., D$

$\tilde{w}_j = 0,\ j = 1, 2, ..., D$

Repeat for $j=D+1,\ldots, p$

$w_j = h_j(\{w_k\}_{k \in B_j \subseteq \{1,2,..,j-1\}})$

$\tilde{w}_j = 0$

$y = w_p$

$$\tilde{y} = 1$$

$$\tilde{w}_p = \tilde{y}$$

Repeat for j=p, p-1,…, D+1

$$\tilde{w}_k = \tilde{w}_k + \frac{\partial h_j}{\partial w_k}\tilde{w}_j, k \in B_j \subseteq \{1,2,...,j-1\}$$

$$\tilde{\theta}_j = \tilde{w}_j,\ j = 1,2,...,D$$

where $\tilde{y},\ \tilde{w}_j,\ \tilde{\theta}_j$ denotes the reverse derivatives $\partial y / \partial y,\ \partial y / \partial w_j,\ \partial y / \partial \theta_j$ respectively.

Thus $\mathbf{s}_y = [\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_D]$. The total number of arithmetic operations for the calculation of $\mathbf{s}_y$ are $\sum_{j=D+1}^{p}(KN_j + N_j)$ and that for the calculation of y are $\sum_{j=D+1}^{p} KN_j$. Thus the computational effort required by reverse differentiation is independent of D. It is noted that the approach presented above can be extended to compute higher-order derivatives.

Structural analysis programs usually involve program statements which perform vector and matrix operations and solve implicit linear equations. Higher-dimensional implicit linear equations are involved and the number of elementary intermediate variables required to store information for differentiation is large. Thus, it is more efficient to perform differentiation at the vector or matrix levels.

Recall that in our application, the output function is the scalar $V(\boldsymbol{\theta})$ and the input parameters are $\boldsymbol{\theta}$. For each of the most basic operations found in structural analysis programs, we have derived the corresponding operations necessary for reverse

differentiation at the vector or matrix levels (Appendix 2B). Those operations for the forward differentiation are very straightforward and obvious and no derivation will be given. Table 2.1 summarizes these operations. $\hat{\mathrm{Y}}$ denotes some matrix whose (*i,j*)-th entry is the forward partial derivative $\partial \mathrm{Y}_{ij} / \partial \theta_k$ of the (*i,j*)-th entry of a matrix Y with respect to some θ_k and $\tilde{\mathrm{Y}}$ denote some matrix whose (*i,j*)-th entry is the reverse partial derivative $\partial V / \partial \mathrm{Y}_{ij}$ of the output function *V* with respect to the (*i,j*)-th entry of Y. In the first column of Table 1, each equation carries out a certain operation inside the program. The left hand side of the equation in each of the row except the last row gives the intermediate output corresponding to the inputs on the right hand side which can in turn be the intermediate output resulting from the previous program statement. The last row shows an implicit equation for solving a certain intermediate output **v** given U and **w**. The second column shows the forward differentiation operations. The derivatives of the intermediate output with respect to some variable θ_k are computed given the values of the derivatives of the input with respect to the same variable, which are obtained from previous steps in the program. The third column shows the reverse differentiation operations. All the reverse partial derivatives are initialized to be zero at the beginning of the reverse differentiation. The reverse partial derivative of the output function *V* with respect to the intermediate input is incremented by the amount shown in the table given the values of the derivatives of the output function *V* with respect to the intermediate output that the input affects. For example, consider the two consecutive operations in the middle of a program:

$$\mathbf{w} = \mathbf{u} + \mathbf{v}$$
$$\mathbf{z} = \alpha \mathbf{u}$$

where α, **u** and **v** are the input vectors and **w** and **z** are the intermediate output vectors. Given $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{w}}$, we need to update $\tilde{\mathbf{u}}$ and $\tilde{\mathbf{v}}$. The corresponding reverse differentiation codes are as follows:

$$\tilde{\mathbf{u}} = \tilde{\mathbf{u}} + \alpha \tilde{\mathbf{z}}; \alpha = \alpha + \mathbf{u}^T \tilde{\mathbf{z}}$$
$$\tilde{\mathbf{u}} = \tilde{\mathbf{u}} + \tilde{\mathbf{w}}; \tilde{\mathbf{v}} = \tilde{\mathbf{v}} + \tilde{\mathbf{w}};$$

Based on the results developed above, a very efficient reverse differentiation code has been obtained for the case involving linear dynamical systems (Appendix 2B).

The idea of algorithmic differentiation can be extended to treat the case with nonsmooth intrinsic elementary functions (for example, those functions involving absolute signs and those problems involving hysteretic models). The ideas presented above could be incorporated in commercial structural analysis softwares to create a program code for a more accurate and efficient sensitivity analysis accompanying response analysis. The coding needs only one time effort, which can be made automatic by writing a program with the rules for "algorithm differentiation" developed above using object oriented programs such as Fortran, C, C++ or Matlab such that the code for sensitivity analysis can be created automatically given the original program code for response analysis. The idea is to write a command code to read the code for response analysis and then do the "translation" and creation of the differentiation code. It should be noted that the above methods can be easily extended if the sensitivity of a vector function is of interest.

Table 2.1 Some Basic operations of structural analysis program and the corresponding forward differentiation (FD) and reverse differentiation (RD) operations

Basic operations	FD operations	RD operations
$\mathbf{v} = \alpha \mathbf{u},\ \alpha \in \mathbb{R}; \mathbf{u}, \mathbf{v} \in \mathbb{R}^m$	$\hat{\mathbf{v}} = \alpha \hat{\mathbf{u}} + \hat{\alpha} \mathbf{u}$	$\tilde{\mathbf{u}} += \alpha \tilde{\mathbf{v}}$, $\tilde{\alpha} += \mathbf{u}^T \tilde{\mathbf{v}}$
$\mathbf{w} = \mathbf{u} + \mathbf{v},\ \mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathbb{R}^m$	$\hat{\mathbf{w}} = \hat{\mathbf{u}} + \hat{\mathbf{v}}$	$\tilde{\mathbf{u}} += \tilde{\mathbf{w}}, \tilde{\mathbf{v}} += \tilde{\mathbf{w}}$
$w = \mathbf{u}^T \mathbf{v},\ \mathbf{u}, \mathbf{v} \in \mathbb{R}^m; w \in \mathbb{R}$	$\hat{w} = \mathbf{u}^T \hat{\mathbf{v}} + \hat{\mathbf{u}}^T \mathbf{v}$	$\tilde{\mathbf{u}} += \tilde{w} \mathbf{v},\ \tilde{\mathbf{v}} += \tilde{w} \mathbf{u}$
$V = \alpha U,\ U, V \in \mathbb{R}^{p \times q};\ \alpha \in \mathbb{R}$	$\hat{V} = \hat{\alpha} U + \alpha \hat{U}$	$\tilde{\alpha} +=$ sum(sum(U.*V)), $\tilde{U} += \alpha \tilde{V}$ ***
$W = U + V,\ U, V, W \in \mathbb{R}^{p \times q}$	$\hat{W} = \hat{U} + \hat{V}$	$\tilde{U} += \tilde{W}, \tilde{V} += \tilde{W}$
$W = UV,\ U \in \mathbb{R}^{p \times q}, V \in \mathbb{R}^{q \times r}$	$\hat{W} = \hat{U}V + U\hat{V}$	$\tilde{U} += \tilde{W}V^T, \tilde{V} += U^T \tilde{W}$
* $\mathbf{w} = U\mathbf{v},\ U \in \mathbb{R}^{p \times q}, \mathbf{v} \in \mathbb{R}^{q \times 1}$	$\hat{\mathbf{w}} = \hat{U}\mathbf{v} + U\hat{\mathbf{v}}$	$\tilde{U} += \tilde{\mathbf{w}}\mathbf{v}^T, \tilde{\mathbf{v}} += U^T \tilde{\mathbf{w}}$
** $\mathbf{w} = U\mathbf{v},\ U \in \mathbb{R}^{p \times p}, \mathbf{v} \in \mathbb{R}^{p \times 1}$	$\hat{\mathbf{v}}$ is the solution of: $U\hat{\mathbf{v}} = \hat{\mathbf{w}} - \hat{U}\mathbf{v}$	$U^T \mathbf{y} = \tilde{\mathbf{v}}, \tilde{\mathbf{w}} += \mathbf{y}$, $\tilde{U} += -\mathbf{y}\mathbf{v}^T$

* Explicit equation for solving **w**

** Implicit equation for solving **v**

*** sum(sum(U.*V)) is a Matlab command where U.*V calculates a new matrix W whose (*i,j*) entry is the product of the (*i,j*) entries of U and V and sum(sum(W)) calculates the sum of all the elements in the matrix

2.3.2 Control of δt

The acceptance probability of a candidate sample at the end of the (**θ, p**) trajectory for the Hamiltonian dynamics of Equations (2.8) and (2.9) is influenced by the discretization errors introduced by the integration algorithm. The distance *d* moved in the (**θ, p**) space after one evolution depends on δt. In HMCM, δt should be chosen small enough so that the average rejection rate due to the Metropolis step is not too large. On the other hand, larger δt facilitates a bigger movement from the existing samples and so a better exploration of the phase space. Therefore, we want to choose δt which is as large as possible while at the same time maintaining a reasonable acceptance rate of the Metropolis step. This can be achieved by maximizing the expected distance $d(\delta t)$ moved by a sample with respect to δt:

$$d(\delta t) = (\delta t)\overline{P}_{\text{acc}}(\delta t) \tag{2.22}$$

where the average acceptance probability in HMCM, $\overline{P}_{\text{acc}}$, can be estimated by counting the proportion of distinct samples out of the amount of samples simulated. To do the above maximization, one can use a small number of samples and empirically explore different δt's to achieve maximum $d(\delta t)$ with δt chosen such that $\overline{P}_{\text{acc}} \geq p_0$ (say $p_0 = 0.1$).

2.3.3 Increasing the acceptance probability of samples

If the acceptance probability is increased for a fixed δt, then it will produce a reduction in the repetition of samples, thus improving the efficiency of exploration of the posterior PDF by the HMCM samples. In very high dimensions, one way to further increase the acceptance probability is to use more accurate higher-order symplectic integrators, such as

those in Forest and Ruth (1990), but at the expense of increased computational effort. Another variant is to utilize information in the trajectory samples when moving from ($\boldsymbol{\theta}_{i\text{-}1}$, $\mathbf{p}_{i\text{-}1}$) to ($\boldsymbol{\theta}_i$, $\mathbf{p}_i$) in Step 2 of HMCM (Neal 1994; Cheung and Beck 2007c) as follows.

When generating a trajectory from Hamiltonian equations, the original HMCM only considers the state generated in the last step (the L-th time step) as a candidate for a new sample. Therefore, another way to improve the acceptance probability is to consider most of the states along the trajectory generated by a symplectic integrator as possible candidates. Here we construct a new acceptance procedure for HMCM, which is a modification of that proposed by Neal (1994). The main idea is to consider two equal-sized windows of states in which there are W states, one around the current state $\mathbf{x}(0)$ and the other close to the end of the trajectory. One of the states in these windows will be the new sample $\tilde{\mathbf{x}}$. To maintain the invariance of $\pi(\boldsymbol{\theta})$, the position of $\mathbf{x}(0)=(\boldsymbol{\theta}(0), \mathbf{p}(0))$ within the window has to be randomly selected. To achieve this, an offset parameter K which is simulated from some fixed distribution is required. The modified acceptance procedure for a particular trajectory in the k-th iteration of HMCM is as follows:

1. Randomly draw a window size $W \in \mathbb{Z}^+$ from some fixed distribution (e.g., uniform distribution) such that $1 \le W \le L+1$ or simply fix W. Simulate an offset K uniformly from $\{0, 1, 2, \ldots, W-1\}$. Denote $\mathbf{x}(i)=(\boldsymbol{\theta}(i\delta t), \mathbf{p}(i\delta t))$. Simulate the direction λ for the trajectory with $\lambda = 1$ and $\lambda = -1$ being equally likely or simply fix λ at 1. Define index sets $\mathcal{V}_1$ and $\mathcal{V}_2$: $\mathcal{V}_1 = \{\lambda(L-K-W+1), \ldots, \lambda(L-K)\}$, $\mathcal{V}_2 = \{\lambda(-K), \ldots, \lambda(-K+W-1)\}$. Compute a trajectory $\mathcal{T}$ of length L: $\{\mathbf{x}(-\lambda K)), \ldots, \mathbf{x}(0), \ldots, \mathbf{x}(\lambda(L-K))\}$ and save the total energy values H_i corresponding to $\mathbf{x}(i)$ for $i \in \mathcal{V}_1 \cup \mathcal{V}_2$.

2. Let $H_{\mathcal{T}} = \min \{H_i\}$ for all $i \in \mathcal{V}_1 \cup \mathcal{V}_2$. The new sample $\tilde{\mathbf{x}}$ is equal to $\mathbf{x}(i)$ where i is drawn from the set $\mathcal{V}_1 \cup \mathcal{V}_2$ according to the probability mass function $p(i)$ as follows:

$$p(i) = \frac{(\mathrm{I}(i \in \mathcal{V}_1) + \mathrm{I}(i \in \mathcal{V}_2))\exp(-H_i + H_{\mathcal{T}})}{S_{\mathcal{T}}} \tag{2.23}$$

where $\mathrm{I}(.)$ is an indicator function which gives the value of 1 if the condition inside the parenthesis is true and gives 0 otherwise and $S_{\mathcal{T}}$ is the normalizing constant given by:

$$S_{\mathcal{T}} = \sum_{i \in \mathcal{V}_1 \cup \mathcal{V}_2} (\mathrm{I}(i \in \mathcal{V}_1) + \mathrm{I}(i \in \mathcal{V}_2))\exp(-H_i + H_{\mathcal{T}}) \tag{2.24}$$

It should be noted that the two windows will overlap if $W>(L+2)/2$ and $\mathrm{I}(i \in \mathcal{V}_1) + \mathrm{I}(i \in \mathcal{V}_2)$ will be equal to 2. When $W=1$, the above procedure reduces to the original HMCM algorithm which considers only the last state along the trajectory. When $W=L+1$, the above procedure reduces to a procedure which considers all the states along $\mathcal{T}$.

2.3.4 Starting Markov Chain in high probability region of posterior PDF

Starting the Markov chain with an initial point $\boldsymbol{\theta}_0$ closer to the important region of the posterior PDF can lead to more efficient exploration of this region. The following has been found to be effective:

The optimization of $V(\boldsymbol{\theta})$ (equivalently $\pi(\boldsymbol{\theta})$) to select $\boldsymbol{\theta}_0$ can be performed using an efficient SPSA (simultaneous perturbation stochastic approximation) optimization algorithm (Spall 1998a) with the use of common random numbers (Kleinman 1999). $\boldsymbol{\theta}_0$ is taken as the approximate optimal solution $\boldsymbol{\theta}^*$ obtained by the optimization algorithm. This method relies on the approximation of $\nabla V(\boldsymbol{\theta})$ using a two-sided perturbation as follows:

$$\frac{\partial V}{\partial \theta_i} \approx \frac{V(\boldsymbol{\theta} + h\boldsymbol{\Delta}) - V(\boldsymbol{\theta} - h\boldsymbol{\Delta})}{2h\Delta_i} \tag{2.25}$$

where $\boldsymbol{\Delta}$= [Δ_1, Δ_2,…, Δ_D] is the perturbation vector, the distribution of which is user-specified and h is a scalar which dictates the size of the perturbation of $\boldsymbol{\theta}$. A simple and valid choice for $\boldsymbol{\Delta}$ (Spall 1998b, Sadegh and Spall 1998) is to use a symmetric Bernoulli distribution: $P(\Delta_i = 1) = P(\Delta_i = -1) = 0.5$, for i=1,2,...,D.

In SPSA, all components of $\boldsymbol{\theta}$ are perturbed randomly and simultaneously and only 2 evaluations of the function V are required (instead of 2D evaluations required in the finite central difference method) to estimate the whole gradient vector ∇V. The optimization algorithm for determining an optimal point $\boldsymbol{\theta}^*$ is done by running the following recursive equation, starting with some initial guess $\boldsymbol{\theta}_0$:

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - a_k g_k(\boldsymbol{\theta}_k) \tag{2.26}$$

where $g_k(\boldsymbol{\theta}_k)$ is the estimate of the gradient of V evaluated at $\boldsymbol{\theta}_k$:

$$g_k(\boldsymbol{\theta}_k) = \frac{V(\boldsymbol{\theta}_k + b_k\boldsymbol{\Delta}_k) - V(\boldsymbol{\theta}_k - b_k\boldsymbol{\Delta}_k)}{2b_k\Delta_k} \tag{2.27}$$

$\boldsymbol{\Delta}_k$= [Δ_{k1}, Δ_{k2},…, Δ_{kD}] is the perturbation vector generated in the k-th iteration using the Bernoulli distribution as before; $a_k = a_0/(A+k+1)^{\alpha}$ and $b_k = b_0/(k+1)^{\gamma}$ are gain sequences which are critical to the performance of SPSA based optimization. Normalization of $\boldsymbol{\theta}$ is performed so that each component of $\boldsymbol{\theta}$ is of comparable scale. Some guidelines for the selection of the non-negative coefficients a_0, b_0, A, α and γ are provided in Spall (1998b). Common random numbers can be used to further improve the convergence of the above SPSA optimization algorithm (Kleinman et al. 1999). Another improvement is to use a second-order stochastic algorithm analogous to the deterministic Newton-Raphson algorithm (Spall 1997).

It should be noted that the approach presented in this section cannot solve the case involving well-separated regions with high probability content of the posterior PDF. On the other hand, with enough samples in previous levels, TMCMC can potentially provide initial points in different regions of high probability content of the posterior PDF by making use of multiple chains. However, the inherent convergence and efficiency problems of the random walk MH algorithm in higher dimensions still exists. One can incorporate HMCM proposed in this paper into TMCMC by replacing the random walk Metropolis algorithm in simulating from the whole sequence of PDFs or just the last PDF in the sequence. In practice, the case involving well-separated regions with high probability content of the posterior PDF is relatively rare.

2.3.5 Assessment of Markov Chain reaching stationarity

Given a finite set of N samples $\boldsymbol{\theta}^{(k)}$, $k = 1,2,\ldots, N$, from a Markov Chain distributed according to its stationary PDF $\pi(\boldsymbol{\theta})$, the estimate for the expectation of any function $\mathbf{g}(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$ is as follows:

$$\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})] = \int \mathbf{g}(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \approx \frac{1}{N}\sum_{k=1}^{N}\mathbf{g}(\boldsymbol{\theta}^{(k)}) \tag{2.28}$$

For example, if $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$, then $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})]$ will become the expected value of $\boldsymbol{\theta}$, i.e., $\mathrm{E}[\boldsymbol{\theta}]$. If the Markov chain is ergodic, the right-hand side of (2.28) converges almost surely to the left hand side for samples simulated using MCMC procedures such as the one presented in this paper (Tierney 1994). In this section, we first present a new approach to assess whether the samples $\boldsymbol{\theta}^{(k)}$, $k = 1,2,\ldots, N$, simulated using an MCMC algorithm, have converged to samples from its stationary PDF $\pi(\boldsymbol{\theta})$. Then, we examine how the accuracy of the estimator in (2.28) depends on the number of samples N.

A common existing approach for convergence assessment is based on observing whether the sample estimate of a certain $\mathrm{E}[\mathbf{g}(\boldsymbol{\theta})]$ stabilizes for some chosen function $\mathbf{g}$. However,

this can give misleading results since the stabilization can be a result of the chain of samples being trapped in some neighborhood of the parameter space (but the Markov Chain has not yet converged to the stationary distribution). Another major drawback of this approach is that it is hard to judge how far the underlying Markov Chain is away from reaching stationarity or convergence since one does not know a priori what value the estimate for E[$\mathbf{g}(\boldsymbol{\theta})$] should converge to.

To solve the above issues, we establish a known quantity depending on $\pi(\boldsymbol{\theta})$ which can also be estimated from the samples, then we check how far the estimate is from the exact value of the chosen quantity. Consider the quantity:

$$I_i = \mathrm{E}[g(\theta_i)] = \int g(\theta_i)\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{2.29}$$

where $g(\theta_i)$ is such that there exists some differentiable function $G(\theta_i)$ with $G'(\theta_i)=g(\theta_i)$. Recall that $\pi(\boldsymbol{\theta})=c^{-1}\exp(-V(\boldsymbol{\theta}))$. Denote $\boldsymbol{\theta}_{-i}$ as a vector containing all elements of $\boldsymbol{\theta}$ except θ_i; $\pi(\theta_i)$ as the marginal distribution of θ_i; and θ_i^u and θ_i^l as the upper limit and lower limit of the domain of integration with respect to θ_i, respectively. After performing integration by parts on I_i with respect to θ_i, an alternative expression for E[$g(\theta_i)$] can be obtained. If we divide this alternative expression by I_i as follows, Q_i should be equal to1:

$$Q_i = \frac{\int G(\theta_i)\dfrac{\partial V(\boldsymbol{\theta})}{\partial \theta_i}\pi(\boldsymbol{\theta})d\boldsymbol{\theta} + \int [G(\theta_i)\pi(\boldsymbol{\theta})]_{\theta_i\to\theta_i^l}^{\theta_i\to\theta_i^u} d\boldsymbol{\theta}_{-i}}{\int g(\theta_i)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} = 1 \tag{2.30}$$

The second term in the nominator can be expressed in terms of $\pi(\theta_i)$ as follows:

$$\int [G(\theta_i)\pi(\boldsymbol{\theta})]_{\theta_i\to\theta_i^l}^{\theta_i\to\theta_i^u} d\boldsymbol{\theta}_{-i} = [G(\theta_i)\int \pi(\boldsymbol{\theta})d\boldsymbol{\theta}_{-i}]_{\theta_i\to\theta_i^l}^{\theta_i\to\theta_i^u} = [G(\theta_i)\pi(\theta_i)]_{\theta_i\to\theta_i^l}^{\theta_i\to\theta_i^u} \tag{2.31}$$

Thus, (2.30) becomes:

$$Q_i = \frac{\int G(\theta_i)\frac{\partial V(\boldsymbol{\theta})}{\partial \theta_i}\pi(\boldsymbol{\theta})d\boldsymbol{\theta} + [G(\theta_i)\pi(\theta_i)]_{\theta_i \to \theta_i^l}^{\theta_i \to \theta_i^u}}{\int g(\theta_i)\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} = 1 \tag{2.32}$$

Denote $\theta_i^{(k)}$ as the *i*-th component of the *k*-th sample $\boldsymbol{\theta}^{(k)}$ from $\pi(\boldsymbol{\theta})$. The sample estimate $\tilde{Q}_i$ for Q_i is given by:

$$\tilde{Q}_i = \frac{\frac{1}{N}\sum_{k=1}^{N} G(\theta_i^{(k)})\frac{\partial V(\boldsymbol{\theta})}{\partial \theta_i}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}} + [G(\theta_i)\pi(\theta_i)]_{\theta_i \to \theta_i^l}^{\theta_i \to \theta_i^u}}{\frac{1}{N}\sum_{k=1}^{N} g(\theta_i^{(k)})} \tag{2.33}$$

It is convenient to choose $g(\theta_i)$=1 and thus (2.33) becomes:

$$\tilde{Q}_i = \frac{1}{N}\sum_{k=1}^{N} \theta_i^{(k)} \frac{\partial V(\boldsymbol{\theta})}{\partial \theta_i}|_{\boldsymbol{\theta}=\boldsymbol{\theta}^{(k)}} \tag{2.34}$$

where the second term in the numerator of (2.33) is dropped because usually, for model updating problems, $\pi(\theta_i)$ decays exponentially as θ_i approaches the limit of domain of integration. Asymptotically, all $\tilde{Q}_i$'s should converge to 1 with increasing *N*. With the above construction, we can define a quantity which averages over all Q_i's:

$$\bar{Q} = \sum_{i=1}^{D} Q_i / D \tag{2.35}$$

The exact value of $\bar{Q}$ is 1. The estimate $\tilde{\bar{Q}}$ for $\bar{Q}$ by simulation is obtained by averaging all $\tilde{Q}_i$'s:

$$\tilde{\bar{Q}} = \sum_{i=1}^{D} \tilde{Q}_i / D \tag{2.36}$$

In the example in this chapter, we assume that the Markov Chain is close enough to stationarity if the error of $\tilde{\bar{Q}}$ is less than a certain acceptable threshold, i.e, $|\tilde{\bar{Q}} - 1| < \varepsilon$.

2.3.6 Statistical accuracy of sample estimator

Now, let $\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]$ denote the estimator of $\mathrm{E}[g(\boldsymbol{\theta})]$ as in (2.28) for some function g. Let $\boldsymbol{\theta}^{(k)}$, $k = 1,2,\ldots, N$, denote samples from the stationary PDF $\pi(\boldsymbol{\theta})$. The statistical accuracy of the sample estimator $\tilde{\mathrm{E}}[g(\boldsymbol{\theta})] = \frac{1}{N}\sum_{k=1}^{N} g(\boldsymbol{\theta}^{(k)})$ can be assessed by evaluating the corresponding coefficient of variation (c.o.v.) $\tilde{\delta}_g$ which can be estimated using the following:

$$\tilde{\delta}_g = \frac{\sqrt{\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})])}}{\mathrm{E}[\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]]} \tag{2.37}$$

where the mean $\mathrm{E}[\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]]$ and variance $\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})])$ of the sample estimate can be estimated using the following (the derivation of $\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})])$ is shown in Appendix 2C):

$$\mathrm{E}[\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]] = \mathrm{E}[g(\boldsymbol{\theta})] \approx \tilde{\mathrm{E}}[g(\boldsymbol{\theta})] = \frac{1}{N}\sum_{k=1}^{N} g(\boldsymbol{\theta}^{(k)}) \tag{2.38}$$

$$\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]) = \frac{\rho(0)}{N}(1+\lambda) \tag{2.39}$$

$$\lambda = 2\sum_{\tau=1}^{N-1}(1-\frac{\tau}{N})\frac{\rho(\tau)}{\rho(0)} \in [0, N-1] \tag{2.40}$$

$$\begin{aligned}\rho(\tau) &= \mathrm{E}[(g(\boldsymbol{\theta}^{(k+\tau)}) - \mathrm{E}[g(\boldsymbol{\theta})])(g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])] \\ &\approx \frac{1}{N-\tau}\sum_{k=1}^{N-\tau}(g(\boldsymbol{\theta}^{(k+\tau)}) - \tilde{\mathrm{E}}[g(\boldsymbol{\theta})])(g(\boldsymbol{\theta}^{(k)}) - \tilde{\mathrm{E}}[g(\boldsymbol{\theta})])\end{aligned} \tag{2.41}$$

$\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})])$ is equal to the lower bound $\rho(0)/N$ (corresponding to λ=0) when the samples are independent (such as when using standard Monte Carlo simulation) while $\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})])$ is equal to the upper bound $\rho(0)$ (corresponding to λ=N-1) when the samples are perfectly correlated. The closer the value of λ is to zero, the less correlated the samples are. In fact, $N/(1+\lambda)$ can be viewed as the effective number of independent samples. Equations (2.38)–(2.41) can be used to estimate the c.o.v. for the estimator of $\mathrm{E}[g(\boldsymbol{\theta})]$ from N MCMC samples.

2.4 Illustrative example: Ten-story building

Suppose that noisy accelerometer data (simulated here) are available from a 10-story building excited by an earthquake. Two sets of data are considered: Dataset 1 are the acceleration data that are contaminated by a typical amount of noise (10% rms noise-to-signal ratio) used in published simulated data studies; Dataset 2 are the acceleration data that are contaminated by a large amount of noise (100% rms noise-to-signal ratio) to examine the robustness of the Bayesian procedure to extreme noise levels. System identification is to be performed using a 10-story linear lumped-mass shear-building model and so we estimate the mass m_i, damping coefficient c_i, and stiffness parameter k_i for each story, i=1,…,10. A duration of 10s (with a sample interval of 0.01s) of the total acceleration at the base, the first floor and the roof are measured. The measurements corresponding to dataset 1 and dataset 2 are shown in Figures 2.1 and 2.2 respectively. Let N_o=2 denote the number of observed degrees of freedom (first floor and roof) and N_T =1000 denote the

length of the discrete time history data. Let $\hat{y}_n(t_j)$ denote the corresponding measured output and $y_n(t_j;\boldsymbol{\theta})$, which satisfies the following equation of motion, denote the output at time t_j at the n-th observed degree of freedom predicted by the proposed structural model:

$$\mathbf{M_s}\ddot{\mathbf{y}}(t)+\mathbf{C_s}\dot{\mathbf{y}}(t)+\mathbf{K_s}\mathbf{y}(t)=-\mathbf{M}_s\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}a_g(t) \tag{2.42}$$

where the mass matrix $\mathbf{M_s}$, is a diagonal matrix diag(m_1 ,…, m_{10}); damping matrix $\mathbf{C_s}$ and stiffness matrix $\mathbf{K_s}$ are given by the following (the empty entries of $\mathbf{C_s}$ and $\mathbf{K_s}$ are zero):

$$\mathbf{C_s}=\begin{bmatrix} c_1+c_2 & -c_2 & & & & & \\ -c_2 & c_2+c_3 & -c_3 & & & & \\ & -c_3 & c_3+c_4 & -c_4 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -c_9 & c_9+c_{10} & -c_{10} & \\ & & & & -c_{10} & c_{10} & \end{bmatrix} \tag{2.43}$$

$$\mathbf{K_s}=\begin{bmatrix} k_1+k_2 & -k_2 & & & & & \\ -k_2 & k_2+k_3 & -k_3 & & & & \\ & -k_3 & k_3+k_4 & -k_4 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & -k_9 & k_9+k_{10} & -k_{10} & \\ & & & & -k_{10} & k_{10} & \end{bmatrix} \tag{2.44}$$

The prediction and measurement errors $\varepsilon_n(t_j)=\hat{y}_n(t_j)-y_n(t_j;\boldsymbol{\theta})$ for n=1,2,…,N_o and j=1,2,…, N_T, are modeled as independent and identically distributed Gaussian variables with mean zero and some unknown variance σ^2, based on the Principle of Maximum Entropy (Jaynes 2003). Altogether, we need to estimate 31 model parameters with σ included.

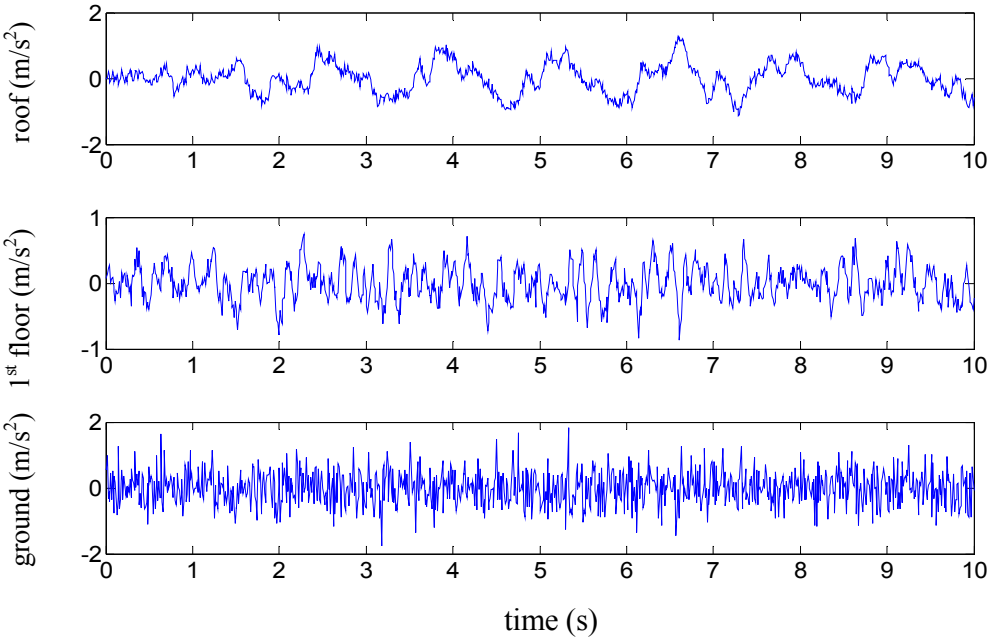

Figure 2.1: The acceleration dataset 1 in ten-story building

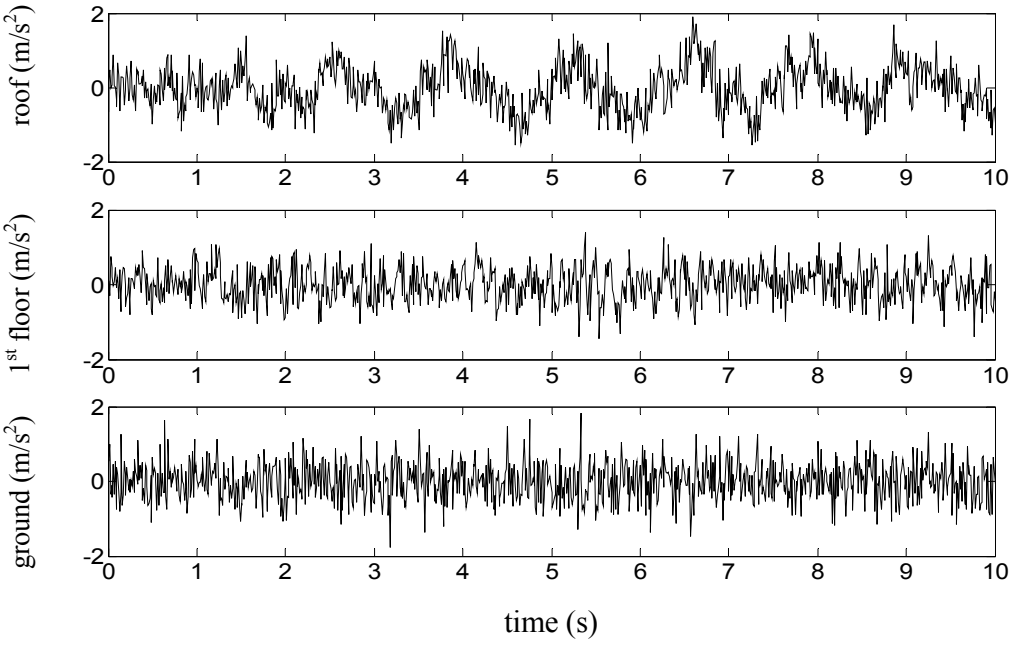

Figure 2.2: The acceleration dataset 2 in ten-story building

The likelihood function $p(\mathcal{D}|\,\boldsymbol{\theta})$ for this problem is:

$$p(\mathcal{D}\,|\,\boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{N_o N_T/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N_o}\sum_{j=1}^{N_T}[\hat{y}_n(t_j) - y_n(t_j;\boldsymbol{\theta})]^2\right) \tag{2.45}$$

Note that this updating problem is unidentifiable because the mass, stiffness and damping parameters can be uniformly scaled without changing the $y_n(t_j;\boldsymbol{\theta})$. The prior PDF for $\boldsymbol{\theta}$ is chosen to be independent distributions, that is, m_i, c_i, k_i follow a Gaussian distribution with means equal to their nominal values $m_0=2\times10^4$kg; $c_0=6\times10^4$ Nm^{-1}s, $k_0=2\times10^7$Nm^{-1}, and the corresponding coefficients of variation (c.o.v.) of 10%, 30%, 30% and σ follows a lognormal distribution with median $\sigma_0=1.0$ms^{-2} and a logarithmic standard deviation of $s_0=0.3$ (the c.o.v. is about 30%). These nominal values are not equal to the exact values, which are assumed to be unknown. For the mass parameters, relatively smaller values of c.o.v. are assumed since these parameters can usually be more accurately determined from the structural drawings than the other parameters. For each of the other parameters that are not so well known a priori, a larger c.o.v. is assumed. It should be noted the objective of the prior PDFs is to allow prior information to be incorporated when performing model updating. For those parameters where there is little prior information, prior PDFs that reflect higher uncertainty (i.e., in this case, larger c.o.v.) are used. Under such circumstances, the updated uncertainties for these parameters depend mostly on the data and are often insensitive to the prior PDFs. Here we define the dimensionless uncertain parameters θ_i, i=1,2…30, as the original parameters divided by their nominal values: $\theta_i=m_i/m_0$ for i=1,…,10; $\theta_i=c_{i-10}/c_0$ for i=11,…,20; $\theta_i=k_{i-20}/k_0$ for i=21,…,30 and $\theta_{31}=\sigma/\sigma_0$.

HMCM is applied by first doing 3000 evaluations of $\pi(\boldsymbol{\theta})$ for dataset 1 and 4000 evaluations of $\pi(\boldsymbol{\theta})$ for dataset 2 to find the initial point via the SPSA algorithm. The SPSA stopping criteria is such that each component of $\boldsymbol{\theta}$ and $\ln\pi(\boldsymbol{\theta})$ of the current iteration and the previous iteration differ by less than a prescribed threshold of 1%. Then 3000 HMCM samples are generated which are sufficient to reduce the error of $\tilde{\tilde{Q}}$ time-stepping to less

than ε=0.1, where $\tilde{\tilde{Q}}$ is evaluated using (2.34) and (2.36). In the HMCM, L is chosen to be an integer selected from a uniform distribution over the interval [0,40] and δt=0.0005 for dataset 1 and δt=0.0075 for dataset 2 to give an average probability of accepting candidate samples of about 0.8-0.9. The upper limit of L is chosen such that the correlation between the neighboring samples for each component is small (in this case, the correlation coefficient of the neighboring samples is less than 0.2).

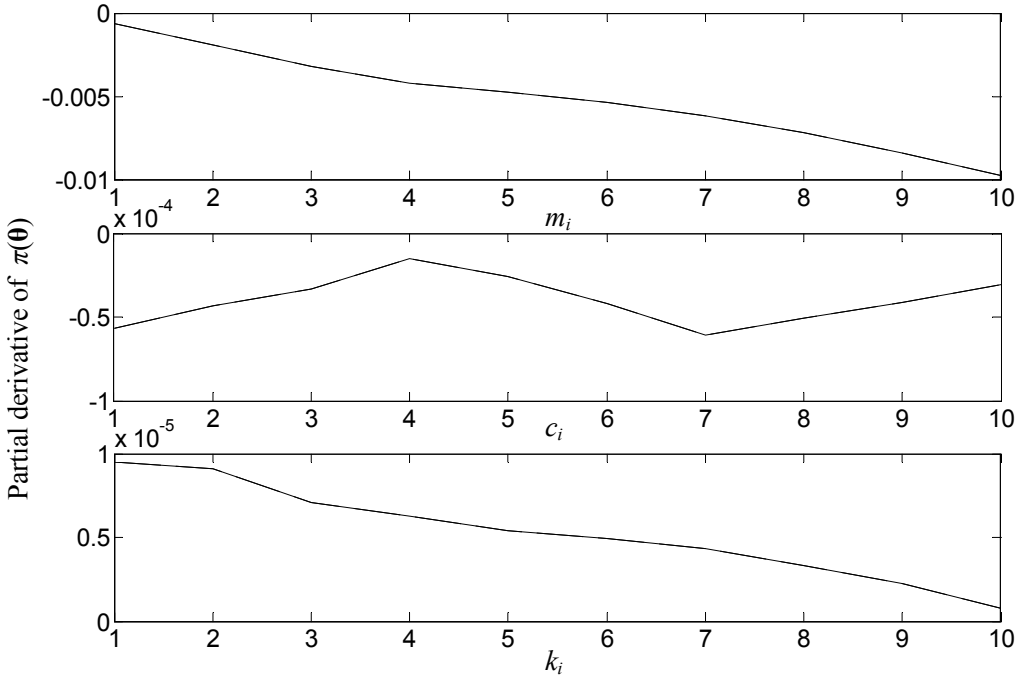

Figure 2.3: Gradient using two different methods: reverse algorithmic differentiation and central finite difference for mass parameters (top figure), damping parameters (middle figure) and stiffness parameters (bottom figure); the curves are indistinguishable

The partial derivative of $V(\boldsymbol{\theta})$ in HMCM with respect to θ_{31} can be determined analytically:

$$\frac{\partial V}{\partial \theta_{31}} = \frac{\partial V}{\partial \sigma}\frac{\partial \sigma}{\partial \theta_{31}} = \sigma_0 \left\{ \frac{N_o N_T}{\sigma} - \frac{1}{\sigma^3} \sum_{n=1}^{N_o} \sum_{j=1}^{N_T} [\hat{y}_n(t_j) - y_n(t_j;\boldsymbol{\theta})]^2 + \frac{1}{\sigma}\left(1 + \frac{\ln\sigma - \ln\sigma_0}{{s_0}^2}\right) \right\} \quad (2.46)$$

The remaining 30 components of the gradient of V with respect to $\boldsymbol{\theta}$ are calculated using the efficient reverse algorithmic differentiation code which was developed in this study. Figure 2.3 shows that the gradient computation using the reverse algorithmic differentiation (RD) overlaps with that obtained by central finite difference (CFD) with optimum perturbation size. It should be noted that the amount of computations required by CFD to calculate a gradient vector is 30 times that required by RD.

Table 2.2 shows the sample mean (column 3), sample c.o.v. (column 4) and estimation error (column 5) of the structural parameters, along with the exact values (column 2) of the parameters used to generate dataset 1. Compared with the prior uncertainty in the parameters, the posterior (updated) uncertainty is reduced since the data provide information about these parameters. There is a smaller degree of reduction in the uncertainty in the mass parameters than that in the damping and stiffness parameters. This is because the prior PDF for the mass parameters is closer to the corresponding posterior PDF than that for the other parameters. As expected, there is a higher uncertainty in the damping parameters than in the mass and stiffness parameters. This is because the modal contributions to the response are more sensitive to the mass and stiffness than to the damping. It can be seen that the estimation error is reasonably small: 0.3%-10.0% for mass parameters; 0.4%-13.7% for damping parameters; 0.75%-7.0% for stiffness parameters. Column 6 shows the magnitude of the error in terms of the number of standard deviations. It can be seen that the magnitude of error is less than 2 standard deviations for almost all parameters.

Table 2.3 shows the results using dataset 2, which is the large noise case. It can be seen that even in this case, the performance of Bayesian system identification is still good. In most cases, the errors in the stiffness parameters are significantly larger than dataset 1. The

results for the stiffness parameters are highly correlated with one another and are not jointly Gaussian. Figure 2.4 shows the samples plots for some pairs of θ_i corresponding to the stiffness. It can be seen clearly that the stiffness parameters are not jointly Gaussian. Figures 2.5 and 2.6 show plots where posterior samples for some k_i/k_0 and $\ln(k_i/k_0)$, respectively, are plotted on Gaussian probability paper. If the samples essentially lie on a straight line in these plots, the posterior marginal distribution of θ_i can be taken to be approximately Gaussian or lognormal, respectively. From the figures, it can be seen that the marginal distribution for some stiffness parameters (for example, k_2, k_8, k_9, k_{10}, etc.) are non-Gaussian and also not log-normal. The multivariate Gaussian approximation of the posterior PDF that is effectively assumed in Bayesian updating using Laplace's asymptotic approximation (Beck and Katafygiotis, 1991 and 1998), is not so good here because the few observed locations (N_o=2), high noise-to-signal ratio (100%) and many parameters (31) make the problem unidentifiable (e.g. see Figure 2.4). Being able to capture the non-Gaussian behavior of the posterior PDF is essential for robust prediction of the future response and reliability of structures (Cheung and Beck 2007a).

To illustrate the predictive power and robustness of the Bayesian model updating approach using HMCM, we compare the exact time histories of the total acceleration (Figure 2.7), the displacement (Figure 2.8) and the interstory drift (Figure 2.9) of some *unobserved* floors with the corresponding mean response from the robust predictive PDF given by equation (1.5). The solid curve shows the exact values of the response; the dashed curve shows the mean robust response estimated by averaging over the mean responses from each of the posterior samples. The two dotted curves give the responses that are twice the standard deviation of the predicted robust response from the mean robust response. The curves for the exact and the mean total acceleration, displacement and drift responses are almost indistinguishable. Also, all figures show that the exact response lies almost always between the two dotted-dashed curves. It can be seen that Bayesian robust analyses are able to give robust prediction of the response even at the unobserved degrees of freedom, despite the fact that the model is unidentifiable based on very noisy data. The total

acceleration being a linear combination of displacements and velocities has its uncertainty contributed by both displacements and velocities while the interstory drift being the difference of the displacement of the two neighboring floors has its uncertainty contributed by the displacement of the two floors. Thus higher uncertainties can be found in predicting the total acceleration and interstory drift than predicting the floor displacement.

Table 2.2 Statistical results for structural parameter estimates for 10% noise-to-signal ratio [Dataset 1]

Parameter	Exact Value β_i	μ_i=Mean estimate of parameter	σ_i / μ_i =c.o.v estimate of parameter	Error= $\lvert\beta_i - \mu_i\rvert/\beta_i$	$\lvert\mu_i-\beta_i\rvert/\sigma_i$
1 m_1	1.92×10^4	2.00×10^4	3.2%	3.8%	1.16
2 m_2	1.97×10^4	2.06×10^4	5.2%	4.4%	0.82
3 m_3	1.95×10^4	1.95×10^4	7.2%	0.3%	0.04
4 m_4	2.06×10^4	2.00×10^4	5.9%	3.0%	0.52
5 m_5	2.05×10^4	2.02×10^4	5.4%	1.1%	0.21
6 m_6	1.98×10^4	2.01×10^4	6.3%	1.8%	0.29
7 m_7	1.94×10^4	1.91×10^4	6.8%	1.0%	0.14
8 m_8	2.06×10^4	2.00×10^4	9.1%	2.7%	0.30
9 m_9	1.90×10^4	2.08×10^4	7.3%	9.9%	1.23
10 m_{10}	2.01×10^4	2.18×10^4	5.4%	8.6%	1.47
11 c_1	7.70×10^4	8.62×10^4	5.9%	12.0%	1.81
12 c_2	7.78×10^4	8.20×10^4	7.9%	5.4%	0.66
13 c_3	7.86×10^4	7.70×10^4	12.0%	2.0%	0.17
14 c_4	7.28×10^4	7.46×10^4	8.8%	2.4%	0.27
15 c_5	7.19×10^4	8.18×10^4	5.6%	13.7%	2.15
16 c_6	7.37×10^4	7.07×10^4	8.4%	4.0%	0.50
17 c_7	7.10×10^4	7.77×10^4	10.4%	9.3%	0.82
18 c_8	7.11×10^4	6.20×10^4	10.1%	12.8%	1.46
19 c_9	6.90×10^4	6.93×10^4	13.8%	0.4%	0.03
20 c_{10}	7.57×10^4	6.63×10^4	7.2%	12.4%	1.97
21 k_1	2.16×10^7	2.24×10^7	3.4%	4.0%	1.14
22 k_2	1.74×10^7	1.76×10^7	4.6%	0.8%	0.16
23 k_3	2.04×10^7	2.07×10^7	7.4%	1.7%	0.22
24 k_4	1.99×10^7	2.09×10^7	4.7%	5.0%	1.00
25 k_5	1.74×10^7	1.86×10^7	5.5%	6.5%	1.11
26 k_6	1.68×10^7	1.74×10^7	6.8%	3.3%	0.48
27 k_7	1.87×10^7	1.89×10^7	7.3%	0.9%	0.12
28 k_8	1.77×10^7	1.89×10^7	9.8%	7.0%	0.66
29 k_9	1.84×10^7	1.86×10^7	8.7%	1.0%	0.11
30 k_{10}	1.72×10^7	1.64×10^7	5.3%	4.6%	0.92
31 σ	0.040	0.041	1.6%	2.5%	1.49

Table 2.3 Statistical results for structural parameter estimates for 100% noise-to-signal ratio [Dataset 2]

Parameter	Exact Value β_i	μ_i=Mean estimate of parameter	σ_i / μ_i =c.o.v estimate of parameter	Error= $\lvert\beta_i - \mu_i\rvert/\beta_i$	$\lvert\mu_i-\beta_i\rvert/\sigma_i$
1 m_1	1.92×10^4	1.95×10^4	7.3%	1.2%	0.17
2 m_2	1.97×10^4	2.02×10^4	9.3%	2.3%	0.24
3 m_3	1.95×10^4	1.95×10^4	9.0%	0.2%	0.02
4 m_4	2.06×10^4	2.07×10^4	9.5%	0.4%	0.04
5 m_5	2.05×10^4	1.95×10^4	9.4%	5.0%	0.53
6 m_6	1.98×10^4	2.04×10^4	9.5%	3.0%	0.31
7 m_7	1.94×10^4	2.00×10^4	9.6%	3.2%	0.32
8 m_8	2.06×10^4	1.98×10^4	10.3%	3.7%	0.37
9 m_9	1.90×10^4	1.91×10^4	10.1%	1.1%	0.08
10 m_{10}	2.01×10^4	2.05×10^4	10.1%	2.4%	0.23
11 c_1	7.70×10^4	7.45×10^4	20.0%	3.2%	0.17
12 c_2	7.78×10^4	6.86×10^4	22.3%	12.0%	0.61
13 c_3	7.86×10^4	6.82×10^4	23.7%	13.3%	0.65
14 c_4	7.28×10^4	5.92×10^4	27.9%	18.7%	0.83
15 c_5	7.19×10^4	5.96×10^4	30.3%	17.2%	0.68
16 c_6	7.37×10^4	6.13×10^4	27.4%	16.9%	0.74
17 c_7	7.10×10^4	7.14×10^4	25.6%	1.0%	0.02
18 c_8	7.11×10^4	6.67×10^4	26.2%	6.3%	0.25
19 c_9	6.90×10^4	6.06×10^4	28.5%	12.2%	0.49
20 c_{10}	7.57×10^4	6.79×10^4	24.6%	10.4%	0.47
21 k_1	2.16×10^7	2.05×10^7	9.1%	4.8%	0.56
22 k_2	1.74×10^7	1.50×10^7	10.3%	13.7%	1.54
23 k_3	2.04×10^7	1.97×10^7	14.9%	3.2%	0.22
24 k_4	1.99×10^7	2.29×10^7	14.5%	15.1%	0.90
25 k_5	1.74×10^7	2.24×10^7	17.4%	28.7%	1.28
26 k_6	1.68×10^7	1.99×10^7	20.0%	18.3%	0.77
27 k_7	1.87×10^7	1.93×10^7	21.0%	3.1%	0.14
28 k_8	1.77×10^7	1.99×10^7	19.5%	12.4%	0.57
29 k_9	1.84×10^7	1.82×10^7	20.3%	1.5%	0.07
30 k_{10}	1.72×10^7	1.74×10^7	31.8%	2.3%	0.02
31 σ	0.400	0.395	1.6%	1.1%	0.73

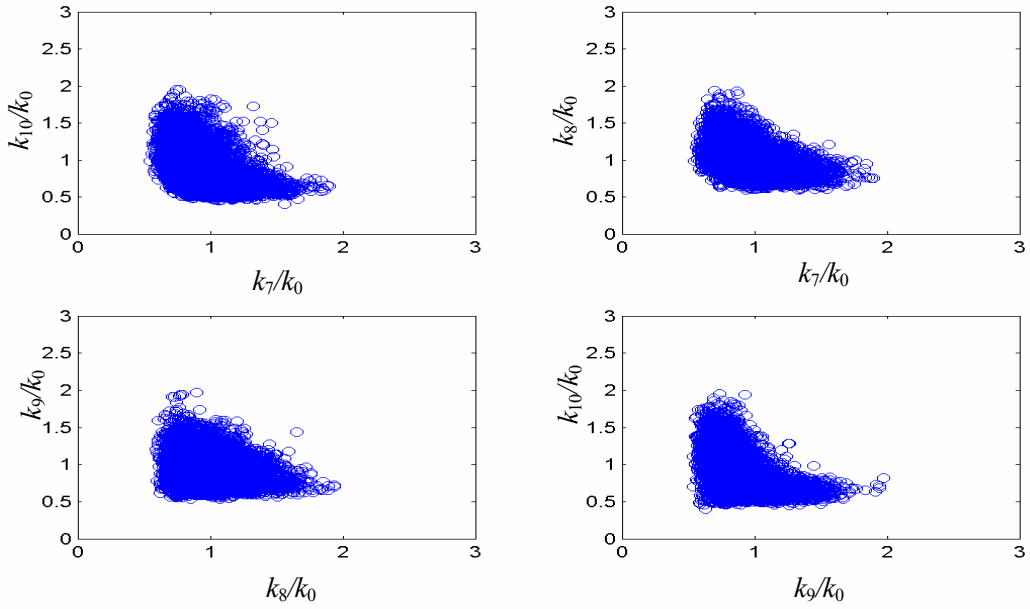

Figure 2.4: Pairwise posterior sample plots for some stiffness parameters

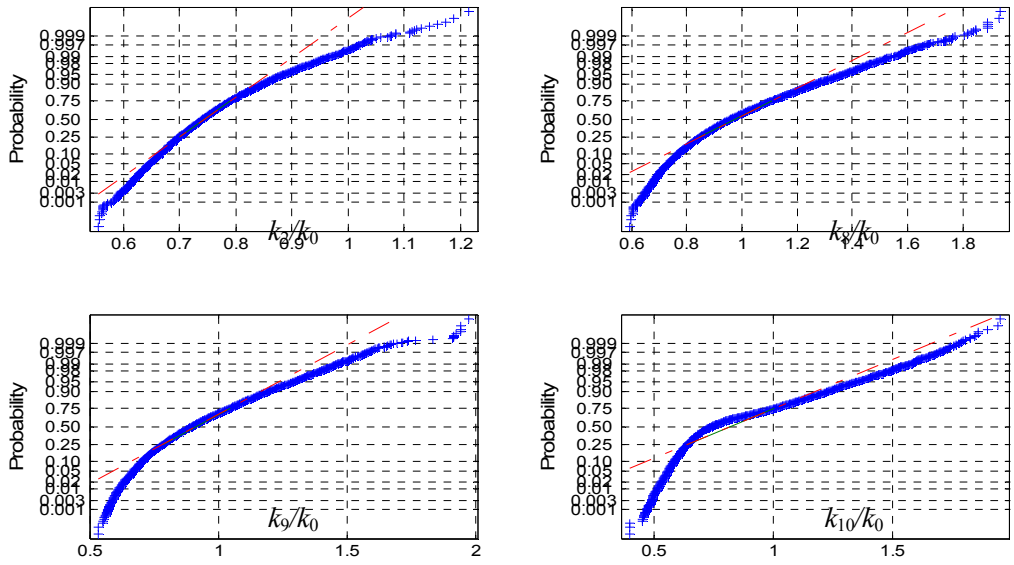

Figure 2.5: Gaussian probability paper plots for some k_i

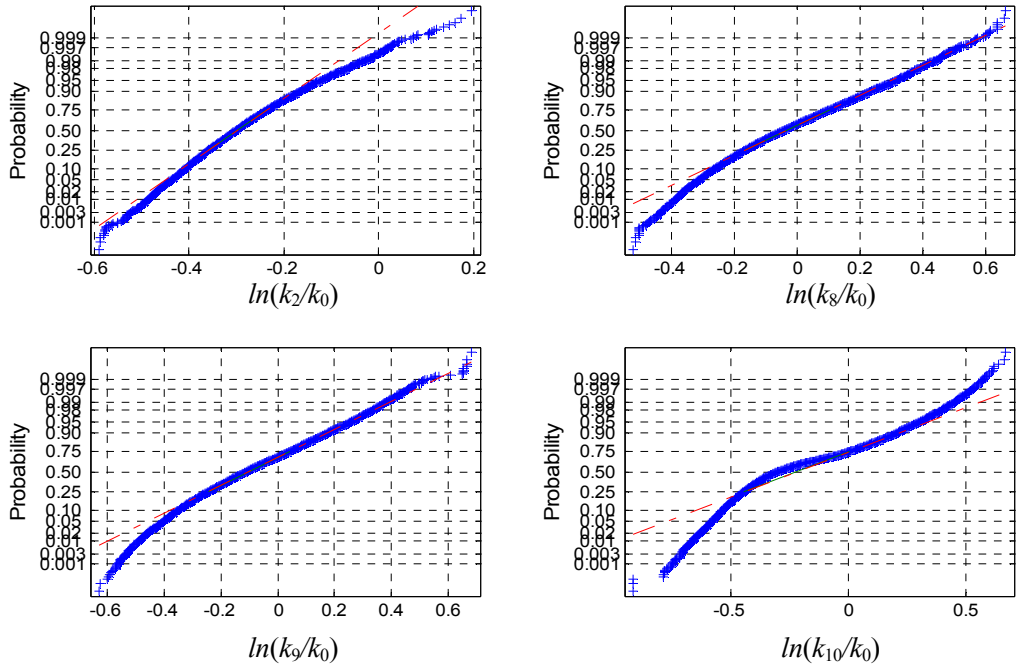

Figure 2.6: Gaussian probability paper plots for some lnk_i

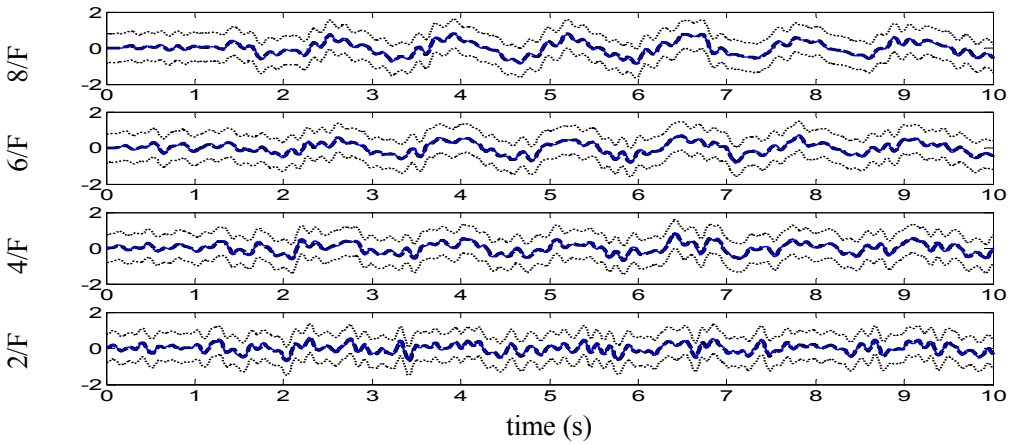

Figure 2.7: The exact (solid) and mean predicted (dashed) time histories of the total acceleration (m/s^2) at some unobserved floors together with time histories of the total acceleration that are twice the standard deviation of the predicted robust response from the mean robust response (dotted) [Dataset 2]

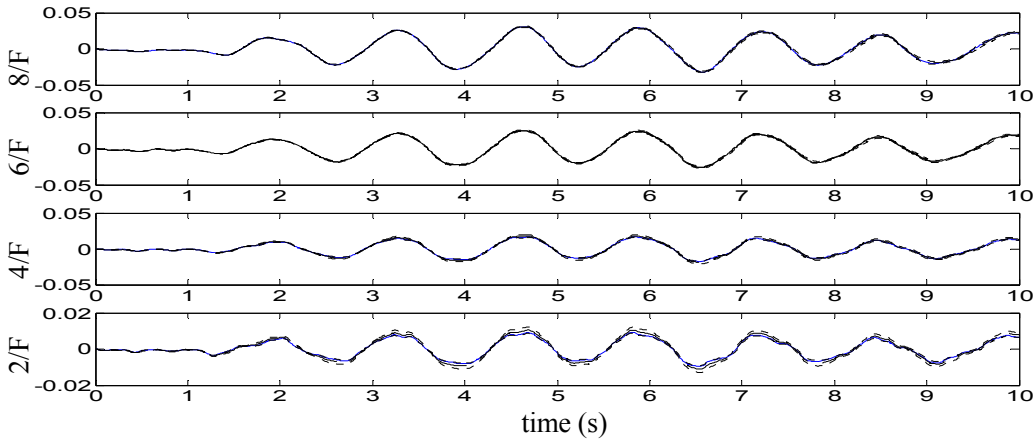

Figure 2.8: The exact (solid) and mean (dashed) time histories of the displacement (m) at some unobserved floors together with time histories of the displacement that are twice the standard deviation of the predicted robust response from the mean robust response (dotted) [Dataset 2]

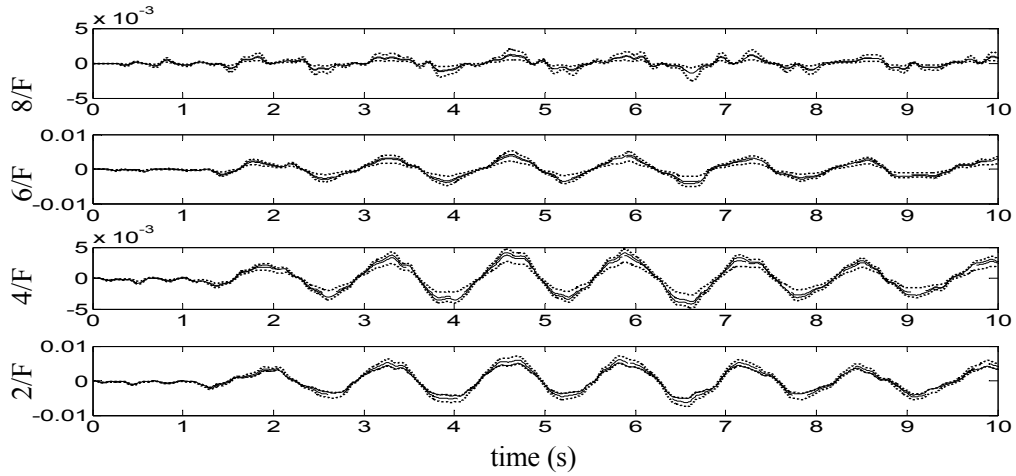

Figure 2.9: The exact (solid) and mean (dashed) time histories of the interstory drift (m) at some unobserved floors together with time histories of the interstory drift that are twice the standard deviation of the predicted robust response from the mean robust response (dotted) [Dataset 2]

The building considered in this example has nonclassical damping and thus possesses complex modes. Table 2.4 gives the sample mean (with sample c.o.v. inside the parenthesis) of the natural frequency (column 4) and damping ratio (column 5) for each complex mode along with the exact values of the natural frequency and damping ratio (columns 2 and 3). It can be seen that the Bayesian analysis is able to give robust estimates for these modal parameters of the underlying structure despite the large noise and lack of identifiability in the structural parameters. As expected, the estimates for the lower modes are better than those for the higher modes (as can be seen from the higher sample c.o.v. for the parameters corresponding to higher modes) because only the first few complex modes of the structure are excited significantly by the earthquake ground motion, so it is the information from these modes that are primarily utilized in the estimation of the shear-building model parameters. However, the higher-mode frequencies and damping parameters are still quite accurately estimated, presumably because the tridiagonal shear-building stiffness and damping matrices induce strong constraints on the modal parameters.

Table 2.4 The exact natural frequency and damping ratio for each complex mode [Dataset 2]

Complex Mode	Natural frequency (Hz)	Damping ratio (%)	Natural frequency (Hz) from Bayesian updating	Damping ratio (%) from Bayesian updating
1	0.735	0.92	0.734(0.2%)	0.85(8.0%)
2	2.158	2.71	2.149(0.3%)	2.60(7.1%)
3	3.562	4.45	3.600(0.7%)	4.03(9.5%)
4	4.891	6.03	4.878(0.8%)	5.83(8.6%)
5	6.047	7.65	6.022(1.8%)	7.33(8.8%)
6	7.106	9.11	7.214(2.3%)	8.42(10.1%)
7	8.049	10.13	7.990(2.4%)	9.17(11.5%)
8	8.620	11.11	8.828(2.7%)	9.56(13.1%)
9	9.306	11.58	9.661(3.2%)	9.60(13.5%)
10	9.631	11.92	10.519(4.5%)	9.26(15.5%)

2.5 Multiple-Group MCMC

Assume $\boldsymbol{\theta}$ is divided into G groups, i.e. $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_G]$. Given a current sample $\boldsymbol{\theta}$, a new sample $\boldsymbol{\theta}^* = [\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_G^*]$ from a target distribution $\pi(\boldsymbol{\theta})$ is generated by repeating the following starting with j=1 until j=G:

1. Generate the j-th group $\boldsymbol{\theta}_j^*$ of the new sample from transition PDF $K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$ with the corresponding stationary PDF $\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$ where $\{\boldsymbol{\theta}^*\}_{1:j-1} = [\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_{j-1}^*]$, $\{\boldsymbol{\theta}\}_{j+1:G} = [\boldsymbol{\theta}_{j+1}, \ldots, \boldsymbol{\theta}_G]$, $\{\boldsymbol{\theta}^*\}_{1:j-1} = \{\}$ if j-1<1 and $\{\boldsymbol{\theta}\}_{j+1:G} = \{\}$ if j+1>G;
2. j=j+1.

The above procedure is valid as soon as $K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$ satisfies the local stationarity condition:

$$\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) = \int K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) d\boldsymbol{\theta}_j \tag{2.47}$$

The validity of the above procedures is proved by showing the satisfaction of the stationarity condition in Appendix 2D.

Special case 1:

$$\begin{aligned} &K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \\ &= T_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) + [1 - a_j(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \delta(\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_j) \end{aligned} \tag{2.48}$$

where

$$T_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) = r_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) q_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \tag{2.49}$$

$$
\begin{aligned}
& r_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \\
&= \min\{1, \frac{\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) q_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G})}{\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) q_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})}\} \\
&= \min\{1, \frac{\pi(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G}) q_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G})}{\pi(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) q_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})}\}
\end{aligned}
\tag{2.50}
$$

$$
a_j(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) = \int T_j(\widetilde{\boldsymbol{\theta}_j} \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) d\widetilde{\boldsymbol{\theta}_j} \tag{2.51}
$$

The above transition PDF corresponds to a Metropolis Hasting (MH) algorithm with proposal PDF $q_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$ with stationary PDF $\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$. This transition PDF in (2.48) is shown to satisfy (2.47) in Appendix 2E.

The algorithm for simulating the *j*-the group $\boldsymbol{\theta}_j^*$ is:

1. Draw a candidate $\boldsymbol{\theta}_j^c$ from the proposal PDF $q_j(\boldsymbol{\theta}_j^c \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$ and accept $\boldsymbol{\theta}_j^* = \boldsymbol{\theta}_j^c$ with probability $r_j(\boldsymbol{\theta}_j^c \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$ given in (2.50);
2. If rejected, then $\boldsymbol{\theta}_j^* = \boldsymbol{\theta}_j$.

It should be noted that the original Gibbs sampling is a special case of the above where $q_j(\boldsymbol{\theta}_j^c \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) = \pi(\boldsymbol{\theta}_j^c \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$ and $r(\boldsymbol{\theta}_j^c \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) = 1$. $\boldsymbol{\theta}_j^* = \boldsymbol{\theta}_j^c$ is always accepted for all *j*. The proposal PDF $q_j(\boldsymbol{\theta}_j^c \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) = q_j(\boldsymbol{\theta}_j^c \mid \boldsymbol{\theta}_j)$ and $q_j(\boldsymbol{\theta}_j^c \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) = q_j(\boldsymbol{\theta}_j^c)$ are some of the simple special cases.

Simulation procedures such as Gibbs sampling, MH or HMCM applied to each group of uncertain parameters as above is valid since the corresponding transition PDFs for each group satisfy (2.47). If MH, Gibbs sampling and HMCM are used for each group, reversibility (detailed balanced condition) is satisfied for each group of uncertain

parameters. However, as shown in Appendix 2F, in general, reversibility is not satisfied for the whole uncertain parameter vector $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_G]$ even if reversibility is satisfied for each group (i.e., the Markov Chain samples from $\pi(\boldsymbol{\theta})$ generated using the above procedures is not reversible) and $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_G$ are statistically independent, i.e. the target PDF can be expressed as, $\pi(\boldsymbol{\theta}) = \prod_{j=1}^{G} \pi_j(\boldsymbol{\theta}_j)$

2.6 Transitional multiple-group hybrid MCMC

For a general case which may involve i) well-separated high probability regions; ii) high-dimensional uncertain parameters or iii) may be unidentiable, a powerful stochastic simulation algorithm for generating samples from the posterior PDF $\pi(\boldsymbol{\theta})$ can be obtained by combining TMCMC and multi-group MCMC algorithms as follows. This hybrid algorithm is applied to problems in later chapters. Consider a sequence of intermediate PDFs $\pi_l(\boldsymbol{\theta})$ for l=0,1,…, L, such that the first and last PDFs, $\pi_0(\boldsymbol{\theta})$ and $\pi_L(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})$, in the sequence are the prior $p(\boldsymbol{\theta}|\mathcal{M}_j)$ and posterior $p(\boldsymbol{\theta}|D,\mathcal{M}_j)$, respectively:

$$\pi_l(\boldsymbol{\theta}) \propto p^{\tau_l}(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j)\,p(\boldsymbol{\theta}\,|\,\mathcal{M}_j) \tag{2.52}$$

where $0=\tau_0<\tau_1<\ldots<\tau_L=1$.

First, N_0 samples are generated from the prior $p(\boldsymbol{\theta}|\mathcal{M}_j)$. Then do the following procedures for l=1,…,L. At the beginning of the l-th level, we have the samples $\boldsymbol{\theta}_{l-1}^{(m)}$, m=1,2,…,$N_{l\text{-}1}$, from $\pi_{l\text{-}1}(\boldsymbol{\theta})$. First, select τ_l such that the effective sample size $1/\sum_{s=1}^{N_{l-1}} \tilde{w}_s^{\,2}$ = some threshold (e.g., 0.9 $N_{l\text{-}1}$) (Cheung and Beck 2008d), where $\tilde{w}_s = w_s / \sum_{s=1}^{N_{l-1}} w_s$ and w_s = $p^{\tau_l-\tau_{l-1}}(\mathcal{D}\,|\,\boldsymbol{\theta}_{l-1}^{(s)},\mathcal{M}_j)$, s=1,2,…,$N_{l\text{-}1}$. If $\tau_l>1$, then set $L=l$ and $\tau_l=1$, then recompute w_s and

$\tilde{w}_s$. Then the N_l samples $\boldsymbol{\theta}_l^{(n)}$ from $\pi_l(\boldsymbol{\theta})$ are generated by doing the following for n=1,2,…,N_l:

1. Draw a number s' from a discrete distribution $p(S=s)=\tilde{w}_s$, s=1,2,…,N_{l-1};
2. Using $\boldsymbol{\theta}_{l-1}^{(s')}$ as the current sample, generate a sample $\boldsymbol{\theta}_l^{(n)}$ for $\boldsymbol{\theta}$ by multi-group MCMC algorithms. Set $\boldsymbol{\theta}_{l-1}^{(s')}=\boldsymbol{\theta}_l^{(n)}$.

Appendix 2A

The Hamiltonian equations (2.8) and (2.9) are equivalent to the following diffusionless Itô stochastic differential equation:

$$d\mathbf{x}(t)=\mathbf{v}(\mathbf{x}(t),t)dt \tag{A2.1}$$

where the state is formed by augmenting the displacement vector with the momentum vector:

$$\mathbf{x}(t)=\begin{bmatrix}\boldsymbol{\theta}(t)\\ \mathbf{p}(t)\end{bmatrix} \tag{A2.2}$$

and the drift term $\mathbf{v}$ ($\mathbf{x}(t)$,t) of the corresponding Fokker-Planck Equation (FPE) is given by:

$$\mathbf{v}(\mathbf{x}(t),t)=\begin{bmatrix}\dfrac{\partial H}{\partial \mathbf{p}}\\ -\dfrac{\partial H}{\partial \boldsymbol{\theta}}\end{bmatrix} \tag{A2.3}$$

Here we will show that the probability density function $f(\boldsymbol{\theta},\mathbf{p})$ as defined in (2.6) is the stationary distribution for the evolution in (A2.1). Consider

$$
\begin{aligned}
\nabla\bullet(f\mathbf{v}(\mathbf{x}(t),t)) &= f\nabla\bullet(\mathbf{v}(\mathbf{x}(t),t)) + \mathbf{v}(\mathbf{x}(t),t)\bullet\nabla f \\
&= f\nabla\bullet(\mathbf{v}(\mathbf{x}(t),t)) - f\mathbf{v}(\mathbf{x}(t),t)\bullet\nabla H \ (\because f = C^{-1}\exp(-H) \Rightarrow \nabla f = -f\nabla H) \\
&= f(\nabla\bullet(\mathbf{v}(\mathbf{x}(t),t)) - \mathbf{v}(\mathbf{x}(t),t)\bullet\nabla H) \\
&= f(\begin{bmatrix} \frac{\partial}{\partial\boldsymbol{\theta}} \\ \frac{\partial}{\partial\mathbf{p}} \end{bmatrix}\bullet\begin{bmatrix} \frac{\partial H}{\partial\mathbf{p}} \\ -\frac{\partial H}{\partial\boldsymbol{\theta}} \end{bmatrix} - \begin{bmatrix} \frac{\partial H}{\partial\mathbf{p}} \\ -\frac{\partial H}{\partial\boldsymbol{\theta}} \end{bmatrix}\bullet\begin{bmatrix} \frac{\partial H}{\partial\boldsymbol{\theta}} \\ \frac{\partial H}{\partial\mathbf{p}} \end{bmatrix}) \\
&= f[\sum_{i=1}^{D}\frac{\partial H^2}{\partial\theta_i\partial p_i} - \frac{\partial H^2}{\partial p_i\partial\theta_i} - (\sum_{i=1}^{D}\frac{\partial H}{\partial p_i}\frac{\partial H}{\partial\theta_i} - \frac{\partial H}{\partial\theta_i}\frac{\partial H}{\partial p_i})] \\
&= 0
\end{aligned}
$$

Thus $f(\boldsymbol{\theta}, \mathbf{p})$ is the stationary distribution for (A2.1) (equivalently for (2.8) and (2.9)) since it satisfies the corresponding stationary diffusionless FPE (Liouville's equation):

$$\nabla\bullet(f\mathbf{v}(\mathbf{x}(t),t)) = 0 \tag{A2.4}$$

Appendix 2B

For the operations shown in Table 2.1, the derivation of the corresponding reverse differentiation (RD) rules are given as follows.

For $\mathbf{v} = \alpha\mathbf{u} \in \mathbb{R}^m$,

$$\tilde{u}_i = \frac{\partial V}{\partial u_i} += \frac{\partial V}{\partial v_i}\frac{\partial v_i}{\partial u_i} = \alpha\frac{\partial V}{\partial v_i} \Rightarrow \tilde{\mathbf{u}} += \alpha\tilde{\mathbf{v}} \tag{B2.1}$$

$$\tilde{\alpha} = \frac{\partial V}{\partial\alpha} += \sum_{i=1}^{m}\frac{\partial V}{\partial v_i}\frac{\partial v_i}{\partial\alpha} = \sum_{i=1}^{m}\tilde{v}_i u_i = \mathbf{u}^T\tilde{\mathbf{v}} \tag{B2.2}$$

For $\mathbf{w} = \mathbf{u} + \mathbf{v} \in \mathbb{R}^m$,

$$\tilde{u}_i = \frac{\partial V}{\partial u_i} += \frac{\partial V}{\partial w_i}\frac{\partial w_i}{\partial u_i} = \widetilde{w_i} \Rightarrow \tilde{\mathbf{u}} += \widetilde{\mathbf{w}} \tag{B2.3}$$

Similarly, we also have $\tilde{\mathbf{v}} += \widetilde{\mathbf{w}}$

For $w = \mathbf{u}^T\mathbf{v}$,

$$\tilde{u}_i = \frac{\partial V}{\partial u_i} += \frac{\partial V}{\partial w}\frac{\partial w}{\partial u_i} = \tilde{w}v_i \Rightarrow \tilde{\mathbf{u}} += \tilde{w}\mathbf{v} \tag{B2.4}$$

Similarly we have, $\tilde{\mathbf{v}} += \tilde{w}\mathbf{u}$

For $\mathrm{V} = \alpha\mathrm{U},\ \mathrm{U}, \mathrm{V} \in \mathbb{R}^{p\times q};\ \alpha \in \mathbb{R}$,

$$\tilde{\alpha} = \frac{\partial V}{\partial \alpha} += \sum_{i,j}\frac{\partial V}{\partial \mathrm{V}_{ij}}\frac{\partial \mathrm{V}_{ij}}{\partial \alpha} = \sum_{i,j}\widetilde{\mathrm{V}_{ij}}\mathrm{U}_{ij} = \mathrm{sum}(\mathrm{sum}(\mathrm{U}.*\mathrm{V})) \tag{B2.5}$$

$$\widetilde{\mathrm{U}}_{ij} = \frac{\partial V}{\partial \mathrm{U}_{ij}} += \frac{\partial V}{\partial \mathrm{V}_{ij}}\frac{\partial \mathrm{V}_{ij}}{\partial \mathrm{U}_{ij}} = \widetilde{\mathrm{V}}_{ij}\alpha \Rightarrow \widetilde{\mathrm{U}} += \alpha\widetilde{\mathrm{V}} \tag{B2.6}$$

For $\mathrm{W} = \mathrm{U} + \mathrm{V},\ \mathrm{U}, \mathrm{V}, \mathrm{W} \in \mathbb{R}^{p\times q}$,

$$\tilde{\mathrm{U}} += \tilde{\mathrm{W}},\ \tilde{\mathrm{V}} += \tilde{\mathrm{W}} \tag{B2.7}$$

The above follows the same proof as the vector case.

For $\mathrm{W} = \mathrm{UV},\ \mathrm{U} \in \mathbb{R}^{p\times q}, \mathrm{V} \in \mathbb{R}^{q\times r}$,

$$\widetilde{\mathrm{U}}_{ij} = \frac{\partial V}{\partial \mathrm{U}_{ij}} += \sum_{k=1}^{r} \frac{\partial V}{\partial \mathrm{W}_{ik}} \frac{\partial \mathrm{W}_{ik}}{\partial \mathrm{U}_{ij}} \quad (\because \mathrm{U}_{ij} \text{ affects } \mathrm{W}_{ik})$$

$$= \sum_{k=1}^{r} \widetilde{\mathrm{W}}_{ik} \mathrm{V}_{jk} \qquad (\because \mathrm{W}_{ik} = \sum_{l=1}^{q} \mathrm{U}_{il} \mathrm{V}_{lk} \Rightarrow \frac{\partial \mathrm{W}_{ik}}{\partial \mathrm{U}_{ij}} = \mathrm{V}_{jk})$$

$$\Rightarrow \widetilde{\mathrm{U}} += \widetilde{\mathrm{W}} \mathrm{V}^{\mathrm{T}}$$

$$\widetilde{\mathrm{V}}_{ij} = \frac{\partial V}{\partial \mathrm{V}_{ij}} += \sum_{k=1}^{p} \frac{\partial V}{\partial \mathrm{W}_{kj}} \frac{\partial \mathrm{W}_{kj}}{\partial \mathrm{V}_{ij}} \quad (\because \mathrm{V}_{ij} \text{ affects } \mathrm{W}_{kj})$$

$$= \sum_{k=1}^{p} \widetilde{\mathrm{W}}_{kj} \mathrm{U}_{ki} \qquad (\because \mathrm{W}_{kj} = \sum_{l=1}^{q} \mathrm{U}_{kl} \mathrm{V}_{lj} \Rightarrow \frac{\partial \mathrm{W}_{kj}}{\partial \mathrm{V}_{ij}} = \mathrm{U}_{ki})$$

$$\Rightarrow \widetilde{\mathrm{V}} += \mathrm{U}^{\mathrm{T}} \widetilde{\mathrm{W}}$$

For $\mathbf{w} = \mathrm{U}\mathbf{v}$, $\mathrm{U} \in \mathbb{R}^{p \times q}$, $\mathbf{v} \in \mathbb{R}^{q \times 1}$,

$$\tilde{\mathrm{U}} += \widetilde{\mathbf{w}} \mathbf{v}^{\mathrm{T}}, \ \tilde{\mathbf{v}} += \mathrm{U}^{\mathrm{T}} \widetilde{\mathbf{w}}$$

This follows directly from the previous case when r=1.

For an implicit equation for $\mathbf{v}$: $\mathbf{w} = \mathrm{U}\mathbf{v}$, $\mathrm{U} \in \mathbb{R}^{p \times p}$, $\mathbf{v} \in \mathbb{R}^{p \times 1}$, assume U is invertible and denote V as its inverse.

$$\mathbf{v} = \mathrm{R}\mathbf{w} \quad (\because \mathbf{w} = \mathrm{U}\mathbf{v})$$

$$v_i = \sum_{j=1}^{p} \mathrm{R}_{ij} w_j$$

$$\widetilde{w}_i = \frac{\partial V}{\partial w_i} += \sum_{j=1}^{p} \frac{\partial V}{\partial v_j} \frac{\partial v_j}{\partial w_i} = \sum_{j=1}^{p} \tilde{v}_j \mathrm{R}_{ji} = y_i$$

$$\Rightarrow \widetilde{\mathbf{w}} += \mathbf{y} = \mathrm{R}^{\mathrm{T}} \tilde{\mathbf{v}} = \mathrm{U}^{-\mathrm{T}} \tilde{\mathbf{v}}$$

$$w_i = \sum_{j=1}^{p} \mathrm{U}_{ij} v_j \Rightarrow 0 = \sum_{j=1}^{p} \mathrm{U}_{ij} \frac{\partial v_j}{\partial \mathrm{U}_{ij}} + v_j$$

$$\Rightarrow 0 = \sum_{j=1}^{p} \mathrm{U}_{ij} \frac{\partial v_j}{\partial \mathrm{U}_{kl}} + \sum_{j=1}^{p} \frac{\partial \mathrm{U}_{ij}}{\partial \mathrm{U}_{kl}} v_j$$

$$\Rightarrow 0 = \sum_{j=1}^{p} \mathrm{U}_{ij} \frac{\partial v_j}{\partial \mathrm{U}_{kl}} + \delta_{ik} v_l$$

$$\Rightarrow -\delta_{ik} v_l = \sum_{j=1}^{p} \mathrm{U}_{ij} \frac{\partial v_j}{\partial \mathrm{U}_{kl}}$$

$$\Rightarrow -v_l \mathbf{e}_k = \mathrm{U} \frac{\partial \mathbf{v}}{\partial \mathrm{U}_{kl}}$$

$$\Rightarrow \frac{\partial \mathbf{v}}{\partial \mathrm{U}_{kl}} = -v_l \mathrm{U}^{-1} \mathbf{e}_k = -v_l \mathrm{R} \mathbf{e}_k$$

$$\Rightarrow \frac{\partial v_j}{\partial \mathrm{U}_{kl}} = -v_l \sum_{i=1}^{p} \mathrm{R}_{ji} \delta_{ik} = -v_l \mathrm{R}_{jk}$$

$$\widetilde{\mathrm{U}}_{kl} = \frac{\partial V}{\partial \mathrm{U}_{kl}} + = \sum_{j=1}^{p} \frac{\partial V}{\partial v_j} \frac{\partial v_j}{\partial \mathrm{U}_{kl}} = -v_l \sum_{j=1}^{p} \tilde{v}_j \mathrm{R}_{jk} = -v_l y_k$$

$$\Rightarrow \widetilde{\mathrm{U}} + = -\mathbf{y}\mathbf{v}^T$$

Appendix 2C

$$\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]) = \mathrm{E}[(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})] - \mathrm{E}[\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]])^2] = \mathrm{E}[(\frac{1}{N}\sum_{k=1}^{N} g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])^2]$$

$$= \mathrm{E}[(\frac{1}{N}\sum_{k=1}^{N} g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])(\frac{1}{N}\sum_{j=1}^{N} g(\boldsymbol{\theta}^{(j)}) - \mathrm{E}[g(\boldsymbol{\theta})])]$$

$$= \frac{1}{N^2}\mathrm{E}[\sum_{k=1}^{N}\sum_{j=1}^{N}(g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])(g(\boldsymbol{\theta}^{(\tau)}) - \mathrm{E}[g(\boldsymbol{\theta})])]$$

$$= \frac{1}{N^2}\sum_{k=1}^{N}\sum_{\tau=1}^{N}\mathrm{E}[(g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])(g(\boldsymbol{\theta}^{(\tau)}) - \mathrm{E}[g(\boldsymbol{\theta})])]$$

$$= \frac{1}{N^2}\{\sum_{k=1}^{N}\mathrm{E}[(g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])^2]$$

$$+2\sum_{\tau=1}^{N-1}\sum_{k=1}^{N-\tau}\mathrm{E}[(g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])(g(\boldsymbol{\theta}^{(k+\tau)}) - \mathrm{E}[g(\boldsymbol{\theta})])]\}$$

where $E[(g(\boldsymbol{\theta}^{(k+\tau)}) - \mathrm{E}[g(\boldsymbol{\theta})])(g(\boldsymbol{\theta}^{(k)}) - \mathrm{E}[g(\boldsymbol{\theta})])] = \rho(\tau)$, for all *k*.

Thus $\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})])$ becomes:

$$\begin{aligned}
\mathrm{Var}(\tilde{\mathrm{E}}[g(\boldsymbol{\theta})]) &= \frac{1}{N^2}\sum_{k=1}^{N}\rho(0) + \frac{2}{N^2}\sum_{\tau=1}^{N-1}\sum_{k=1}^{N-\tau}\rho(\tau) \\
&= \frac{\rho(0)}{N} + \frac{2}{N^2}\sum_{\tau=1}^{N-1}(N-\tau)\rho(\tau) \\
&= \frac{\rho(0)}{N}(1 + 2\sum_{\tau=1}^{N-1}(1-\frac{\tau}{N})\frac{\rho(\tau)}{\rho(0)}) \\
&= \frac{\rho(0)}{N}(1+\lambda)
\end{aligned}$$

where $\lambda = 2\sum_{\tau=1}^{N-1}(1-\frac{\tau}{N})\frac{\rho(\tau)}{\rho(0)}$.

Appendix 2D

The transition PDF for $\boldsymbol{\theta}^*$ given $\boldsymbol{\theta}$ for the multiple group MCMC presented is:

$$K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) = \prod_{j=1}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}).$$

The PDF of $\boldsymbol{\theta}^*$ with the above transition PDF is given by:

$$\begin{aligned}
p(\boldsymbol{\theta}^*) &= \int K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} = \int \prod_{j=1}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \\
&= \int \prod_{j=1}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\{\boldsymbol{\theta}\}_{2:G} \mid \boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_1)d\boldsymbol{\theta} \\
&= \int \prod_{j=1}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta}_1 \mid \{\boldsymbol{\theta}\}_{2:G})\pi(\{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_{2:G} \\
&= \int [\prod_{j=2}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})]K_1(\boldsymbol{\theta}_1^* \mid \boldsymbol{\theta}_1, \{\boldsymbol{\theta}\}_{2:G})\pi(\boldsymbol{\theta}_1 \mid \{\boldsymbol{\theta}\}_{2:G})\pi(\{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_1 d\boldsymbol{\theta}_{2:G} \\
&= \int [\prod_{j=2}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})]\pi(\{\boldsymbol{\theta}\}_{2:G})\{\underbrace{\int K_1(\boldsymbol{\theta}_1^* \mid \boldsymbol{\theta}_1, \{\boldsymbol{\theta}\}_{2:G})\pi(\boldsymbol{\theta}_1 \mid \{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_1}_{=\pi(\boldsymbol{\theta}_1^* \mid \{\boldsymbol{\theta}\}_{2:G})}\}d\boldsymbol{\theta}_{2:G}
\end{aligned}$$

$$= \int [\prod_{j=2}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \pi(\{\boldsymbol{\theta}\}_{2:G}) \pi(\boldsymbol{\theta}_1^* | \{\boldsymbol{\theta}\}_{2:G}) d\boldsymbol{\theta}_{2:G}$$

$$= \int [\prod_{j=2}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \pi(\boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{2:G}) d\boldsymbol{\theta}_{2:G} \quad \text{.....(D2.1)}$$

$$= \int [\prod_{j=2}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \pi(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) \pi(\boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) d\boldsymbol{\theta}_{2:G}$$

$$= \int [\prod_{j=2}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \pi(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) \pi(\boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) d\boldsymbol{\theta}_2 d\boldsymbol{\theta}_{3:G}$$

$$= \int [\prod_{j=3}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \pi(\boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) \{ \underbrace{\int K_2(\boldsymbol{\theta}_2^* \mid \boldsymbol{\theta}^*_1, \boldsymbol{\theta}_2, \{\boldsymbol{\theta}\}_{3:G}) \pi(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) d\boldsymbol{\theta}_2}_{= \pi(\boldsymbol{\theta}_2^* | \boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G})} \} d\boldsymbol{\theta}_{3:G}$$

$$= \int [\prod_{j=3}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \pi(\boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) \pi(\boldsymbol{\theta}_2^* | \boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) d\boldsymbol{\theta}_{3:G}$$

$$= \int [\prod_{j=3}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})] \pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \{\boldsymbol{\theta}\}_{3:G}) d\boldsymbol{\theta}_{3:G} \quad \text{.....(D2.2)}$$

$\vdots$

(observe the patterns in (D2.1) and (D2.2) and keep repeating each time reducing the dimension of integration by 1 group)

$$= \int K_G(\boldsymbol{\theta}_G^* \mid \{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G) \pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, ..., \boldsymbol{\theta}_{G-1}^*, \boldsymbol{\theta}_G) d\boldsymbol{\theta}_G$$

$$= \int K_G(\boldsymbol{\theta}_G^* \mid \{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G) \pi(\boldsymbol{\theta}_G | \{\boldsymbol{\theta}^*\}_{1:G\text{-}1}) \pi\{\boldsymbol{\theta}^*\}_{1:G\text{-}1} d\boldsymbol{\theta}_G$$

$$= \pi\{\boldsymbol{\theta}^*\}_{1:G\text{-}1} \underbrace{\int K_G(\boldsymbol{\theta}_G^* \mid \{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G) \pi(\boldsymbol{\theta}_G | \{\boldsymbol{\theta}^*\}_{1:G\text{-}1}) d\boldsymbol{\theta}_G}_{\pi(\boldsymbol{\theta}_G^* | \{\boldsymbol{\theta}^*\}_{1:G-1}) \text{ (by 1.1)}}$$

$$= \pi\{\boldsymbol{\theta}^*\}_{1:G\text{-}1} \pi(\boldsymbol{\theta}_G^* | \{\boldsymbol{\theta}^*\}_{1:G-1})$$

$$= \pi(\{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G^*)$$

$$= \pi(\boldsymbol{\theta}^*)$$

Appendix 2E

For simpler proof, it can be shown that the transition function T_j for the j-th group components satisfies the following by making use of (2.49)and (2.50):

$$\begin{aligned} & T_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) \\ & = T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G}) \pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) \end{aligned} \quad \text{(E2.1)}$$

Proof:

$$T_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$= \min\{\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})q_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}),$$
$$\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})q_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G})\}$$
$$T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$= \min\{\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})q_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G}),$$
$$\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})q_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})\}$$

Similar to regular MH, the relation (E2.1) leads $K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$ to satisfy the reversibility condition which is sufficient to guarantee for it to satisfy the stationary condition (2.47). Alternatively, we can also check directly as follows:

$$\int K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})d\boldsymbol{\theta}_j$$
$$= \int T_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})d\boldsymbol{\theta}_j$$
$$+(1-a_j(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j:G}))\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})d\boldsymbol{\theta}_j$$
$$+\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) - a_j(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j:G})\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j+1:G})d\boldsymbol{\theta}_j\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$+\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) - a_j(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j:G})\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$= a_j(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j:G})\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$+\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) - a_j(\{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}\}_{j:G})\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$
$$= \pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$$

Appendix 2F

Consider the case where G=2, and Gibbs sampling which is a special case of multi-group MH,

$$K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}_2^* \mid \boldsymbol{\theta}_1^*)\pi(\boldsymbol{\theta}_1^* \mid \boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \frac{\pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)\pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\pi(\boldsymbol{\theta}_1^*)\pi(\boldsymbol{\theta}_2)}$$

$$K(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*) = \pi(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_1\,|\,\boldsymbol{\theta}_2^*)\pi(\boldsymbol{\theta}_1^*,\boldsymbol{\theta}_2^*) = \frac{\pi(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^*)\pi(\boldsymbol{\theta}_1^*,\boldsymbol{\theta}_2^*)}{\pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2^*)}$$

Therefore in this case, for the Markov chain to be reversible, i.e., satisfying the detailed balance(reversibility condition), we need:

$$\frac{\pi(\boldsymbol{\theta}_1^*,\boldsymbol{\theta}_2)}{\pi(\boldsymbol{\theta}_1^*)\pi(\boldsymbol{\theta}_2)} = \frac{\pi(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^*)}{\pi(\boldsymbol{\theta}_1)\pi(\boldsymbol{\theta}_2^*)}$$

In general, the above is not true. Thus, one can expect that in general, for any G>1, multi-group MCMC does not satisfy reversibility condition even reversibility condition can be satisfied for each group of parameters.

However, when $\pi(\boldsymbol{\theta}) = \pi_1(\boldsymbol{\theta}_1)\pi_2(\boldsymbol{\theta}_2)$, the reversibility condition is satisfied. It can be shown that if $\pi(\boldsymbol{\theta})$ can be written as $\pi(\boldsymbol{\theta}) = \prod_{j=1}^{G}\pi_j(\boldsymbol{\theta}_j)$, then reversibility condition is satisfied and the Markov Chain is reversible. As an example, consider Gibbs sampling where each group is simulated from its own conditional.

$$K(\boldsymbol{\theta}^*\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta}) = \prod_{j=1}^{G}\pi_j(\boldsymbol{\theta}_j^*)\prod_{j=1}^{G}\pi_j(\boldsymbol{\theta}_j) = \prod_{j=1}^{G}\pi_j(\boldsymbol{\theta}_j)\prod_{j=1}^{G}\pi_j(\boldsymbol{\theta}_j^*) = K(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)$$

Of course, in this case, Gibbs sampling is just the same as standard Monte Carlo simulation.

Now consider the case with the target PDF of this type and transition PDF being independent from the other groups, the Markov Chain of samples simulated using multi-group MH is in general not reversible, (not satisfying the reversibility condition).

The transition PDF for this case is:

$$K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) = \prod_{j=1}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) = \prod_{j=1}^{G} [T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) + (1 - a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_j)]$$

where the transition function $T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)$ and $a(\boldsymbol{\theta}_j)$ are:

$$T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) = q_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) r_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) = q_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) \min\{1, \frac{\pi(\boldsymbol{\theta}^*_j) q_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_j^*)}{\pi(\boldsymbol{\theta}_j) q_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)}\}$$

$$a(\boldsymbol{\theta}_j) = \int T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) d\boldsymbol{\theta}_j^*$$

It is obvious we have the following:

$$T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j) = T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_j^*)\pi(\boldsymbol{\theta}_j^*) \tag{F2.1}$$

As a special case for multi-group MCMC, as shown before, this transition PDF satisfies stationarity condition:

$$\begin{aligned}
\int K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta} &= \int \prod_{j=1}^{G} K(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j) d\boldsymbol{\theta} \\
&= \prod_{j=1}^{G} \int K(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j) d\boldsymbol{\theta}_j \\
&= \prod_{j=1}^{G} \pi(\boldsymbol{\theta}_j^*) \\
&= \pi(\boldsymbol{\theta}^*)
\end{aligned}$$

For transition PDF having parts involving different combination of delta functions, special care has to be taken to prove or disprove the reversibility condition. To deal with this issue, our trick here is to consider the following for any nonnegative $h(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$:

$$\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}d\boldsymbol{\theta}^* = \int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}d\boldsymbol{\theta}^*, \forall h(\boldsymbol{\theta},\boldsymbol{\theta}^*) \geq 0$$
$$\Leftrightarrow \int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) - K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)]d\boldsymbol{\theta}d\boldsymbol{\theta}^* = 0, \forall h(\boldsymbol{\theta},\boldsymbol{\theta}^*) \geq 0 \quad \text{(F2.2)}$$
$$\Leftrightarrow K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) = K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)$$

First, let's expand the transition PDF into the sum of terms (here there will be 2^G) since the integration will depend on the number of delta functions involved in the term. It can be seen that the number of terms which involves the product of k delta functions and G-k transition functions is equal to $C_k^G = G!/[(G-k)!k!]$.

$$\begin{aligned}
&K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) \\
&= \prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) + (1 - a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_j)] \\
&= \prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)] + \prod_{j=1}^{G}(1 - a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_j) \\
&+ \sum_{k=1}^{G-1}\sum_{m=1}^{C_k^n}[(1 - a(\boldsymbol{\theta}_{n_1^m}))\delta(\boldsymbol{\theta}_{n_1^m}^* - \boldsymbol{\theta}_{n_1^m})]...[(1 - a(\boldsymbol{\theta}_{n_k^m}))\delta(\boldsymbol{\theta}_{n_k^m}^* - \boldsymbol{\theta}_{n_k^m})] \\
&\quad T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}^* \mid \boldsymbol{\theta}_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}_{i_2^m}^* \mid \boldsymbol{\theta}_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}^* \mid \boldsymbol{\theta}_{i_{G-k}^m})
\end{aligned} \quad \text{(F2.3)}$$

where $\{n_1^m, n_2^m ..., n_k^m\}$ is the m-th combination of k numbers drawn from the set {1,2,…,G} and $n_1^m < n_2^m < ... < n_k^m$; $\{i_1^m, i_2^m ..., i_{G-k}^m\}$ = $\{1,2,\ldots,G\} \setminus \{n_1^m, n_2^m ..., n_k^m\}$ (i.e.,=the G-k numbers that are not in $\{n_1^m, n_2^m ..., n_k^m\}$ but in $\{1,2,\ldots,G\}$) and $i_1^m < i_2^m < ... < i_{G-k}^m$ (actually this ordering is not necessary, just for clarity for presentation).

So similarly,

$$
\begin{aligned}
&K(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)\\
&=\prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j\,|\,\boldsymbol{\theta}^*_j)+(1-a(\boldsymbol{\theta}^*_j))\delta(\boldsymbol{\theta}_j-\boldsymbol{\theta}^*_j)]\\
&=\prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j\,|\,\boldsymbol{\theta}^*_j)]+\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}^*_j))\delta(\boldsymbol{\theta}_j-\boldsymbol{\theta}^*_j)\\
&+\sum_{k=1}^{G-1}\sum_{m=1}^{C_k^n}[(1-a(\boldsymbol{\theta}^*_{n_1^m}))\delta(\boldsymbol{\theta}_{n_1^m}-\boldsymbol{\theta}^*_{n_1^m})]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))\delta(\boldsymbol{\theta}_{n_k^m}-\boldsymbol{\theta}^*_{n_k^m})]\\
&+T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}\,|\,\boldsymbol{\theta}^*_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}_{i_2^m}\,|\,\boldsymbol{\theta}^*_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}\,|\,\boldsymbol{\theta}^*_{i_{G-k}^m})
\end{aligned}
\tag{F2.4}
$$

$$
\begin{aligned}
&\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}^*_j\,|\,\boldsymbol{\theta}_j)]\pi(\boldsymbol{\theta})d\boldsymbol{\theta}d\boldsymbol{\theta}^*=\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}^*_j\,|\,\boldsymbol{\theta}_j)]\prod_{j=1}^{G}\pi(\boldsymbol{\theta}_j)d\boldsymbol{\theta}d\boldsymbol{\theta}^*\\
&=\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}^*_j\,|\,\boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j)]d\boldsymbol{\theta}d\boldsymbol{\theta}^*\\
&=\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}_j\,|\,\boldsymbol{\theta}^*_j)\pi(\boldsymbol{\theta}^*_j)]d\boldsymbol{\theta}d\boldsymbol{\theta}^*\\
&=\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}_j\,|\,\boldsymbol{\theta}^*_j)]\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}d\boldsymbol{\theta}^*
\end{aligned}
\tag{F2.5}
$$

$$\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}^*_j-\boldsymbol{\theta}_j)]\pi(\boldsymbol{\theta})d\boldsymbol{\theta}d\boldsymbol{\theta}^*=\int h(\boldsymbol{\theta}^*,\boldsymbol{\theta}^*)\{\prod_{j=1}^{G}[1-a(\boldsymbol{\theta}^*_j)]\}\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^*$$

$$\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}^*_j))\delta(\boldsymbol{\theta}_j-\boldsymbol{\theta}^*_j)]\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}d\boldsymbol{\theta}^*=\int h(\boldsymbol{\theta}\ ,\boldsymbol{\theta}\ )\{\prod_{j=1}^{G}[1-a(\boldsymbol{\theta}_j)]\}\pi(\boldsymbol{\theta}\ )d\boldsymbol{\theta}$$

Thus combining the above two, we have,

$$
\begin{aligned}
&\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}^*_j-\boldsymbol{\theta}_j)]\pi(\boldsymbol{\theta})d\boldsymbol{\theta}d\boldsymbol{\theta}^*\\
&=\int h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}^*_j))\delta(\boldsymbol{\theta}_j-\boldsymbol{\theta}^*_j)]\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}d\boldsymbol{\theta}^*
\end{aligned}
\tag{F2.6}
$$

$$I_1 = \int \frac{h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}))\delta(\boldsymbol{\theta}^*_{n_1^m}-\boldsymbol{\theta}_{n_1^m})]...[(1-a(\boldsymbol{\theta}_{n_k^m}))\delta(\boldsymbol{\theta}^*_{n_k^m}-\boldsymbol{\theta}_{n_k^m})]}{T_{i_1^m}(\boldsymbol{\theta}^*_{i_1^m}|\boldsymbol{\theta}_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}^*_{i_2^m}|\boldsymbol{\theta}_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}^*_{i_{G-k}^m}|\boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}d\boldsymbol{\theta}^*}$$

$$= \int \frac{h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}))\delta(\boldsymbol{\theta}^*_{n_1^m}-\boldsymbol{\theta}_{n_1^m})]...[(1-a(\boldsymbol{\theta}_{n_k^m}))\delta(\boldsymbol{\theta}^*_{n_k^m}-\boldsymbol{\theta}_{n_k^m})]}{T_{i_1^m}(\boldsymbol{\theta}^*_{i_1^m}|\boldsymbol{\theta}_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}^*_{i_2^m}|\boldsymbol{\theta}_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}^*_{i_{G-k}^m}|\boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta}_{n_1^m})...\pi(\boldsymbol{\theta}_{n_k^m})\pi(\boldsymbol{\theta}_{i_1^m})...\pi(\boldsymbol{\theta}_{i_{G-k}^m})d\boldsymbol{\theta}d\boldsymbol{\theta}^*}$$

$$= \int \frac{h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}))\delta(\boldsymbol{\theta}^*_{n_1^m}-\boldsymbol{\theta}_{n_1^m})]...[(1-a(\boldsymbol{\theta}_{n_k^m}))\delta(\boldsymbol{\theta}^*_{n_k^m}-\boldsymbol{\theta}_{n_k^m})]}{[T_{i_1^m}(\boldsymbol{\theta}^*_{i_1^m}|\boldsymbol{\theta}_{i_1^m})\pi(\boldsymbol{\theta}_{i_1^m})]...[T_{i_{G-k}^m}(\boldsymbol{\theta}^*_{i_{G-k}^m}|\boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta}_{i_{G-k}^m})]\pi(\boldsymbol{\theta}_{n_1^m})...\pi(\boldsymbol{\theta}_{n_k^m})d\boldsymbol{\theta}d\boldsymbol{\theta}^*}$$

$$= \int \frac{h(\boldsymbol{\theta}_{n_1^m}=\boldsymbol{\theta}^*_{n_1^m},...,\boldsymbol{\theta}_{n_k^m}=\boldsymbol{\theta}^*_{n_k^m},\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}^*_{n_1^m}))]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))]}{[T_{i_1^m}(\boldsymbol{\theta}^*_{i_1^m}|\boldsymbol{\theta}_{i_1^m})\pi(\boldsymbol{\theta}_{i_1^m})]...[T_{i_{G-k}^m}(\boldsymbol{\theta}^*_{i_{G-k}^m}|\boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta}_{i_{G-k}^m})]..\pi(\boldsymbol{\theta}^*_{n_1^m})...\pi(\boldsymbol{\theta}^*_{n_k^m})d\boldsymbol{\theta}_{i_1^m}d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m}d\boldsymbol{\theta}^*}$$

$$= \int \frac{h(\boldsymbol{\theta}^*_{n_1^m},...,\boldsymbol{\theta}^*_{n_k^m},\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}^*_{n_1^m}))]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))]}{[T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}|\boldsymbol{\theta}^*_{i_1^m})\pi(\boldsymbol{\theta}^*_{i_1^m})]...[T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}|\boldsymbol{\theta}^*_{i_{G-k}^m})\pi(\boldsymbol{\theta}^*_{i_{G-k}^m})]..\pi(\boldsymbol{\theta}^*_{n_1^m})...\pi(\boldsymbol{\theta}^*_{n_k^m})d\boldsymbol{\theta}_{i_1^m}d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m}d\boldsymbol{\theta}^*}$$

$$= \int \frac{h(\boldsymbol{\theta}^*_{n_1^m},...,\boldsymbol{\theta}^*_{n_k^m},\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}^*_{n_1^m}))]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))]}{T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}|\boldsymbol{\theta}^*_{i_1^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}|\boldsymbol{\theta}^*_{i_{G-k}^m})\pi(\boldsymbol{\theta}^*_{i_1^m})...\pi(\boldsymbol{\theta}^*_{i_{G-k}^m})\pi(\boldsymbol{\theta}^*_{n_1^m})...\pi(\boldsymbol{\theta}^*_{n_k^m})d\boldsymbol{\theta}_{i_1^m}d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m}d\boldsymbol{\theta}^*}$$

$$= \int \frac{h(\boldsymbol{\theta}^*_{n_1^m},...,\boldsymbol{\theta}^*_{n_k^m},\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}^*_{n_1^m}))]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))]}{T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}|\boldsymbol{\theta}^*_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}_{i_2^m}|\boldsymbol{\theta}^*_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}|\boldsymbol{\theta}^*_{i_{G-k}^m})\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}_{i_1^m}d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m}d\boldsymbol{\theta}^*}$$

Similarly,

$$I_2 = \int \frac{h(\boldsymbol{\theta},\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}^*_{n_1^m}))\delta(\boldsymbol{\theta}_{n_1^m}-\boldsymbol{\theta}^*_{n_1^m})]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))\delta(\boldsymbol{\theta}_{n_k^m}-\boldsymbol{\theta}^*_{n_k^m})]}{T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}|\boldsymbol{\theta}^*_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}_{i_2^m}|\boldsymbol{\theta}^*_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}|\boldsymbol{\theta}^*_{i_{G-k}^m})\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}d\boldsymbol{\theta}^*}$$

$$= \int \frac{h(\tilde{\boldsymbol{\theta}},\tilde{\boldsymbol{\theta}}_{n_1^m},...,\tilde{\boldsymbol{\theta}}_{n_k^m},\boldsymbol{\theta}^*_{i_1^m},\boldsymbol{\theta}^*_{i_2^m},...,\boldsymbol{\theta}^*_{i_{G-k}^m})[(1-a(\tilde{\boldsymbol{\theta}}_{n_1^m}))]...[(1-a(\tilde{\boldsymbol{\theta}}_{n_k^m}))]}{T_{i_1^m}(\boldsymbol{\theta}^*_{i_1^m}|\tilde{\boldsymbol{\theta}}_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}^*_{i_2^m}|\tilde{\boldsymbol{\theta}}_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}^*_{i_{G-k}^m}|\tilde{\boldsymbol{\theta}}_{i_{G-k}^m})\pi(\tilde{\boldsymbol{\theta}}\ )d\boldsymbol{\theta}^*_{i_1^m}d\boldsymbol{\theta}^*_{i_2^m}...d\boldsymbol{\theta}^*_{i_{G-k}^m}d\tilde{\boldsymbol{\theta}}}$$

Thus it can be seen that I_1 in general is not equal to I_2. Thus the transition PDF does not satisfy the reversibility condition.

CHAPTER 3

Algorithms for stochastic system model class comparison and averaging

The computation of the evidence in (1.3) required for model class comparison and averaging is highly nontrivial. Laplace's method of asymptotic approximation (Beck and Katafygiotis 1991, 1998) has been proposed by researchers such as Mackay (1992) and Beck and Yuen (2004), which, in effect, utilizes a Gaussian sum approximation of the posterior PDF. However, the accuracy of such an approximation is questionable when (i) the amount of data is small, or (ii) the chosen class of models turns out to be unidentifiable based on the available data. It should be noted that variational methods (Beal 2003) can provide a lower bound to the log evidence that is required for Bayesian model class selection. For a comparison between two model classes, we need to consider the difference of the corresponding log evidences. Approximating the difference of the log evidences by the difference of the corresponding lower bounds can lead to misleading results in model comparison and based on such approximation, one may get an erroneous result for the posterior probability of each model class, $P(\mathcal{M}_j \mid \mathcal{D}, M)$. Fortunately, stochastic simulation methods to evaluate the evidence are practical and applicable to all cases; they are discussed in the next section.

3.1 Stochastic simulation methods for calculating model class evidence

3.1.1 Method based on samples from the prior

The most direct way to calculate the evidence $p(\mathcal{D}|\mathcal{M}_j)$ is to apply the standard Monte Carlo method to (1.3) based on samples $\boldsymbol{\theta}^{(k)}$, $k = 1,2,\ldots, N$, from the prior $p(\boldsymbol{\theta}|\mathcal{M}_j)$ as follows:

$$p(\mathcal{D} \mid \mathcal{M}_j) \approx \frac{1}{N} \sum_{k=1}^{N} p(\mathcal{D} \mid \boldsymbol{\theta}^{(k)}, \mathcal{M}_j) \tag{3.1}$$

However, this is usually a highly inefficient method to estimate $p(\mathcal{D}|\mathcal{M}_j)$. The region of high-probability content of $p(\boldsymbol{\theta}|\mathcal{M}_j)$ is often very different from the region where $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}_j)$ has its largest values, implying that it is very rare for the samples from $p(\boldsymbol{\theta}|\mathcal{M}_j)$ to fall into this latter region. This usually leads to the Monte Carlo estimator having an extremely large variance and so it produces a poor estimate of the evidence unless a huge amount of samples are employed. For higher-dimensional problems encountered in practice, this method is often computationally prohibitive.

3.1.2 Multi-level methods

Ching and Chen (2007) evaluate the evidence by considering a sequence of intermediate PDFs $\pi_i(\boldsymbol{\theta})$ for i=0,1,…, l, such that the first and last PDFs, $\pi_0(\boldsymbol{\theta})$ and $\pi_l(\boldsymbol{\theta})$, in the sequence are the prior $p(\boldsymbol{\theta}|\mathcal{M}_j)$ and posterior $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$, respectively:

$$\pi_i(\boldsymbol{\theta}) \propto p^{\alpha_i}(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_j)\, p(\boldsymbol{\theta} \mid \mathcal{M}_j) \tag{3.2}$$

where $0=\alpha_0<\alpha_1<\ldots<\alpha_l=1$. In their approach, the evidence $p(\mathcal{D}|\mathcal{M}_j)$ is then estimated as follows:

$$p(\mathcal{D}\,|\,\mathcal{M}_j) \approx \prod_{i=1}^{l} \frac{1}{N_{i-1}} \sum_{m=1}^{N_{i-1}} p^{\alpha_i - \alpha_{i-1}}(\boldsymbol{\theta}_{i-1}^{(m)}\,|\,\mathcal{D}, \mathcal{M}_j) \tag{3.3}$$

where $\boldsymbol{\theta}_{i-1}^{(m)}$, m=1,2,…,$N_{i\text{-}1}$, are the samples distributed according to $\pi_{i\text{-}1}(\boldsymbol{\theta})$ which are generated by the TMCMC (Transition Markov Chain Monte Carlo) method. This approach is similar to Annealed Importance Sampling (AIS) (Neal 2001) and Linked Importance Sampling (LIS) (Neal 2005); the main differences lie on the way samples are propagated from one level to the next level and the use of bridge sampling (Meng and Wong 1996) in LIS.

Cheung and Beck (2007b) introduce an alternative method by showing that the logarithm of the evidence can be expressed as the following one-dimensional integration of expected log-likelihood from α=0 to 1:

$$\ln p(\mathcal{D}\,|\,\mathcal{M}_j) = \int_0^1 E[\ln p(\mathcal{D}\,|\,\boldsymbol{\theta}, \mathcal{M}_j)\,|\,\mathcal{D}, \alpha, \mathcal{M}_j]\, d\alpha \tag{3.4}$$

where $E[\ln p(\mathcal{D}\,|\,\boldsymbol{\theta}, \mathcal{M}_j)\,|\,\mathcal{D}, \alpha, \mathcal{M}_j]$ is the expectation with respect to the PDF:

$$p(\boldsymbol{\theta}\,|\,\mathcal{D}, \alpha, \mathcal{M}_j) \propto p^{\alpha}(\mathcal{D}\,|\,\boldsymbol{\theta}, \mathcal{M}_j)\, p(\boldsymbol{\theta}\,|\,\mathcal{M}_j) \tag{3.5}$$

and the integrand in (3.4) can be estimated as follows:

$$E[\ln p(\mathcal{D}\,|\,\boldsymbol{\theta}, \mathcal{M}_j)\,|\,\mathcal{D}, \alpha, \mathcal{M}_j] \approx \frac{1}{N_\alpha} \sum_{m=1}^{N_\alpha} \ln p(\mathcal{D}\,|\,\boldsymbol{\theta}_\alpha^{(m)}, \mathcal{M}_j) \tag{3.6}$$

where $\boldsymbol{\theta}_\alpha^{(m)}$, m=1,2,…,N_α, are samples distributed according to $p(\boldsymbol{\theta}|\mathcal{D},\alpha,\mathcal{M}_j)$ that are generated by an MCMC algorithm such as TMCMC. A one-dimensional numerical integration scheme can be applied to calculate the integral in (3.4).

The major drawback of this method and the one proposed in Ching and Chen (2007) is the fact that both methods rely on the availability of samples distributed according to PDFs in a sequence depending on α in (3.2) and (3.5). For example, in Ching and Chen (2007), the samples for $\pi_i(\boldsymbol{\theta})$ for each i are generated using sampling and re-sampling and MCMC simulation methods. If the number of samples for each level i are not large enough for convergence to the stationary PDF $\pi_i(\boldsymbol{\theta})$ of the Markov Chain, the samples may not be distributed according to the underlying target distribution $\pi_i(\boldsymbol{\theta})$. Thus, the resulting estimate of the evidence in (3.3) will be biased since the samples $\boldsymbol{\theta}_{i-1}^{(m)}$ are not distributed according to $\pi_{i-1}(\boldsymbol{\theta})$. While in very low dimensions the number of samples required for convergence to $\pi_i(\boldsymbol{\theta})$ at all levels may be affordable, this will not be the case in higher dimensions.

3.1.3 Methods based on samples from the posterior

The key idea here is to calculate the evidence for model class $\mathcal{M}_j$ based on samples from the posterior $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ which have already been obtained from an MCMC Bayesian updating procedure. One possible approach is to estimate the evidence $p(\mathcal{D}|\mathcal{M}_j)$ by importance sampling, which modifies (3.1) as follows:

$$p(\mathcal{D}\,|\,\mathcal{M}_j) \approx \frac{1}{N}\sum_{k=1}^{N}\frac{p(\mathcal{D}\,|\,\boldsymbol{\theta}^{(k)},\mathcal{M}_j)\,p(\boldsymbol{\theta}^{(k)}\,|\,\mathcal{M}_j)}{g(\boldsymbol{\theta}^{(k)})} \tag{3.7}$$

where the $\boldsymbol{\theta}^{(k)}$, k=1,2,…,N, are samples drawn from an importance sampling density $g(\boldsymbol{\theta})$ that is constructed using the samples from the posterior $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$; for example, by finding a kernel density estimate for $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ (e.g., Silverman 1986, Au and Beck 1999). However, unless $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ is approximately Gaussian, it is known that such an importance sampling density may lead to very poor results (i.e. large variance) in higher dimensions, especially when $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ has a heavier tail than the importance sampling density (Au and Beck 2003).

Gelfand and Dey (1994) proposed the following method by using Bayes' Theorem and the fact that the integration of any probability density function $h(\boldsymbol{\theta})$ over the whole domain equals 1:

$$\frac{1}{p(\mathcal{D}\,|\,\mathcal{M}_j)} = \int \frac{h(\boldsymbol{\theta})}{p(\mathcal{D}\,|\,\mathcal{M}_j)}\,d\boldsymbol{\theta} = \int \frac{h(\boldsymbol{\theta})\,p(\boldsymbol{\theta}\,|\,\mathcal{D},\mathcal{M}_j)}{p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j)\,p(\boldsymbol{\theta}\,|\,\mathcal{M}_j)}\,d\boldsymbol{\theta} \approx \frac{1}{N}\sum_{k=1}^{N} \frac{h(\boldsymbol{\theta}^{(k)})}{p(\mathcal{D}\,|\,\boldsymbol{\theta}^{(k)},\mathcal{M}_j)\,p(\boldsymbol{\theta}^{(k)}\,|\,\mathcal{M}_j)}$$

where $\boldsymbol{\theta}^{(k)}$, $k = 1,2,\ldots, N$, are N samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. Thus, the estimate for the evidence is given as follows:

$$p(\mathcal{D}\,|\,\mathcal{M}_j) \approx N\Big(\sum_{k=1}^{N} \frac{h(\boldsymbol{\theta}^{(k)})}{p(\mathcal{D}\,|\,\boldsymbol{\theta}^{(k)},\mathcal{M}_j)\,p(\boldsymbol{\theta}^{(k)}\,|\,\mathcal{M}_j)}\Big)^{-1} \tag{3.8}$$

The above is a generalization of the special case proposed by Newton and Raftery (1994) where $h(\boldsymbol{\theta}) = p(\boldsymbol{\theta}|\mathcal{M}_j)$. The main advantages of this estimator are: (i) except for the calculation of $h(\boldsymbol{\theta})$ when $h(\boldsymbol{\theta}) \neq p(\boldsymbol{\theta}|\mathcal{M}_j)$, no additional computational effort is required since the values of $p(\mathcal{D}|\boldsymbol{\theta}^{(k)},\mathcal{M}_j)$ have already been obtained during the simulation of $\boldsymbol{\theta}^{(k)}$, $k = 1,2,\ldots, N$, from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$; and (ii) the estimator is consistent, i.e, as N approaches infinity, the estimator converges to the exact value of the evidence $p(\mathcal{D}|\mathcal{M}_j)$. However, a serious drawback of this estimator is that it can be quite unstable due to the occurrence of samples with small $h(\boldsymbol{\theta})/[p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}_j)\ p(\boldsymbol{\theta}|\mathcal{M}_j)]$ which may have a significant effect on the estimate if the sample size is not large enough and, in fact, may give infinite variance for the estimator in (3.8).

3.2 Proposed method based on posterior samples

Here we derive an alternative to the above methods for calculating the evidence from posterior samples. In Step 1, we derive an approximate analytical expression for the posterior and then in Step 2, we use this approximate posterior to approximate the evidence.

3.2.1 Step 1: Analytical approximation for the posterior PDF

Consider any MCMC algorithm with transition PDF $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ that is constructed to generate posterior samples from its stationary PDF $\pi(\boldsymbol{\theta})=p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. The key idea is to observe that $\pi(\boldsymbol{\theta})$ satisfies the following stationarity condition:

$$\pi(\boldsymbol{\theta}) = \int K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* \tag{3.9}$$

We use (3.9) to derive an approximate analytical expression for the posterior PDF $\pi(\boldsymbol{\theta})$ (we will soon illustrate this using special cases).

Consider a general choice of $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ that includes many MCMC algorithms:

$$K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*) = T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*) + (1 - a(\boldsymbol{\theta}^*))\delta(\boldsymbol{\theta} - \boldsymbol{\theta}^*) \tag{3.10}$$

where $T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)$ is a smooth function that does not contain delta functions and $a(\boldsymbol{\theta}^*)$ is the acceptance probability given by the following integral $a(\boldsymbol{\theta}^*) = \int T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)d\boldsymbol{\theta} \leq 1$ so that $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ is correctly normalized. Then we have,

$$\begin{aligned}
\pi(\boldsymbol{\theta}) &= \int K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* \\
&= \int T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* + \int (1 - a(\boldsymbol{\theta}^*))\delta(\boldsymbol{\theta} - \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* \\
&= \int T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* + \pi(\boldsymbol{\theta}) - a(\boldsymbol{\theta})\pi(\boldsymbol{\theta})
\end{aligned}$$

Thus, the proposed analytical approximation for $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ in terms of posterior samples is:

$$p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}^{(j)}) = \pi(\boldsymbol{\theta}) = \frac{\int T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^*}{a(\boldsymbol{\theta})} \approx \frac{1}{a(\boldsymbol{\theta})N}\sum_{k=1}^{N} T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) \tag{3.11}$$

where $\boldsymbol{\theta}^{(k)}$ are samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. This equation is also valid in the case where $a(\boldsymbol{\theta})=1$, $\forall\, \boldsymbol{\theta}$, so $K(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)=T(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)$. It is also worth noting that (3.11) can be used to give a kernel density estimate of the posterior PDF which can be used in the multi-level MCMC method of Beck and Au (2002). We consider three choices for $K(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)$ here.

3.2.1.1 $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ from Metropolis-Hastings algorithm

Consider the Metropolis-Hastings algorithm (Metropolis et al. 1953; Hastings 1970; Robert and Casella 2004) with a proposal distribution $q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$, then:

$$T(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*) = r(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)q(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*) \tag{3.12}$$

where $r(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)$ is given by

$$r(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*) = \min\{1, \frac{p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j)p(\boldsymbol{\theta}\,|\,\mathcal{M}_j)q(\boldsymbol{\theta}^*\,|\,\boldsymbol{\theta})}{p(\mathcal{D}\,|\,\boldsymbol{\theta}^*,\mathcal{M}_j)p(\boldsymbol{\theta}^*\,|\,\mathcal{M}_j)q(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^*)}\} \tag{3.13}$$

Equation (3.11) can be used to give an analytical approximation to the posterior where the denominator in (3.11) is estimated as follows:

$$a(\boldsymbol{\theta}) = \int T(\boldsymbol{\theta}^*\,|\,\boldsymbol{\theta})d\boldsymbol{\theta}^* = \int r(\boldsymbol{\theta}^*\,|\,\boldsymbol{\theta})q(\boldsymbol{\theta}^*\,|\,\boldsymbol{\theta})d\boldsymbol{\theta}^* \approx \frac{1}{N_2}\sum_{k=1}^{N_2} r(\boldsymbol{\theta}^{*(k)}\,|\,\boldsymbol{\theta}) \tag{3.14}$$

where the $\boldsymbol{\theta}^{*(k)}$ are N_2 samples from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ for fixed $\boldsymbol{\theta}$. Note that the posterior samples used in (3.11) need not be generated using the Metropolis-Hastings algorithm but it is often convenient to do so.

Chib and Jeliazkov (2001) considered the special case where one can obtain posterior samples from the Metropolis-Hastings algorithm and obtained the same results as in (3.11)-

(3.14) by making use of a specific property of the Metropolis-Hastings algorithm known as the reversibility of the transition PDF:

$$K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta}) = K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*) \tag{3.15}$$

Any Markov chain with transition PDF $K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*)$ satisfying (3.15) also satisfies (3.9) but not conversely. The approach presented in this paper is a generalization to any MCMC algorithm since it only requires the stationarity condition (3.9) to hold.

3.2.1.2 $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ from Gibbs sampling algorithm

Suppose that $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ are divided into G groups of uncertain parameter vectors, i.e. $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_G]$ and $\boldsymbol{\theta}^* = [\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_G^*]$, then the Gibbs sampling algorithm (Geman and Geman 1984) has transition PDF:

$$K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) = T(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) = \prod_{j=1}^{G} \pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) \tag{3.16}$$

where $\{\boldsymbol{\theta}^*\}_{1:j-1} = [\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_{j-1}^*]$, $\{\boldsymbol{\theta}\}_{j+1:G} = [\boldsymbol{\theta}_{j+1}, \ldots, \boldsymbol{\theta}_G]$ and for j=1, $\{\boldsymbol{\theta}^*\}_{1:j-1}$ is dropped and for j=G, $\{\boldsymbol{\theta}\}_{j+1:G}$ is dropped. Since $\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G})$ is the target conditional PDF of $\boldsymbol{\theta}_j^*$ given all the other components:

$$\pi(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \{\boldsymbol{\theta}\}_{j+1:G}) = \frac{\pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_j^*, \boldsymbol{\theta}_{j+1}, \ldots, \boldsymbol{\theta}_G)}{\int \pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_{j-1}^*, \tilde{\boldsymbol{\theta}}_j, \boldsymbol{\theta}_{j+1}, \ldots, \boldsymbol{\theta}_G) d\tilde{\boldsymbol{\theta}}_j} \tag{3.17}$$

In this case, $a(\boldsymbol{\theta})$ is always 1. Thus, the analytical approximation to the posterior PDF in terms of posterior samples $\boldsymbol{\theta}^{(k)}$ in (3.11) becomes:

$$p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) \approx \frac{1}{N} \sum_{k=1}^{N} T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) \tag{3.18}$$

Note that this case is only appropriate when the conditional PDFs in (3.17) can be evaluated analytically. Also, the posterior samples used in (3.18) need not come from Gibbs sampling.

Chib (1995) considered this special case where one can obtain posterior samples using Gibbs sampling. His approach requires additional simulation of samples with the amount of computational effort increasing linearly with the number of groups. The approach presented here for this case results in an estimator for the posterior which does not require additional simulation of samples once the posterior samples have been obtained.

3.2.1.3 $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ from hybrid MCMC-Gibbs sampling algorithm

A hybrid approach of simulating samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ is proposed where $\boldsymbol{\theta}$ is split into several groups of uncertain parameters where the conditional distribution of almost every group of uncertain parameters given the other groups of uncertain parameters is such that direct simulation is possible, facilitating the use of Gibbs sampling. MCMC methods such as Metropolis-Hastings algorithm or advanced MCMC methods such as those presented in Beck and Au (2002), Ching and Chen (2007), and Cheung and Beck (2007c, 2008a) and Chapter 2, can be used to simulate samples from the conditional distribution of the groups of uncertain parameters that cannot be done by the standard MCS (Monte Carlo simulation) procedure. This approach is especially effective for a case that often occurs in applications where the sum of the dimensions of the groups of parameters whose conditional distributions allow direct MCS simulation is large and the correlation induced by the data between different groups of parameters is small.

This Gibbs sampling in groups naturally leads to the choice for the transition PDF $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ similar to the one in (3.16). Suppose that $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ are divided into G groups, i.e. $\boldsymbol{\theta} = [\boldsymbol{\theta}_1,$

$\boldsymbol{\theta}_2,\ldots, \boldsymbol{\theta}_G]$, $\boldsymbol{\theta}^* = [\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \ldots, \boldsymbol{\theta}_G^*]$. Imagine a sample $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2,\ldots, \boldsymbol{\theta}_G]$ is generated given $\boldsymbol{\theta}^*$ such that the j-th group is generated by using a Markov chain with transition PDF $K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j:G})$ with the corresponding stationary PDF $\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G})$. This implies the following:

$$K(\boldsymbol{\theta} \mid \boldsymbol{\theta}^*) = \prod_{j=1}^{G} K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j:G}) \tag{3.19}$$

where the K_j satisfy the stationarity condition

$$\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G}) = \int K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j \tag{3.20}$$

It is shown in Appendix 3A that the transition PDF $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ defined by (3.19) and (3.20) satisfies the stationarity condition (3.9).

***Illustrative special case*: G=2**

Suppose that there are G=2 groups and that the transition PDFs for each group are as follows:

$$K_1(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*) = \pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2^*) \tag{3.21}$$

$$K_2(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) = T_2(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) + (1 - a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*))\delta(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_2^*) \tag{3.22}$$

where

$$T_2(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) = r(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) q(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) \tag{3.23}$$

$$r(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) = \min\{1, \frac{\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) q(\boldsymbol{\theta}_2^* \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}{\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) q(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*)}\} \tag{3.24}$$

$$a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) = \int T_2(\tilde{\boldsymbol{\theta}}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*) d\tilde{\boldsymbol{\theta}}_2 \tag{3.25}$$

These choices of the transition PDFs for $\boldsymbol{\theta}_1$ and for $\boldsymbol{\theta}_2$ correspond to generating samples of ($\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$) by first sampling $\boldsymbol{\theta}_1$ from its corresponding conditional PDF as in Gibbs sampling and then sampling the second group $\boldsymbol{\theta}_2$ by Metropolis-Hastings sampling with a proposal PDF $q(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^*)$ and the stationary PDF being the corresponding full conditional PDF $\pi(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1)$. These choices are appropriate if $\pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)$ is of the form that allows direct sampling for $\boldsymbol{\theta}_1$, given $\boldsymbol{\theta}_2$, but $\pi(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1)$ does not allow direct sampling of $\boldsymbol{\theta}_2$, given $\boldsymbol{\theta}_1$. In this special case with G=2, the expression for $\pi(\boldsymbol{\theta})$ which is necessary for the calculation of the evidence evaluated at $\boldsymbol{\theta}$ can be derived as follows:

$$\begin{aligned}
\pi(\boldsymbol{\theta}) &= \int K(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}) \pi(\tilde{\boldsymbol{\theta}}) d\tilde{\boldsymbol{\theta}} \\
&= \int \pi(\boldsymbol{\theta}_1 \mid \tilde{\boldsymbol{\theta}}_2)[T_2(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_2) + (1 - a_2(\boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_2))\delta(\boldsymbol{\theta}_2 - \tilde{\boldsymbol{\theta}}_2)]\pi(\tilde{\boldsymbol{\theta}}) d\tilde{\boldsymbol{\theta}} \\
&= I_1 + \int \pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)(1 - a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2))\pi(\tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) d\tilde{\boldsymbol{\theta}}_1
\end{aligned}$$

where $I_1 = \int \pi(\boldsymbol{\theta}_1 \mid \tilde{\boldsymbol{\theta}}_2) T_2(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_2) \pi(\tilde{\boldsymbol{\theta}}_2) d\tilde{\boldsymbol{\theta}}_2$ and $\pi(\boldsymbol{\theta}_2)$ is the marginal PDF of $\boldsymbol{\theta}_2$ from $\pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Thus,

$$\begin{aligned}
\pi(\boldsymbol{\theta}) &= I_1 + \pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)(1 - a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)) \int \pi(\tilde{\boldsymbol{\theta}}_1, \boldsymbol{\theta}_2) d\tilde{\boldsymbol{\theta}}_1 \\
&= I_1 + \pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)(1 - a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2))\pi(\boldsymbol{\theta}_2) \\
&= I_1 + (1 - a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2))\pi(\boldsymbol{\theta}) \qquad (\pi(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)\pi(\boldsymbol{\theta}_2)) \\
&= I_1 + \pi(\boldsymbol{\theta}) - \pi(\boldsymbol{\theta}) a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)
\end{aligned}$$

Finally,

$$p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) = \pi(\boldsymbol{\theta}) = \frac{I_1}{a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)} \tag{3.26}$$

where the numerator and denominator can be estimated by:

$$I_1 = \int \pi(\boldsymbol{\theta}_1 \mid \tilde{\boldsymbol{\theta}}_2) T(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \tilde{\boldsymbol{\theta}}_2) \pi(\tilde{\boldsymbol{\theta}}_2) d\tilde{\boldsymbol{\theta}}_2 \approx \frac{1}{N_1} \sum_{i=1}^{N_1} \pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2^{(i)}) T(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(i)}) \tag{3.27}$$

where $\boldsymbol{\theta}_2^{(i)}$ are marginal samples of $\boldsymbol{\theta}_2$ obtained from posterior samples ($\boldsymbol{\theta}_1^{(i)}$, $\boldsymbol{\theta}_2^{(i)}$) corresponding to $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}^{(j)})$ where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, and:

$$a_2(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \approx \frac{1}{N_2} \sum_{i=1}^{N_2} r(\tilde{\boldsymbol{\theta}}_2^{(i)} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \tag{3.28}$$

where $\tilde{\boldsymbol{\theta}}_2^{(i)}$ are samples from $q(\tilde{\boldsymbol{\theta}}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ for fixed $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$.

More general case*: *A case with G>2

Consider a generalization of the above case where the transition PDFs for each component are taken as follows. For the first J groups of parameters, $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$,…, $\boldsymbol{\theta}_J$, use a Gibbs sampling transition PDF:

$$K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j:G}) = \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G}), j = 1, ..., J \tag{3.29}$$

and for the remaining (G-J) groups, $\boldsymbol{\theta}_{J+1}$, $\boldsymbol{\theta}_{J+2}$,…, $\boldsymbol{\theta}_G$, use a Metropolis-Hastings transition PDFs, so for $j = J+1$,…,$G > 2$,

$$\begin{aligned} K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j:G}) &= T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}^*\}_{j+1:G}) \\ &\quad + (1 - a_j(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j^*, \{\boldsymbol{\theta}^*\}_{j+1:G})) \delta(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) \end{aligned} \tag{3.30}$$

where

$$T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j:G}) = r_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\ \}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) q_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \quad (3.31)$$

$$r_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) = \min\{1, \frac{\pi(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) q_j(\boldsymbol{\theta}^*_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G})}{\pi(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) q_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G})}\} \quad (3.32)$$

$$a_j(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) = \int T_j(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j \quad (3.33)$$

Here, the Metropolis-Hastings algorithm for group $\boldsymbol{\theta}_j$, j=J+1,…, G, has proposal PDF $q_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}^*_j, \{\boldsymbol{\theta}^*\}_{j+1:G})$ and the stationary PDF is the corresponding conditional PDF $\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G})$.

First we consider a very important general case where J=G-1 and $\pi(\boldsymbol{\theta})$ can be derived as follows using both global and local stationarity conditions in (3.9) and (3.20) respectively:

$$\begin{aligned}
\pi(\boldsymbol{\theta}) &= \int K(\boldsymbol{\theta} \mid \tilde{\boldsymbol{\theta}}) \pi(\tilde{\boldsymbol{\theta}}) d\tilde{\boldsymbol{\theta}} \\
&= \int [\prod_{j=1}^{G-1} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:G})][T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G) + (1 - a(\{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)) \delta(\boldsymbol{\theta}_G - \tilde{\boldsymbol{\theta}}_G)] \pi(\tilde{\boldsymbol{\theta}}) d\tilde{\boldsymbol{\theta}} \\
&= I_1 + \int [\prod_{j=1}^{G-1} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:G-1,} \boldsymbol{\theta}_G)](1 - a(\{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G)) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1,}, \boldsymbol{\theta}_G) d\{\tilde{\boldsymbol{\theta}}\}_{1:G-1}
\end{aligned}$$

where I_1 is given in (3.34) and thus,

$$\begin{aligned}
\pi(\boldsymbol{\theta}) &= I_1 + \pi(\boldsymbol{\theta}_G)(1 - a(\{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G)) \int [\prod_{j=1}^{G-1} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:G-1,} \boldsymbol{\theta}_G)] \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \mid \boldsymbol{\theta}_G) d\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \\
&= I_1 + \pi(\boldsymbol{\theta}_G)(1 - I_2) I_3
\end{aligned}$$

where I_1, I_2 and I_3 can be estimated by:

$$
\begin{aligned}
I_1 &= \int [\prod_{j=1}^{G-1} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:G})] T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G) \pi(\tilde{\boldsymbol{\theta}}) d\tilde{\boldsymbol{\theta}} \\
&\approx \frac{1}{N_1} \sum_{k=1}^{N_1} [\prod_{j=1}^{G-1} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}^{(k)}\}_{j+1:G})] T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G^{(k)})
\end{aligned}
\tag{3.34}
$$

where $\tilde{\boldsymbol{\theta}}^{(k)}$ are samples from $p(\tilde{\boldsymbol{\theta}} \mid \mathcal{D}, \mathcal{M}^{(j)})$ and $\tilde{\boldsymbol{\theta}}_j^{(k)}$ is the j-th group of $\tilde{\boldsymbol{\theta}}^{(k)}$.

$$
\begin{aligned}
I_2 &= a(\{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G) = \int T_G(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G) d\tilde{\boldsymbol{\theta}}_G \\
&= \int r_G(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G) q_G(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G) d\tilde{\boldsymbol{\theta}}_G \approx \frac{1}{N_2} \sum_{m=1}^{N_2} r_G(\tilde{\boldsymbol{\theta}}_G^{(m)} \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G)
\end{aligned}
\tag{3.35}
$$

where $\tilde{\boldsymbol{\theta}}_G^{(m)}$ are samples from $q_G(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G)$.

$$
\begin{aligned}
I_3 &= \int [\prod_{j=1}^{G-1} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:G-1,} \boldsymbol{\theta}_G)] \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \mid \boldsymbol{\theta}_G) d\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \\
&\approx \frac{1}{N_3} \sum_{i=1}^{N_3} [\prod_{j=1}^{G-1} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{j+1:G-1,} \boldsymbol{\theta}_G)]
\end{aligned}
\tag{3.36}
$$

where $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}$ are samples from $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \mid \boldsymbol{\theta}_G)$ which can be generated one group after another using Gibbs sampling as follows: With $\tilde{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}$, generate the first group $\tilde{\boldsymbol{\theta}}_1^{(i)}$ of $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}$ from $\pi(\tilde{\boldsymbol{\theta}}_1 \mid \{\tilde{\boldsymbol{\theta}}^{(i-1)}\}_{2:G-1}, \boldsymbol{\theta}_G)$ and for m=2,…,G-2, the m-th group $\tilde{\boldsymbol{\theta}}_m^{(i)}$ of $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}$ from $\pi(\tilde{\boldsymbol{\theta}}_m \mid \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:m-1}, \{\tilde{\boldsymbol{\theta}}^{(i-1)}\}_{m+1:G-1}, \boldsymbol{\theta}_G)$ and the (G-1)-th group $\tilde{\boldsymbol{\theta}}_{G-1}^{(i)}$ of $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}$ from $\pi(\tilde{\boldsymbol{\theta}}_{G-1} \mid \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-2}, \boldsymbol{\theta}_G)$. As soon as we pick $\boldsymbol{\theta}$ for one of the samples from $p(\tilde{\boldsymbol{\theta}} \mid \mathcal{D}, \mathcal{M}^{(j)})$, $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}$ generated using the above procedures will follow $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \mid \boldsymbol{\theta}_G)$.

Finally, the expression for estimating $\pi(\boldsymbol{\theta}_G)$ which is the marginal PDF of $\boldsymbol{\theta}_G$ from $\pi(\boldsymbol{\theta})$ is (see in Appendix 3B for derivation):

$$\pi(\boldsymbol{\theta}_G) = \frac{II_1}{II_2} \tag{3.37}$$

where the numerator and denominator can be estimated by

$$II_1 = \int T_G(\boldsymbol{\theta}_G \mid \{\hat{\boldsymbol{\theta}}\}_{1:G-1}, \hat{\boldsymbol{\theta}}_G)\pi(\hat{\boldsymbol{\theta}})d\hat{\boldsymbol{\theta}}_G d\{\hat{\boldsymbol{\theta}}\}_{1:G-1} \approx \frac{1}{M_1}\sum_{k=1}^{M_1} T_G(\boldsymbol{\theta}_G \mid \{\hat{\boldsymbol{\theta}}^{(k)}\}_{1:G-1}, \hat{\boldsymbol{\theta}}_G{}^{(k)}) \tag{3.38}$$

$$II_2 = \int r_G(\tilde{\boldsymbol{\theta}}_G \mid \{\tilde{\boldsymbol{\theta}}\}_{1:G-1}, \boldsymbol{\theta}_G)\tilde{q}(\{\tilde{\boldsymbol{\theta}}\}_{1:G} \mid \boldsymbol{\theta}_G)d\{\tilde{\boldsymbol{\theta}}\}_{1:G} \approx \frac{1}{M_2}\sum_{i=1}^{M_2} r_G(\tilde{\boldsymbol{\theta}}_G^{(i)} \mid \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}, \boldsymbol{\theta}_G) \tag{3.39}$$

where $\hat{\boldsymbol{\theta}}^{(k)}$ are samples from $p(\hat{\boldsymbol{\theta}} \mid \mathcal{D}, \mathcal{M}^{(j)})$ and $\hat{\boldsymbol{\theta}}_j^{(k)}$ is the j-th group of $\hat{\boldsymbol{\theta}}^{(k)}$; $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}$ are samples from $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \mid \boldsymbol{\theta}_G)$ which have already been generated when estimating I_3 in (3.36) and $\tilde{\boldsymbol{\theta}}_G^{(i)}$ is generated from $q_G(\tilde{\boldsymbol{\theta}}_G \mid \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}, \boldsymbol{\theta}_G)$. Thus one can see that $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:G-1}$ and $\tilde{\boldsymbol{\theta}}_G^{(i)}$ jointly follow the probability distribution given below:

$$\tilde{q}(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G \mid \boldsymbol{\theta}_G) = q_G(\tilde{\boldsymbol{\theta}}_G \mid \{\tilde{\boldsymbol{\theta}}\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \mid \boldsymbol{\theta}_G) \tag{3.40}$$

It should be noted that for any number G of groups of parameter vectors, to calculate $\pi(\boldsymbol{\theta})$, only I_1, I_2, II_1 and II_2 need to be calculated. The above derivation is quite general without requiring reversibility or detailed balance conditions.

***A general case*:**

Now we consider an even more general case with any J and G. We show in Appendix 3C that the joint transition PDF corresponding to those groups of parameters simulated from their conditional distributions (i.e., $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$,…, $\boldsymbol{\theta}_J$) conditioned on the other groups (i.e., $\boldsymbol{\theta}_{J+1}$,

$\boldsymbol{\theta}_{J+2},\ldots,\boldsymbol{\theta}_G$) satisfies the following stationarity condition with the conditional stationary PDF :

$$\pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) = \int \prod_{j=1}^{J} K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_{1:J} \quad (3.41)$$

The above is true for all values of $\{\boldsymbol{\theta}\}_{1:J}$ and $\{\tilde{\boldsymbol{\theta}}\}_{J+1:G}$ and thus $\{\tilde{\boldsymbol{\theta}}\}_{J+1:G}$ can be simply replaced by $\{\boldsymbol{\theta}\}_{J+1:G}$. Using the above, $\pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G})$ can be estimated as follows:

$$\begin{aligned}
&\pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G}) = \int \prod_{j=1}^{J} K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j:J}, \{\boldsymbol{\theta}\}_{J+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_{1:J} \\
&= \int \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\boldsymbol{\theta}\}_{J+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_{1:J} \qquad (3.42) \\
&\approx \frac{1}{N_1} \sum_{i=1}^{N_1} [\prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{j+1:J}, \{\boldsymbol{\theta}\}_{J+1:G})]
\end{aligned}$$

where $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:J}$, i=1,2,…,N_1 are samples from $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G})$ which can be generated one group after another using Gibbs sampling as follows: With $\{\tilde{\boldsymbol{\theta}}^{(0)}\}_{1:J} = \{\boldsymbol{\theta}\}_{1:J}$, generate the first group $\tilde{\boldsymbol{\theta}}_1^{(i)}$ of $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:J}$ from $\pi(\tilde{\boldsymbol{\theta}}_1 \mid \{\tilde{\boldsymbol{\theta}}^{(i-1)}\}_{2:J}, \{\boldsymbol{\theta}\}_{J+1:G})$ and for m=2,…,J-1, the m-th group $\tilde{\boldsymbol{\theta}}_m^{(i)}$ of $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:J}$ from $\pi(\tilde{\boldsymbol{\theta}}_m \mid \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:m-1}, \{\tilde{\boldsymbol{\theta}}^{(i-1)}\}_{m+1:J}, \{\boldsymbol{\theta}\}_{J+1:G})$ and the J-th group $\tilde{\boldsymbol{\theta}}_J^{(i)}$ of $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:J}$ from $\pi(\tilde{\boldsymbol{\theta}}_J \mid \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:J-1}, \{\boldsymbol{\theta}\}_{J+1:G})$. As soon as we pick $\{\tilde{\boldsymbol{\theta}}^{(0)}\}_{1:J} = \{\boldsymbol{\theta}\}_{1:J}$, the first J groups of one of the samples from $p(\tilde{\boldsymbol{\theta}} \mid \mathcal{D}, \mathcal{M}^{(j)})$, $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:J}$ generated using the above procedures will follow $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G})$.

Since $\pi(\boldsymbol{\theta}) = \pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G}) \pi(\{\boldsymbol{\theta}\}_{J+1:G})$, to estimate $\pi(\boldsymbol{\theta})$ for the calculation of the evidence evaluated at $\boldsymbol{\theta}$, it remains to calculate $\pi(\{\boldsymbol{\theta}\}_{J+1:G})$ as follows. Using the local

stationarity conditions in (3.20) for groups $\boldsymbol{\theta}_{J+1}$, $\boldsymbol{\theta}_{J+2},\ldots,\boldsymbol{\theta}_G$, it is shown in Appendix 3D that the following is true for $j=J+1,\ldots,G$:

$$\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{j+1:G}) = \frac{\int T_j(\boldsymbol{\theta}_j \mid \{\tilde{\boldsymbol{\theta}}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j d\{\tilde{\boldsymbol{\theta}}\}_{1:j-1}}{\int T_j(\tilde{\boldsymbol{\theta}}_j \mid \{\tilde{\boldsymbol{\theta}}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j d\{\tilde{\boldsymbol{\theta}}\}_{1:j-1}} = \frac{I_1}{I_2} \quad (3.43)$$

where the numerator I_1 and denominator I_2 can be estimated by:

$$I_1 \approx \frac{1}{N_1} \sum_{i=1}^{N_1} T_j(\boldsymbol{\theta}_j \mid \{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j^{(i)}, \{\boldsymbol{\theta}\}_{j+1:G}) \quad (3.44)$$

$$\begin{aligned} I_2 &= \int r_j(\hat{\boldsymbol{\theta}}_j \mid \{\hat{\boldsymbol{\theta}}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \tilde{q}_j(\{\hat{\boldsymbol{\theta}}\}_{1:j-1}, \hat{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{j+1:G}) d\hat{\boldsymbol{\theta}}_j d\{\hat{\boldsymbol{\theta}}\}_{1:j-1} \\ &\approx \frac{1}{N_2} \sum_{i=1}^{N_2} r_j(\hat{\boldsymbol{\theta}}_j^{(i)} \mid \{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \end{aligned} \quad (3.45)$$

where $\{\boldsymbol{\theta}\}_{j+1:G}$ is empty when $j \geq G$. One can easily verify that $\{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}$ and $\hat{\boldsymbol{\theta}}_j^{(i)}$ jointly follow the probability distribution:

$$\tilde{q}_j(\{\hat{\boldsymbol{\theta}}\}_{1:j-1}, \hat{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{j+1:G}) = q_j(\hat{\boldsymbol{\theta}}_j \mid \{\hat{\boldsymbol{\theta}}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \pi(\{\hat{\boldsymbol{\theta}}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G}) \quad (3.46)$$

Thus, $\{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}$, $i=1,2,\ldots,N_2$, are samples from $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$ which can be generated one group after another using a sampling procedure as follows: With $\tilde{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}$, generate the first group $\hat{\boldsymbol{\theta}}_1^{(i)}$ of $\{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}$ from $\pi(\hat{\boldsymbol{\theta}}_1^i \mid \{\hat{\boldsymbol{\theta}}^{(i-1)}\}_{2:j-1}, \{\boldsymbol{\theta}\}_{j:G})$ and for $m=2,\ldots,J-1$, the m-th group $\hat{\boldsymbol{\theta}}_m^{(i)}$ of $\{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}$ from $\pi(\hat{\boldsymbol{\theta}}_m \mid \{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:m-1}, \{\hat{\boldsymbol{\theta}}^{(i-1)}\}_{m+1:j}, \{\boldsymbol{\theta}\}_{j+1:G})$ and the J-th group $\hat{\boldsymbol{\theta}}_J^{(i)}$ of $\{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}$ from $\pi(\hat{\boldsymbol{\theta}}_J \mid \{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:J-1}, \{\hat{\boldsymbol{\theta}}^{(i-1)}\}_{J+1:j}, \{\boldsymbol{\theta}\}_{j+1:G})$. For $j>J+1$, the following procedures are required: the groups, $\hat{\boldsymbol{\theta}}_{J+1}^{(i)}$, $\hat{\boldsymbol{\theta}}_{J+2}^{(i)},\ldots,\hat{\boldsymbol{\theta}}_{j-1}^{(i)}$ of $\{\hat{\boldsymbol{\theta}}^{(i)}\}_{1:j-1}$ are generated by the

Metropolis-Hastings algorithm with a proposal PDF $q_m(\boldsymbol{\theta}_m \mid \{\widehat{\boldsymbol{\theta}}^{(i)}\}_{1:m-1}, \{\widehat{\boldsymbol{\theta}}^{(i-1)}\}_{m:j}, \{\boldsymbol{\theta}\}_{j+1:G})$ for m=J+1,…, j-1, and a stationary PDF equal to the corresponding full conditional PDF $\pi(\widehat{\boldsymbol{\theta}}_m \mid \{\widehat{\boldsymbol{\theta}}^{(i)}\}_{1:m-1}, \{\widehat{\boldsymbol{\theta}}^{(i-1)}\}_{m+1:j}, \{\boldsymbol{\theta}\}_{j+1:G})$.

Finally, $\widehat{\boldsymbol{\theta}}_j^{(i)}$ is generated from $q_j(\widehat{\boldsymbol{\theta}}_j \mid \{\widehat{\boldsymbol{\theta}}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$. As soon as we pick $\{\widehat{\boldsymbol{\theta}}^{(0)}\}_{1:j} = \{\boldsymbol{\theta}\}_{1:j}$, the first j groups of one of the samples from $p(\tilde{\boldsymbol{\theta}} \mid \mathcal{D}, \mathcal{M}^{(j)})$, $\{\widehat{\boldsymbol{\theta}}^{(i)}\}_{1:j}$ that is generated by using the above procedure will follow $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})$. The proof for the validity of these procedures is omitted out for brevity.

The samples $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:j}$ in (3.44) for i=1,2,…,N_1 are from $\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{j+1:G})$ which, for $j \le G$-1, have already been generated when estimating the denominator of $\pi(\boldsymbol{\theta}_{j+1} \mid \{\boldsymbol{\theta}\}_{j+2:G})$ and for j=G, samples $\{\tilde{\boldsymbol{\theta}}^{(i)}\}_{1:j}$ in (3.44), for i=1,2,…,N_1 are from $\pi(\tilde{\boldsymbol{\theta}})$ and they have also already been generated. It should be noted that $\{\boldsymbol{\theta}\}_{j+2:G}$ is empty when $j \ge G$-1. In short, $\pi(\boldsymbol{\theta})$ can be estimated as follows:

$$\pi(\boldsymbol{\theta}) = \pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G})\pi(\{\boldsymbol{\theta}\}_{J+1:G}) = \pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G}) \prod_{j=J+1}^{G} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{j+1:G}) \quad (3.47)$$

where the estimate for $\pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\boldsymbol{\theta}\}_{J+1:G})$ is given by (3.42) and those for $\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{j+1:G})$ are given by (3.43)-(3.45) for j=J+1,…, G. For the special case when J=0, the above result will be similar to that presented in Chib and Jeliazkov (2001) with some reordering of the groups of parameters.

3.2.2 Step 2: Approximation of log evidence

By Bayes' Theorem, the log evidence is given by:

$$\ln p(\mathcal{D}\,|\,\mathcal{M}_j) = \ln p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j) + \ln p(\boldsymbol{\theta}\,|\,\mathcal{M}_j) - \ln p(\boldsymbol{\theta}\,|\,\mathcal{D},\mathcal{M}_j) \tag{3.48}$$

The above is true for all $\boldsymbol{\theta}$. The last term can be calculated using the method presented in Step 1 and the first two terms can be computed directly from the given likelihood and prior for the model class. The biggest advantage of the proposed method for evaluating the evidence is that it is valid for any MCMC method that is constructed to generate posterior samples from the stationary PDF of the corresponding Markov chain.

By (3.48), it can be seen that the accuracy of the estimate for the evidence depends only on the accuracy of the estimate for the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ evaluated at $\boldsymbol{\theta}$ since the first two terms are known exactly. The method for calculating the statistical accuracy of the proposed evidence estimator is presented in the next section. All of the estimates for $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ given previously are unbiased and consistent and the dependence of their accuracy on the number of samples depends on the choice of $\boldsymbol{\theta}$ and which $K(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ is adopted. A more accurate estimate for the log evidence can be obtained by averaging the estimates from (3.48) using different $\boldsymbol{\theta}$:

$$\ln p(\mathcal{D}\,|\,\mathcal{M}_j) = \frac{1}{Q}\sum_{q=1}^{Q}[\ln p(\mathcal{D}\,|\,\boldsymbol{\theta}^{(q)},\mathcal{M}_j) + \ln p(\boldsymbol{\theta}^{(q)}\,|\,\mathcal{M}_j) - \ln p(\boldsymbol{\theta}^{(q)}\,|\,\mathcal{D},\mathcal{M}_j)] \tag{3.49}$$

For instance, the $\boldsymbol{\theta}^{(q)}$'s can be chosen to be those samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ that give the Q largest values of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$, or, equivalently, of $[\ln p(\mathcal{D}|\boldsymbol{\theta}^{(k)},\mathcal{M}_j)+\ln p(\boldsymbol{\theta}|\mathcal{M}_j)]$.

If only one $\boldsymbol{\theta}$ is used, it could be chosen to be, for example, the mean of the available samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$, or it could be chosen to be the sample from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ which gives the maximum value of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$, or, equivalently, the one which gives the maximum value of $[\ln p(\mathcal{D}|\boldsymbol{\theta}^{(k)},\mathcal{M}_j)+\ln p(\boldsymbol{\theta}|\mathcal{M}_j)]$. Recall that for the analytical approximation of the evidence, it is valid to use the transition PDF of any MCMC method; it need not be the same as the MCMC method that one uses to generate the posterior samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$.

If of interest, after the evidence has been calculated the expected information gain from the data using $\mathcal{M}_j$ can be obtained from (1.4) as follows:

$$E[\ln \frac{p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j)}{p(\boldsymbol{\theta} \mid \mathcal{M}_j)}] \approx \frac{1}{N} \sum_{k=1}^{N} \ln(p(\mathcal{D} \mid \boldsymbol{\theta}^{(k)}, \mathcal{M}_j) - \ln[p(\mathcal{D} \mid \mathcal{M}_j)] \tag{3.50}$$

where the $\boldsymbol{\theta}^{(k)}$ are N posterior samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. The information entropy (Cover and Thomas 2001) of the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ can also be obtained as follows:

$$\begin{aligned} & -\int \ln p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) d\boldsymbol{\theta} \\ & = -\int [\ln p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_j) + \ln p(\boldsymbol{\theta} \mid \mathcal{M}_j)] p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) d\boldsymbol{\theta} + \ln p(\mathcal{D} \mid \mathcal{M}_j) \\ & \approx -\frac{1}{N} \sum_{k=1}^{N} [\ln(p(\mathcal{D} \mid \boldsymbol{\theta}^{(k)}, \mathcal{M}_j) + \ln(p(\boldsymbol{\theta}^{(k)} \mid \mathcal{M}_j)] + \ln[p(\mathcal{D} \mid \mathcal{M}_j)] \end{aligned} \tag{3.51}$$

where the $\boldsymbol{\theta}^{(k)}$ are N posterior samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$.

3.2.3 Statistical accuracy of the proposed evidence estimators

The statistical accuracy of the estimators e for the log evidence $\ln p(\mathcal{D}|\mathcal{M}_j)$ that are given in (3.48) can be assessed by estimating their coefficient of variation (c.o.v.) $\tilde{\delta}_e$:

$$\tilde{\delta}_e = \frac{\sqrt{\mathrm{Var}(e)}}{\mathrm{E}[e]} \tag{3.52}$$

where $\mathrm{E}[e] = \ln p(\mathcal{D}|\mathcal{M}_j)$ because the estimator e for the log evidence obtained using the proposed method (e.g. (3.11) with (3.14)), along with (3.48), is unbiased; and $\mathrm{Var}(e)$ is equal to the variance $\mathrm{Var}(\ln \tilde{p}(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j))$ of the natural log of the estimator $\tilde{p}(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j)$ for $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. All the proposed estimators for $\ln p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ presented in the previous subsection are of the form:

$$\ln \tilde{p}(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) = \ln \tilde{I}_1 - \ln \tilde{I}_2 \tag{3.53}$$

where $\tilde{I}_1$ and $\tilde{I}_2$ are always positive and have the following forms:

$$\tilde{I}_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} g_1(\boldsymbol{\theta}^{(i)}) \tag{3.54}$$

$$\tilde{I}_2 = \frac{1}{N_2} \sum_{i=1}^{N_2} g_2(\hat{\boldsymbol{\theta}}^{(i)}) \tag{3.55}$$

where $\boldsymbol{\theta}^{(k)}$ are samples or marginal samples from the posterior PDF and $\hat{\boldsymbol{\theta}}^{(k)}$ are samples from some 'artificial' proposal PDF. For example, for $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ from hybrid MCMC-Gibbs sampling algorithm, $\tilde{I}_1$ and $\tilde{I}_2$ are given by (3.27) and (3.28) respectively, $g_1(\boldsymbol{\theta}^{(i)})$ is equal to $\pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2^{(i)}) T(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{(i)})$ in (3.27) where $\boldsymbol{\theta}_2^{(i)}$ are marginal samples of $\boldsymbol{\theta}_2$ obtained from posterior samples $(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)})$ corresponding to $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}^{(j)})$ where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$; and $g_2(\hat{\boldsymbol{\theta}}^{(i)})$ is equal to $r(\hat{\boldsymbol{\theta}}_2^{(i)} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ in (3.28) where $\hat{\boldsymbol{\theta}}_2^{(i)}$ are samples from a chosen MCMC proposal PDF $q(\hat{\boldsymbol{\theta}}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ for fixed $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$. Since $\boldsymbol{\theta}^{(k)}$ and $\hat{\boldsymbol{\theta}}^{(k)}$ are independent of each other, Var($\ln \tilde{p}(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j)$) is equal to the sum of the variances of $\ln \tilde{I}_1$ and $\ln \tilde{I}_2$:

$$\text{Var}(\ln \tilde{p}(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j)) = \text{Var}(\ln \tilde{I}_1) + \text{Var}(\ln \tilde{I}_2) \tag{3.56}$$

If $K(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$ from Gibbs sampling algorithm is used, $\tilde{I}_2$ is always equal to 1 and $\text{Var}(\ln \tilde{I}_2)$ is always equal to 0.

To estimate the c.o.v. $\tilde{\delta}_e$ of the log evidence estimator e in (3.52) from one simulation run, E[e] is replaced by the log evidence estimate and Var(e) is replaced by the sum of the

estimates of the variances of $\ln \tilde{I}_1$ and $\ln \tilde{I}_2$ according to (3.56). To estimate these latter variances, first we estimate the means and variances of $\tilde{I}_1$ and $\tilde{I}_2$ as follows:

$$\mathrm{E}[\tilde{I}_1] = \mathrm{E}[g_1(\boldsymbol{\theta})] \approx \frac{1}{N_1}\sum_{i=1}^{N_1} g_1(\boldsymbol{\theta}^{(i)}), \mathrm{E}[\tilde{I}_2] = \mathrm{E}[g_2(\hat{\boldsymbol{\theta}})] \approx \frac{1}{N_2}\sum_{i=1}^{N_2} g_2(\hat{\boldsymbol{\theta}}^{(i)}) \tag{3.57}$$

$$\mathrm{Var}(\tilde{I}_2) = \frac{\mathrm{Var}(g_2(\hat{\boldsymbol{\theta}}))}{N_2} \approx \frac{1}{N_2{}^2}\sum_{i=1}^{N_2} (g_2(\hat{\boldsymbol{\theta}}^{(i)}) - \mathrm{E}[g_2(\hat{\boldsymbol{\theta}})])^2 \tag{3.58}$$

$$\mathrm{Var}(\tilde{I}_1) = \frac{\rho(0)}{N_1}(1+\lambda) \tag{3.59}$$

$$\lambda = 2\sum_{\tau=1}^{N_1-1}(1-\frac{\tau}{N_1})\frac{\rho(\tau)}{\rho(0)} \in [0, N_1 - 1] \tag{3.60}$$

$$\begin{aligned} \rho(\tau) &= E[(g_1(\boldsymbol{\theta}^{(i+\tau)}) - \mathrm{E}[g_1(\boldsymbol{\theta})])(g_1(\boldsymbol{\theta}^{(i)}) - \mathrm{E}[g_1(\boldsymbol{\theta})])] \\ &\approx \frac{1}{N_1 - \tau}\sum_{i=1}^{N_1-\tau} (g(\boldsymbol{\theta}^{(i+\tau)}) - \mathrm{E}[g_1(\boldsymbol{\theta})])(g(\boldsymbol{\theta}^{(i)}) - \mathrm{E}[g_1(\boldsymbol{\theta})]) \end{aligned} \tag{3.61}$$

From (3.54) and (3.55), it can be seen that, by the Central Limit Theorem, $\tilde{I}_1$ and $\tilde{I}_2$ approach Gaussian PDFs as the sample sizes N_1 and N_2 become sufficiently large. Thus, using the above estimates of the means and variances of $\tilde{I}_1$ and $\tilde{I}_2$, one way to evaluate the means and variances of $\ln \tilde{I}_k$, k=1,2, is to simulate samples $\tilde{I}_k^{(i)}$ (using MCS) from the PDF which can be approximated by a Gaussian PDF with mean $\mathrm{E}[\tilde{I}_k]$ and variance $\mathrm{Var}[\tilde{I}_k]$ and the estimates for the mean and variance of $\ln \tilde{I}_k$ are then equal to the sample mean and variance of $\ln \tilde{I}_k^{(i)}$. An alternative way to estimate the means and variances of $\ln \tilde{I}_1$ and $\ln \tilde{I}_2$ is by Gaussian quadrature integration with 3-point Hermite-Gauss rule (since $\tilde{I}_1$ and $\tilde{I}_2$ are approximately Gaussian):

$$\mathrm{E}[\ln \tilde{I}_k] \approx \frac{1}{6}\ln(\mathrm{E}[\tilde{I}_k] - \sqrt{3}\sqrt{\mathrm{Var}[\tilde{I}_k]}) + \frac{2}{3}\ln(\mathrm{E}[\tilde{I}_k]) + \frac{1}{6}\ln(\mathrm{E}[\tilde{I}_k] + \sqrt{3}\sqrt{\mathrm{Var}[\tilde{I}_k]}) \quad (3.62)$$

$$\begin{aligned}\mathrm{E}[(\ln \tilde{I}_k)^2] \approx{}& \frac{1}{6}(\ln(\mathrm{E}[\tilde{I}_k] - \sqrt{3}\sqrt{\mathrm{Var}[\tilde{I}_k]}))^2 + \\ & \frac{2}{3}(\ln(\mathrm{E}[\tilde{I}_k]))^2 + \frac{1}{6}(\ln(\mathrm{E}[\tilde{I}_k] + \sqrt{3}\sqrt{\mathrm{Var}[\tilde{I}_k]}))^2\end{aligned} \quad (3.63)$$

$$\mathrm{Var}(\ln \tilde{I}_k) = \mathrm{E}[(\ln \tilde{I}_k)^2] - (\mathrm{E}[\ln \tilde{I}_k])^2 \quad (3.64)$$

for k=1, 2 and $\mathrm{E}[\tilde{I}_k] - \sqrt{3}\sqrt{\mathrm{Var}[\tilde{I}_k]} > 0$. Obviously, finding the variances of $\ln \tilde{I}_1$ and $\ln \tilde{I}_2$ using (3.62)-(3.64) requires fewer computations than MCS. It is found that for the illustrative examples, this method gives similar results for the estimates of the variances of $\ln \tilde{I}_1$ and $\ln \tilde{I}_2$ as those obtained by MCS with a large number of samples.

The statistical accuracy of the estimator $f = \exp(e)$ for the evidence $p(\mathcal{D}|\mathcal{M}_j)$ can be assessed by evaluating the corresponding c.o.v. $\tilde{\delta}_f$ which can be estimated using (3.52) with e replaced by f where the estimate for E[f] is equal to the estimate for the evidence $p(\mathcal{D}|\mathcal{M}_j)$ obtained using the proposed method and (3.48), and Var(f) can be estimated using (3.62)-(3.64) by replacing $\tilde{I}_k$ by e, ln by exp and $\ln \tilde{I}_k$ by f.

In order to avoid numerical overflow when calculating a certain quantity, one should first calculate the logarithm of such quantity and exponentiate at the end. For example, when calculating $\tilde{\delta}_f$, a numerical overflow may occur due to a possible numerical overflow when calculating E[f] and Var(f). Thus, one should calculate $\ln \tilde{\delta}_f$, which is equal to (lnVar(f))/2-lnE[f] where lnE[f] is equal to $\ln p(\mathcal{D}|\mathcal{M}_j)$.

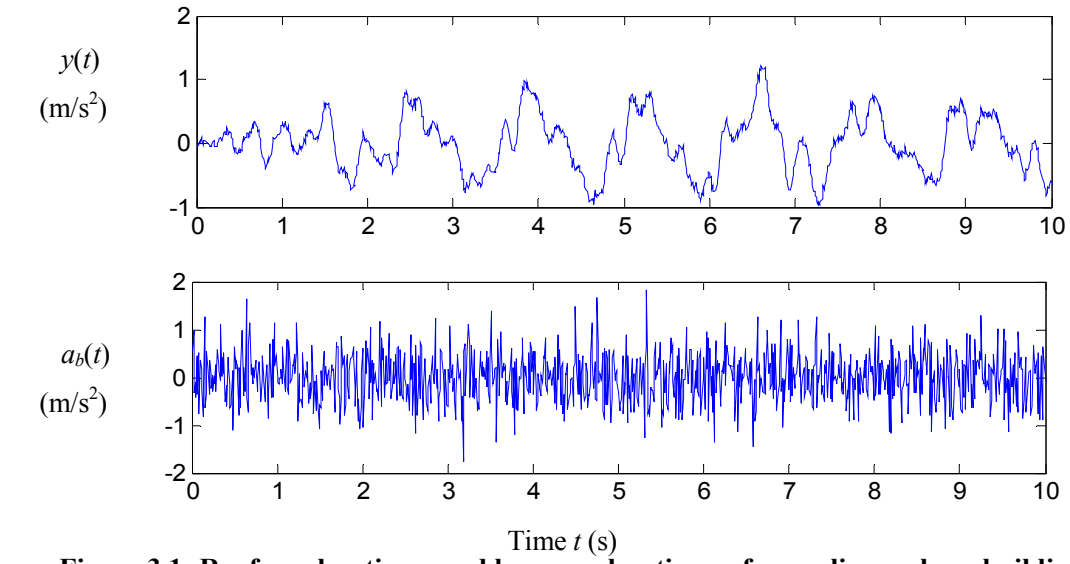

Figure 3.1: Roof acceleration y and base acceleration a_b from a linear shear building with nonclassical damping

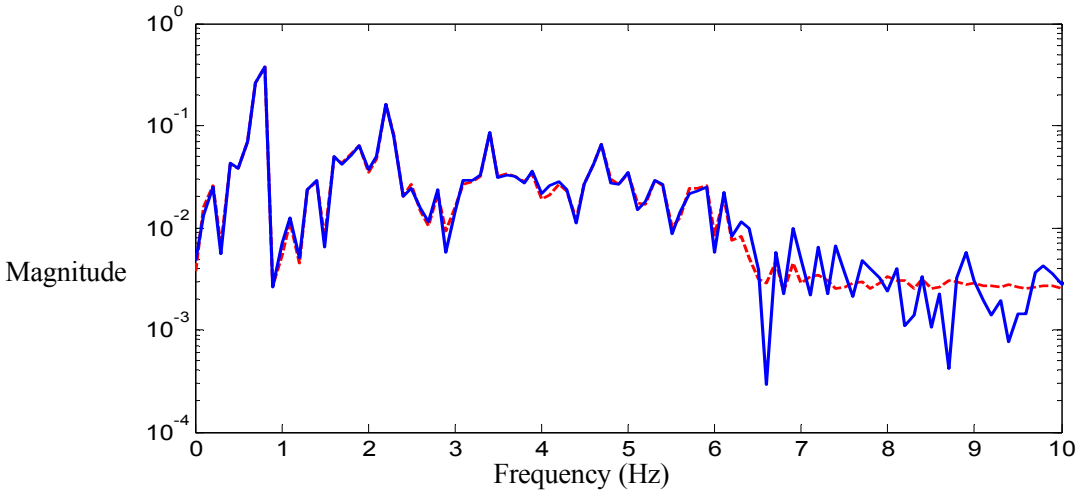

Figure 3.2: Magnitude of the FFT estimated from the measured roof acceleration data (solid curve) and mean of magnitude of the FFT from the roof acceleration estimated using posterior samples from the most probable model class $\mathcal{M}_5$ (dashed curve)

3.3 Illustrative examples

3.3.1 Example 1: Modal identification for ten-story building

In this example, the linear seismic response of a 10-story shear building with nonclassical damping is considered. The simulated dynamic data $\mathcal{D}$ consist of 10s (with a sample interval Δt of 0.01s) of the acceleration of the base a_b, and at the roof contaminated by Gaussian white noise of 10% rms noise-to-signal ratio (Figure 3.1). Here we consider a set $M=\{\mathcal{M}_j: j=1,2,\ldots,6\}$ consisting of 6 candidate model classes where $\mathcal{M}_j$ includes the linear modal model with classical damping consisting of j modes and the corresponding uncertain parameters are the modal frequencies $f_1,\ldots,f_j$, damping ratios $\xi_1,\ldots,\xi_j$, modal participation factors $\rho_1,\ldots,\rho_j$ and the prediction-error variance σ^2. Thus, for $\mathcal{M}_j$, the uncertain parameter vector $\boldsymbol{\theta}$ consists of $3j+1$ parameters (e.g. $\mathcal{M}_5$ has 16 parameters). The prior PDF for $\boldsymbol{\theta}$ is chosen as the product of independent distributions with f_j, ξ_j and σ each following a lognormal distribution with median equal to the nominal values $(2j-1)$ Hz, 0.05, and 0.1m/s^2, respectively, and with the corresponding c.o.v. of $(20+10j)$%, $(20+10j)$% and 55%, respectively; ρ_j is uniformly distributed over the range of [-3 3]. Let $N_T=1000$ be the number of sampling intervals of the measured time history data. Let $q_j(n;\boldsymbol{\theta})$ denote the roof absolute acceleration at the n-th sampled time instant predicted by the proposed linear modal model and let $y(n)$ denote the corresponding measured output. The combined prediction and measurement errors $\varepsilon(n)=y(n)-q_j(n;\boldsymbol{\theta})$, $n=1,2,\ldots,N_T$, are modeled as independently and identically distributed Gaussian variables with mean zero and some unknown prediction error variance σ^2 (this is the maximum entropy PDF, that is, it has the largest amount of uncertainty among all PDFs of unbounded variables with the same means and variances). Thus, the likelihood function $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}_j)$ is given by:

$$p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j)=\frac{1}{(2\pi\sigma^2)^{N_T/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{n=1}^{N_T}[y(n)-q_j(n;\boldsymbol{\theta})]^2\right) \tag{3.65}$$

where $q_j(n;\boldsymbol{\theta})$ is given by the sum of the base acceleration $a_b(n)$ and modal accelerations at time $t_n = n\Delta t$:

$$q_j(n;\boldsymbol{\theta}) = a_b(n) + \sum_{m=1}^{j} \zeta_m(n;\boldsymbol{\theta}) \tag{3.66}$$

where the m-th mode acceleration $\zeta_m(n;\boldsymbol{\theta})$ satisfies the SDOF (single degree of freedom) linear oscillator equation:

$$\ddot{\zeta}_m(t) + 2\xi_m\omega_m\dot{\zeta}_m(t) + {\omega_m}^2\zeta_m(t) = -\rho_m a_b(t) \tag{3.67}$$

Note that none of the candidate model classes correspond to the one used to generate the data. Our goal is to find the probability of each candidate model class given the dynamic data $\mathcal{D}$. A new variant of HMCM (Cheung and Beck 2007c, 2008a) and Chapter 2 is applied to simulate 2500 samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. The competing candidate models are taken to be equally plausible before getting any data from the system, i.e., $P(\mathcal{M}_j|M)=1/6$. The evidence and the updated model class probability $P(\mathcal{M}_j|\mathcal{D},M)$ are calculated using the proposed method. For convenience, instead of using the transition PDF corresponding to HMCM, which is rather complex (Cheung and Beck 2008a and Chapter 2), the transition PDF corresponding to the Metropolis-Hastings algorithm is adopted to calculate the evidence where the 'proposal' PDF $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ is chosen to be a multivariate Gaussian with mean $\boldsymbol{\theta}$ and a covariance matrix αC where α is some positive scaling factor and C is equal to the sample covariance matrix estimated using the samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. Here, we have $q(\boldsymbol{\theta}^*|\boldsymbol{\theta}) = q(\boldsymbol{\theta}|\boldsymbol{\theta}^*)$. Using Equations (3.11)-(3.14), the value of the posterior PDF at $\boldsymbol{\theta}$ can be estimated using samples $\boldsymbol{\theta}^{(k)}$, k=1,2,…, N, from the posterior $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ and samples $\boldsymbol{\theta}^{*(k)}$, k=1,2,…, N_2, from $q(\boldsymbol{\theta}^*|\boldsymbol{\theta})$ for some chosen $\boldsymbol{\theta}$:

$$p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) \approx \frac{\frac{1}{N}\sum_{k=1}^{N} T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})}{a(\boldsymbol{\theta})} \tag{3.68}$$

where

$$T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) = \min\{1, \frac{p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_j) p(\boldsymbol{\theta} \mid \mathcal{M}_j)}{p(\mathcal{D} \mid \boldsymbol{\theta}^{(k)}, \mathcal{M}_j) p(\boldsymbol{\theta}^{(k)} \mid \mathcal{M}_j)}\} q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)}) \tag{3.69}$$

$$a(\boldsymbol{\theta}) \approx \frac{1}{N_2}\sum_{k=1}^{N_2} r(\boldsymbol{\theta}^{*(k)} \mid \boldsymbol{\theta}) = \frac{1}{N_2}\sum_{k=1}^{N_2} \min\{1, \frac{p(\mathcal{D} \mid \boldsymbol{\theta}^{*(k)}, \mathcal{M}_j) p(\boldsymbol{\theta}^{*(k)} \mid \mathcal{M}_j)}{p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_j) p(\boldsymbol{\theta} \mid \mathcal{M}_j)}\} \tag{3.70}$$

It should be noted that all $T(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ can be calculated easily since the values of $p(\mathcal{D} \mid \boldsymbol{\theta}^{(k)}, \mathcal{M}_j) p(\boldsymbol{\theta}^{(k)} \mid \mathcal{M}_j)$ for all k have already been calculated during the simulation of samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ and $q(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(k)})$ can also be calculated very efficiently. Also, when evaluating ratios such as in (3.69) and (3.70), one should first calculate the logarithm of such ratios and exponentiate at the end in order to avoid numerical overflow.

Denote the numerator of (3.68) by $I(\boldsymbol{\theta})$. Given a particular choice of $\boldsymbol{\theta}$, the variance of the evidence depends only on the variance of the estimate $\hat{p}(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j)$ of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ which further depends on the variance of $I(\boldsymbol{\theta})$ and the variance of the estimate $\hat{a}(\boldsymbol{\theta})$ for $a(\boldsymbol{\theta})$. The scaling factor α in the proposal PDF q can always be chosen to be small enough such that $\hat{a}(\boldsymbol{\theta})$ is closer to 1 and has very small variance; however, the trade-off is that $I(\boldsymbol{\theta})$ will have a larger variance when α is smaller. To decrease the variance of $I(\boldsymbol{\theta})$, one should choose larger α. Thus, one can expect that there exists an optimal choice for α that leads to the smallest variance of the estimate for $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. An iterative process is used to select α so that the c.o.v. of the estimator of the log evidence given in the previous subsection is approximately minimized. A natural starting choice is α=1. During the trial and error process to pick a good α, only a rough estimate for the c.o.v. is needed. Thus, Q=1 and only

small N_2 (e.g, N_2=20) is used and $\boldsymbol{\theta}$ is chosen to be the sample from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ which maximizes $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. It should be noted that with this choice of $\boldsymbol{\theta}$, from (3.69), $T(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(k)}) = q(\boldsymbol{\theta}\,|\,\boldsymbol{\theta}^{(k)})$ is always true. For this example, good choices for α are α=1 for j=1,2,3 and α=2 for j=4,5,6.

The estimate for the log evidence $\ln p(\mathcal{D}|\mathcal{M}_j)$, the posterior mean of the log likelihood function $E[\ln(p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j)]$ (a data-fit measure), the expected information gain EIG (a model class complexity measure given in (3.50)) and the posterior probability $P(\mathcal{M}_j|\mathcal{D},M)$ of the model classes obtained using the proposed methods with N=2500 in Equation (3.68) and N_2=2000 in Equation (3.70) are shown in Table 3.1. Here, equation (3.49) is used with Q=1 and $\boldsymbol{\theta}=\boldsymbol{\theta}_{\mathbf{max}}$, the posterior sample that gives the maximum value of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. The c.o.v. of the evidence estimate is given by the number in the parenthesis next to the log evidence estimate. It can be seen that the c.o.v. is quite small. It can also be seen that a model class consisting of a larger number of modes has a larger posterior mean of the log likelihood function which shows that it gives a better fit to the data on average, as expected. However, it also has a larger expected information gain and thus a model class consisting of a larger number of modes is not necessarily the more plausible one. Bayesian model class selection shows that model class $\mathcal{M}_5$ is the most probable model class based on the data, i.e., the model class consisting of 5 classical modes gives the best balance between the data fit and the information gain from the data based on the identity in (1.4).

Table 3.2 shows the sample posterior means for the natural frequency, damping ratio and roof participation factor for each mode in $\mathcal{M}_5$. The numbers in bold give the values for the exact model. Note that there are no exact counterparts for the classically-damped model's participation factors ρ_i in Table 3.2 since the actual system is non-classically damped. It can be seen that the modal frequencies and damping ratios in $\mathcal{M}_5$, on average, are very close to those corresponding to the exact model, except for the damping ratio of the highest mode which makes a small contribution to the roof response. Also, the sum of the posterior mean

participation factors in $\mathcal{M}_5$ has mean 1.002 (very close to the theoretical value of unity for the sum over all 10 modes for the classically-damped linear dynamic model) with c.o.v. 0.268%.

Figure 3.2 shows the magnitude of the FFT of the roof acceleration data (solid curve) and the mean of magnitude of the FFT of the roof acceleration estimated using posterior samples from $\mathcal{M}_5$ (dashed curve). It can be seen that $\mathcal{M}_5$ with the first 5 modes up to about 6 Hz gives a very good match of the magnitude of the FFT over a dynamic range of 40 db.

Table 3.1 Results obtained for Example 1 using the proposed method with θ_{max} and Q=1 in Equation (3.49)

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$	$\mathcal{M}_5$	$\mathcal{M}_6$
$E[\ln(p(\mathcal{D}\mid\boldsymbol{\theta},\mathcal{M}_j))]$	35.22	507.19	809.52	1337.32	1674.46	1707.68
EIG	39.67	41.61	59.70	74.86	98.47	147.58
$\ln p(\mathcal{D}\mid\mathcal{M}_j)$	-4.45 (6.8%)	465.58 (8.4%)	749.82 (13.1%)	1262.46 (18.3%)	1575.99 (18.7%)	1560.10 (17.5%)
$P(\mathcal{M}_j\mid\mathcal{D},M)$	0	0	0	0	0.9999997	$3X10^{-7}$

Table 3.2 Posterior means for the natural frequencies, modal damping ratios and roof participation factors for the most probable model class $\mathcal{M}_5$ in Example 1 (exact values in bold)

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5
f_i (Hz)	0.74 **0.74**	2.15 **2.16**	3.55 **3.56**	4.85 **4.89**	5.93 **6.05**
ξ_i	0.92% **0.92%**	2.72% **2.71%**	4.30% **4.45%**	5.63% **6.03%**	4.84% **7.65%**
ρ_i	1.273	-0.415	0.226	-0.139	0.057

3.3.2 Example 2: Nonlinear response of four-story building

In this example, the nonlinear seismic response of a four-story building is considered. The simulated noisy accelerometer data $\mathcal{D}$ consist of 10s (with a sample interval Δt of 0.01s) of the total acceleration at the base and at all the floors (Figure 3.3). The simulated Gaussian white noise has a noise-to-signal ratio of 10% rms of the roof acceleration. The data $\mathcal{D}$ are

generated from a shear building model with Rayleigh damping and hysteretic bilinear interstory restoring forces. Here we consider a set $M=\{\mathcal{M}_j: j=1,2,3\}$ consisting of 3 candidate model classes which involve an inelastic shear building model as follows:

Model class $\mathcal{M}_1$: A inelastic shear building model with viscous damping and bilinear hysteretic bilinear restoring force model (Figure 3.4). The lumped masses m_i, i=1, 2, 3, 4, on each floor are assumed fixed at 2×10^4kg for all floors. The vector $\boldsymbol{\theta}$ to be updated by the dynamic data $\mathcal{D}$ consists of D=17 parameters with the first component θ_1 equal to the prediction error variance σ^2 and for s=2,…,D, $\theta_s = \log(\varphi_{s-1}/l_{s-1})$ where φ_{s-1}'s are comprised of the following 16 structural parameters: for i=1,2,3,4, the initial stiffness k_i, post-yield stiffness reduction factor r_i, yield displacement u_i and the damping coefficient c_i of the viscous damper of the i-th floor and the l_{s-1}'s are the corresponding nominal values given later.

Let $q_i(n;\theta_2,...,\theta_D)$ denote the output at time $t_n= n\Delta t$ (Δt=0.01s) at the i-th observed degree of freedom predicted by the proposed structural model and $y_i(n)$ denote the corresponding measured output. The combined prediction and measurement errors $\varepsilon_i(n) = y_i(n) - q_i(n;\boldsymbol{\theta})$ for n=1,…, N_T =1000 and i=1,…,N_o = 4 are modeled as independently and identically distributed Gaussian variables with mean zero and some unknown prediction-error variance σ^2. Thus the likelihood function $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}_1)$ is given by:

$$p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_1) = \frac{1}{(2\pi\sigma^2)^{N_o N_T/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N_o}\sum_{n=1}^{N_T}[y_i(n) - q_i(n;\theta_2,...,\theta_D)]^2\right) \qquad (3.71)$$

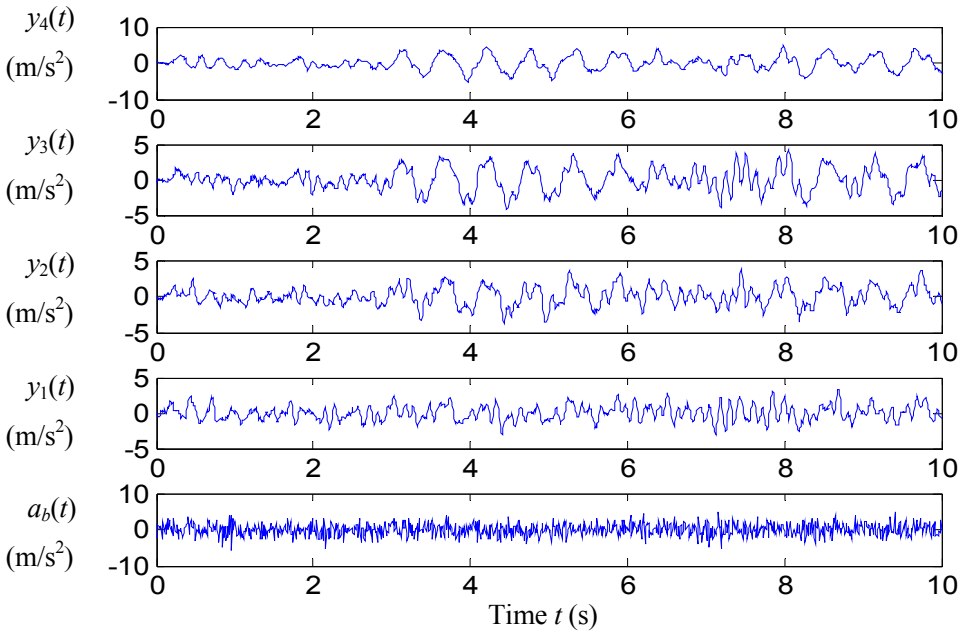

Figure 3.3: Floor accelerations and base acceleration from a nonlinear four-story building response (y_i(t): total acceleration at the i-th floor; $a_b(t)$: total acceleration at the base)

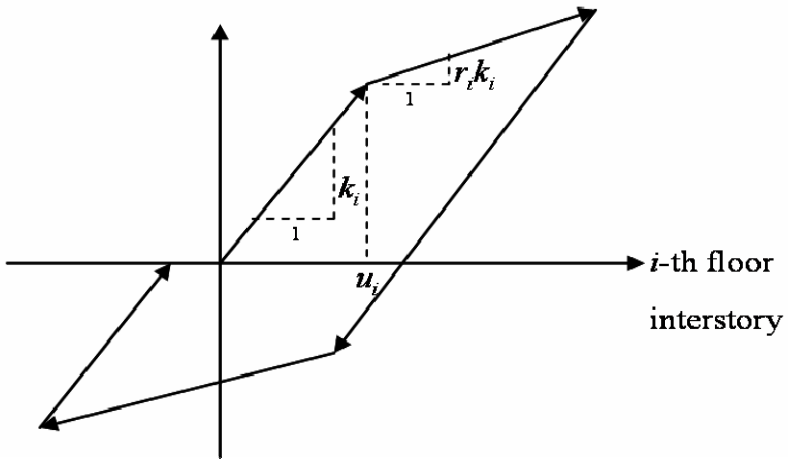

Figure 3.4: The hysteretic restoring force model

For $\mathcal{M}_1$, the prior PDF for $\boldsymbol{\theta}$ is chosen as the product of independent distributions: the structural parameters φ_{s-1} including k_i, r_i, u_i, c_i; follow a lognormal distribution with median equal to the corresponding nominal values l_{s-1} and the corresponding log standard deviations equal to 0.6 and thus the θ_s, for s=2,…,D, follow a Gaussian distribution with zero mean and standard deviation of 0.6; $\theta_1=\sigma^2$ follows an inverse gamma distribution with mean μ equal to its nominal value and c.o.v. δ =1.0, i.e., $p(\sigma^2) \propto (\sigma^2)^{-\alpha-1}\exp(-\beta/\sigma^2)$ where $\alpha=\delta^{-2}+2, \beta=\mu(\alpha-1)$. The nominal values for the structural parameters k_1, k_2, k_3, k_4 are 2.2, 2.0, 1.7, 1.45 (10^7Nm^{-1}) respectively; the nominal values for r_i are 0.1 for all i; the nominal values for u_i are 8mm for i=1,2 and 7mm for i=3,4; the nominal values for c_1, c_2, c_3, c_4 are 6.93, 6.45, 5.73, 5.13 (10^4Nm^{-1}s) respectively. The nominal modal damping ratios are 2% and 5% for the first and second modes respectively. The nominal value for σ^2 is the square of 10% of the maximum of the r.m.s of the total accelerations measured at each of the 4 floors. $q_i(n;\boldsymbol{\theta})$ is the i-th component at time t_n of $\mathbf{q}(t_n)$ which satisfies the following equation of motion:

$$\mathbf{M_s}\ddot{\mathbf{q}}(t)+\mathbf{C_s}\dot{\mathbf{q}}(t)+\mathbf{F}(\mathbf{Q}(t),\dot{\mathbf{Q}}(t))=-\mathbf{M}_s\begin{bmatrix}1\\ \vdots \\ 1\end{bmatrix}a_g(t) \tag{3.72}$$

where the mass matrix $\mathbf{M_s}$, is a diagonal matrix diag(m_1, m_2, m_3, m_4); and the damping matrix $\mathbf{C_s}$ is given as follows:

$$\mathbf{C_s}=\begin{bmatrix} c_1+c_2 & -c_2 & 0 & 0 \\ -c_2 & c_2+c_3 & -c_3 & 0 \\ 0 & -c_3 & c_3+c_4 & -c_4 \\ 0 & 0 & -c_4 & c_4 \end{bmatrix} \tag{3.73}$$

The hysteretic restoring force $\mathbf{F}(\mathbf{Q}(t),\dot{\mathbf{Q}}(t))$, which depends on the whole time history $[\mathbf{Q}(t),\dot{\mathbf{Q}}(t)]$ of responses from time=0 up to time τ, i.e., $\mathbf{q}(\tau)$ and $\dot{\mathbf{q}}(\tau)$ for all $\tau \in [0,t]$, is modeled by the bilinear hysteretic model mentioned above.

Model class $\mathcal{M}_2$: Same as $\mathcal{M}_1$ except that the damping matrix is replaced by a Rayleigh damping matrix $\mathbf{C_s}$, i.e. $\mathbf{C_s}=\rho\mathbf{M_s}+\gamma\mathbf{K_s}$ where $\mathbf{M_s}$ and $\mathbf{K_s}$ are the mass and stiffness matrix of the shear building model in $\mathcal{M}_1$, respectively, and ρ, γ are some uncertain positive scalars (such that a higher mode has the same or larger modal damping ratio than a lower mode).

$$\mathbf{K_s}=\begin{bmatrix} k_1+k_2 & -k_2 & 0 & 0 \\ -k_2 & k_2+k_3 & -k_3 & 0 \\ 0 & -k_3 & k_3+k_4 & -k_4 \\ 0 & 0 & -k_4 & k_4 \end{bmatrix} \tag{3.74}$$

This model class contains the system used to generate the simulated noisy data $\mathcal{D}$. For this case, the uncertain parameter vector $\boldsymbol{\theta}$ to be updated by the dynamic data $\mathcal{D}$ consists of D=15 parameters. The prior PDF for $\boldsymbol{\theta}$ is modeled with the same independent distributions as $\mathcal{M}_1$ except the prior for the c_i's is replaced by one for ρ, γ which are independent lognormal distributions with medians equal to the corresponding nominal values and the corresponding log standard deviations equal to 0.6. The nominal values for ρ, γ are 0.7959 and 2.50×10^{-3} so that the corresponding nominal modal damping ratios for the first 2 modes are 5%.

Model class $\mathcal{M}_3$: Same as $\mathcal{M}_2$ except that the hysteretic force model is an elastic-perfectly plastic model, i.e., r_i =0, i=1,2,3,4. The number of uncertain parameters to be updated by the dynamic data is 11.

The three competing candidate models are taken as equally plausible a priori, i.e., $P(\mathcal{M}_j|M)$= 1/3, j=1,2,3. At the end of simulation, the N posterior samples for the structural parameters $\varphi_{s\text{-}1}^{(k)}$, s=2,…,D, k=1,2,…,N, can be obtained by $\varphi_{s\text{-}1}^{(k)}=l_{s\text{-}1}\exp(\theta_s^{(k)})$ where $\boldsymbol{\theta}^{(k)}=[\theta_1^{(k)}\ \theta_2^{(k)}...\theta_D^{(k)}]^{\mathrm{T}}$, k=1,2,…,N, are samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. All the structural parameters $\varphi_{s\text{-}1}$ are constrained to be positive. This is the reason for the transformation between θ_s's and $\varphi_{s\text{-}1}$'s. If samples for $\varphi_{s\text{-}1}$'s are directly generated by MCMC methods such as the Metropolis-Hastings algorithm, or advanced MCMC methods such as those

presented in Beck and Au (2002), Ching and Chen (2007), Cheung and Beck (2008a) and Chapter 2, then they are not constrained to be positive. Therefore, performing the simulation in the $[\sigma^2\ \varphi_1 \ldots \varphi_{D-1}]^T$ space for the posterior samples can result in increased rejection of samples and thus increased computational effort. Performing the simulation in the transformed $\boldsymbol{\theta}$ space, as done here, guarantees samples for the φ_{s-1}'s are always positive. The way the samples for σ^2 are simulated also guarantees that they are always positive.

Here, a hybrid approach making use of the TMCMC multi-level method (Ching and Chen 2007) and Gibbs sampling is adopted to generate 7000 samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. For TMCMC, 1000 samples are generated for each of the intermediate levels but 7000 samples are generated in the last level corresponding to the posterior (the first 2000 samples are discarded to allow for "burn-in" to the stationary state). During the l-th tempering level with tempering parameter τ_l and the target PDF proportional to $p^{\tau_l}(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j)\,p(\boldsymbol{\theta}\,|\,\mathcal{M}_j)$, a new sample $\boldsymbol{\theta}'$ is generated as follows. First, a sample $\boldsymbol{\theta}^{(k)}$ is picked using re-sampling as in the TMCMC method among those samples that have been generated. Second, we perform Gibbs sampling by fixing the value of θ_1 at $\theta_1^{(k)}$, the first component of $\boldsymbol{\theta}^{(k)}$ (the prediction error variance), while the remaining $D-1$ components of $\boldsymbol{\theta}'$ are generated using the Metropolis-Hastings algorithm applied to the PDF of these components conditional on $\theta_1=\theta_1^{(k)}$, as in the TMCMC method, and finally the first component θ_1' of $\boldsymbol{\theta}'$ is generated from its PDF conditional on the previously-generated $D-1$ components, which is an inverse gamma distribution, proportional to $(\theta_1')^{-\alpha'-1}\exp(-\beta'/\theta_1')$ where $\alpha'=\alpha+\tau_l N_o N_T/2$ and β' is given by:

$$\beta' = \beta + \frac{\tau_l}{2}\sum_{i=1}^{N_o=4}\sum_{n=1}^{N_T=1000}[y_i(n) - q_i(n;\theta_2,\ldots,\theta_D)]^2 \tag{3.75}$$

The HMCM method in Cheung and Beck (2008a) and Chapter 2 is applied in the last level of the TMCMC method in place of the Metropolis Hastings algorithm for more effective sampling of the posterior PDF. It should be noted that to obtain a more accurate estimate

for the evidence using the proposed method, one just needs to simulate more samples from the posterior PDF in the last level of TMCMC while if the TMCMC method is used exclusively to calculate the evidence, one will need to perform an additional simulation run with increased samples at all levels of the TMCMC method.

For convenience and illustration of the proposed method, in this example, instead of using the transition PDF corresponding to the one that we use to generate the posterior samples, the transition PDF of the type presented in (3.19)-(3.22) is used to approximate the evidence where $\boldsymbol{\theta}$ is divided into 2 groups: $\boldsymbol{\theta}_1=\theta_1$ and $\boldsymbol{\theta}_2=[\theta_2,\ldots,\theta_D]$. Thus, the estimate for $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ with $\boldsymbol{\theta}$ chosen to be the sample from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}^{(j)})$ that gives the maximum value of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$, i.e., the one which gives the maximum value of $p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}_j)p(\boldsymbol{\theta}\,|\,\mathcal{M}_j)$, is given by (3.21)-(3.28):

$$p(\boldsymbol{\theta}\,|\,\mathcal{D},\mathcal{M}_j)=\pi(\boldsymbol{\theta})=\frac{I_1}{a_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)} \tag{3.76}$$

where the numerator and denominator can be estimated by:

$$I_1=\int \pi(\boldsymbol{\theta}_1\,|\,\tilde{\boldsymbol{\theta}}_2)T(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1,\tilde{\boldsymbol{\theta}}_2)\pi(\tilde{\boldsymbol{\theta}}_2)d\tilde{\boldsymbol{\theta}}_2 \approx \frac{1}{N_1}\sum_{i=1}^{N_1}\pi(\boldsymbol{\theta}_1\,|\,\boldsymbol{\theta}_2{}^{(i)})T(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2{}^{(i)}) \tag{3.77}$$

$$a_2(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \approx \frac{1}{N_2}\sum_{i=1}^{N_2} r(\boldsymbol{\theta}_2^{**(i)}\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2) \tag{3.78}$$

$$T_2(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(i)})=r(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(i)})q(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(i)}) \tag{3.79}$$

$$r(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(i)})=\min\{1,\frac{p(\mathcal{D}\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2,\mathcal{M}_j)p(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2\,|\,\mathcal{M}_j)q(\boldsymbol{\theta}_2^{(i)}\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)}{p(\mathcal{D}\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(i)},\mathcal{M}_j)p(\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(i)}\,|\,\mathcal{M}_j)q(\boldsymbol{\theta}_2\,|\,\boldsymbol{\theta}_1,\boldsymbol{\theta}_2^{(i)})}\} \tag{3.80}$$

$$r(\boldsymbol{\theta}_2^{**(i)} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \min\{1, \frac{p(\mathcal{D} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{**(i)}, \mathcal{M}_j)\, p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{**(i)} \mid \mathcal{M}_j)\, q(\boldsymbol{\theta}_2 \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^{**(i)})}{p(\mathcal{D} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \mathcal{M}_j)\, p(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \mid \mathcal{M}_j)\, q(\boldsymbol{\theta}_2^{**(i)} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)}\} \quad (3.81)$$

where $\boldsymbol{\theta}_2^{(i)}$ are marginal samples of $\boldsymbol{\theta}_2$ obtained from posterior samples $(\boldsymbol{\theta}_1^{(i)}, \boldsymbol{\theta}_2^{(i)})$ corresponding to $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, and $\boldsymbol{\theta}_2^{**(i)}$ are samples from $q(\boldsymbol{\theta}_2^{**} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ for fixed $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$. Also, $\pi(\boldsymbol{\theta}_1 \mid \boldsymbol{\theta}_2)$ is the value of the inverse gamma PDF $\propto (\theta_1)^{-\alpha''-1}\exp(-\beta''/\theta_1)$ evaluated at θ_1 where $\alpha''=\alpha+N_o N_T/2$ and β'' is given by:

$$\beta'' = \beta + \frac{1}{2}\sum_{i=1}^{N_o}\sum_{n=1}^{N_T}[y_i(n) - q_i(n;\theta_2,...,\theta_D)]^2 \quad (3.82)$$

and the artificial 'proposal' PDF $q(\boldsymbol{\theta}_2^{**}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is chosen to be a global/independent proposal PDF given by a weighted sum of PDFs as follows:

$$q(\boldsymbol{\theta}_2^{**}|\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = q(\boldsymbol{\theta}_2^{**}) = \frac{1}{N_1}\sum_{s=1}^{N_1} \mathrm{N}(\boldsymbol{\theta}_2^{**}; \boldsymbol{\theta}_2^{(s)}, \Sigma) \quad (3.83)$$

where $\mathrm{N}(\boldsymbol{\theta}_2^{**}; \boldsymbol{\theta}_2^{(s)}, \Sigma)$ is a multivariate Gaussian PDF with mean $\boldsymbol{\theta}_2^{(s)}$ and covariance matrix Σ; $\boldsymbol{\theta}_2^{(s)}$, s =1,2,…,$N_1$= 5000, are the marginal samples of $\boldsymbol{\theta}_2^*$ obtained from posterior samples $(\boldsymbol{\theta}_1^{(s)}, \boldsymbol{\theta}_2^{(s)})$ corresponding to $p(\boldsymbol{\theta}^*|\mathcal{D},\mathcal{M}_j)$ where $\boldsymbol{\theta}^* = (\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*)$, and Σ is equal to some positive number κ times the sample covariance matrix of the samples $\boldsymbol{\theta}_2^{(s)}$, s =1,2,…,N_1= 5000. For all of the three model classes, κ is 0.2^2 (this is a reasonable choice which can be obtained readily when we simulate the posterior samples using TMCMC). It should be noted that all required quantities for I_1 in (3.77) can be calculated easily since the values of $p(\mathcal{D} \mid \boldsymbol{\theta}^{(s)}, \mathcal{M}_j)$ and $p(\boldsymbol{\theta}^{(s)} \mid \mathcal{M}_j)$ for all s have already been calculated during the simulation of samples from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. To calculate the quantity in (3.28), samples $\boldsymbol{\theta}_2^{**(s)}$ are

generated from $q(\boldsymbol{\theta}_2^{**} \mid \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ as follows: for s=1,2,…,5000, generate $\boldsymbol{\theta}_2^{**(s)}$ from $\mathrm{N}(\boldsymbol{\theta}_2^{**}; \boldsymbol{\theta}_2^{(s)}, \Sigma)$.

The estimate for the log evidence $\ln p(\mathcal{D}|\mathcal{M}_j)$, the posterior mean of the log likelihood function $E[\ln(p(\mathcal{D} \mid \boldsymbol{\theta}, \mathcal{M}_j)]$(the data-fit measure), the expected information gain EIG (the complexity measure) and the posterior probability $P(\mathcal{M}_j|\mathcal{D},M)$ of the model classes obtained using the proposed method are shown in Table 3.3 (rows 1-3 and 5) based on equation (3.49) with Q=1 and $\boldsymbol{\theta}=\boldsymbol{\theta}_{\mathbf{max}}$, the sample from $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ which gives the maximum value of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$. The c.o.v. of the evidence estimate is given by the number in the parenthesis next to the log evidence estimate. It can be seen that the c.o.v. is very small. It is interesting to note that the expected information gain from the data by model class $\mathcal{M}_2$ is less than that by the model class $\mathcal{M}_1$ with more parameters and more than that by the model class $\mathcal{M}_3$ with fewer parameters, as might be expected. Model class $\mathcal{M}_2$ has the largest posterior mean of the log likelihood function which shows that it gives the best fit to the data. Bayesian model class selection shows that model class $\mathcal{M}_2$ also gives the best balance between the data fit and the information gain from the data and is thus the most probable model class, consistent with the fact that it is the model class containing the system from which the noisy dynamic data $\mathcal{D}$ is generated. The estimate for the log evidence $\ln p(\mathcal{D}|\mathcal{M}_j)$ obtained using the TMCMC method is given in row 4 of Table 3.3 for comparison. Based on the results from one simulation run, it can be seen that $\ln p(\mathcal{D}|\mathcal{M}_j)$ obtained by the TMCMC method is different from that obtained by the proposed method (the more accurate one) by 53.7%, 0.7% and 55.5% respectively.

Table 3.3 Results obtained for Example 2 using the proposed method with θ_{max} and Q=1 in Equation (3.49)

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$
$E[\ln(p(\mathcal{D}\mid\boldsymbol{\theta},\mathcal{M}_j)]$	249.2	682.1	368.1
EIG	122.5	77.6	65.0
$\ln p(\mathcal{D}\mid\mathcal{M}_j)$ (by the proposed method)	126.7(4.7%)	604.5(8.6%)	303.1(8.8%)
$\ln p(\mathcal{D}\mid\mathcal{M}_j)$ (by TMCMC)	194.8	608.7	195.6
$P(\mathcal{M}_j\mid\mathcal{D},M)$	0.0	1.0	0.0

Appendix 3A

Here we show that the transition PDF in (3.19) and (3.20) satisfies (3.9). If the Markov chain is in a state with PDF $\pi(\boldsymbol{\theta})$, the PDF for the state at the next step is given by:

$$p(\boldsymbol{\theta}^*)=\int K(\boldsymbol{\theta}^*\mid\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}=\int\prod_{j=1}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$=\int\prod_{j=1}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})\pi(\boldsymbol{\theta}_1\mid\{\boldsymbol{\theta}\}_{2:G})\pi(\{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_1d\boldsymbol{\theta}_{2:G}$$

$$=\int[\prod_{j=2}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})]K_1(\boldsymbol{\theta}_1^*\mid\boldsymbol{\theta}_1,\{\boldsymbol{\theta}\}_{2:G})\pi(\boldsymbol{\theta}_1\mid\{\boldsymbol{\theta}\}_{2:G})\pi(\{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_1d\boldsymbol{\theta}_{2:G}$$

$$=\int[\prod_{j=2}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})]\pi(\{\boldsymbol{\theta}\}_{2:G})\{\underbrace{\int K_1(\boldsymbol{\theta}_1^*\mid\boldsymbol{\theta}_1,\{\boldsymbol{\theta}\}_{2:G})\pi(\boldsymbol{\theta}_1\mid\{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_1}_{=\pi(\boldsymbol{\theta}_1^*\mid\{\boldsymbol{\theta}\}_{2:G})\text{ (by (3.20))}}\}d\boldsymbol{\theta}_{2:G}$$

$$=\int[\prod_{j=2}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})]\pi(\{\boldsymbol{\theta}\}_{2:G})\pi(\boldsymbol{\theta}_1^*\mid\{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_{2:G}$$

$$=\int[\prod_{j=2}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})]\pi(\boldsymbol{\theta}_1^*,\{\boldsymbol{\theta}\}_{2:G})d\boldsymbol{\theta}_{2:G}\quad\text{.....(A3.1)}$$

$$=\int[\prod_{j=2}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})]\pi(\boldsymbol{\theta}_2\mid\boldsymbol{\theta}_1^*,\{\boldsymbol{\theta}\}_{3:G})\pi(\boldsymbol{\theta}_1^*,\{\boldsymbol{\theta}\}_{3:G})d\boldsymbol{\theta}_2d\boldsymbol{\theta}_{3:G}$$

$$=\int[\prod_{j=3}^{G}K_j(\boldsymbol{\theta}_j^*\mid\{\boldsymbol{\theta}^*\}_{1:j-1},\boldsymbol{\theta}_j,\{\boldsymbol{\theta}\}_{j+1:G})]\pi(\boldsymbol{\theta}_1^*,\{\boldsymbol{\theta}\}_{3:G})\{\underbrace{\int K_2(\boldsymbol{\theta}_2^*\mid\boldsymbol{\theta}^*{}_1,\boldsymbol{\theta}_2,\{\boldsymbol{\theta}\}_{3:G})\pi(\boldsymbol{\theta}_2\mid\boldsymbol{\theta}_1^*,\{\boldsymbol{\theta}\}_{3:G})d\boldsymbol{\theta}_2}_{=\pi(\boldsymbol{\theta}_2^*\mid\boldsymbol{\theta}_1^*,\{\boldsymbol{\theta}\}_{3:G})\text{ (by (3.20))}}\}d\boldsymbol{\theta}_{3:G}$$

$$= \int [\prod_{j=3}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})]\pi(\boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G})\pi(\boldsymbol{\theta}_2^* | \boldsymbol{\theta}_1^*, \{\boldsymbol{\theta}\}_{3:G}) d\boldsymbol{\theta}_{3:G}$$

$$= \int [\prod_{j=3}^{G} K_j(\boldsymbol{\theta}_j^* \mid \{\boldsymbol{\theta}^*\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}\}_{j+1:G})]\pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, \{\boldsymbol{\theta}\}_{3:G}) d\boldsymbol{\theta}_{3:G} \text{.....(A3.2)}$$

Repeat steps from (A3.1) to (A3.2) to reduce the integration dimension by 1 group each time.

$$\begin{aligned}
p(\boldsymbol{\theta}^*) &= \int K_G(\boldsymbol{\theta}_G^* \mid \{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*, ..., \boldsymbol{\theta}_{G-1}^*, \boldsymbol{\theta}_G) d\boldsymbol{\theta}_G \\
&= \int K_G(\boldsymbol{\theta}_G^* \mid \{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_G | \{\boldsymbol{\theta}^*\}_{1:G-1})\pi\{\boldsymbol{\theta}^*\}_{1:G-1} d\boldsymbol{\theta}_G \\
&= \pi\{\boldsymbol{\theta}^*\}_{1:G-1} \underbrace{\int K_G(\boldsymbol{\theta}_G^* \mid \{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\boldsymbol{\theta}_G | \{\boldsymbol{\theta}^*\}_{1:G-1}) d\boldsymbol{\theta}_G}_{\pi(\boldsymbol{\theta}_G^* | \{\boldsymbol{\theta}^*\}_{1:G-1}) \text{ (by 1.1)}} \\
&= \pi\{\boldsymbol{\theta}^*\}_{1:G-1}\pi(\boldsymbol{\theta}_G^* | \{\boldsymbol{\theta}^*\}_{1:G-1}) \\
&= \pi(\{\boldsymbol{\theta}^*\}_{1:G-1}, \boldsymbol{\theta}_G^*) \\
&= \pi(\boldsymbol{\theta}^*)
\end{aligned}$$

Reversing the roles of $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, one sees that the transition PDF in (3.19) and (3.20) satisfies (3.9).

Appendix 3B

$$\pi(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}) = \int K_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}) d\tilde{\boldsymbol{\theta}}_G$$

$$= \int T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}) d\tilde{\boldsymbol{\theta}}_G + (1 - a_G(\{\boldsymbol{\theta}\}_{1:G}))\pi(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1})$$

$$\Rightarrow a_G(\{\boldsymbol{\theta}\}_{1:G})\pi(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}) = \int T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}) d\tilde{\boldsymbol{\theta}}_G$$

Thus,

$$a_G(\{\boldsymbol{\theta}\}_{1:G})\pi(\{\boldsymbol{\theta}\}_{1:G}) = \int T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)d\tilde{\boldsymbol{\theta}}_G$$

$$\Rightarrow \pi(\boldsymbol{\theta}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1} \mid \boldsymbol{\theta}_G)a_G(\{\boldsymbol{\theta}\}_{1:G}) = \int T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)d\tilde{\boldsymbol{\theta}}_G$$

$$\Rightarrow \pi(\boldsymbol{\theta}_G)\int T_G(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1} \mid \boldsymbol{\theta}_G)d\tilde{\boldsymbol{\theta}}_G = \int T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)d\tilde{\boldsymbol{\theta}}_G$$

$$\Rightarrow \pi(\boldsymbol{\theta}_G)\int T_G(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1} \mid \boldsymbol{\theta}_G)d\tilde{\boldsymbol{\theta}}_G d\{\boldsymbol{\theta}\}_{1:G-1}$$

$$= \int T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)d\tilde{\boldsymbol{\theta}}_G d\{\boldsymbol{\theta}\}_{1:G-1}$$

$$\Rightarrow \pi(\boldsymbol{\theta}_G) = \frac{\int T_G(\boldsymbol{\theta}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1}, \tilde{\boldsymbol{\theta}}_G)d\tilde{\boldsymbol{\theta}}_G d\{\boldsymbol{\theta}\}_{1:G-1}}{\int T_G(\tilde{\boldsymbol{\theta}}_G \mid \{\boldsymbol{\theta}\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\{\boldsymbol{\theta}\}_{1:G-1} \mid \boldsymbol{\theta}_G)d\tilde{\boldsymbol{\theta}}_G d\{\boldsymbol{\theta}\}_{1:G-1}}$$

$$= \frac{\int T_G(\boldsymbol{\theta}_G \mid \{\hat{\boldsymbol{\theta}}\}_{1:G-1}, \hat{\boldsymbol{\theta}}_G)\pi(\{\hat{\boldsymbol{\theta}}\}_{1:G-1}, \hat{\boldsymbol{\theta}}_G)d\hat{\boldsymbol{\theta}}_G d\{\hat{\boldsymbol{\theta}}\}_{1:G-1}}{\int T_G(\tilde{\boldsymbol{\theta}}_G \mid \{\tilde{\boldsymbol{\theta}}\}_{1:G-1}, \boldsymbol{\theta}_G)\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G-1} \mid \boldsymbol{\theta}_G)d\tilde{\boldsymbol{\theta}}_G d\{\tilde{\boldsymbol{\theta}}\}_{1:G-1}}$$

Appendix 3C

Here we prove that $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$,…, $\boldsymbol{\theta}_J$ conditioned on the other components (i.e., $\boldsymbol{\theta}_{J+1}$, $\boldsymbol{\theta}_{J+2}$,…,$\boldsymbol{\theta}_G$) satisfy the stationarity condition in (3.41).

$$\int \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_{1:J}$$

$$= \int \prod_{j=2}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\boldsymbol{\theta}_1 \mid \{\tilde{\boldsymbol{\theta}}\}_{2:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\tilde{\boldsymbol{\theta}}_1 d\{\tilde{\boldsymbol{\theta}}\}_{2:J}$$

$$= \int \prod_{j=2}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \frac{\pi(\boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{2:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})}{\pi(\{\tilde{\boldsymbol{\theta}}\}_{2:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} \frac{\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G})}{\pi(\{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} d\tilde{\boldsymbol{\theta}}_1 d\{\tilde{\boldsymbol{\theta}}\}_{2:J}$$

$$= \int \prod_{j=2}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \frac{\pi(\boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{2:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})}{\pi(\{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} \frac{\pi(\{\tilde{\boldsymbol{\theta}}\}_{1:G})}{\pi(\{\tilde{\boldsymbol{\theta}}\}_{2:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} d\tilde{\boldsymbol{\theta}}_1 d\{\tilde{\boldsymbol{\theta}}\}_{2:J}$$

$$= \int \prod_{j=2}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{2:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\tilde{\boldsymbol{\theta}}_1 \mid \{\tilde{\boldsymbol{\theta}}\}_{2:G}) d\tilde{\boldsymbol{\theta}}_1 d\{\tilde{\boldsymbol{\theta}}\}_{2:J}$$

$$= \int \prod_{j=2}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{2:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \int \pi(\tilde{\boldsymbol{\theta}}_1 \mid \{\tilde{\boldsymbol{\theta}}\}_{2:G}) d\tilde{\boldsymbol{\theta}}_1 d\{\tilde{\boldsymbol{\theta}}\}_{2:J}$$

$$= \int \prod_{j=2}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{2:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_{2:J}$$

$$= \int \prod_{j=3}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\boldsymbol{\theta}_2 \mid \{\boldsymbol{\theta}\}_1, \{\tilde{\boldsymbol{\theta}}\}_{3:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{2:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\tilde{\boldsymbol{\theta}}_2 d\{\tilde{\boldsymbol{\theta}}\}_{3:J}$$

$$= \int \prod_{j=3}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \frac{\pi(\{\boldsymbol{\theta}\}_{1:2}, \{\tilde{\boldsymbol{\theta}}\}_{3:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})}{\pi(\{\boldsymbol{\theta}\}_1, \{\tilde{\boldsymbol{\theta}}\}_{3:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} \frac{\pi(\boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{2:G})}{\pi(\{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} d\tilde{\boldsymbol{\theta}}_2 d\{\tilde{\boldsymbol{\theta}}\}_{3:J}$$

$$= \int \prod_{j=3}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:2}, \{\tilde{\boldsymbol{\theta}}\}_{3:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\tilde{\boldsymbol{\theta}}_2 \mid \boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{3:G}) d\tilde{\boldsymbol{\theta}}_2 d\{\tilde{\boldsymbol{\theta}}\}_{3:J}$$

$$= \int \prod_{j=3}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:2}, \{\tilde{\boldsymbol{\theta}}\}_{3:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) [\int \pi(\tilde{\boldsymbol{\theta}}_2 \mid \boldsymbol{\theta}_1, \{\tilde{\boldsymbol{\theta}}\}_{3:G}) d\tilde{\boldsymbol{\theta}}_2] d\{\tilde{\boldsymbol{\theta}}\}_{3:J}$$

$$= \int \prod_{j=3}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:2}, \{\tilde{\boldsymbol{\theta}}\}_{3:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_{3:J}$$

Observe the patterns and keep repeating each time reducing the dimension of integration by 1 group:

$$\int \prod_{j=1}^{J} \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\tilde{\boldsymbol{\theta}}\}_{j+1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\tilde{\boldsymbol{\theta}}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_{1:J}$$

$$= \int \pi(\boldsymbol{\theta}_J \mid \{\boldsymbol{\theta}\}_{1:J-1}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:J-1}, \{\tilde{\boldsymbol{\theta}}\}_J \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_J$$

$$= \int \frac{\pi(\{\boldsymbol{\theta}\}_{1:J}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})}{\pi(\{\boldsymbol{\theta}\}_{1:J-1}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} \frac{\pi(\{\boldsymbol{\theta}\}_{1:J-1}, \{\tilde{\boldsymbol{\theta}}\}_{J:G})}{\pi(\{\tilde{\boldsymbol{\theta}}\}_{J+1:G})} d\{\tilde{\boldsymbol{\theta}}\}_J$$

$$= \int \pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \pi(\tilde{\boldsymbol{\theta}}_J \mid \{\boldsymbol{\theta}\}_{1:J-1}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_J$$

$$= \pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) \underbrace{\int \pi(\tilde{\boldsymbol{\theta}}_J \mid \{\boldsymbol{\theta}\}_{1:J-1}, \{\tilde{\boldsymbol{\theta}}\}_{J+1:G}) d\{\tilde{\boldsymbol{\theta}}\}_J}_{=1}$$

$$= \pi(\{\boldsymbol{\theta}\}_{1:J} \mid \{\tilde{\boldsymbol{\theta}}\}_{J+1:G})$$

Appendix 3D

$$\pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G}) = \int K_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j$$

$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j$$

$$+ (1 - a_j(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G})) \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G})$$

$$\Rightarrow a_j(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G})$$

$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j$$

Thus,

$$a_j(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G})$$

$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j$$

$$\Rightarrow \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) a_j(\{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G})$$

$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j$$

$$\Rightarrow \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) \int T_j(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j$$

$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j$$

$$\Rightarrow \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) \int T_j(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j d\{\boldsymbol{\theta}\}_{1:j-1}$$

$$= \int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j d\{\boldsymbol{\theta}\}_{1:j-1}$$

$$\Rightarrow \pi(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) = \frac{\int T_j(\boldsymbol{\theta}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1}, \tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j d\{\boldsymbol{\theta}\}_{1:j-1}}{\int T_j(\tilde{\boldsymbol{\theta}}_j \mid \{\boldsymbol{\theta}\}_{1:j-1}, \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) \pi(\{\boldsymbol{\theta}\}_{1:j-1} \mid \boldsymbol{\theta}_j, \{\boldsymbol{\theta}^*\}_{j+1:G}) d\tilde{\boldsymbol{\theta}}_j d\{\boldsymbol{\theta}\}_{1:j-1}}$$

CHAPTER 4

Comparison of different model classes for Bayesian updating and robust predictions using stochastic state-space system models

Past applications of model updating of dynamic systems focus on model classes which consider an uncertain prediction error as the difference between the real system output and the model output and model it probabilistically using Jaynes' Principle of Maximum Information Entropy. In this chapter, in addition to these model classes, we also consider an extension of such model classes to allow more flexibility in treating modeling uncertainties when updating state space models and making robust predictions; this is done by introducing prediction errors in the state vector equation in addition to those in the system output vector equation. The extended model classes allow for interactions between the model parameters and the prediction errors in both the state vector equation and the system output equation to give more robust predictions at unobserved DOFs. In this chapter, we investigate the difference of these model classes and their effect on the robust predictions. Tools developed in the previous chapters are used here to solve the computational problems. Here, only the methodology corresponding to linear dynamic systems with input measurements is presented. The material in this chapter is also presented in Cheung and Beck (2009b). The methodology corresponding to nonlinear dynamic systems is presented in Cheung and Beck (2009a).

4.1 The proposed method

4.1.1 General formulation for model classes

Consider a deterministic state-space model of a linear dynamic system:

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{A}_c(t;\boldsymbol{\theta}_\mathbf{s})\mathbf{x}(t) + \mathbf{B}_c(t;\boldsymbol{\theta}_\mathbf{s})\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}(t;\boldsymbol{\theta}_\mathbf{s})\mathbf{x}(t) + \mathbf{D}(t;\boldsymbol{\theta}_\mathbf{s})\mathbf{u}(t) \\ \mathbf{x}(0) &= \mathbf{x}_0\end{aligned} \tag{4.1}$$

For a given system model, $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}$ and $\mathbf{D}$ are specified functions of parameters $\boldsymbol{\theta}_\mathbf{s}$ and t. The corresponding discrete-time state-space model with a time interval Δt is:

$$\begin{aligned}\mathbf{x}_n &= \mathbf{A}_{n-1}(\boldsymbol{\theta}_\mathbf{s})\mathbf{x}_{n-1} + \mathbf{B}_{n-1}(\boldsymbol{\theta}_\mathbf{s})\mathbf{u}_{n-1}, \quad n \in \mathbb{Z}^+ \\ \mathbf{y}_n &= \mathbf{C}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{x}_n + \mathbf{D}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{u}_n, \quad n \in \{0, \mathbb{Z}^+\}\end{aligned} \tag{4.2}$$

where $\mathbf{x}_n=\mathbf{x}(n\Delta t) \in \mathbb{R}^{N_S}$, $\mathbf{u}_n=\mathbf{u}(n\Delta t)) \in \mathbb{R}^{N_I}$ and $\mathbf{y}_n=\mathbf{y}(n\Delta t) \in \mathbb{R}^{N_O}$ denote the model state, the observed system input and the model output at time $n\Delta t$ respectively. If $\mathbf{A}_c$ and $\mathbf{B}_c$ are time-varying, by the coefficient matrices $\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})$ and $\mathbf{B}_n(\boldsymbol{\theta}_\mathbf{s})$ can be obtained using numerical integration. If $\mathbf{A}_c$ and $\mathbf{B}_c$ are time-invariant, the coefficient matrices $\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})= \mathbf{A}(\boldsymbol{\theta}_\mathbf{s})$ and $\mathbf{B}_n(\boldsymbol{\theta}_\mathbf{s})= \mathbf{B}(\boldsymbol{\theta}_\mathbf{s})$ are related to $\mathbf{A}_c(\boldsymbol{\theta}_\mathbf{s})$ and $\mathbf{B}_c(\boldsymbol{\theta}_\mathbf{s})$ by:

$$\begin{aligned}\mathbf{A}(\boldsymbol{\theta}_\mathbf{s}) &= \exp(\Delta t \mathbf{A}_c(\boldsymbol{\theta}_\mathbf{s})) \\ \mathbf{B}(\boldsymbol{\theta}_\mathbf{s}) &= \mathbf{A}_c^{-1}(\boldsymbol{\theta}_\mathbf{s})(\mathbf{I} - \mathbf{A}(\boldsymbol{\theta}_\mathbf{s}))\mathbf{B}_c(\boldsymbol{\theta}_\mathbf{s})\end{aligned} \tag{4.3}$$

Thus, $\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})$ and $\mathbf{B}_n(\boldsymbol{\theta}_\mathbf{s})$ are nonlinear in the parameters $\boldsymbol{\theta}_\mathbf{s}$ even if $\mathbf{A}_c(t;\boldsymbol{\theta}_\mathbf{s})$ and $\mathbf{B}_c(t;\boldsymbol{\theta}_\mathbf{s})$ can be expressed as a linear function of $\boldsymbol{\theta}_\mathbf{s}$.

As in past applications of the stochastic system-based framework, a model class can be constructed from the deterministic state-space model by stochastic embedding. In this

process, the parameters $\boldsymbol{\theta}_{\mathbf{s}}$ for the coefficient matrices in the discrete-time state-space model are treated as uncertain and an uncertain prediction-error term $\mathbf{v}_n$ is added on the right hand side of the output vector equation in (4.2) so that the model equations become:

$$\begin{aligned} \mathbf{x}_n &= \mathbf{A}_{n-1}(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{x}_{n-1} + \mathbf{B}_{n-1}(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{u}_{n-1}, \quad n \in \mathbb{Z}^+ \\ \mathbf{y}_n &= \mathbf{C}_n(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{x}_n + \mathbf{D}_n(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{u}_n + \mathbf{v}_n, \quad n \in \{0, \mathbb{Z}^+\} \end{aligned} \tag{4.4}$$

where the $\mathbf{v}_n$ at different times are modeled as independent Gaussian PDFs based on the Principle of Maximum Information Entropy (Jaynes 2003).

These model classes can be extended by also adding an uncertain prediction-error term $\mathbf{w}_n$ on the right hand side of the state vector equation as follows:

$$\begin{aligned} \mathbf{x}_n &= \mathbf{A}_{n-1}(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{x}_{n-1} + \mathbf{B}_{n-1}(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{u}_{n-1} + \mathbf{w}_n, \quad n \in \mathbb{Z}^+ \\ \mathbf{y}_n &= \mathbf{C}_n(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{x}_n + \mathbf{D}_n(\boldsymbol{\theta}_{\mathbf{s}})\mathbf{u}_n + \mathbf{v}_n, \quad n \in \{0, \mathbb{Z}^+\} \end{aligned} \tag{4.5}$$

Here the probability models for $\mathbf{w}_n$ and $\mathbf{v}_n$ are taken as independent Gaussian PDFs, again based on the Principle of Maximum Information Entropy: $\mathbf{w}_n \sim N(\mathbf{0}, \mathbf{Q}_n(\boldsymbol{\theta}_{\boldsymbol{w}}))$ and $\mathbf{v}_n \sim N(\mathbf{0}, \mathbf{R}_n(\boldsymbol{\theta}_{\boldsymbol{v}}))$ where $\mathbf{w}_n$ and $\mathbf{v}_n$ at all times are independently distributed. $\mathbf{Q}_n$ and $\mathbf{R}_n$ are specified functions of the uncertain parameters $\boldsymbol{\theta}_{\boldsymbol{w}}$ and $\boldsymbol{\theta}_{\boldsymbol{v}}$, respectively. In the case of uncertain initial conditions, $\mathbf{x}_0$ can be treated as uncertain parameters**.**

The specification of these probability models, along with the two fundamental system probabilistic models, $p(\mathbf{x}_n|\mathbf{x}_{n-1},\mathbf{u}_{n-1},\boldsymbol{\theta}_{\mathbf{s}},\boldsymbol{\theta}_{\boldsymbol{w}})$ and $p(\mathbf{y}_n|\mathbf{x}_n,\mathbf{u}_n,\boldsymbol{\theta}_{\mathbf{s}},\boldsymbol{\theta}_{\boldsymbol{v}})$ implied by (4.5), completely defines the stochastic model of the system dynamics. These, along with the specification of the prior distribution of the uncertain parameters, define a model class $\mathcal{M}$.

Let $\mathbf{U}_n=[\mathbf{u}_0{}^T\ \mathbf{u}_1{}^T \ldots \mathbf{u}_n{}^T]^T$, $\mathbf{Y}_n=[\mathbf{y}_0{}^T\ \mathbf{y}_1{}^T \ldots \mathbf{y}_n{}^T]^T$, $\boldsymbol{\theta}=[\boldsymbol{\theta}_s{}^T\ \boldsymbol{\theta}_{\boldsymbol{w}}{}^T\ \boldsymbol{\theta}_{\boldsymbol{v}}{}^T]^T$. In the case of uncertain initial conditions, $\mathbf{x}_0$ is included as part of $\boldsymbol{\theta}$. Given $\boldsymbol{\theta}$ and the measured system input $\mathbf{U}_n$, the predictive PDF for the system output $\mathbf{Y}_N$ can be written as follows**:**

$$p(\mathbf{Y}_N \mid \boldsymbol{\theta}) = p(\mathbf{y}_0 \mid \boldsymbol{\theta}) \prod_{n=1}^{N} p(\mathbf{y}_n \mid \mathbf{Y}_{n-1}, \boldsymbol{\theta}) \tag{4.6}$$

Here, for convenience, the conditioning of the PDF on $\mathbf{U}_N$ and the model class $\mathcal{M}$ is left implicit, although later when there is conditioning on different model classes, it will be made explicit. The conditional PDF $p(\mathbf{y}_n|\mathbf{Y}_{n-1},\boldsymbol{\theta})$ in (4.6) is a Gaussian PDF with mean $\mathrm{E}(\mathbf{y}_n|\mathbf{Y}_{n-1},\boldsymbol{\theta})= \mathbf{y}_{n|n-1}$ and covariance matrix $\mathrm{Cov}(\mathbf{y}_n|\mathbf{Y}_{n-1},\boldsymbol{\theta})=\mathbf{S}_{n|n-1}$ which are given later, while $p(\mathbf{y}_0|\boldsymbol{\theta})$ is a Gaussian PDF with mean $\mathrm{E}(\mathbf{y}_0|\boldsymbol{\theta})= \mathbf{y}_{0|-1}$ and covariance matrix $\mathrm{Cov}(\mathbf{y}_0|\boldsymbol{\theta})= \mathbf{S}_{0|-1}$ where:

$$\mathbf{y}_{0|-1} = \mathbf{C}_0(\boldsymbol{\theta}_s)\mathbf{x}_0 + \mathbf{D}_0(\boldsymbol{\theta}_s)\mathbf{u}_0 \tag{4.7}$$

$$\mathbf{S}_{0|-1} = \mathbf{R}_0(\boldsymbol{\theta}_v) \tag{4.8}$$

Thus, $p(\mathbf{y}_0|\boldsymbol{\theta})$ and $p(\mathbf{y}_n|\mathbf{Y}_{n-1},\boldsymbol{\theta})$ are given by:

$$p(\mathbf{y}_0 \mid \boldsymbol{\theta}) = \frac{1}{(2\pi)^{No/2} \mid \mathbf{S}_{0|-1} \mid^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}_0 - \mathbf{y}_{0|-1})^T \mathbf{S}_{0|-1}{}^{-1}(\mathbf{y}_0 - \mathbf{y}_{0|-1})\right) \tag{4.9}$$

$$p(\mathbf{y}_n \mid \mathbf{Y}_{n-1}, \boldsymbol{\theta}) = \frac{1}{(2\pi)^{No/2} \mid \mathbf{S}_{n|n-1} \mid^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{y}_n - \mathbf{y}_{n|n-1})^T \mathbf{S}_{n|n-1}{}^{-1}(\mathbf{y}_n - \mathbf{y}_{n|n-1})\right) \tag{4.10}$$

and $p(\mathbf{Y}_N|\boldsymbol{\theta})$ in (4.6) is given by:

$$p(\mathbf{Y}_N \mid \boldsymbol{\theta}) = \frac{1}{(2\pi)^{No(N+1)/2} \prod_{n=0}^{N} \mid \mathbf{S}_{n|n-1} \mid^{1/2}} \exp\left(-\frac{1}{2}\sum_{n=0}^{N}(\mathbf{y}_n - \mathbf{y}_{n|n-1})^T \mathbf{S}_{n|n-1}{}^{-1}(\mathbf{y}_n - \mathbf{y}_{n|n-1})\right) \tag{4.11}$$

For a given $\boldsymbol{\theta}$, $\mathbf{y}_{n|n\text{-}1}$ and $\mathbf{S}_{n|n\text{-}1}$ can be calculated by (4.14) and (4.15) and the following Kalman filter equations which come from Bayesian sequential state updating with $\mathbf{x}_{0|0}=\mathbf{x}_0$ and $\mathbf{P}_{0|0}=\mathbf{O}$:

$$\mathbf{x}_{n|n-1} = \mathbf{A}_{n-1}(\boldsymbol{\theta}_s)\mathbf{x}_{n-1|n-1} + \mathbf{B}_{n-1}(\boldsymbol{\theta}_s)\mathbf{u}_{n-1} \tag{4.12}$$

$$\mathbf{P}_{n|n-1} = \mathbf{A}_{n-1}(\boldsymbol{\theta}_s)\mathbf{P}_{n-1|n-1}\mathbf{A}_{n-1}(\boldsymbol{\theta}_s)^T + \mathbf{Q}_n(\boldsymbol{\theta}_w) \tag{4.13}$$

$$\mathbf{y}_{n|n-1} = \mathbf{C}_n(\boldsymbol{\theta}_s)\mathbf{x}_{n|n-1} + \mathbf{D}_n(\boldsymbol{\theta}_s)\mathbf{u}_n \tag{4.14}$$

$$\mathbf{S}_{n|n-1} = \mathbf{C}_n(\boldsymbol{\theta}_s)\mathbf{P}_{n|n-1}\mathbf{C}_n(\boldsymbol{\theta}_s)^T + \mathbf{R}_n(\boldsymbol{\theta}_v) \tag{4.15}$$

$$\mathbf{x}_{n|n} = \mathbf{x}_{n|n-1} + \mathbf{P}_{n|n-1}\mathbf{C}_n(\boldsymbol{\theta}_s)^T \mathbf{S}_{n|n-1}^{-1}(\mathbf{y}_n - \mathbf{y}_{n|n-1}) \tag{4.16}$$

$$\mathbf{P}_{n|n} = \mathbf{P}_{n|n-1} - \mathbf{P}_{n|n-1}\mathbf{C}_n(\boldsymbol{\theta}_s)^T \mathbf{S}_{n|n-1}^{-1}\mathbf{C}_n(\boldsymbol{\theta}_s)\mathbf{P}_{n|n-1} \tag{4.17}$$

The posterior PDF of $\boldsymbol{\theta}$ is then given by (1.1) where $\mathcal{D} = \hat{\mathbf{Y}}_N$, the measurements for the system output $\mathbf{Y}_N$. The model classes resulting from (4.4) can be viewed as a special case of the extended ones resulting from (4.5) where $\mathbf{Q}_n(\boldsymbol{\theta}_w)= \mathbf{O}$ and thus: $\mathbf{y}_{n|n\text{-}1}=\mathbf{C}_n(\boldsymbol{\theta}_s)\mathbf{x}_n+\mathbf{D}_n(\boldsymbol{\theta}_s)\mathbf{u}_n$, $\mathbf{S}_{n|n\text{-}1}= \mathbf{R}_n(\boldsymbol{\theta}_v)$ where in (4.12), $\mathbf{x}_{n|n\text{-}1} = \mathbf{x}_n$ and $\mathbf{x}_{n\text{-}1|n\text{-}1} = \mathbf{x}_{n\text{-}1}$ and no Kalman filtering needs to be performed**.**

From (4.5), it can be shown that:

$$\begin{aligned}\mathbf{y}_n = {} & \mathbf{C}_n(\boldsymbol{\theta}_s)\{[\prod_{j=0}^{n-1}\mathbf{A}_j(\boldsymbol{\theta}_s)]\mathbf{x}_0 + \sum_{i=0}^{n-2}[\prod_{j=1}^{n-1-i}\mathbf{A}_{n-j}(\boldsymbol{\theta}_s)]\mathbf{B}_i(\boldsymbol{\theta}_s)\mathbf{u}_i + \mathbf{B}_{n-1}(\boldsymbol{\theta}_s)\mathbf{u}_{n-1}\} + \mathbf{D}_n(\boldsymbol{\theta}_s)\mathbf{u}_n \\ & + \mathbf{C}_n(\boldsymbol{\theta}_s)\{\sum_{j=1}^{n-1}[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_s)]\mathbf{w}_j + \mathbf{w}_n\} + \mathbf{v}_n\end{aligned} \tag{4.18}$$

For the case with time-invariant coefficient matrices, we have:

$$\mathbf{y}_n = \mathbf{C}(\boldsymbol{\theta}_s)[\mathbf{A}^n(\boldsymbol{\theta}_s)\mathbf{x}_0 + \sum_{i=0}^{n-1}\mathbf{A}^{n-i-1}(\boldsymbol{\theta}_s)\mathbf{B}(\boldsymbol{\theta}_s)\mathbf{u}_i] + \mathbf{D}(\boldsymbol{\theta}_s)\mathbf{u}_n + \mathbf{C}(\boldsymbol{\theta}_s)\sum_{j=1}^{n}\mathbf{A}^{n-j}(\boldsymbol{\theta}_s)\mathbf{w}_j + \mathbf{v}_n \quad (4.19)$$

Notice that both model classes resulting from (4.4)and (4.5) have the same mean predicted output, given $\boldsymbol{\theta}$. For the extended model class, the prediction errors for the system output are accounted for by both the prediction errors in the state vector equation and output vector equation (the last two terms in (4.18) for the case with time-varying coefficient matrices or (4.19) for the case with time-invariant coefficient matrices). The measurements of the system output also provide information about the prediction errors in the state vector equation, thereby allowing more flexibility in treating modeling uncertainties in the response predictions; this is especially useful for predictions at unobserved DOFs of quantities physically different from the measured quantities.

Given $\boldsymbol{\theta}$, the covariance of the prediction error for the system output at time $n\Delta t$ for the original model class derived from (4.4) is $\mathbf{R}_n(\boldsymbol{\theta}_v)$. For the extended model class derived from (4.5), the covariance of the prediction error for the system output at time $n\Delta t$ (denoted by $\boldsymbol{\Sigma}(n)$) is given by ($\boldsymbol{\Sigma}(0)= \mathbf{R}_0(\boldsymbol{\theta}_v)$):

$$\boldsymbol{\Sigma}(n) = \mathbf{C}_n(\boldsymbol{\theta}_s)\{\sum_{j=1}^{n-1}[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_s)]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_s)]^T + \mathbf{Q}_n(\boldsymbol{\theta}_w)\}\mathbf{C}_n(\boldsymbol{\theta}_s)^T + \mathbf{R}_n(\boldsymbol{\theta}_v) \quad (4.20)$$

For the case with time-invariant coefficient matrices,

$$\boldsymbol{\Sigma}(n) = \mathbf{C}(\boldsymbol{\theta}_s)[\sum_{j=1}^{n}\mathbf{A}^{n-j}(\boldsymbol{\theta}_s)\mathbf{Q}(\boldsymbol{\theta}_w)\mathbf{A}^{n-j}(\boldsymbol{\theta}_s)^T]\mathbf{C}(\boldsymbol{\theta}_s)^T + \mathbf{R}(\boldsymbol{\theta}_v) \quad (4.21)$$

For computational efficiency, it is shown in Appendix 4A that $\boldsymbol{\Sigma}(n)$ can be obtained using the following iterative formula. For (4.20), we have:

$$\begin{aligned}
&\mathbf{S}(1) = \mathbf{Q}_1(\boldsymbol{\theta}_w) \\
&\Sigma(n) = \mathbf{C}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{S}(n)\mathbf{C}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{R}_n(\boldsymbol{\theta}_v) \\
&\mathbf{S}(n+1) = \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{S}(n)\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)
\end{aligned} \tag{4.22}$$

For the case with time-invariant coefficient matrices, we have

$$\begin{aligned}
&\mathbf{S}(1) = \mathbf{Q}(\boldsymbol{\theta}_w) \\
&\Sigma(n) = \mathbf{C}(\boldsymbol{\theta}_\mathbf{s})\mathbf{S}(n)\mathbf{C}(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{R}(\boldsymbol{\theta}_v) \\
&\mathbf{S}(n+1) = \mathbf{A}(\boldsymbol{\theta}_\mathbf{s})\mathbf{S}(n)\mathbf{A}(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}(\boldsymbol{\theta}_w)
\end{aligned} \tag{4.23}$$

For the extended model class, for a given $\boldsymbol{\theta}$, the stochastic system output at one time is stochastically dependent of those at the other times due to the introduction of the prediction errors $\mathbf{w}_n$'s in the state vector equation. Also, the parameters $\boldsymbol{\theta}_\mathbf{s}$ and $\boldsymbol{\theta}_w$ and $\boldsymbol{\theta}_v$ for the prediction errors $\mathbf{w}_n$'s and $\mathbf{v}_n$'s, are stochastically coupled given the data.

It is noted that a regular Kalman filter considers the stochastic state-space model in (4.5) with fixed $\boldsymbol{\theta}_\mathbf{s}$, $\boldsymbol{\theta}_w$ and $\boldsymbol{\theta}_v$ (and also $\mathbf{x}_0$ chosen to follow a Gaussian PDF or being fixed). One important result of the proposed framework is therefore a posterior robust Kalman filter which treats modeling uncertainties and so can give more robust predictions of future responses. The predicted future responses are obtained by the sum of the prediction of the Kalman filter of each model specified by $\boldsymbol{\theta}$ weighted by its posterior probability $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})d\boldsymbol{\theta}$, according to the Theorem of Total Probability.

4.1.2 Model class comparison, averaging and robust system response and failure probability predictions

Let $M=\{\mathcal{M}_j: j=1,2,\ldots N_M\}$ denote the set of candidate model classes comprised of the extended ones and the original ones considered for a system. Bayesian model class selection is used to evaluate the posterior probability $P(\mathcal{M}_j|\mathcal{D},M)$ of an extended model class and the original one to allow a data-based comparison. The posterior probability of the

candidate model classes is evaluated from (1.2) for comparison of these model classes. For this purpose, the evidence $p(\mathcal{D}|\mathcal{M}_j)$ needs to be calculated. The method proposed in Chapter 3 is used to calculate this quantity.

As can be seen later in the illustrative example, the failure probability for the system (the probability of unsatisfactory system performance) is very sensitive to the choice of model classes. *Posterior hyper-robust* predictions as in (1.7) are essential to alleviate such sensitivity. All the probabilistic information for the prediction of future responses **X** is contained in the *hyper-robust predictive* PDF based on *M*, which is given by the Total Probability Theorem:

$$p(\mathbf{X} \mid \mathcal{D}, M) = \sum_{j=1}^{N_M} p(\mathbf{X} \mid \mathcal{D}, \mathcal{M}_j) P(\mathcal{M}_j \mid \mathcal{D}, M) \tag{4.24}$$

where the robust predictive PDF for each model class $\mathcal{M}_j$ is weighted by its posterior probability $P(\mathcal{M}_j|\mathcal{D},M)$.

4.1.2.1 Calculation of hyper-robust system failure probability for the set of candidate model classes

Let *F* denote the events or conditions leading to system failure (unsatisfactory system performance). Here, our interest is primarily the system failure subjected to uncertain future dynamic excitations/inputs **U** modeled by model classes U_j, j=1,2,…N_M. The model parameters $\boldsymbol{\theta}_U$ for U_j can be comprised of 1) model parameters $\boldsymbol{\theta}_u$ (with uncertainty quantified by $p(\boldsymbol{\theta}_u|U_j)$) which is not part of $\boldsymbol{\theta}$ and not updated by $\mathcal{D}$, and 2) $\boldsymbol{\theta}_p$ which are some components of $\boldsymbol{\theta}$ for $\mathcal{M}_j$ (with uncertainty quantified by $p(\boldsymbol{\theta}_p|\mathcal{D},\mathcal{M}_j)$ which has already been obtained during a Bayesian update), i.e. $\boldsymbol{\theta}_U = [\boldsymbol{\theta}_u^T \; \boldsymbol{\theta}_p^T]^T$. The uncertainty in $\boldsymbol{\theta}_U$ is quantified by $p(\boldsymbol{\theta}_U|\mathcal{D},U_j)$ as follows:

$$p(\boldsymbol{\theta}_U \mid \mathcal{D}, U_j) = p(\boldsymbol{\theta}_u \mid U_j) p(\boldsymbol{\theta}_p \mid \mathcal{D}, \mathcal{M}_j) \tag{4.25}$$

This model class can be viewed as a special case of hierarchical model classes covered later in Chapter 5. The uncertainty in **U** is thus quantified by $p(\mathbf{U}|\mathcal{D},U_j)$. For illustration, here we consider one very common case when $\boldsymbol{\theta}_u$ is the same and is chosen to follow the same probability distribution for all U_j. Here we consider a set M_e of N_M stochastic system model classes for the prediction of system failure probability. The j-th model class $M_{e,j}$ in M_e is given by $\mathcal{M}_j$ with the stochastic model for the excitation/input given by U_j. The *hyper-robust* system failure probability $P(F|\mathcal{D},M_e)$ based on M_e is then given by:

$$P(F \mid \mathcal{D}, M_e) = \sum_{j=1}^{N_M} P(F \mid \mathcal{D}, M_{e,j}) P(M_{e,j} \mid \mathcal{D}, M_e) \tag{4.26}$$

It can be shown using theorems developed in Chapter 5 that $P(M_{e,j}|\mathcal{D},M_e)$ is equal to $P(\mathcal{M}_j|\mathcal{D},M)$:

$$P(M_{e,j} \mid \mathcal{D}, M_e) = P(\mathcal{M}_j \mid \mathcal{D}, M) \tag{4.27}$$

For calculating the hyper-robust failure probability, besides calculating the evidence, we also need to simulate samples from the posterior PDF for the candidate model classes to calculate the posterior robust failure probability $P(F|\mathcal{D},M_{e,j})$ based on each model class $M_{e,j}$. By the Theorem of Total Probability, $P(F|\mathcal{D},M_{e,j})$ is given by the following multi-dimensional integral:

$$P(F \mid \mathcal{D}, M_{e,j}) = \int P(F \mid \boldsymbol{\theta}, \mathcal{D}, M_{e,j})\, p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) d\boldsymbol{\theta} \tag{4.28}$$

Let $\mathbf{V}=[\mathbf{v}_1{}^T\ \mathbf{v}_2{}^T\ \mathbf{v}_3{}^T\ldots.]^T$ and $\mathbf{W}=[\mathbf{w}_1{}^T\ \mathbf{w}_2{}^T\ \mathbf{w}_3{}^T\ldots.]^T$. Note that the dimension of uncertain parameters which can include $\boldsymbol{\theta}$, $\boldsymbol{\theta}_u$, **U**, **V** and/or **W** is often very high (say of the order of thousands or more) making the problem very challenging.

By the Theorem of Total Probability, $P(F|\boldsymbol{\theta},\mathcal{D},M_{e,j})$ in (4.28) is given by the following multi-dimensional integral:

$$P(F \mid \boldsymbol{\theta}, \mathcal{D}, M_{e,j}) = \int P(F \mid \boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathcal{D}, M_{e,j})\, p(\boldsymbol{\theta}_u \mid U_j) d\boldsymbol{\theta}_u \tag{4.29}$$

If $\mathcal{M}_j$ is an extended model class derived from (4.5), $P(F|\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathcal{D},M_{e,j})$ becomes:

$$\begin{aligned} P(F \mid \boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathcal{D}, M_{e,j}) = \int I_F(\boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathbf{W}, \mathbf{V}, \mathbf{U}) p(\mathbf{W} \mid \boldsymbol{\theta}_w, \mathcal{D}, \mathcal{M}_j)\, p(\mathbf{V} \mid \boldsymbol{\theta}_v, \mathcal{D}, \mathcal{M}_j) \\ p(\mathbf{U} \mid \boldsymbol{\theta}_u, \boldsymbol{\theta}_p, U_j) d\mathbf{U} d\mathbf{W} d\mathbf{V} \end{aligned} \tag{4.30}$$

If $\mathcal{M}_j$ is the model class derived from (4.4), $P(F|\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathcal{D},M_{e,j})$ is given by:

$$\begin{aligned} P(F \mid \boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathcal{D}, M_{e,j}) = \int I_F(\boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathbf{V}, \mathbf{U}) p(\mathbf{V} \mid \boldsymbol{\theta}_v, \mathcal{D}, \mathcal{M}_j) \\ p(\mathbf{U} \mid \boldsymbol{\theta}_u, \boldsymbol{\theta}_p, U_j) d\mathbf{U} d\mathbf{V} \end{aligned} \tag{4.31}$$

Recall that $p(\mathbf{W}|\boldsymbol{\theta}_w^{(k)},\mathcal{D},\mathcal{M}_j) = p(\mathbf{W}|\boldsymbol{\theta}_w^{(k)},\mathcal{M}_j)$ is chosen to be independently and identically distributed Gaussian with a covariance matrix $\mathbf{Q}(\boldsymbol{\theta}_w)$ and mean equal to zero; $p(\mathbf{V}|\boldsymbol{\theta}_v^{(k)},\mathcal{D},\mathcal{M}_j) = p(\mathbf{V}|\boldsymbol{\theta}_v^{(k)},\mathcal{M}_j)$ is chosen to be independently and identically distributed Gaussian with a covariance matrix and $\mathbf{R}(\boldsymbol{\theta}_v)$ and mean equal to zero; $\boldsymbol{\theta}_p$ is contained inside $\boldsymbol{\theta}$ and $p(\mathbf{U}|\boldsymbol{\theta}_u,\boldsymbol{\theta}_p,\mathcal{D},\mathcal{M}_j,U_j) = p(\mathbf{U}|\boldsymbol{\theta}_u,\boldsymbol{\theta}_p,U_j)$.

One way to calculate $P(F|\mathcal{D},M_{e,j})$ is by using (4.28) and (4.29). Using MCS, $P(F|\mathcal{D},M_{e,j})$ can be estimated by:

$$P(F \mid \mathcal{D}, M_{e,j}) \approx \frac{1}{N} \sum_{k=1}^{N} P(F \mid \boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}_u^{(k)}, \mathcal{D}, M_{e,j}) \tag{4.32}$$

where $\boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\theta}_u^{(k)}$ are samples generated according to $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$ and $p(\boldsymbol{\theta}_u|U_j)$ respectively. For the original and the extended model classes, if the performance measures corresponding to F are the states and/or the quantities that are of the same type as the output measurements and $p(\mathbf{U}|\boldsymbol{\theta}_u,\boldsymbol{\theta}_p,U_j)$ is Gaussian (which is the case considered in the illustrative example), $P(F|\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathcal{D},M_{e,j})$ in (4.30) or (4.31) can be calculated using efficient stochastic simulation algorithms such as Importance Sampling Using Elementary Events

(ISEE) (Au and Beck 2001a), Wedge Simulation Method (WSM) (Katafygiotis and Cheung 2004), Domain Decomposition Method (DDM) (Katafygiotis and Cheung 2006).

For the cases involving stochastic nonlinear models, we consider the following integral (4.33) (or (4.34)) $P(F|\mathcal{D},M_{e,j})$ derived from (4.28)-(4.30) (or (4.28)-(4.31)) for $\mathcal{M}_j$ being the extended model class (or the original model class) in the parameter space of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_u$, $\mathbf{V}$, $\mathbf{W}$ and $\mathbf{U}$ (or $\boldsymbol{\theta}$, $\boldsymbol{\theta}_u$, $\mathbf{V}$ and $\mathbf{U}$):

$$P(F \mid \mathcal{D}, \mathcal{M}_j, U_j) = \int I_F(\boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathbf{W}, \mathbf{V}, \mathbf{U}) p(\mathbf{W} \mid \boldsymbol{\theta}_w, \mathcal{M}_j) p(\mathbf{V} \mid \boldsymbol{\theta}_v, \mathcal{M}_j) \\ p(\mathbf{U} \mid \boldsymbol{\theta}_u, \boldsymbol{\theta}_p, U_j) p(\boldsymbol{\theta}_u \mid U_j) p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) d\mathbf{U} d\mathbf{W} d\mathbf{V} d\boldsymbol{\theta} d\boldsymbol{\theta}_u \tag{4.33}$$

$$P(F \mid \mathcal{D}, \mathcal{M}_j, U_j) = \int I_F(\boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathbf{V}, \mathbf{U}) p(\mathbf{V} \mid \boldsymbol{\theta}_v, \mathcal{M}_j) \\ p(\mathbf{U} \mid \boldsymbol{\theta}_u, \boldsymbol{\theta}_p, U_j) p(\boldsymbol{\theta}_u \mid U_j) p(\boldsymbol{\theta} \mid \mathcal{D}, \mathcal{M}_j) d\mathbf{U} d\mathbf{V} d\boldsymbol{\theta} d\boldsymbol{\theta}_u \tag{4.34}$$

By MCS, $P(F|\mathcal{D},\mathcal{M}_j,U_j)$ in (4.33) (or (4.34)) can be estimated by (4.35) (or (4.36)):

$$P(F \mid \mathcal{D}, \mathcal{M}_j, U_j) \approx \frac{1}{N} \sum_{k=1}^{N} I_F(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}_u^{(k)}, \mathbf{W}^{(k)}, \mathbf{V}^{(k)}, \mathbf{U}^{(k)}) \tag{4.35}$$

$$P(F \mid \mathcal{D}, \mathcal{M}_j, U_j) \approx \frac{1}{N} \sum_{k=1}^{N} I_F(\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}_u^{(k)}, \mathbf{V}^{(k)}, \mathbf{U}^{(k)}) \tag{4.36}$$

where $\boldsymbol{\theta}^{(k)}$, $\boldsymbol{\theta}_u^{(k)}$, $\mathbf{W}^{(k)}$, $\mathbf{V}^{(k)}$ and $\mathbf{U}^{(k)}$ are samples generated according to $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_j)$, $p(\boldsymbol{\theta}_u|U_j)$, $p(\mathbf{W}|\boldsymbol{\theta}_w^{(k)},\mathcal{M}_j)$, $p(\mathbf{V}|\boldsymbol{\theta}_v^{(k)},\mathcal{M}_j)$ and $p(\mathbf{U}|\boldsymbol{\theta}_u^{(k)},\boldsymbol{\theta}_p^{(k)},U_j)$ respectively; $\boldsymbol{\theta}_w^{(k)}$, $\boldsymbol{\theta}_v^{(k)}$ and $\boldsymbol{\theta}_p^{(k)}$ are contained inside $\boldsymbol{\theta}^{(k)}$.

For very small $P(F|\mathcal{D},\mathcal{M}_j)$ (say smaller than 0.01), (4.35) or (4.36) is not computationally efficient. An algorithm based on Subset Simulation (Au and Beck 2001b) applied in the parameter space of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_u$, $\mathbf{V}$, $\mathbf{W}$ **and** $\mathbf{U}$ has recently been developed for the evaluation of

$P(F|\mathcal{D},\mathcal{M}_j,U_j)$ and is presented in Cheung and Beck (2009a). In Chapter 6, a new alternative method of calculating the robust failure probability given dynamic data $\mathcal{D}$ is also presented.

4.2 Illustrative example

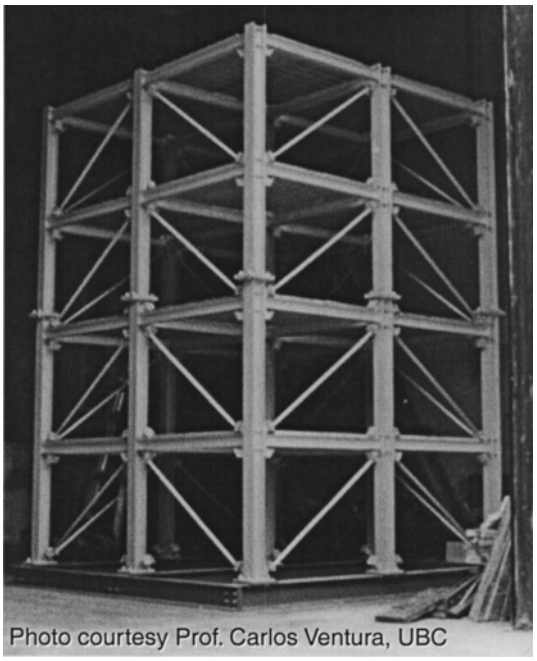

Figure 4.1: IASC-ASCE Structural Health Monitoring Task Group benchmark structure

In this example, the benchmark structure (Figure 4.1) from the IASC-ASCE Structural Health Monitoring Task Group (Johnson et al. 2004) is considered. It is a 4-story, 2 bay by 2 bay steel frame structure built in the Earthquake Engineering Research Laboratory at the University of British Columbia in Canada. A set of simulated dynamic data is used. It consists of 10s (with a sample interval Δt of 0.004s) of the horizontal acceleration $\ddot{y}_{la}$, $\ddot{y}_{lb}$ (in the weak (y) direction), l=1,…,4, of each floor on east and west frames respectively contaminated by Gaussian white noise with noise level of 10% of the maximum over floors of the RMS acceleration responses, which corresponds to input dynamic excitations w_l at

each floor in the y direction (Figure 4.2). These data are generated by a 120-DOF three-dimensional finite element model (Johnson et al. 2004) for the benchmark structure with simulated wind excitations generated by Gaussian white noise processes passed through a 6^{th} order low-pass Butterworth filter with a 100Hz cutoff. The number of observed degrees of freedom is N_o=4 and N =2500 is the length of the discrete time history data.

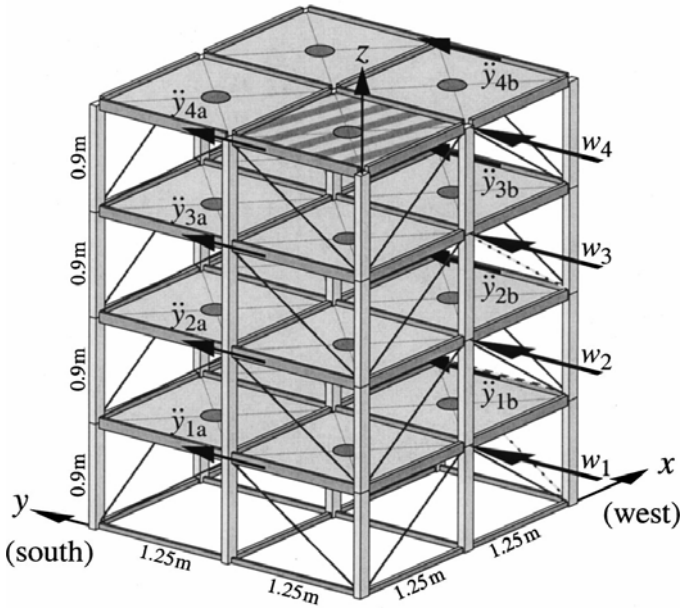

Figure 4.2: Schematic diagram showing the directions of system output measurements and input excitations

Here we consider a set $\boldsymbol{M}=\{\mathcal{M}_i: i=1,2\}$ consisting of 2 candidate model classes with $\mathcal{M}_1$ corresponding to the extended model class derived from (4.5) and $\mathcal{M}_2$ corresponding to the one derived from (4.4). To investigate the effect of introducing the prediction errors in the state vector equation as in the extended model classes, the same type of deterministic state-space model is used for both model classes.

Model class $\mathcal{M}_1$: The deterministic dynamic model consists of a 4-DOF linear lumped-mass shear building model for motion in the y direction with classical damping for the 4 modes. This simple model was selected to produce significant errors in the prediction of the system

response since the data are generated from a more complicated model. The system is assumed to start at rest: $\mathbf{x}_0=\mathbf{0}$. The covariance matrix for the prediction errors $\mathbf{w}_n$ for the state vector equation in (4.5) is modeled as a diagonal matrix:

$$\mathbf{Q}(\boldsymbol{\theta}_w)=\begin{bmatrix}\sigma_{w,1}{}^2\mathbf{I}_{4\times4} & \mathbf{O}_{4\times4}\\ \mathbf{O}_{4\times4} & \sigma_{w,2}{}^2\mathbf{I}_{4\times4}\end{bmatrix},\boldsymbol{\theta}_w=[\sigma_{w,1}{}^2\ \ \sigma_{w,2}{}^2]^T \tag{4.37}$$

and the covariance matrix for the prediction and measurement errors $\mathbf{v}_n$ for the output vector equation is modeled as a diagonal matrix:

$$\mathbf{R}(\boldsymbol{\theta}_v)=\sigma^2\mathbf{I}_{4\times4},\ \boldsymbol{\theta}_v=\sigma^2 \tag{4.38}$$

There is a total of 15 uncertain parameters to be updated**:** lumped mass m_l and stiffness k_l of each story, damping ratio of each mode ξ_l, l=1,…,4 and the variances $\sigma_{w,1}{}^2$, $\sigma_{w,2}{}^2$ and σ^2 for the prediction errors. Note that $\sigma_{w,1}{}^2$ is the variance of the prediction error of the displacement vector equation and $\sigma_{w,2}{}^2$ is the variance of the prediction error of the velocity vector equation. The coefficient matrices $\mathbf{A}_c$, $\mathbf{B}_c$, $\mathbf{C}$ and $\mathbf{D}$ in (4.1) are given as follows in terms of the uncertain mass $\mathbf{M_s}$, damping $\mathbf{C_s}$ and stiffness matrices $\mathbf{K_s}$:

$$\mathbf{A}_c=\begin{bmatrix}\mathbf{O} & \mathbf{I}\\ -\mathbf{M}_s{}^{-1}\mathbf{K}_s & -\mathbf{M}_s{}^{-1}\mathbf{C}_s\end{bmatrix} \tag{4.39}$$

$$\mathbf{B}_c=\begin{bmatrix}\mathbf{O}\\ \mathbf{M}_s{}^{-1}\end{bmatrix} \tag{4.40}$$

$$\mathbf{C}=\begin{bmatrix}-\mathbf{M}_s{}^{-1}\mathbf{K}_s & -\mathbf{M}_s{}^{-1}\mathbf{C}_s\end{bmatrix} \tag{4.41}$$

$$\mathbf{D}=\mathbf{M}_s{}^{-1} \tag{4.42}$$

$$\mathbf{M}_s = \begin{bmatrix} m_1 & 0 & 0 & 0 \\ 0 & m_2 & 0 & 0 \\ 0 & 0 & m_3 & 0 \\ 0 & 0 & 0 & m_4 \end{bmatrix} \tag{4.43}$$

$$\mathbf{K}_s = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 & 0 \\ -k_2 & k_2 + k_3 & -k_3 & 0 \\ 0 & -k_3 & k_3 + k_4 & -k_4 \\ 0 & 0 & -k_4 & k_4 \end{bmatrix} \tag{4.44}$$

$$\mathbf{C}_s = \mathbf{M}_s \left(\sum_{l=1}^{4} \frac{2\xi_l \omega_l}{\boldsymbol{\varphi}_l^T \mathbf{M}_s \boldsymbol{\varphi}_l} \boldsymbol{\varphi}_l \boldsymbol{\varphi}_l^T \right) \mathbf{M}_s \tag{4.45}$$

$$(\mathbf{M}_s^{-1} \mathbf{K}_s - \omega_l^2 \mathbf{I}) \boldsymbol{\varphi}_l = \mathbf{0} \tag{4.46}$$

The likelihood function for $\boldsymbol{\theta}$ can be obtained using the equations in (4.11)-(4.17).

The prior PDF for $\boldsymbol{\theta}$ is chosen to be the product of independent distributions, where m_l, k_l, ξ_l follow a lognormal distribution with medians equal to their nominal values and the corresponding coefficients of variation (c.o.v.) of 10%, 30%, 50% respectively; $\sigma_{w,1}^2$, $\sigma_{w,2}^2$ and σ^2 follow a uniform distribution on the interval [0 $\sigma_{w,1}^2{}_{\max}$] , [0 $\sigma_{w,2}^2{}_{\max}$] and [0 $\sigma^2{}_{\max}$], respectively, where $\sigma^2{}_{\max}$ is equal to the square of the maximum over floors of the RMS of acceleration data; $\sigma_{w,2}^2{}_{\max}$ is equal to the square of the maximum over floors of the RMS of the 'velocity data' obtained by numerically integrating the acceleration data using the trapezoidal rule; $\sigma_{w,1}^2{}_{\max}$ is equal to the square of the maximum over floors of the RMS of the 'displacement data' obtained by numerically integrating the acceleration data twice using the trapezoidal rule. It is well known that the 'velocity data' and 'displacement data' obtained by an integration of the acceleration data give a very poor estimate of the system velocity and displacement. Here these pseudo 'velocity data' and 'displacement data' are only used to choose the maxima for the prior PDF for $\sigma_{w,2}^2$ and $\sigma_{w,1}^2$. During the Bayesian update, if it is observed that the prediction error variance parameter has a large probability

clustering around the upper limit of the uniform prior, the upper limit can be made larger so that the high probability region of the posterior PDF of the parameter is within the range of the uniform prior.

The nominal values for k_l and ξ_l are 67.9MNm^{-1} (Johnson et al. 2004) and 1% respectively, for l=1,…,4 and those for m_1, m_2, m_3 and m_4 are 3246kg, 2652kg, 2652kg and 1809kg respectively. For the mass parameters, relatively smaller values of c.o.v. are chosen since these parameters can usually be more precisely determined from the structural drawings than the other model parameters. For the latter parameters, a larger c.o.v. is chosen. It should be noted that the objective of the prior PDFs is to allow prior information to be incorporated when performing model updating. For those parameters where there is little prior information, prior PDFs that reflect higher uncertainty (i.e., in this case, larger c.o.v.) are used. Under such circumstances, the updated uncertainties for these parameters depend mostly on the data and are often insensitive to the prior PDFs. Here we define the 12 dimensionless uncertain parameters $\boldsymbol{\theta}_s$ corresponding to the physical parameters (including the mass, stiffness and damping ratio parameters) as the original parameters divided by their nominal values.

Model class $\mathcal{M}_2$: The differences between this model class and $\mathcal{M}_1$ are: 1) the prior PDF is the same as $\mathcal{M}_1$ except that $\mathcal{M}_2$ does not include the uncertain parameters $\sigma_{w,1}^2$, $\sigma_{w,2}^2$; and 2) $\mathbf{Q}(\boldsymbol{\theta}_w)= \mathbf{O}$. Thus, the likelihood function is simpler than $\mathcal{M}_1$ and does not require Kalman filtering. Let $\mathbf{y}_n(\boldsymbol{\theta}_s)$ denote the output at time t_n at the l-th observed degree of freedom predicted by the 4-DOF shear building model and $\hat{\mathbf{y}}_n$ denote the corresponding measured output. The prediction and measurement errors for the system output equation is given by: $\mathbf{v}_n = \hat{\mathbf{y}}_n - \mathbf{y}_n(\boldsymbol{\theta}_s)$ for n=0,1,…,N=2500, whose componenents are modeled as independent and identically distributed Gaussian variables with mean zero and some unknown prediction-error variance σ^2, based on the Principle of Maximum Information Entropy (Jaynes 2003). The likelihood function $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}_2)$ for this model class is:

$$p(\hat{\mathbf{Y}}_N \mid \boldsymbol{\theta}, \mathcal{M}_2) = \frac{1}{(2\pi\sigma^2)^{\frac{N_O(N+1)}{2}}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N}[\hat{\mathbf{y}}_n - \mathbf{y}_n(\boldsymbol{\theta}_s)]^T[\hat{\mathbf{y}}_n - \mathbf{y}_n(\boldsymbol{\theta}_s)]\right) \tag{4.47}$$

A hybrid method based on TMCMC and Hybrid Monte Carlo Method presented in Chapter 2 is used to generate samples from the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}_i)$. Table 4.1 shows the sample posterior means (outside the parenthesis) and c.o.v. (coefficient of variation) in % (inside the parenthesis) for the uncertain parameters $\boldsymbol{\theta}_s$ of the underlying deterministic state-space model, the parameters $\boldsymbol{\theta}_w$ for the covariance matrix of the state-vector equation prediction error, and the parameters $\boldsymbol{\theta}_v$ for the covariance matrix of the output-equation prediction error. θ_l, θ_{l+4} and θ_{l+8} are the dimensionless parameters corresponding to m_l, k_l and ξ_l respectively for l=1,…,4. The first number in the second row and second column of Table 4.1 gives the posterior mean of θ_1. The number inside the parenthesis next to this number gives the posterior c.o.v. of θ_1 as a %. The next row gives the result corresponding to θ_2 and so on. The results for $\mathcal{M}_2$ are presented in a similar fashion. For both model classes, the posterior c.o.v. for the parameters related to the damping ratio is larger than those related to the mass and stiffness parameters, showing that there is a larger uncertainty in the damping parameters, as can be expected.

Table 4.1 Posterior means and c.o.v. for the uncertain parameters

	$\mathcal{M}_1$	$\mathcal{M}_2$
θ_1	0.97(0.5),	1.12(0.9),
θ_2	0.98(0.5),	1.13(1.0),
θ_3	0.99(0.5),	1.04(0.9),
θ_4	1.07(0.5),	1.21(1.0),
θ_5	0.76(0.7),	0.81(0.9),
θ_6	0.94(0.6),	1.10(1.0),
θ_7	0.90(0.7),	1.03(0.9),
θ_8	0.92(0.5),	0.95(0.9),
θ_9	1.11(14.8),	0.88(2.7),
θ_{10}	1.42(6.9),	0.86(1.6),
θ_{11}	1.89(4.9),	0.86(1.4),
θ_{12}	1.23(7.2)	1.40(2.1)
$\boldsymbol{\theta}_w$	$5.80\text{x}10^{-11}$(3.7), $2.26\text{x}10^{-6}$(10.1)	Not applicable
$\boldsymbol{\theta}_v$	0.103(2.4)	3.26(1.4)

The exact measurement noise variance is 0.1972 ms^{-2}. It can be seen that the posterior mean of the output-equation prediction-error variance $\boldsymbol{\theta}_v$ for $\mathcal{M}_2$ is about 16 times the exact measurement noise variance, or 4 times if we look at the prediction-error standard deviation, in order to account for modeling errors. This prediction-error standard deviation is about 40% of the maximum over floors of the RMS of acceleration data showing that the models in $\mathcal{M}_2$ have significant modeling error. It can be seen that the posterior mean of $\boldsymbol{\theta}_v$ for $\mathcal{M}_1$ is about 52% of the exact measurement noise variance (about 72% if we look at the prediction-error standard deviation) and is significantly smaller than that for $\mathcal{M}_2$. The prediction-error term in the output-equation for $\mathcal{M}_1$ mostly accounts for the measurement noise while the prediction-error term in the state vector equation accounts for the modeling errors. The prediction-error term in the output vector equation for $\mathcal{M}_2$ has to account for both the measurement noise and the modeling uncertainties and thus its variance is larger than that for $\mathcal{M}_1$. For both model classes, modeling uncertainties are also accounted for by allowing uncertainty in the value of $\boldsymbol{\theta}_s$. Given $\boldsymbol{\theta}$, the covariance matrix $\boldsymbol{\Sigma}(n)$ of the prediction errors for the system output in the output equation for $\mathcal{M}_2$ is $\mathbf{R}(\boldsymbol{\theta}_v)=\sigma^2\mathbf{I}_{4x4}$ for all time while that for $\mathcal{M}_1$ at each time can be obtained using (4.23) as follows:

$$\begin{aligned}
&\mathbf{S}(1) = \mathbf{Q}(\boldsymbol{\theta}_w) \\
&\Sigma(n) = \mathbf{C}(\boldsymbol{\theta}_s)\mathbf{S}(n)\mathbf{C}(\boldsymbol{\theta}_s)^T + \mathbf{R}(\boldsymbol{\theta}_v) \\
&\mathbf{S}(n+1) = \mathbf{A}(\boldsymbol{\theta}_s)\mathbf{S}(n)\mathbf{A}(\boldsymbol{\theta}_s)^T + \mathbf{Q}(\boldsymbol{\theta}_w)
\end{aligned} \tag{4.48}$$

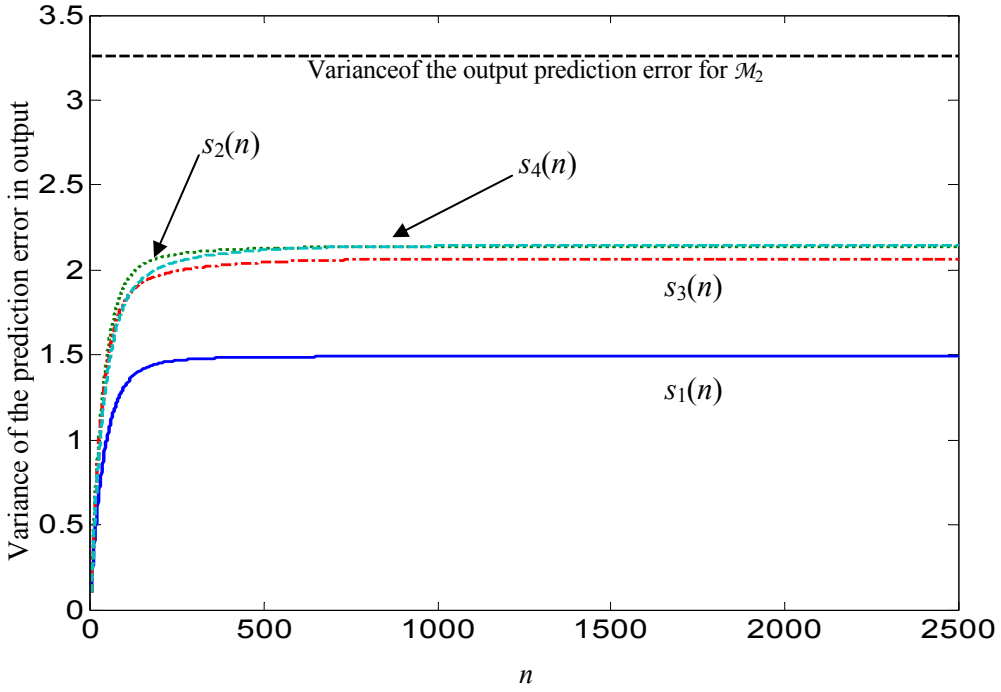

Figure 4.3: The variance of the prediction error for system output in the output equation against time instant (n) given θ=posterior mean of θ

Let $e_l(n)$ denote the prediction error for the l-th system output in the output equation at time $n\Delta t$ for $\mathcal{M}_1$; let $s_l(n)$ denote the variance of $e_l(n)$ and let $\rho_{lm}(n)$ denote the correlation coefficient between $e_l(n)$ and $e_m(n)$. From the covariance matrix $\mathbf{\Sigma}(n)$, we can obtain the variance (diagonal entries of the covariance) of the prediction errors for each system output in the output equation for each time as shown in Figure 4.3. From this figure, it can be seen that at each time, the variance of the prediction error for each system output in the output equation is not the same and they are all smaller than that for $\mathcal{M}_2$, shown as a dashed line in Figure 4.3. Figure 4.4 shows the 6 possible correlation coefficient $\rho_{lm}(n)$ at each time $n\Delta t$, n=1,…,N=2500, given $\boldsymbol{\theta}$=posterior mean of $\boldsymbol{\theta}$ for $\mathcal{M}_1$, i.e., $\rho_{12}(n)$, $\rho_{13}(n)$, $\rho_{14}(n)$, $\rho_{23}(n)$, $\rho_{24}(n)$ and $\rho_{34}(n)$. It can be seen that for $\mathcal{M}_1$, the prediction errors for the system output in the output-equation are correlated, particularly for $e_1(n)$ and $e_2(n)$, $e_2(n)$ and $e_3(n)$, $e_3(n)$ and $e_4(n)$ and $e_1(n)$ and $e_4(n)$. The correlation is higher especially between the prediction errors

for the system output in neighboring floors (i.e., between the first floor and second floor, between the second floor and third floor, between the third floor and the roof), which agrees with intuition. After a transient period, as n increases, the variance $s_l(n)$, l =1,…,4 and correlation coefficients $\rho_{12}(n)$, $\rho_{13}(n)$, $\rho_{14}(n)$, $\rho_{23}(n)$, $\rho_{24}(n)$ and $\rho_{34}(n)$ all converge. Unlike $\mathcal{M}_2$ whose prediction errors for the system output in the output equation are uncorrelated and have the same variance at all time, $\mathcal{M}_1$ allows more flexibility to accommodate the modeling errors by introducing correlation between the prediction errors for different system output in the output equation and allowing different system output prediction error variances through the structure of the stochastic system model in $\mathcal{M}_1$.

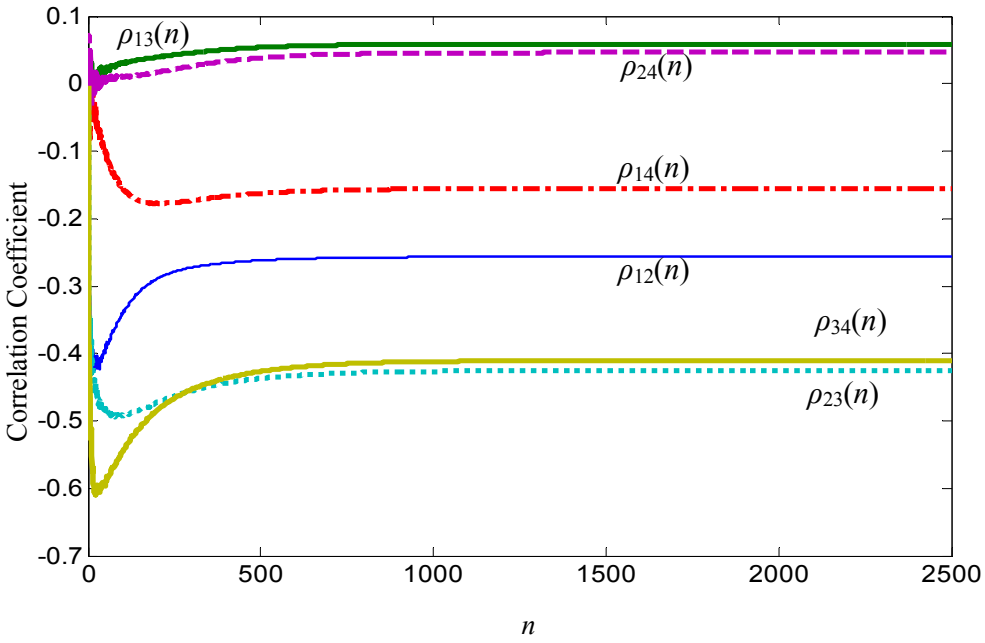

Figure 4.4: The correlation coefficient between prediction errors for different pair of system outputs in the output equation against time instant (n) given θ=posterior mean of θ for $\mathcal{M}_1$

The posterior robust failure probability of the benchmark structure subjected to future uncertain horizontal ground acceleration is calculated for $\mathcal{M}_1$ and $\mathcal{M}_2$ for different threshold levels. Here we assume the structure is subjected to a nonstationary, nonwhite horizontal

ground acceleration $\mathbf{U}_G=[u_0\ u_1\ u_2\ u_3 \ldots\ u_G]^T$ of duration of $G\Delta t$=10s with a sampling time interval of Δt=0.004s. The stochastic model U for the earthquake is given in Schueller and Pradlwalter (2007):

$$\begin{cases} \ddot{a}_g(t)+2\zeta_g\omega_g\dot{a}_g(t)+{\omega_g}^2 a_g(t)=\lambda(t)W(t) \\ \ddot{a}_f(t)+2\zeta_f\omega_f\dot{a}_f(t)+{\omega_f}^2 a_f(t)=2\zeta_g\omega_g\dot{a}_g(t)+{\omega_g}^2 a_g(t) \end{cases} \tag{4.49}$$

where $u_n=\ddot{a}_f(n\Delta t)$, n=0,1,.., G=2500; ζ_g=0.8, ω_g=15rad/s, ζ_f=0.995, ω_f=0.3rad/s; $W(t)$ is a white noise with spectral density I_g=0.02/(2π) m^2s^{-2}, i.e., the corresponding discrete white noise signal is $W(n\Delta t)=W_n=\sqrt{2\pi I_g/\Delta t}Z_n$ where Z_n is a standard Gaussian variable; the initial conditions for each of the equations in (4.49) are taken as zero. The nonstationarity of the ground acceleration is modeled through a time-envelope function $\lambda(t)$ given as follows:

$$\lambda(t)=\begin{cases} t/2, 0\le t\le 2s \\ 1, 2s\le t\le 10s \\ e^{-0.1(t-10)}, t\ge 10s \end{cases} \tag{4.50}$$

For $\mathcal{M}_1$ and $\mathcal{M}_2$, the total number of uncertain parameters involved in calculating the robust failure probability is 8x2501+4x2501+15+2501=32528 and 32526, respectively. First, we consider the case where the structure 'fails' if the maximum interstory displacement of all the stories exceeds some threshold value during a future earthquake. Thus, F can be written as follows in terms of the displacement of all the stories (the first four states in $\mathbf{x}(t)$):

$$\begin{aligned} F &= \bigcup_{n=0}^{G=2500}\bigcup_{l=1}^{4}\{|x_l(t_n)-x_{l-1}(t_n)|\ge b_l \cup |x_1(t_n)|\ge b_1\} \\ &= \max_{\substack{n\in\{0,1,\ldots,2500\}\\ l\in\{1,\ldots,4\}}}\{\frac{|x_l(t_n)-x_{l-1}(t_n)|}{b_l},\frac{|x_1(t_n)|}{b_1}\}\ge 1 \end{aligned} \tag{4.51}$$

where the threshold b_l for all the stories is the same, i.e., $b_l=b$.

Figure 4.5 shows the posterior robust failure probability of the structure for $\mathcal{M}_1$ (solid curve) and $\mathcal{M}_2$ (dashed curve) for different threshold levels. It can be seen that the posterior robust failure probability for $\mathcal{M}_1$ is quite different from that for $\mathcal{M}_2$. As the threshold level increases, the difference becomes even more pronounced.

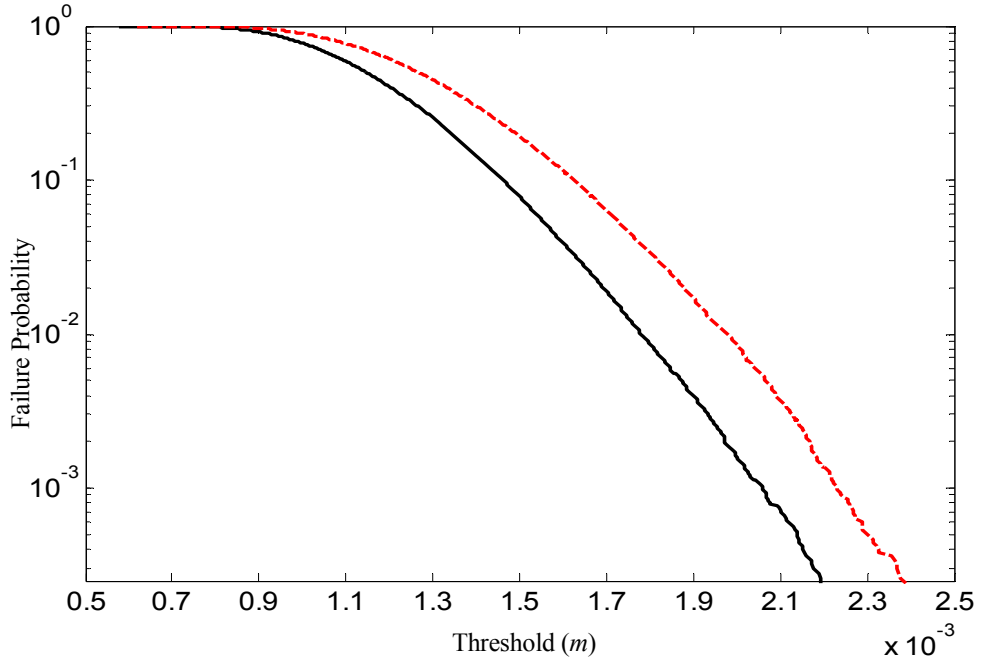

Figure 4.5: Posterior robust failure probability against the threshold of maximum interstory displacements of all floors for $\mathcal{M}_1$ (solid curve) and $\mathcal{M}_2$ (dashed curve)

Figure 4.6 shows the posterior (solid curve) robust failure probability of the structure for $\mathcal{M}_1$ and the nominal (dashed curve) structural failure probability for different threshold levels.

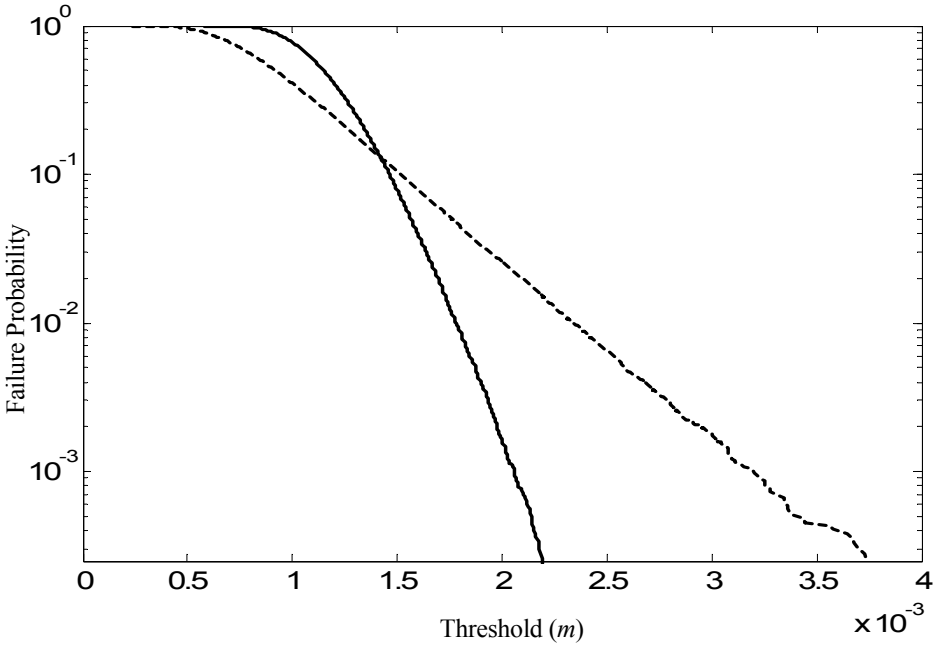

Figure 4.6: Posterior (solid curve) robust (for $\mathcal{M}_1$) and nominal (dashed) failure probability against the threshold of maximum interstory displacements of all floors

Figure 4.7 shows the prior robust failure probability of the structure for $\mathcal{M}_1$ for different threshold levels. It can be seen that the prior robust failure probality is a lot larger due to larger model uncertainties. Thus, for this model class, it is critical to collect data to reduce the model uncertainties. Figure 4.8 shows the prior robust failure probability of the structure for $\mathcal{M}_2$ (dashed curve), the posterior robust failure probability for $\mathcal{M}_2$ (solid curve), and the nominal structural failure probability (dot-dashed curve) for different threshold levels. For both model classes, it can be seen that the posterior robust failure probability is quite different from the nominal structural failure probability and the prior robust failure probability, showing the importance of using data to reduce model uncertainties and obtain more robust predictions.

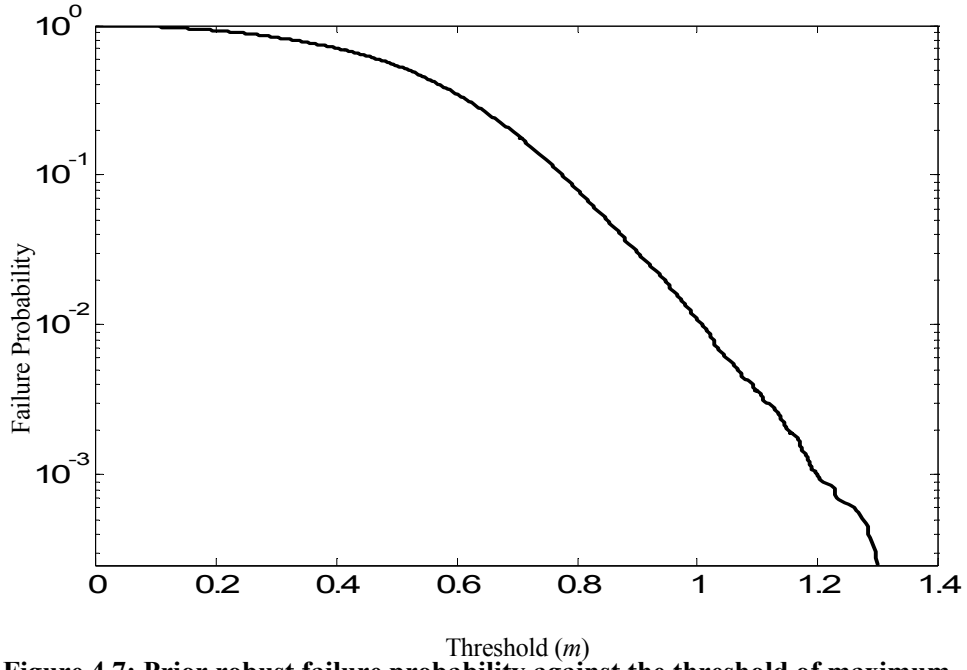

Figure 4.7: Prior robust failure probability against the threshold of maximum interstory displacements of all floors for $\mathcal{M}_1$

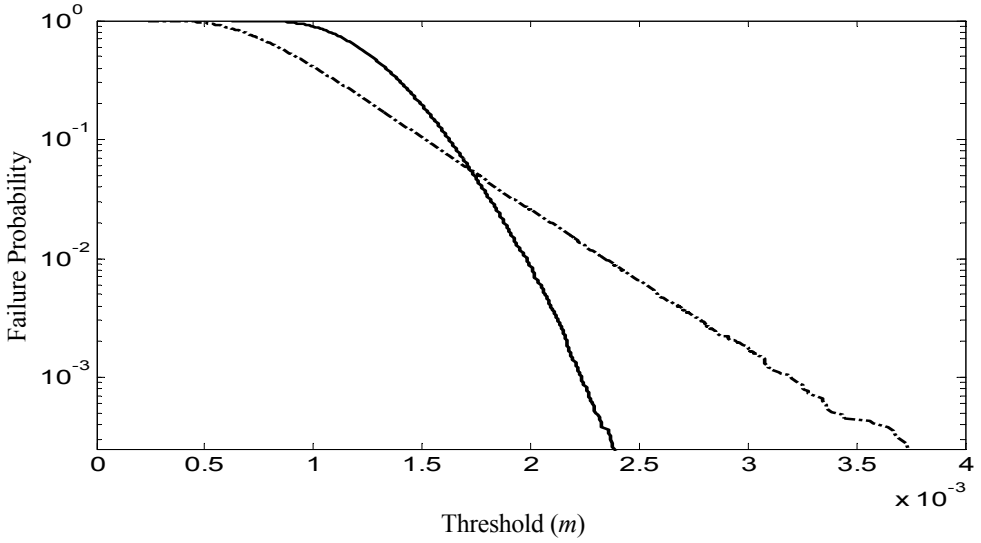

Figure 4.8: Posterior (solid curve) and prior (dashed) robust (for $\mathcal{M}_2$) and nominal (dot-dashed) failure probability against the threshold of maximum interstory displacements of all floors

Next, we consider the case where the structure 'fails' if the maximum absolute acceleration of all the stories exceeds some threshold value.

$$F = \bigcup_{n=0}^{G=2500} \bigcup_{l=1}^{4} \{ | a_l(t_n) | \geq b_l \} = \max_{\substack{n \in \{0,1,\ldots,2500\} \\ l \in \{1,\ldots,4\}}} | \frac{a_l(t_n)}{b_l} | \geq 1 \tag{4.52}$$

where the threshold b_l for all the stories is the same, i.e., $b_l=b$; $a_l(t)$ denotes the l-th story absolute acceleration at time t. Figure 4.9 shows the posterior robust failure probability (y-axis) of the structure for $\mathcal{M}_1$ (solid curve) and $\mathcal{M}_2$ (dashed curve) for different threshold levels. It can be seen that once again the posterior robust failure probability for $\mathcal{M}_1$ is significantly different from that for $\mathcal{M}_2$. As the threshold level increases, the difference becomes even more pronounced.

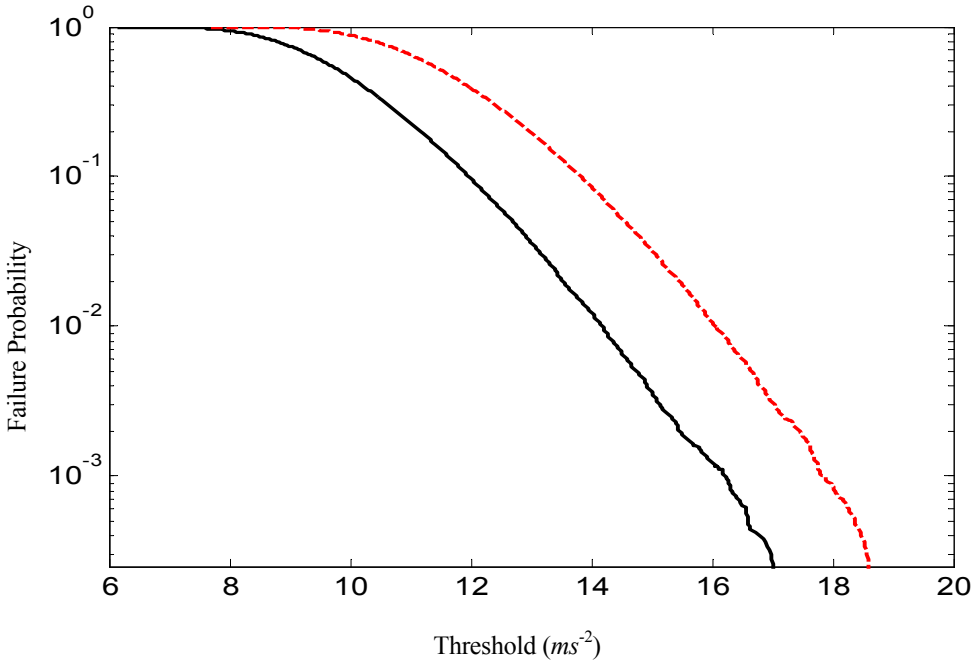

Figure 4.9: Posterior robust failure probability against the threshold of maximum absolute accelerations of all floors for $\mathcal{M}_1$ (solid curve) and $\mathcal{M}_2$ (dashed curve)

The above results show that the posterior failure probability (especially for the tail of the posterior probability distribution of the response of interest) is sensitive to the choice of the model class and hence to the way that model uncertainties are treated. One concern here is how to combine the results (quite different in this problem) obtained for different candidate model classes. The solution to this is to calculate the posterior hyper-robust failure probability using Bayesian model averaging as in (4.26), which requires calculating the posterior probability of the candidate model classes.

Table 4.2 Results for model class comparison

	$\mathcal{M}_1$	$\mathcal{M}_2$
$E[\ln(p(\mathcal{D}\|\boldsymbol{\theta},\mathcal{M}_i))]$	$-1.5762\text{x}10^4$	$-2.0251\text{x}10^4$
EIG	76.12	63.52
$\ln p(\mathcal{D}\|\mathcal{M}_i)$	$-1.5838\text{x}10^4$	$-2.0315\text{x}10^4$
$P(\mathcal{M}_i\|\mathcal{D},M)$	1.00	0.00

First we perform model class comparision. The estimates, obtained using the method presented in Chapter 3, for the log evidence $\ln p(\mathcal{D}|\mathcal{M}_j)$, the posterior mean of the log likelihood function $E[\ln(p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}_i))]$ (a data-fit measure), the expected information gain EIG (a model class complexity measure given in (1.4)) and the posterior probability $P(\mathcal{M}_i|\mathcal{D},M)$ of the model classes are shown in Table 4.2. Model class comparison shows that the proposed extended model class $\mathcal{M}_1$ is substantially more probable than $\mathcal{M}_2$ based on the data, implying that it gives the better balance between the data fit and the information gain from the data. The posterior probability for $\mathcal{M}_1$ based on the data is essentially 1. It can be seen that $\mathcal{M}_1$ has a much larger posterior mean of the log likelihood function than $\mathcal{M}_2$ which shows that $\mathcal{M}_1$ gives a much better fit to the data on average. Although $\mathcal{M}_1$ has a larger expected information gain, showing that it is more "complex" than $\mathcal{M}_2$, the difference between the EIG of $\mathcal{M}_1$ and $\mathcal{M}_2$ is relatively very small compared with the difference of the posterior mean of the log likelihood of the two model classes. Thus, the mean data-fit is dominant in making $\mathcal{M}_1$ have a larger evidence and thus be the more plausible model class.

From the results in Table 4.2 and Figures 4.5 and 4.9, it can be seen that $P(F|\mathcal{D},\mathcal{M}_2)P(\mathcal{M}_2|\mathcal{D},M)$ is negligible and so the contribution of $\mathcal{M}_2$ can be dropped when calculating the posterior hyper-robust failure probability of the structure. Also, having a posterior probability $P(\mathcal{M}_2|\mathcal{D},M)$ that is much smaller than for $\mathcal{M}_1$ implies $\mathcal{M}_2$ is relatively improbable conditioned on the data $\mathcal{D}$ and so $\mathcal{M}_2$ may be dropped when making robust prediction of any response of the structure.

Appendix 4A

$$\begin{aligned}\boldsymbol{\Sigma}(n) &= \mathbf{C}_n(\boldsymbol{\theta}_\mathbf{s})\{\sum_{j=1}^{n-1}[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]^T + \mathbf{Q}_n(\boldsymbol{\theta}_w)\}\mathbf{C}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{R}_n(\boldsymbol{\theta}_v)\\ &= \mathbf{C}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{S}(n)\mathbf{C}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{R}_n(\boldsymbol{\theta}_v)\end{aligned}$$

where $\mathbf{S}(n) = \sum_{j=1}^{n-1}[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]^T + \mathbf{Q}_n(\boldsymbol{\theta}_w)$

$$\begin{aligned}\mathbf{S}(n+1) &= \sum_{j=1}^{n}[\prod_{i=1}^{n+1-j}\mathbf{A}_{n+1-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n+1-j}\mathbf{A}_{n+1-i}(\boldsymbol{\theta}_\mathbf{s})]^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)\\ &= \sum_{j=1}^{n-1}[\prod_{i=1}^{n+1-j}\mathbf{A}_{n+1-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n+1-j}\mathbf{A}_{n+1-i}(\boldsymbol{\theta}_\mathbf{s})]^T + \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{Q}_n(\boldsymbol{\theta}_w)\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)\\ &= \sum_{j=1}^{n-1}[\prod_{i=0}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=0}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]^T + \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{Q}_n(\boldsymbol{\theta}_w)\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)\\ &= \sum_{j=1}^{n-1}\{\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]^T\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T\}\\ &\quad + \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{Q}_n(\boldsymbol{\theta}_w)\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)\\ &= \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\sum_{j=1}^{n-1}\{[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]^T\}\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T\\ &\quad + \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{Q}_n(\boldsymbol{\theta}_w)\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)\\ &= \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\{\sum_{j=1}^{n-1}\{[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]\mathbf{Q}_j(\boldsymbol{\theta}_w)[\prod_{i=1}^{n-j}\mathbf{A}_{n-i}(\boldsymbol{\theta}_\mathbf{s})]^T\} + \mathbf{Q}_n(\boldsymbol{\theta}_w)\}\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)\\ &= \mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})\mathbf{S}(n)\mathbf{A}_n(\boldsymbol{\theta}_\mathbf{s})^T + \mathbf{Q}_{n+1}(\boldsymbol{\theta}_w)\end{aligned}$$

CHAPTER 5

New Bayesian updating methodology for model validation and robust predictions of a target system based on hierarchical subsystem tests

In this chapter, the problem of model validation for a system is considered. Superficially, the problem of how to validate a model seems solvable but it is still not settled; indeed, it is clear that a model that has given good predictions in tests so far might perform poorly under different circumstances, such as an excitation with different characteristics.

The material in this chapter is based on Cheung and Beck (2008b, g). Our philosophy when predicting the behavior of a system of interest is that one should develop candidate sets of probabilistic predictive input-output models to give robust predictions that explicitly address errors due to imperfect models and uncertainties due to incomplete information. For model validation, it is then desirable to check based on system test data whether any of the proposed candidate model sets are highly probable and whether they provide high quality predictions of the system behavior of interest.

Sometimes the full system cannot be readily tested because it is too expensive or too large, or due to other limitations, but some of its subsystems may be tested. Here we introduce the concept of hierarchical stochastic system model classes and then propose a Bayesian methodology using them to treat modeling and input uncertainties in model validation,

uncertainty propagation and robust predictions of the response of the full system. The Sandia static-frame validation problem is used to illustrate the proposed methodology. The results of other researchers' studies of this problem are presented in a special issue of the journal Computer Methods in Applied Mechanics and Engineering (Chleboun 2008; Babuška et al. 2008; Grigoriu and Field 2008; Pradlwarter and Schuëller 2008; Rebba and Cafeo 2008).

5.1 Hierarchical stochastic system model classes and model validation

In this section, a novel model validation methodology based on a new concept of hierarchical stochastic system model classes is proposed (building on the theoretical foundations presented in previous chapters) so that a rational decision can be made regarding which proposed model classes should be used for predicting the response of a target system. The proposed methodology is based on using full Bayesian updating to investigate multiple important aspects of the performance of the candidate model classes, including their quality of prediction, their posterior probabilities and their contribution to response predictions of the final system. We do not make a binary reject/accept step but instead provide the decision maker with information about these important aspects, which can be combined with other considerations when making a decision related to the target system; for example, should the current target system design be accepted or modified?

Suppose during construction of the system, a series of I experiments are conducted where data $\mathcal{D}_i$, i=1,…, I, are collected from each of I similarly complex, or successively more complex, subsystems and these data are to be used to predict the response of the more complex target system. The i-th level subsystem is either a standalone subsystem (especially in lower levels) or one comprised of a combination of some (or all) tested subsystems from the previous levels, together, possibly, with new untested subsystems.

5.1.1 Analysis and full Bayesian updating of i-th subsystem

The presentation in this subsection is very general and the reader may find it helpful to look at the example illustrating the hierarchical concepts in the last subsection of this section. We assume that a set $M_i = \{\mathcal{M}_j^{(i)}: j=1,2,\ldots N_i\}$ of model classes is proposed for the i-th subsystem which are either newly defined or built-up by extending the model classes for some (or all) tested subsystems in the previous levels. In the latter case, a model class for the i-th subsystem is built-up by extending at most one model class for each relevant lower-level subsystem since candidate model classes for each such subsystem are supposed to be competing. Denote uncertain model parameters for the model class $\mathcal{M}_j^{(i)}$ by $\boldsymbol{\theta}^{(i,j)}=[\boldsymbol{\varphi}^{(i,j)}, \boldsymbol{\xi}^{(i,j)}]$ where $\boldsymbol{\varphi}^{(i,j)}$, if any, are the new uncertain model parameters and $\boldsymbol{\xi}^{(i,j)}$, if any, are the uncertain model parameters corresponding to a model class for some subsystems in the previous levels, that is, these parameters of $\mathcal{M}_j^{(i)}$ are also in model classes of subsystems of the ith subsystem. In the proposed hierarchical approach, the model class $\mathcal{M}_j^{(i)}$ is based on the "prior" (prior to the i^{th} subsystem test but posterior to all previous tests):

$$p(\boldsymbol{\theta}^{(i,j)} \mid \mathcal{D}_1,...,\mathcal{D}_{i-1},\mathcal{M}_j^{(i)}) = p(\boldsymbol{\varphi}^{(i,j)} \mid \mathcal{M}_j^{(i)})\, p(\boldsymbol{\xi}^{(i,j)} \mid \mathcal{D}_1,...,\mathcal{D}_{i-1}) \tag{5.1}$$

where $p(\boldsymbol{\varphi}^{(i,j)}|\mathcal{M}_j^{(i)})$ quantifies the prior uncertainties in the new parameters $\boldsymbol{\varphi}^{(i,j)}$ in model class $\mathcal{M}_j^{(i)}$ and $p(\boldsymbol{\xi}^{(i,j)}|\mathcal{D}_1,\ldots, \mathcal{D}_{i-1})$ is the most updated PDF of $\boldsymbol{\xi}^{(i,j)}$ given data collected from all subsystems in the previous levels. For simplicity, the conditioning of $p(\boldsymbol{\xi}^{(i,j)}|\mathcal{D}_1,\ldots, \mathcal{D}_{i-1})$ on the model classes previously considered which contain components of $\boldsymbol{\xi}^{(i,j)}$ are left implicit. For $i=1$, $p(\boldsymbol{\theta}^{(i,j)}|\mathcal{D}_1,\ldots, \mathcal{D}_{i-1},\mathcal{M}_j^{(i)}) = p(\boldsymbol{\theta}^{(1,j)}|\mathcal{M}_j^{(1)})$.

At the end of the experiments on the i-th subsystem where data $\mathcal{D}_i$ are collected, the following procedure is used to check the prediction quality of each candidate model class being considered for the i-th subsystem. For each model class $\mathcal{M}_j^{(i)}$ in M_i and for each measured quantity in $\mathcal{D}_i$, the *consistency* of the predicted response is first investigated by calculating the difference of the measured quantity in $\mathcal{D}_i$ and the mean of the corresponding

prior robust predicted response. The robust predicted response given by $\mathcal{M}_j^{(i)}$ is consistent if this difference is no more than a certain number of standard deviations (e.g., no more than 2 to 3 standard deviations). An alternative way of investigating the consistency is to check whether each measured quantity in $\mathcal{D}_i$ is within q percentile and (100-q) percentile of the robust predicted response (e.g., q can be 1). The mean and standard deviation of the prior robust predicted response can be calculated using (1.5) and (1.6) but with samples drawn from the prior in (5.1).

Next, the *accuracy* of the prediction is investigated by calculating the probability that the prior robust predicted response using $\mathcal{M}_j^{(i)}$ (again based on $p(\boldsymbol{\theta}^{(i,j)}|\mathcal{D}_1,\ldots,\mathcal{D}_{i-1},\mathcal{M}_j^{(i)})$ in (5.1)) is within a certain b% (e.g. 10%) of the measured quantity using (1.5) and (1.6). This probability is related to the prediction error of each model class for the i-th level subsystem and reflects the predictability of these models before being updated using data $\mathcal{D}_i$. Note that a model class may give consistent predictions but not accurate ones because, for example, it has a relatively large standard deviation.

Next, for each model class $\mathcal{M}_j^{(i)}$ in M_i, the uncertainties in the model parameters $\boldsymbol{\theta}^{(i,j)}$ are updated using all the available data, as quantified by $p(\boldsymbol{\theta}^{(i,j)}|\mathcal{D}_1,\ldots,\mathcal{D}_i,\mathcal{M}_j^{(i)})$ through Bayes' Theorem:

$$p(\boldsymbol{\theta}^{(i,j)} \mid \mathcal{D}_1,...,\mathcal{D}_i,\mathcal{M}_j^{(i)}) = c_{i,j}^{-1}\, p(\mathcal{D}_i \mid \boldsymbol{\theta}^{(i,j)},\mathcal{M}_j^{(i)})\, p(\boldsymbol{\theta}^{(i,j)} \mid \mathcal{D}_1,...,\mathcal{D}_{i-1},\mathcal{M}_j^{(i)}) \qquad (5.2)$$

where the data $\mathcal{D}_1,\ldots,\mathcal{D}_{i-1}$ are modeled as irrelevant to the probability of getting $\mathcal{D}_i$ when $\boldsymbol{\theta}^{(i,j)}$ is given since this parameter vector defines the predictive probability model for the model class $\mathcal{M}_j^{(i)}$. Recall that $\boldsymbol{\xi}^{(i,j)}$ are the uncertain model parameters corresponding to some model classes of subsystems already considered in the previous levels. A subtle point to be noted is that sometimes uncertainties for some other model parameters $\boldsymbol{\Phi}^{(i,j)}$ corresponding to the model classes containing components of $\boldsymbol{\xi}^{(i,j)}$ will also be updated when updating uncertainties in $\boldsymbol{\xi}^{(i,j)}$ using $\mathcal{D}_1,\ldots,\mathcal{D}_{i-1}$. Since $\boldsymbol{\Phi}^{(i,j)}$ and $\boldsymbol{\xi}^{(i,j)}$ are not

stochastically independent given $\mathcal{D}_1,\ldots,\mathcal{D}_i$, the uncertainties in both $\boldsymbol{\theta}^{(i,j)}$ and $\boldsymbol{\Phi}^{(i,j)}$ need to be updated together from Bayes' Theorem:

$$\begin{aligned} &p(\boldsymbol{\theta}^{(i,j)},\Phi^{(i,j)} \mid \mathcal{D}_1,...,\mathcal{D}_i,\mathcal{M}_j^{(i)}) \\ &= \tilde{c}_{i,j}^{-1}\, p(\mathcal{D}_i \mid \boldsymbol{\theta}^{(i,j)},\mathcal{M}_j^{(i)})\, p(\boldsymbol{\xi}^{(i,j)},\boldsymbol{\Phi}^{(i,j)} \mid \mathcal{D}_1,...,\mathcal{D}_{i-1})\, p(\boldsymbol{\varphi}^{(i,j)} \mid \mathcal{M}_j^{(i)}) \end{aligned} \tag{5.3}$$

where $\boldsymbol{\theta}^{(i,j)}=[\boldsymbol{\varphi}^{(i,j)}, \boldsymbol{\xi}^{(i,j)}]$ and the data $\mathcal{D}_1,\ldots,\mathcal{D}_{i-1}$ are modeled as irrelevant to the probability of getting $\mathcal{D}_i$ given $\boldsymbol{\theta}^{(i,j)}$, as before. Finally, $p(\boldsymbol{\theta}^{(i,j)}|\mathcal{D}_1,\ldots,\mathcal{D}_i,\mathcal{M}_j^{(i)})$ can be obtained as the marginal PDF of $p(\boldsymbol{\theta}^{(i,j)},\boldsymbol{\Phi}^{(i,j)}|\mathcal{D}_1,\ldots,\mathcal{D}_i,\mathcal{M}_j^{(i)})$.

The posterior probability $P(\mathcal{M}_j^{(i)}|\mathcal{D}_1,\ldots,\mathcal{D}_i,M_i)$ of each model class in M_i can be calculated as follows to evaluate the relative plausibility of each model class. If a model class $\mathcal{M}_j^{(i)}$ is built-up by extending or using model classes which have been updated using data from subsystems in the previous levels $k_1, k_2,\ldots, k_m$ where $k_1< k_2<\ldots< k_m$ and $1\le m<i$, $P(\mathcal{M}_j^{(i)}|\mathcal{D}_1,\ldots,\mathcal{D}_i,M_i)$ is equal to $P(\mathcal{M}_j^{(i)}|\mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{D}_i,M_i)$. The most up-to-date evidence $p(\mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{D}_i|\mathcal{M}_j^{(i)})$ for $\mathcal{M}_j^{(i)}$ that is provided by the data $\mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{D}_i$, and which is required for calculating $P(\mathcal{M}_j^{(i)}|\mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{D}_i,M_i)$, is given by:

$$p(\mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{D}_i \mid \mathcal{M}_j^{(i)}) = p(\mathcal{D}_{k_1},...,\mathcal{D}_{k_m} \mid \mathcal{M}_j^{(i)})\, p(\mathcal{D}_i \mid \mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{M}_j^{(i)}) \tag{5.4}$$

In this equation, $p(\mathcal{D}_i|\mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{M}_j^{(i)})$ is given by:

$$p(\mathcal{D}_i \mid \mathcal{D}_{k_1},...,\mathcal{D}_{k_m},\mathcal{M}_j^{(i)}) = \int p(\mathcal{D}_i \mid \boldsymbol{\theta}^{(i,j)},\mathcal{M}_j^{(i)})\, p(\boldsymbol{\theta}^{(i,j)} \mid \mathcal{D}_1,...,\mathcal{D}_{i-1},\mathcal{M}_j^{(i)})\, d\boldsymbol{\theta}^{(i,j)} \tag{5.5}$$

which can be determined using a stochastic simulation method, such as the Hybrid Gibbs TMCMC method presented in Appendix 5A. The other factor in (5.4), $p(\mathcal{D}_{k_1},...,\mathcal{D}_{k_m} \mid \mathcal{M}_j^{(i)})$, is given by a product of the evidences which have already been

determined at the end of previous experiments. This point will be more clear in the example illustrating the hierarchical concepts in the last subsection of this section or one can refer to Cheung and Beck (2008b) for more details. Based on (5.4), $P(\mathcal{M}_j^{(i)}|\mathcal{D}_1,\ldots,\mathcal{D}_i, M_i)$ $= P(\mathcal{M}_j^{(i)}|\mathcal{D}_{k_1},\ldots,\mathcal{D}_{k_m}, \mathcal{D}_i, M_i)$ can be calculated using (1.2) with $\mathcal{M}_j$ replaced by $\mathcal{M}_j^{(i)}$, M replaced by M_i and $\mathcal{D}$ by $\mathcal{D}_{k_1},\ldots,\mathcal{D}_{k_m}, \mathcal{D}_i$.

In the special case that $\mathcal{M}_j^{(i)}$ is newly defined, i.e., not built-up by extending any model classes for subsystems in the previous levels, the posterior probability $P(\mathcal{M}_j^{(i)}|\mathcal{D}_1,\ldots, \mathcal{D}_i, M_i)$ is given by $P(\mathcal{M}_j^{(i)}|\mathcal{D}_i, M_i)$, which can be calculated using (1.2) with $\mathcal{M}_j$ replaced by $\mathcal{M}_j^{(i)}$, M replaced by M_i and $\mathcal{D}$ by $\mathcal{D}_i$ where the evidence $p(\mathcal{D}_i|\mathcal{M}_j^{(i)})$ for $\mathcal{M}_j^{(i)}$ is given by:

$$p(\mathcal{D}_i \mid \mathcal{M}_j^{(i)}) = \int p(\mathcal{D}_i \mid \boldsymbol{\theta}^{(i,j)}, \mathcal{M}_j^{(i)}) p(\boldsymbol{\theta}^{(i,j)} \mid \mathcal{M}_j^{(i)}) d\boldsymbol{\theta}^{(i,j)} \tag{5.6}$$

which can be determined using a stochastic simulation method.

Based on all the data, $\mathcal{D}_1,\ldots, \mathcal{D}_i$, so far, the posterior robust prediction of the response vector $\mathbf{X}$ for the target system can be calculated using (1.5) and (1.7). If a model class $\mathcal{M}_j^{(i)}$ is very improbable compared to the others in M_i, so that its contribution to the hyper-robust response prediction of the target system is negligible in (1.7), it can be neglected when building the candidate model classes for higher level subsystems in order to save computations. Note that (1.7) allows calculation of the most robust predictions for the *i*-th subsystem based on all the available information and viable model classes.

For each model class $\mathcal{M}_j^{(i)}$ in M_i and for each measured quantity in $\mathcal{D}_i$, the consistency of the predicted response is again investigated by examining the difference of the measured quantity in $\mathcal{D}_i$ and the mean of the corresponding posterior robust predicted response (again judged in terms of the number of standard deviations of the posterior robust predicted response). The robust predicted response is based on the "posterior" $p(\boldsymbol{\theta}^{(i,j)}|\mathcal{D}_1,\ldots, \mathcal{D}_i, \mathcal{M}_j^{(i)})$

given by (5.2) or (5.3) and its mean and standard deviation are calculated using (1.5) and (1.6). One can also check whether each measured quantity in $\mathcal{D}_i$ is within q percentile and (100-q) percentile of the posterior robust predicted response. Next, the accuracy of the prediction is investigated by calculating the probability that the robust predicted response (again based on $p(\boldsymbol{\theta}^{(i,j)}|\mathcal{D}_1,\ldots,\mathcal{D}_i,\mathcal{M}_j^{(i)})$) is within a certain b% (e.g. 10%) of the measured quantity using (1.5) and (1.6).

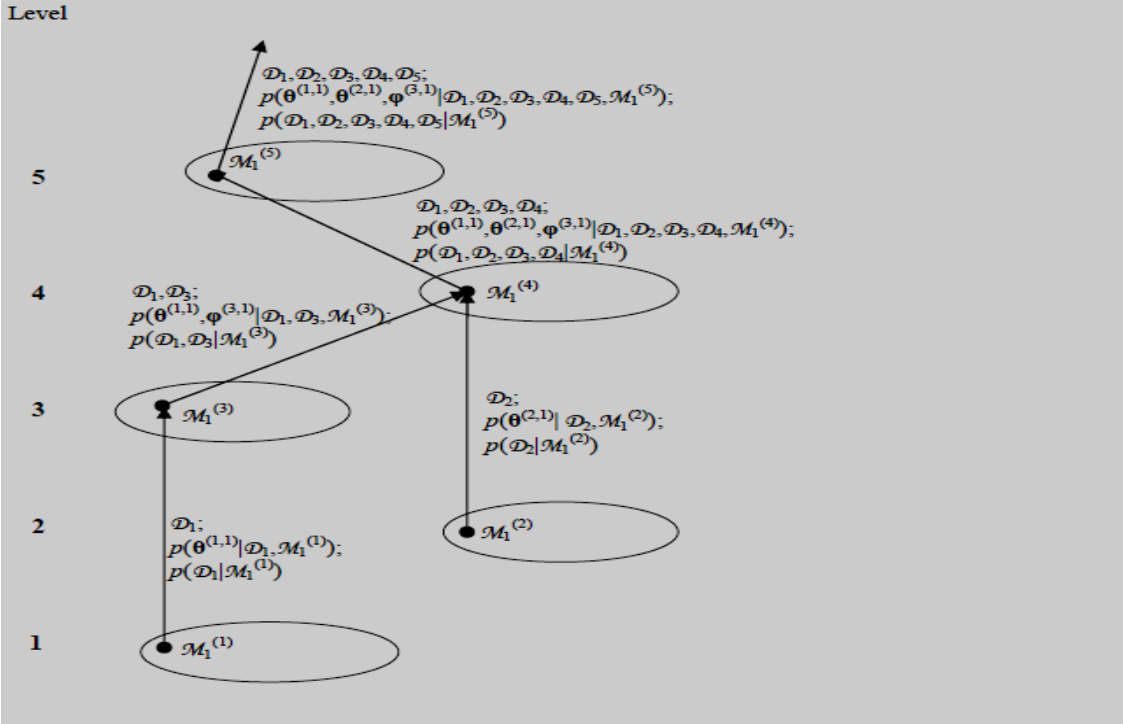

Figure 5.1: Schematic plot for an illustrative example of hierarchical model classes

5.1.2 Example to illustrate hierarchical model classes

The following example is presented to illustrate the above theory on how to propagate uncertainties in parameters and calculate the posterior probability for a hierarchical stochastic system model class. Figure 5.1 shows the hierarchical structure of some of the model classes for the illustrative example. The ellipses show the subsystems for different

levels; a black dot inside an ellipse shows a candidate model class corresponding to that subsystem; the lower end of an arrow points to a model class which is used to build another model class pointed to by the top end of the same arrow. Shown next to an arrow is the set of data used to update the lower level model classes, along with the posterior PDF for the previous model class and the evidence required for calculating the posterior probability of this model class.

Recall that $\mathcal{M}_1^{(1)}$ in M_1 is the first candidate model class with uncertain parameters $\boldsymbol{\theta}^{(1,1)}$ for the first level subsystem from which data $\mathcal{D}_1$ is collected. The posterior PDF $p(\boldsymbol{\theta}^{(1,1)}|\mathcal{D}_1, \mathcal{M}_1^{(1)})$ for $\mathcal{M}_1^{(1)}$ is given by (5.2) with the chosen prior PDF $p(\boldsymbol{\theta}^{(1,1)}|\, \mathcal{M}_1^{(1)})$. The evidence $p(\mathcal{D}_1|\mathcal{M}_1^{(1)})$, which is required for calculating the posterior probability $P(\mathcal{M}_1^{(1)}|\mathcal{D}_1, M_1)$ for $\mathcal{M}_1^{(1)}$, is given by (5.6) with i=1 and j=1.

Suppose that $\mathcal{D}_2$ is collected from a second level subsystem that is independent of the first level subsystem and $\mathcal{M}_1^{(2)}$ in M_2 is a newly defined candidate model class with new uncertain parameters $\boldsymbol{\theta}^{(2,1)}$. The posterior PDF $p(\boldsymbol{\theta}^{(2,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{M}_1^{(2)})=p(\boldsymbol{\theta}^{(2,1)}|\mathcal{D}_2, \mathcal{M}_1^{(2)})$ for $\mathcal{M}_1^{(2)}$ is given by (5.2) with the chosen prior PDF $p(\boldsymbol{\theta}^{(2,1)}|\, \mathcal{M}_1^{(2)})$. The evidence $p(\mathcal{D}_2|\mathcal{M}_1^{(2)})$, which is required for calculating the posterior probability $P(\mathcal{M}_1^{(2)}|\, \mathcal{D}_1,\mathcal{D}_2, M_2) = P(\mathcal{M}_1^{(2)}|\mathcal{D}_2, M_2)$ for $\mathcal{M}_1^{(2)}$, is given by (5.6) with i=2 and j=1.

Suppose that the third level subsystem contains the first level subsystem but not the second level subsystem. Assume that the first candidate model class $\mathcal{M}_1^{(3)}$ in M_3, with uncertain parameters $\boldsymbol{\theta}^{(3,1)}$ for the third level subsystem from which $\mathcal{D}_3$ is collected, is built-up by extending the model class $\mathcal{M}_1^{(1)}$ (i.e., existing parameters $\boldsymbol{\xi}^{(3,1)} = \boldsymbol{\theta}^{(1,1)}$) and $\boldsymbol{\varphi}^{(3,1)}$ are the new uncertain model parameters, so $\boldsymbol{\theta}^{(3,1)} = [\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\varphi}^{(3,1)}]$. The posterior PDF $p(\boldsymbol{\theta}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(3)})$ for $\mathcal{M}_1^{(3)}$ is given by (5.2) with the prior PDF $p(\boldsymbol{\theta}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{M}_1^{(3)})=p(\boldsymbol{\theta}^{(1,1)}|\mathcal{D}_1, \mathcal{M}_1^{(1)})\, p(\boldsymbol{\varphi}^{(3,1)}|\mathcal{M}_1^{(3)})$ and so this posterior is independent of $\mathcal{D}_2$, as expected. The evidence $p(\mathcal{D}_1,\mathcal{D}_3|\mathcal{M}_1^{(3)})$, which is required for calculating the posterior probability $P(\mathcal{M}_1^{(3)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3,$

M_3) = $P(\mathcal{M}_1^{(3)}|\mathcal{D}_1, \mathcal{D}_3, M_3)$ for $\mathcal{M}_1^{(3)}$, is equal to $p(\mathcal{D}_1|\mathcal{M}_1^{(3)})\ p(\mathcal{D}_3|\mathcal{D}_1,\mathcal{M}_1^{(3)})$ by (5.4) where $p(\mathcal{D}_3|\mathcal{D}_1,\mathcal{M}_1^{(3)})$ is given by (5.5) which becomes here:

$$p(\mathcal{D}_3 \mid \mathcal{D}_1, \mathcal{M}_1^{(3)}) = \int p(\mathcal{D}_3 \mid \boldsymbol{\theta}^{(3,1)}, \mathcal{M}_1^{(3)})\, p(\boldsymbol{\theta}^{(3,1)} \mid \mathcal{D}_1, \mathcal{D}_2, \mathcal{M}_1^{(3)})\, d\boldsymbol{\theta}^{(3,1)} \tag{5.7}$$

and $p(\mathcal{D}_1|\mathcal{M}_1^{(3)}) = p(\mathcal{D}_1|\mathcal{M}_1^{(1)})$, since 1) $\mathcal{M}_1^{(3)}$ is built-up by extending $\mathcal{M}_1^{(1)}$; 2) prior to the collection of $\mathcal{D}_3$, $\mathcal{D}_1$ is used to update $\mathcal{M}_1^{(1)}$. Recall that $p(\mathcal{D}_1|\mathcal{M}_1^{(1)})$ has already been determined.

Suppose that the fourth level subsystem is a combination of the first and second level subsystems but not the third one. Assume that the first candidate model class $\mathcal{M}_1^{(4)}$ in M_4, with uncertain parameters $\boldsymbol{\theta}^{(4,1)}$ for the fourth level subsystem from which $\mathcal{D}_4$ is collected, is built-up by using the model classes $\mathcal{M}_1^{(3)}$ and $\mathcal{M}_1^{(2)}$ (i.e., $\boldsymbol{\xi}^{(4,1)} = [\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)}]$) and there are no new uncertain model parameters. Thus $\boldsymbol{\theta}^{(4,1)} = \boldsymbol{\xi}^{(4,1)} = [\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)}]$ and $\boldsymbol{\Phi}^{(4,1)} = \boldsymbol{\varphi}^{(3,1)}$ since when updating $\mathcal{M}_1^{(3)}$, $\boldsymbol{\varphi}^{(3,1)}$ and $\boldsymbol{\theta}^{(1,1)}$ are both updated and $\mathcal{D}_1$ and $\mathcal{D}_3$ are used to update both of them. The posterior PDF $p(\boldsymbol{\theta}^{(4,1)}, \boldsymbol{\Phi}^{(4,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(4)}) = p(\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)}, \boldsymbol{\varphi}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(4)})$ for $\mathcal{M}_1^{(4)}$ is given by (5.3) with the prior PDF $p(\boldsymbol{\theta}^{(4,1)}, \boldsymbol{\Phi}^{(4,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(4)}) = p(\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\varphi}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_3, \mathcal{M}_1^{(3)})\ p(\boldsymbol{\theta}^{(2,1)}|\mathcal{D}_2,\mathcal{M}_1^{(2)})$. The evidence $p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4|\mathcal{M}_1^{(4)})$, which is required for calculating the posterior probability $P(\mathcal{M}_1^{(4)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, M_4)$ for $\mathcal{M}_1^{(4)}$, is equal to $p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3|\mathcal{M}_1^{(4)})\ p(\mathcal{D}_4|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(4)})$ by (5.4) where $p(\mathcal{D}_4|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(4)})$ is given by (5.5) which becomes here:

$$p(\mathcal{D}_4 \mid \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(4)}) = \int p(\mathcal{D}_4 \mid \boldsymbol{\theta}^{(4,1)}, \mathcal{M}_1^{(4)})\, p(\boldsymbol{\theta}^{(4,1)} \mid \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(4)})\, d\boldsymbol{\theta}^{(4,1)} \tag{5.8}$$

where $p(\boldsymbol{\theta}^{(4,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(4)}) = p(\boldsymbol{\theta}^{(1,1)}|\mathcal{D}_1, \mathcal{D}_3, \mathcal{M}_1^{(3)})\ p(\boldsymbol{\theta}^{(2,1)}|\mathcal{D}_2,\mathcal{M}_1^{(2)})$ and $p(\boldsymbol{\theta}^{(1,1)}|\mathcal{D}_1, \mathcal{D}_3, \mathcal{M}_1^{(3)})$ is the marginal PDF of the posterior PDF $p(\boldsymbol{\theta}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{M}_1^{(3)})$ for $\mathcal{M}_1^{(3)}$ while $p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3|\mathcal{M}_1^{(4)}) = p(\mathcal{D}_1, \mathcal{D}_3|\mathcal{M}_1^{(3)})p(\mathcal{D}_2|\mathcal{M}_1^{(2)})$, since 1) $\mathcal{M}_1^{(4)}$ is built-up by using $\mathcal{M}_1^{(3)}$ and

$\mathcal{M}_1^{(2)}$; 2) prior to the collection of $\mathcal{D}_4$, $\mathcal{D}_1$ and $\mathcal{D}_3$ are used to update $\mathcal{M}_1^{(3)}$ and $\mathcal{D}_2$ is used to update $\mathcal{M}_1^{(2)}$. Recall that $p(\mathcal{D}_1, \mathcal{D}_3|\mathcal{M}_1^{(3)})$ and $p(\mathcal{D}_2|\mathcal{M}_1^{(2)})$ have already been determined.

Suppose that the fifth level subsystem contains third and fourth level subsystems. Assume that the first candidate model class $\mathcal{M}_1^{(5)}$ in M_5, with uncertain parameters $\boldsymbol{\theta}^{(5,1)}$ for the fifth level subsystem from which $\mathcal{D}_5$ is collected, is built-up by using the model class $\mathcal{M}_1^{(4)}$ with no new uncertain model parameters. Thus, $\boldsymbol{\theta}^{(5,1)} = \boldsymbol{\xi}^{(5,1)} = [\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)}, \boldsymbol{\varphi}^{(3,1)}]$ since when updating $\mathcal{M}_1^{(4)}$, $\boldsymbol{\theta}^{(1,1)}$, $\boldsymbol{\theta}^{(2,1)}$ and $\boldsymbol{\varphi}^{(3,1)}$ are updated and $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and $\mathcal{D}_4$ are used to update them. The posterior PDF $p(\boldsymbol{\theta}^{(5,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5, \mathcal{M}_1^{(5)}) = p(\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)}, \boldsymbol{\varphi}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5, \mathcal{M}_1^{(5)})$ for $\mathcal{M}_1^{(5)}$ is given by (5.2) with the prior PDF $p(\boldsymbol{\theta}^{(5,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(5)}) = p(\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)}, \boldsymbol{\varphi}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(4)})$. The evidence $p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5|\mathcal{M}_1^{(5)})$, which is required for calculating the posterior model probability $P(\mathcal{M}_1^{(5)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{D}_5, M_5)$ for $\mathcal{M}_1^{(5)}$, is equal to $p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4|\mathcal{M}_1^{(5)})\, p(\mathcal{D}_5|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(5)})$ by (5.4) where $p(\mathcal{D}_5|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(5)})$ is given by (5.5) which becomes here:

$$p(\mathcal{D}_5 \mid \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(5)}) = \int p(\mathcal{D}_5 \mid \boldsymbol{\theta}^{(5,1)}, \mathcal{M}_1^{(5)})\, p(\boldsymbol{\theta}^{(5,1)} \mid \mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(5)})\, d\boldsymbol{\theta}^{(5,1)} \quad (5.9)$$

where $p(\boldsymbol{\theta}^{(5,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(5)}) = p(\boldsymbol{\theta}^{(1,1)}, \boldsymbol{\theta}^{(2,1)}, \boldsymbol{\varphi}^{(3,1)}|\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4, \mathcal{M}_1^{(4)})$ while $p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4|\mathcal{M}_1^{(5)}) = p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4|\mathcal{M}_1^{(4)})$, since 1) $\mathcal{M}_1^{(5)}$ is built-up by using $\mathcal{M}_1^{(4)}$; 2) prior to the collection of $\mathcal{D}_5$, $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ and $\mathcal{D}_4$ are used to update $\mathcal{M}_1^{(4)}$. Recall that $p(\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \mathcal{D}_4|\mathcal{M}_1^{(4)})$ has already been determined.

5. 2 Illustrative example based on a validation challenge problem

For illustration, the static-frame validation challenge problem (Babuška et al. 2008) is considered. It is one of the problems presented at the Validation Challenge Workshop at Sandia National Laboratory on May 27-29, 2006. The purpose of this particular challenge problem is to predict the probability of the event F (*regulatory assessment*): $|w_p| \geq 3$mm, where w_p is the vertical displacement of the midpoint P of beam 4 of the frame structure

(our target system) shown in Figure 1 of Babuška et al (2008) and Figure 1 in Cheung and Beck (2008b). The structure is subjected to a uniform load q = 6kN/m on beam 4. Information regarding the geometry of the frame structure is shown in Table 1 of Babuška et al (2008) and in Tables 1 and 2 in Cheung and Beck (2008b). Also, in the definition of the challenge problem, the structure is given to be linear elastic with a one-dimensional tension model for each of the rods and a one-dimensional Bernoulli beam model for the bending of the beam. The coupling of bending and compression is given to be negligible for beam 4. It is given that all the bars are made of the same inhomogeneous material but come from independent sources and so can have variable material properties; in fact, the only uncertainty considered in this challenge problem is Young's modulus E (or compliance S=1/E) along each of the bars. Given Young's modulus variation along each of the bars, w_p can be predicted using the equations in Babuška et al (2008) and in Appendix I in Cheung and Beck (2008b).

The simulated experiments are set up to resemble a typical situation in which data are collected from a hierarchy of successively more complex subsystems that become "closer" to the final system and the amount of data reduces in the higher levels of the hierarchy. Data from three experiments which involve systems of increasing complexity are presented as part of the challenge problem:

The first experiment is referred to as the *calibration* experiment. It involves N_c bars where each bar has a cross section area A_c =4.0cm^2 and length L_c= 20 cm, is fixed rigidly at one end and is loaded by a tensile axial force F_c=1.2kN at the other end. The available data $\mathcal{D}_1$ from this experiment are the elongation $\delta L_c^{(i)}$, i=1,2…, N_c, of the bars from the initial length and the Young's modulus $E_c^{(i)}(L_c/2)$ at the midpoint of the bars.

The second experiment is referred to as the *validation* experiment. The set-up is similar to the first experiment. The only difference is that the bars have longer length L_v= 80cm and only the total elongation $\delta L_v^{(i)}$, i=1,2…, N_v, is measured. Let $\mathcal{D}_2$ denote the data in this case.

The third experiment is referred to as the *accreditation* experiment. It involves a frame structure (Figure 4 in Babuška et al (2008) and Figure 2 in Cheung and Beck (2008b)) subject to a point load F_a=6kN at the midpoint Q of bar 1. The available data $\mathcal{D}_3$ are the vertical displacement $w_a^{(i)}$, i=1,…, N_a, of the point Q. Information regarding the geometry of the frame is shown in Table 3 in Babuška et al (2008) and Tables 3 and 4 in Cheung and Beck (2008b). Notice that the system here is not a subsystem of the target system.

Data collected from the above three experiments are shown in Babuška et al (2008) and in Tables 5, 6 and 7 respectively in Cheung and Beck (2008b). Three cases of N_c, N_v and N_a, as shown in Table 5.1, are considered. For instance, for case 1, N_c = 5, N_v = 2 and N_a = 1 correspond to the first five, the first two and the first of the measurements listed in Tables 5, 6 and 7 respectively in Cheung and Beck (2008b). A superscript is added to $\mathcal{D}_i$ to denote different data cases. For instance, $\mathcal{D}_1^{(1)}$ denotes data collected from the calibration experiment with N_c = 5, $\mathcal{D}_2^{(1)}$ denotes data collected from the validation experiment with N_v = 2 and $\mathcal{D}_3^{(1)}$ denotes data collected from the accreditation experiment with N_a = 1. Given Young's modulus of each of the bars, the elongation of the bars in the first and second experiment and the vertical displacement in the third experiment can be predicted using the equations in Babuška et al (2008) and in Appendix I in Cheung and Beck (2008b). For convenience, the superscripts in $\boldsymbol{\theta}^{(i,j)}$ are omitted in this section. Also, only the results for data $\mathcal{D}_1^{(3)}$, $\mathcal{D}_2^{(3)}$ and $\mathcal{D}_3^{(3)}$ are presented here; results for data cases 1 and 2 may be found in Cheung and Beck (2008b).

Table 5.1 Number of samples for different cases

Case	N_c	N_v	N_a
1	5	2	1
2	20	4	1
3	30	10	2

5.2.1 Using data $\mathcal{D}_1$ from the calibration experiment

For the quantification of the uncertainties in Young's modulus $E(x)$, $0 \le x \le L$, of a bar of length L using data $\mathcal{D}_1$ from the calibration experiment, a set M_1 of four candidate model classes $\mathcal{M}_j^{(1)}$, j=1,2,3,4, is considered as follows:

Model class $\mathcal{M}_1^{(1)}$: The compliance $S(x)=S=1/E$ is constant along a bar and the value for each bar is assumed to be a sample from a Gaussian distribution with mean μ_s and variance σ_s^2. The elongation δL_c of a bar of length L_c is given by $\delta L_c = F_c L_c S/A + \varepsilon_c$ where ε_c is the prediction error, assumed to follow a Gaussian distribution with mean zero and variance σ_ε^2. The term ε_c is needed since from $\mathcal{D}_1$, it can be seen that δL_c is obviously not proportional to S. The prior PDF for $\boldsymbol{\theta} = [\mu_s \; \sigma_s^2 \; \sigma_\varepsilon^2]^T$ is chosen as three independent probability distributions: μ_s follows a truncated Gaussian distribution (constrained to be positive) which is proportional to a Gaussian distribution with mean equal to the sample mean of measurements of the mid-point compliance $S_c(L_c/2)$ and c.o.v. (coefficient of variation) of 1.0; σ_s^2 follows an inverse gamma distribution with mean μ equal to the sample variance of measurements of $S_c(L_c/2)$ and c.o.v. δ =1.0, i.e., $p(\sigma_s^2) \propto (\sigma_s^2)^{-\alpha-1}\exp(-\beta/\sigma_s^2)$ where $\alpha=\delta^{-2}+2$, $\beta=\mu(\alpha-1)$; l_s follows an inverse gamma distribution with mean equal to 10^{-11} m^2 (slightly more than the mean-square of the elongation measurements) and c.o.v. equal to 1.0. The prior c.o.v. of all of the uncertain parameters is chosen to be 1.0 to reflect a large uncertainty in the values of these parameters. If the type of material of the bars had been known in advance, the prior mean for μ_s could have been chosen to be the nominal value of the compliance obtained from previous tests performed on such material and the prior mean for σ_s^2 could have been chosen to be the prior mean for μ_s multiplied by a coefficient of variation chosen to reflect previously observed variability in the material compliance.

Model class $\mathcal{M}_2^{(1)}$: The compliance $S(x)$ is assumed to follow a stationary Gaussian random field with mean μ_s and correlation function $\mathrm{Cov}(S(x_1),S(x_2)|\sigma_s^2, l_s, r) = \sigma_s^2\exp(-(|x_1-x_2|/l_s)^r)$ where r is equal to 1. The prior PDF for $\boldsymbol{\theta} = [\mu_s \; \sigma_s^2 \; l_s]^T$ is chosen as three independent

distributions: the prior PDFs for the mean μ_s and the variance σ_s^2 follow the same distributions as in $\mathcal{M}_1^{(1)}$; the correlation length l_s follows a uniform distribution on the interval $[10^{-5}L, L]$ where we choose L=0.5m to give a reasonable range.

Model class $\mathcal{M}_3^{(1)}$: Everything is the same as $\mathcal{M}_2^{(1)}$ except r is equal to 2.

Model class $\mathcal{M}_4^{(1)}$: Everything is the same as $\mathcal{M}_2^{(1)}$ and $\mathcal{M}_3^{(1)}$ except that r is uncertain. The prior PDF for $\boldsymbol{\theta} = [\mu_s \; \sigma_s^2 \; l_s \; r]^T$ is chosen as four independent distributions: μ_s, σ_s^2, l_s follow the same distributions as in $\mathcal{M}_2^{(1)}$ and $\mathcal{M}_3^{(1)}$ and r follows a uniform distribution on [0.5, 3].

Babuška et al. (2008) and Grigoriu and Field (2008) also study the static-frame challenge problem using Bayesian updating. The perfectly-correlated Gaussian model for the compliance in $\mathcal{M}_1^{(1)}$ and the partially-correlated stationary Gaussian random field model for the compliance in $\mathcal{M}_3^{(1)}$ are also considered in Babuška et al. (2008). The partially-correlated Gaussian random field model for the compliance in $\mathcal{M}_2^{(1)}$ is considered in Grigoriu and Field (2008). $\mathcal{M}_2^{(1)}$ and $\mathcal{M}_3^{(1)}$ are included here for comparison purposes only. In practice, when r is uncertain, only $\mathcal{M}_4^{(1)}$ needs to be considered. For r=0, the correlation coefficient between the compliance at one position on the bar and that at another position is always equal to e^{-1}. This model is thought to be unreasonably constrained and so it is not considered. This is why the lower bound of r is taken to be positive.

Babuška et al. (2008) find point estimates of μ_s and σ_s^2 in $\mathcal{M}_1^{(1)}$ by matching the first two sample moments of the compliance data $S_c^{(i)}(L_c/2)$, i=1,2…, N_c, and l_s in $\mathcal{M}_3^{(1)}$ by matching the sample variance of the elongation data $\delta L_c^{(i)}$, i=1,2…, N_c, and the sample covariance of $\delta L_c^{(i)}$ and $S_c^{(i)}(L_c/2)$, i=1,2…, N_c. Grigoriu and Field (2008) approximate the uncertain parameters by point estimates by matching the sample moments similar to Babuška et al (2008) except that they do not consider the sample covariance of $\delta L_c^{(i)}$ and $S_c^{(i)}(L_c/2)$, i=1,2…, N_c. In Grigoriu and Field (2008), the uncertainties in the model parameters μ_s, σ_s^2 and l_s are not considered and not directly propagated into the predictions so probabilistic

information in these parameters is not subsequently characterized. Babuška et al. (2008) quantify the uncertainties by using kernel density estimation to reconstruct the joint PDF of δL_c and $S_c(L_c/2)$ from the data for $\delta L_c^{(i)}$ and $S_c^{(i)}(L_c/2)$ and then using the bootstrapping method to generate additional "data".

Appropriate quantification of uncertainties in the parameters (i.e. obtaining complete probabilistic information in terms of the posterior PDF for each model class) is desirable since it significantly affects the effectiveness and robustness of model class updating, comparison and validation, as well as the prediction of the responses and the failure probability of the target structure. Here we use the challenge problem to illustrate how the uncertainties can be quantified appropriately and effectively by exploiting the *full* power of Bayesian analysis using the proposed concept of hierarchical stochastic system model classes and recently-developed computational tools. Later, when we present the analysis results, it will be clear that given the calibration data, the uncertainty in μ_s is quite small but the uncertainties in other parameters and data-induced correlation between the parameters are not negligible; the complete probabilistic information is, however, encapsulated in the samples from the posterior.

To quantify the uncertainties of $\boldsymbol{\theta}$ using Bayesian analysis and $\mathcal{D}_1^{(3)}$, the elongation data $\delta L_c^{(i)}$ and the compliance data $S_c^{(i)}(L_c/2)$, i=1,2,…, N_c should be considered simultaneously since they are correlated to each other given $\boldsymbol{\theta}$ and the proposed model classes.

The posterior PDF for model class $\mathcal{M}_j^{(1)}$, for j=1,2,3,4, is given by Bayes' Theorem: $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{M}_j^{(1)}) = p(\mathcal{D}_1^{(3)}|\boldsymbol{\theta},\mathcal{M}_j^{(1)})p(\boldsymbol{\theta}|\mathcal{M}_j^{(1)})/p(\mathcal{D}_1^{(3)}|\mathcal{M}_j^{(1)})$ where the prior PDF $p(\boldsymbol{\theta}|\mathcal{M}_j^{(1)})$ is described above and the likelihood function p(D1(3)|θ,Mj(1)) is given by the following. The likelihood function for $\mathcal{M}_1^{(1)}$ is:

$$p(\mathcal{D}_1^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_1^{(1)})$$
$$= \frac{1}{[2\pi \mid \mathbf{C}(\sigma_s^{\,2}, \sigma_\varepsilon^{\,2}) \mid^{1/2}]^{N_c}} \exp(-\frac{1}{2}\sum_{i=1}^{N_c}[\mathbf{y}^{(i)} - \boldsymbol{\mu}(\mu_s)]^T \mathbf{C}^{-1}(\sigma_s^{\,2}, \sigma_\varepsilon^{\,2})[\mathbf{y}^{(i)} - \boldsymbol{\mu}(\mu_s)]) \tag{5.10}$$

where

$$\mathbf{y}^{(i)} = \begin{bmatrix} \delta L_c^{\,(i)} \\ S_c^{\,(i)}(L_c/2) \end{bmatrix} \tag{5.11}$$

$$\boldsymbol{\mu}(\mu_s) = \begin{bmatrix} \frac{F_c L_c}{A_c} \\ 1 \end{bmatrix} \mu_s \tag{5.12}$$

$$\mathbf{C}(\sigma_s^{\,2}, \sigma_\varepsilon^{\,2}) = \begin{bmatrix} (\frac{F_c L_c}{A_c})^2 \sigma_s^{\,2} + \sigma_\varepsilon^{\,2} & \frac{F_c L_c}{A_c}\sigma_s^{\,2} \\ \frac{F_c L_c}{A_c}\sigma_s^{\,2} & \sigma_s^{\,2} \end{bmatrix} \tag{5.13}$$

For $\mathcal{M}_2^{(1)}$ and $\mathcal{M}_3^{(1)}$, the likelihood function is the same as that for $\mathcal{M}_4^{(1)}$ with r=1 and 2, respectively. The likelihood function for $\mathcal{M}_4^{(1)}$ is given by:

$$p(\mathcal{D}_1^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_4^{(1)})$$
$$= \frac{1}{[2\pi\sigma_s^{\,2} \mid \mathbf{C}(l_s, r) \mid^{1/2}]^{N_c}} \exp(-\frac{1}{2\sigma_s^{\,2}}\sum_{i=1}^{N_c}[\mathbf{y}^{(i)} - \boldsymbol{\mu}(\mu_s)]^T \mathbf{C}^{-1}(l_s, r)[\mathbf{y}^{(i)} - \boldsymbol{\mu}(\mu_s)]) \tag{5.14}$$

where $\mathbf{y}^{(i)}$ and $\boldsymbol{\mu}(\theta_1)$ are given by (5.11) and (5.12) and $\mathbf{C}(l_s, r)$ is given by:

$$\mathbf{C}(l_s, r) = \begin{bmatrix} C_{11}(l_s, r) & C_{12}(l_s, r) \\ C_{12}(l_s, r) & 1 \end{bmatrix} \tag{5.15}$$

where the entries C_{11} and C_{12} of $\mathbf{C}$ are given by:

$$C_{11}(l_s, r) = (\frac{F_c}{A_c})^2 \int_0^{L_c} \int_0^{L_c} \frac{\mathrm{Cov}(S(x_1), S(x_2) \mid \sigma_s^2, l_s, r)}{\sigma_s^2} dx_1 dx_2 = 2(\frac{F_c}{A_c})^2 \int_0^{L_c} (L_c - x) \exp(-(\frac{x}{l_s})^r) dx \tag{5.16}$$

$$C_{12}(l_s, r) = \frac{F_c}{A_c} \int_0^{L_c} \frac{\mathrm{Cov}(S(x), S(L_c/2) \mid \sigma_s^2, l_s, r)}{\sigma_s^2} dx = \frac{F_c}{A_c} \int_0^{L_c} \exp(-(\frac{|x - L_c/2|}{l_s})^r) dx \tag{5.17}$$

For $\mathcal{M}_2^{(1)}$, r is equal to 1 and thus the above integrals can be evaluated analytically to give:

$$C_{11}(l_s, 1) = 2(\frac{F_c}{A_c})^2 l_s (L_c - l_s + l_s \exp(-\frac{L_c}{l_s})); \; C_{12}(l_s, 1) = 2(\frac{F_c}{A_c}) l_s (1 - \exp(-\frac{L_c}{2l_s})) \tag{5.18}$$

For $\mathcal{M}_3^{(1)}$, r is equal to 2 and thus the above integrals can be expressed in terms of the error function to give:

$$\begin{aligned} C_{11}(l_s, 2) &= (\frac{F_c}{A_c})^2 [L_c l_s \sqrt{\pi} \mathrm{erf}(\frac{L_c}{l_s}) - l_s^2 (1 - \exp(-(\frac{L_c}{l_s})^2))] \\ C_{12}(l_s, 2) &= \frac{F_c}{A_c} l_s \sqrt{\pi} \mathrm{erf}(\frac{L_c}{2l_s}) \end{aligned} \tag{5.19}$$

Since the computer always has a precision limit in representing numbers, when performing the analysis, we make sure l_s is such that $\mathbf{C}(l_s, r)$ is positive definite, i.e., $C_{11}(l_s, r)$ and $|\mathbf{C}(l_s, r)| = C_{11}(l_s, r) - C_{12}^2(l_s, r)$ are both positive. The interval of l_s for its prior PDF in $\mathcal{M}_2^{(1)}$, $\mathcal{M}_3^{(1)}$ and $\mathcal{M}_4^{(1)}$ satisfies this constraint.

Table 5.2 shows the statistical results using the calibration data $\mathcal{D}_1^{(3)}$ where N_c = 30. The (j+1)-th column gives the results obtained using a full Bayesian analysis for model class $\mathcal{M}_j^{(1)}$, j=1,2,3,4. We used the Hybrid Gibbs TMCMC algorithm presented in Appendix 5A for simulating samples from the posterior $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{M}_j^{(1)})$ and for calculating the evidence $p(\mathcal{D}_1^{(3)}|\mathcal{M}_j^{(1)})$ which is required for the calculation of the probability $P(\mathcal{M}_j^{(1)}|\mathcal{D}_1^{(3)},M)$ of each model class conditioned on the data $\mathcal{D}_1^{(3)}$. This algorithm is used for simulating samples from the posterior $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{M}_j^{(1)})$ because of its ability to handle the case where we do not know apriori whether there may be several separated neighborhoods of high probability regions of $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{M}_j^{(1)})$ between which the transition using a Markov chain of samples is not efficient.

The second row of Table 5.2 gives the MAP (maximum a posteriori) estimate $\boldsymbol{\theta}_{\mathbf{MAP}}$ (that is, $\boldsymbol{\theta}$ that globally maximizes the product $p(\mathcal{D}_1^{(3)}|\boldsymbol{\theta},\mathcal{M}_j^{(1)})p(\boldsymbol{\theta}|\mathcal{M}_j^{(1)})$ and so $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{M}_j^{(1)})$). The third row gives the mean (the number before the semicolon), c.o.v. (the number after the semicolon) and the correlation coefficient matrix R from the posterior samples for $\boldsymbol{\theta}$ where the (i,j) entry of R is the correlation coefficient between θ_i and θ_j. Only the upper diagonal entries of R are presented since it is symmetric. Compared with the prior uncertainty in the parameters, the posterior (updated) uncertainty is reduced since the data provide information about these parameters. For all data cases and four model classes, μ_s has a lot smaller uncertainty than the other parameters which have significant uncertainties. It can be seen that the posterior mean of σ_s^2 given data $\mathcal{D}_1^{(3)}$ is quite different from the sample variance of the compliance measurements $S_c^{(i)}$, i=1,…, N_c since the elongation data $\delta L_c^{(i)}$ in $\mathcal{D}_1^{(3)}$ give extra information about this parameter. Because the challenge problem assumes an exact theory for the deformation analysis, prediction errors for each model class are accounted for by the modeling parameters such as σ_s^2. In general, prediction errors can be explicitly accounted for by adding them to the output equation (Beck and Katafygiotis (1998)), as done in $\mathcal{M}_1^{(1)}$.

Table 5.2 Statistical results using data $\mathcal{D}_1^{(3)}$ from the calibration experiment

		$\mathcal{M}_1^{(1)}$	$\mathcal{M}_2^{(1)}$	$\mathcal{M}_3^{(1)}$	$\mathcal{M}_4^{(1)}$
MAP	μ_s (Pa^{-1})	8.64×10^{-11}	8.87×10^{-11}	8.87×10^{-11}	8.87×10^{-11}
	σ_s^2(Pa^{-2})	3.24×10^{-23}	4.87×10^{-23}	4.76×10^{-23}	4.72×10^{-23}
	σ_ε^2*(m^2);l_s(m)	1.11×10^{-11}*	0.0284	0.0307	0.0305
	r				3
Statistics parameters (stochastic Simulation)	μ_s (Pa^{-1})	8.64×10^{-11};1.2%	8.88×10^{-11};0.83%	8.87×10^{-11};0.69%	8.88×10^{-11};0.8%
	σ_s^2(Pa^{-2})	3.69×10^{-23};26.0%	5.19×10^{-23};19.5%	5.37×10^{-23};20.4%	5.20×10^{-23};19.9%
	l_s(m)	1.24×10^{-11};23.7%	0.0319;27.5%	0.0327;23.6%	0.0328;27.8%
	r				1.79;40.5%
	R	$\begin{bmatrix} 1 & 0.09 & -0.11 \\ & 1 & -0.09 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.05 & 0.20 \\ & 1 & -0.10 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.10 & 0.04 \\ & 1 & -0.21 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.01 & 0.15 & 0.07 \\ & 1 & -0.05 & -0.14 \\ & & 1 & -0.01 \\ & & & 1 \end{bmatrix}$
Log evidence		1059.63	1071.34	1071.66	1071.87
E[ln$p(\mathcal{D}_1^{(3)}\|\boldsymbol{\theta},\mathcal{M}_j^{(1)})$]		1064.89	1079.75	1080.15	1079.82
Expected information gain		5.27	8.41	8.49	7.95
$P(\mathcal{M}_j^{(1)}\|\mathcal{D}_1^{(3)},M_1)$		2.01×10^{-6}	0.245	0.338	0.416
$P(F\|\boldsymbol{\theta}_{\mathbf{MAP}}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$		3.61×10^{-2}	3.56×10^{-7}	3.19×10^{-9}	6.70×10^{-12}
$P(F\|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$		9.81×10^{-2}(1.9%)	3.58×10^{-4}(16.1%)	1.30×10^{-4}(26.1%)	2.79×10^{-4}(16.5%)
$P(F\|\mathcal{D}_1^{(3)}, M_1)$		2.48×10^{-4}			

It can be seen from the correlation coefficient matrix that there is only weak correlation between pairs of parameters, although one must be careful since a small correlation coefficient between two uncertain parameters only implies weak *linear* dependence and does not necessarily imply weak dependence between them unless the parameters are jointly Gaussian. A simple example for this is $W=Z^2$ and a standard normal variable Z which are uncorrelated but strongly dependent. To investigate dependence between different pairs of parameters, sample plots of some pairs of the components of $\boldsymbol{\theta}$ from the posterior $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ are shown in Figure 5.2 (for j=2), Figure 5.3 (for j=3) and Figure

5.4 (j=4). Each axis corresponds to an uncertain parameter θ_i divided by its posterior mean μ_i given $\mathcal{D}_1^{(3)}$ and a specific model class $\mathcal{M}_j^{(1)}$, which can be estimated as follows:

$$\mu_i = E[\theta_i \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)}] \approx \frac{1}{K}\sum_{k=1}^{K} \theta_i^{(k)} \tag{5.20}$$

where $[\theta_i^{(1)},...,\theta_i^{(K)}]$ are K posterior samples for θ_i from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$. All the other parameters have significantly larger uncertainties than θ_1. It can be seen that $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_2^{(1)})$ and $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_3^{(1)})$ are not close to a multivariate Gaussian PDF and $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_4^{(1)})$ departs substantially from a multivariate Gaussian. For $\mathcal{M}_4^{(1)}$, the samples for r show truncation due to the choice of truncated uniform priors for r.

Figure 5.5 gives the histogram of posterior samples for r from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_4^{(1)})$. This figure suggests that $p(r|\mathcal{D}_1^{(3)}, \mathcal{M}_4^{(1)})$ is multi-modal and every value of r is of non-negligible plausibility. The above results exhibit the strength of the stochastic simulation method in capturing the full characteristics of the complex posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ represented by the generated posterior samples.

The stochastic simulation estimate for log evidence, posterior mean of the log likelihood function (a datafit measure), expected information gain and the probability $P(\mathcal{M}_j^{(1)}|\mathcal{D}_1^{(3)},M_1)$ of the model classes are shown in the fourth through seventh rows, respectively, of Table 5.2. Based on the calibration data, $\mathcal{M}_1^{(1)}$ is very improbable compared with the other model classes $\mathcal{M}_2^{(1)}$, $\mathcal{M}_3^{(1)}$ and $\mathcal{M}_4^{(1)}$ which have similar posterior probabilities. These latter model classes have essentially the same posterior mean of the log likelihood function which shows that they give a similar fit to the data on average and they also have similar expected information gains.

Grigoriu and Field (2008) perform model selection by calculating the posterior model probabilities of the MLE (maximum likelihood estimate) models (rather than the posterior

probability for the whole model class) in which the modeling parameters are obtained by matching the moments calculated from the data. Such an approach considers the magnitude of the likelihood functions of the MLE models and no uncertainties in the parameters are considered when performing model selection. The fact that there exists many plausible models in a model class is not considered, in contrast to our full Bayesian treatment. In particular, when the evidence for the model class is not employed, there is no automatic downgrading of more "complex" models that extract more information from the data, so this can lead to what is commonly called "data overfitting" (Bishop 2006). Note that one cannot simply count the number of uncertain parameters in a model class to judge reliably its complexity; for example, one should use the evidence for the model class and not the simplified version known as BIC (Bayesian information criterion) for model selection (Beck and Yuen 2004, Muto and Beck 2008).

For each of the four model classes $\mathcal{M}_j^{(1)}$, given $\boldsymbol{\theta}$, it can be shown that the response w_p of interest for the target frame structure follows a Gaussian distribution with mean $\mu_p = K_p \mu_s$ and variance $\sigma_p^2 = \sigma_s^2 V_{p,1}$ for $\mathcal{M}_1^{(1)}$ and $\sigma_s^2 V_{p,j}(l_s)$ for $\mathcal{M}_j^{(1)}$, j=2,3 and $\sigma_s^2 V_{p,j}(l_s, r)$ for $\mathcal{M}_j^{(1)}$, j=4 where the expressions for K_p and $V_{p,j}$ are given in Cheung and Beck (2008b). It should be stressed that w_p is not Gaussian (in this case, it follows a distribution which a weighted infinite sum of Gaussian PDFs) and it is Gaussian only when given $\boldsymbol{\theta}$.

The eighth row in Table 5.2 gives the failure probability $P(F|\boldsymbol{\theta}_{\mathbf{MAP}}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ of the target frame structure with $\boldsymbol{\theta} = \boldsymbol{\theta}_{\mathbf{MAP}}$ based on the calibration data $\mathcal{D}_1^{(3)}$ and each model class, which can be expressed in terms of the CDF of a standard Gaussian random variable $\Phi(\mathbf{z})$:

$$P(F \mid \boldsymbol{\theta}_{\mathbf{MAP}}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)}) = \Phi\left(\frac{-0.003 - \mu_p(\boldsymbol{\theta}_{\mathbf{MAP}})}{\sigma_p(\boldsymbol{\theta}_{\mathbf{MAP}})}\right) + 1 - \Phi\left(\frac{0.003 - \mu_p(\boldsymbol{\theta}_{\mathbf{MAP}})}{\sigma_p(\boldsymbol{\theta}_{\mathbf{MAP}})}\right) \quad (5.21)$$

The ninth row gives the predicted robust failure probability $P(F|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ (the number outside the parenthesis) of the target frame structure with the uncertainty in $\boldsymbol{\theta}$ taken into account for each model class, and it is calculated using:

$$\begin{aligned} P(F \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)}) &= \int P(F \mid \boldsymbol{\theta}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)}) p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)}) d\boldsymbol{\theta} \\ &\approx 1 + \frac{1}{K} \sum_{k=1}^{K} [\Phi(\frac{-0.003 - \mu_p(\boldsymbol{\theta}^{(k)})}{\sigma_p(\boldsymbol{\theta}^{(k)})}) - \Phi(\frac{0.003 - \mu_p(\boldsymbol{\theta}^{(k)})}{\sigma_p(\boldsymbol{\theta}^{(k)})})] \end{aligned} \tag{5.22}$$

where $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K, are posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$. An alternative way to calculate $P(F|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ is by simulating samples of w_p based on posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ and check how many leads to failure. But in the problem considered here, this is always less efficient than (5.22) (Refer to Appendix 5B). It should be noted that a very efficient stochastic simulation method called Subset Simulation (Au and Beck 2001b) can also be used for calculating $P(F|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ using posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$. The number inside the parenthesis gives the estimate of the coefficient of variation (c.o.v.) of the above predicted robust failure probability estimate. It can be seen that $P(F|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ is orders of magnitude different from $P(F|\boldsymbol{\theta}_{\mathbf{MAP}}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ showing that the effects of the uncertainties in the parameters on the failure probabilities is substantial. In fact, ignoring the uncertainty in $\boldsymbol{\theta}$ would be disastrous since $P(F|\boldsymbol{\theta}_{\mathbf{MAP}}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ greatly underestimates the failure probability for all model classes and it varies greatly from one model class to another, in contrast with the robust case $P(F|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$. Figure 5.6 shows $P(F|\boldsymbol{\theta}^{(k)}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ corresponding to each posterior sample model $\boldsymbol{\theta}^{(k)}$, sorted in increasing order. Figure 5.7 shows the CDF of $P(F|\boldsymbol{\theta}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ estimated using posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$. Figures 5.6 and 5.7 confirm that there is a large variability in $P(F|\boldsymbol{\theta}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$ due to the uncertainties in $\boldsymbol{\theta}$.

Posterior model averaging can be carried out to obtain the predicted hyper-robust failure probability $P(F|\mathcal{D}_1^{(3)}, M_1)$ given the set of candidate model classes M_1 (last row of Table 5.2):

$$P(F \mid \mathcal{D}_1^{(3)}, M_1) = \sum_{j=1}^{4} P(F \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)}) P(\mathcal{M}_j^{(1)} \mid \mathcal{D}_1^{(3)}, M_1) = 2.48 \times 10^{-4} \qquad (5.23)$$

Figure 5.8 shows the CDFs of the predicted vertical displacement w_p at point P in the target frame structure corresponding to each sample $\boldsymbol{\theta}^{(k)}$, k=1,2,…,4000, from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_4^{(1)})$. The robust posterior CDF of the response w_p of interest for the target frame structure can be obtained using the Theorem of Total Probability, as in the previous section. Figure 5.9 shows that the robust CDFs for the three model classes are very close to each other in the high probability region but differ somewhat in the tails so the predicted failure probability is quite different (though still within the same order of magnitude), as shown in Table 5.2.

From the results in Table 5.2, it can be seen that $P(F|\mathcal{D}_1^{(3)}, \mathcal{M}_1^{(1)})P(\mathcal{M}_1^{(1)}|\mathcal{D}_1^{(3)}, M_1)$ is negligible compared to $P(F|\mathcal{D}_1^{(3)}, M_1)$ and so the contribution of $\mathcal{M}_1^{(1)}$ is negligible to the prediction of interest, the failure probability of the target frame structure. Also, having a posterior model class probability $P(\mathcal{M}_1^{(1)}|\mathcal{D}_1^{(3)},M_1)$ that is several orders of magnitude smaller than those for the other model classes implies $\mathcal{M}_1^{(1)}$ is relatively improbable conditioned on the data $\mathcal{D}_1^{(3)}$. Thus, $\mathcal{M}_1^{(1)}$ is dropped in the subsequent analyses.

Note that the posterior probability $P(\mathcal{M}_j^{(1)}|\mathcal{D}_1^{(3)},M_1)$ for each model class conditioned on the data $\mathcal{D}_1^{(3)}$ gives the plausibility of each $\mathcal{M}_j^{(1)}$ given the set of candidate model classes $M_1=\{\mathcal{M}_j^{(1)}, j=1,2,3,4\}$ and $P(F|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})P(\mathcal{M}_j^{(1)}|\mathcal{D}_1^{(3)}, M_1)$ gives the contribution of each model class to the desired response prediction. These probabilities do not give information regarding the predictability of each model class for the response of other systems, including the target system. It is shown in the following sections how the data from the validation and

accreditation experiments are used to evaluate the prediction consistency and accuracy of the calibrated model classes.

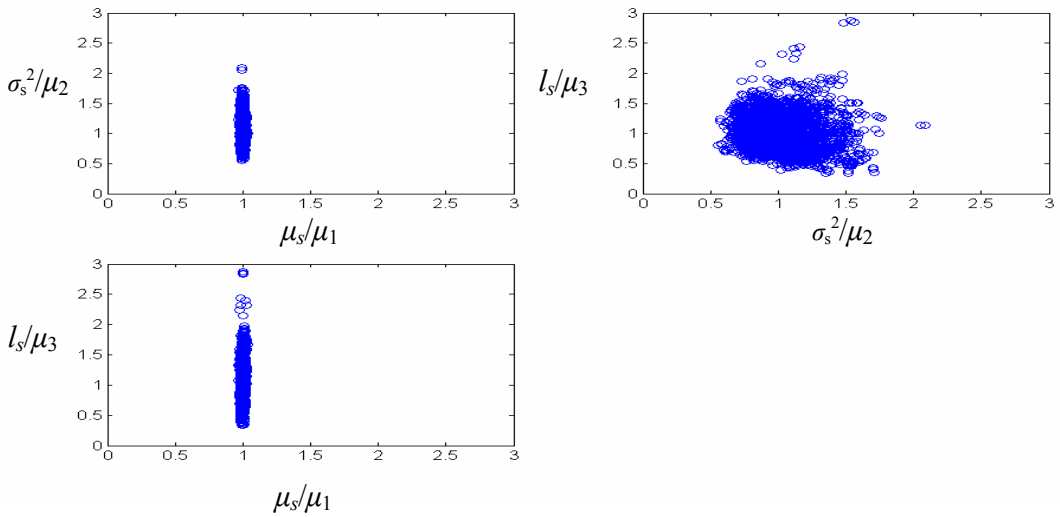

Figure 5.2: Pairwise sample plots of posterior samples for $p(\theta|\, \mathcal{D}_1^{(3)}, \mathcal{M}_2^{(1)})$ normalized by posterior mean

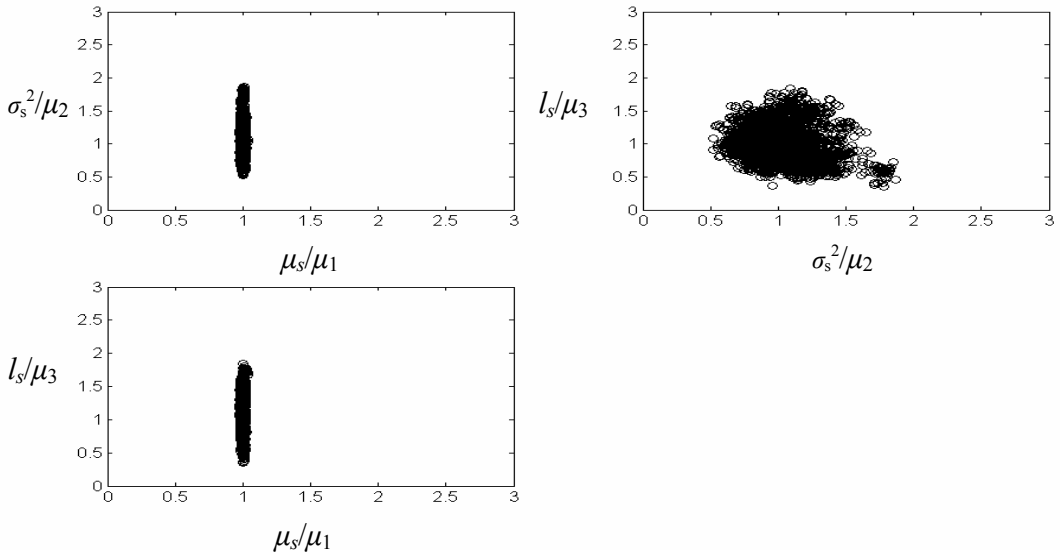

Figure 5.3: Pairwise sample plots of posterior samples for $p(\theta|\mathcal{D}_1^{(3)}, \mathcal{M}_3^{(1)})$ normalized by posterior mean

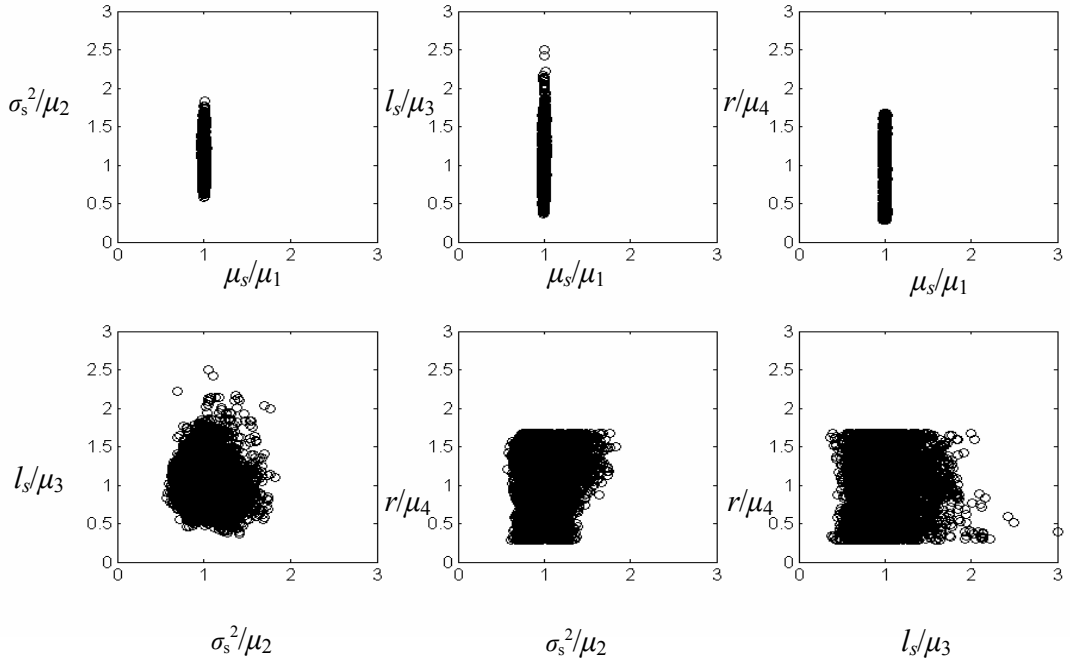

Figure 5.4: Pairwise sample plots of posterior samples for $p(\theta|\mathcal{D}_1^{(3)}, \mathcal{M}_4^{(1)})$ normalized by posterior mean

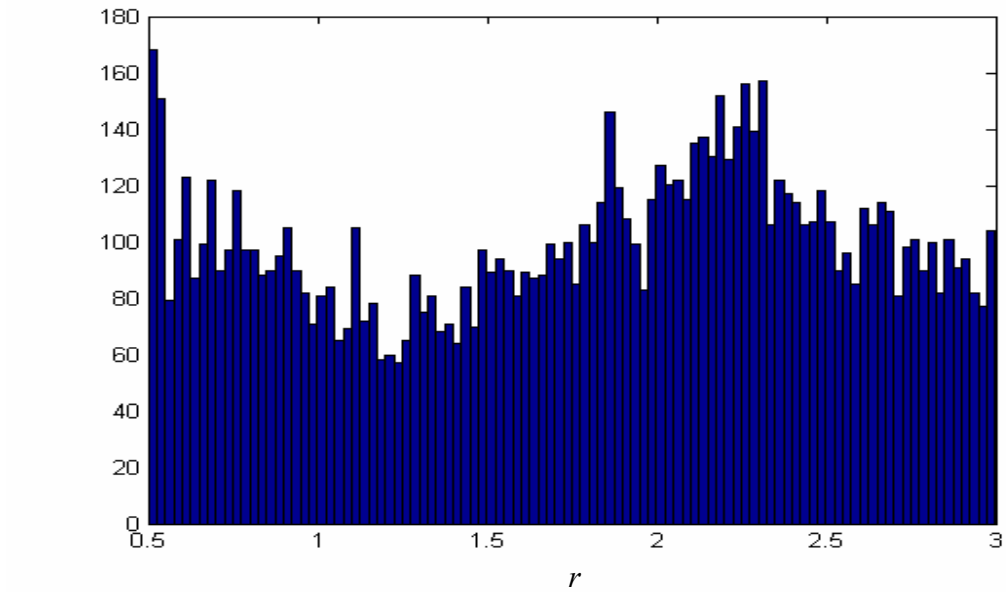

Figure 5.5: Histogram for posterior samples for $p(r|\mathcal{D}_1^{(3)}, \mathcal{M}_4^{(3)})$

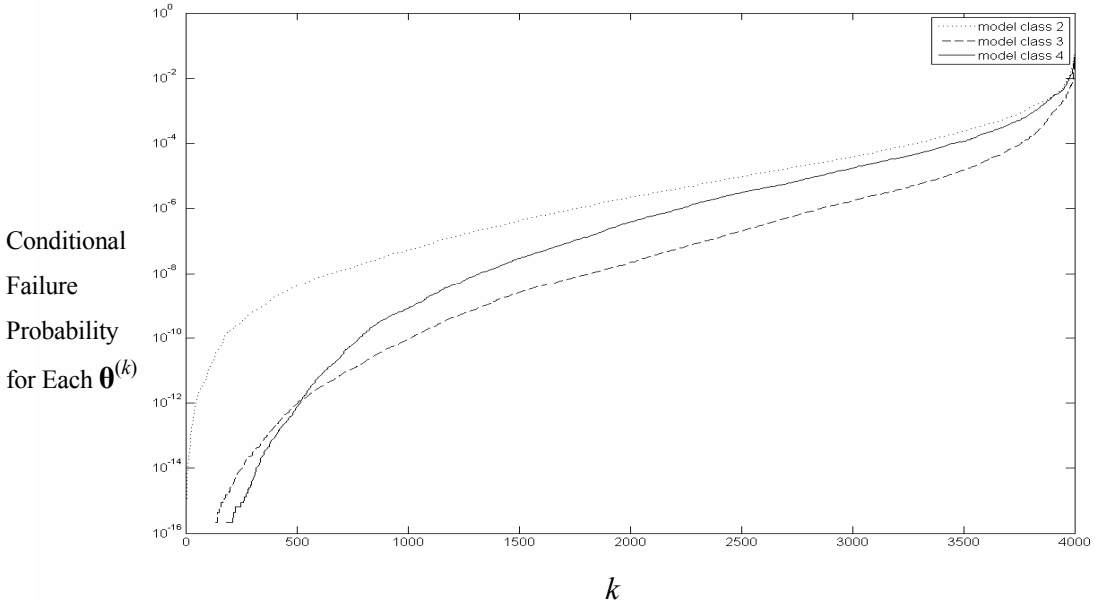

Figure 5.6: The failure probability (sorted in increasing order) conditioned on each posterior sample $\boldsymbol{\theta}^{(k)}$ for model class $\mathcal{M}_j^{(1)}$, i.e. $P(F|\boldsymbol{\theta}^{(k)},\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$, for j=2,3,4

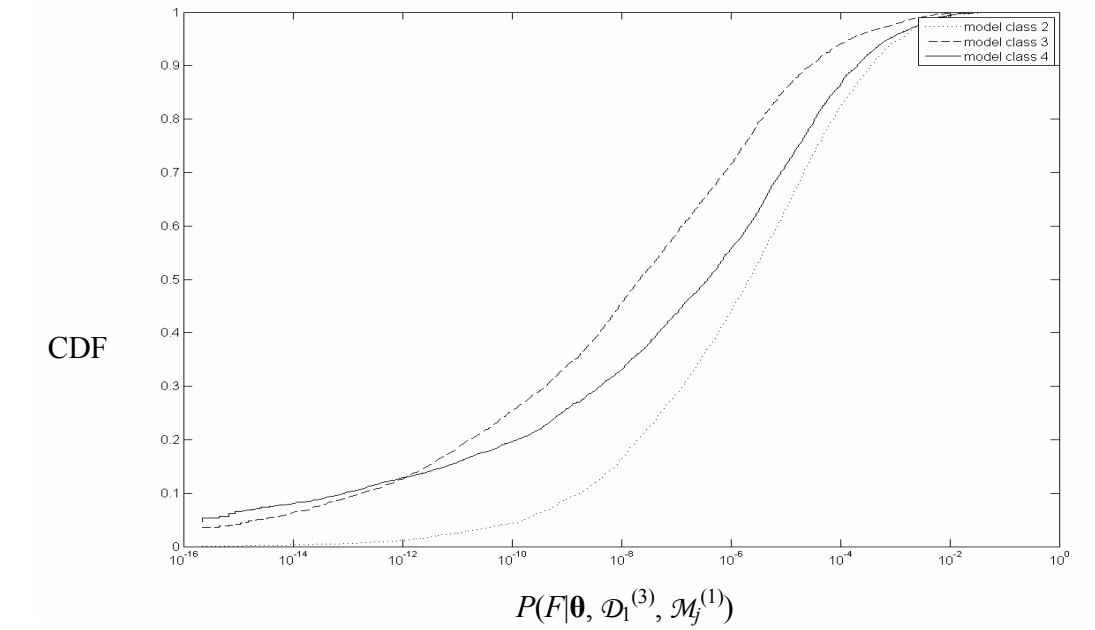

Figure 5.7: CDF of failure probability $P(F|\boldsymbol{\theta}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$, j=2,3,4, estimated using posterior samples for model class $\mathcal{M}_j^{(1)}$

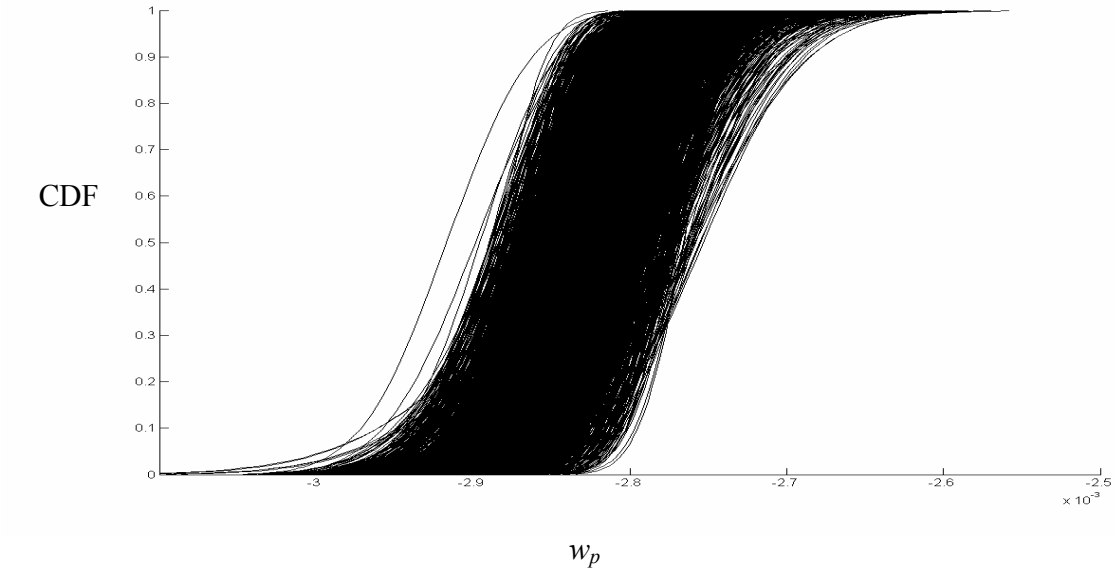

Figure 5.8: CDF of predicted vertical displacement w_p at point P in the target frame structure conditioned on each sample from $p(\theta|\mathcal{D}_1^{(3)}, \mathcal{M}_4^{(1)})$

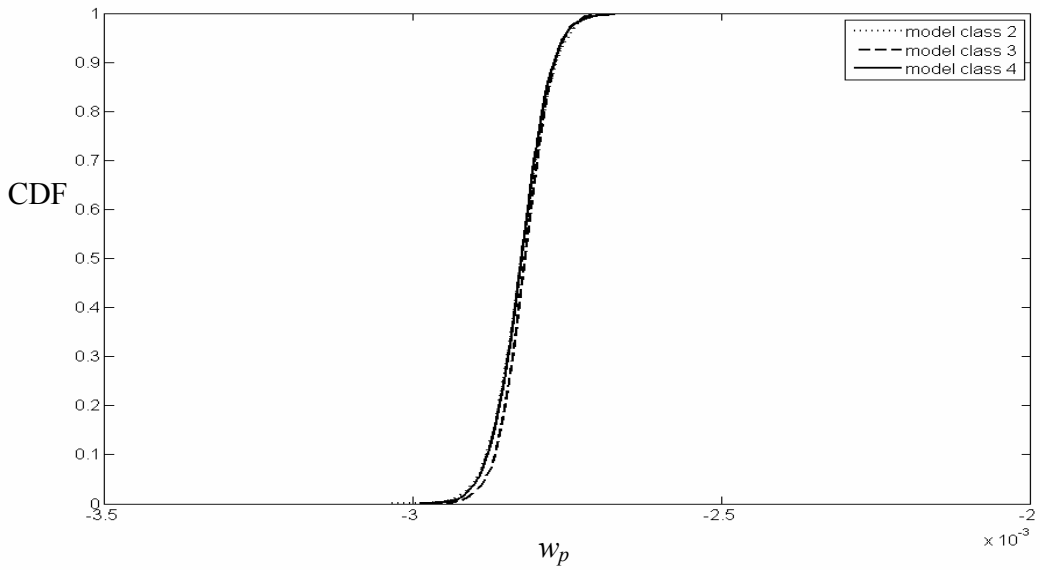

Figure 5.9: Robust posterior CDF of predicted vertical displacement w_p at point P in the target frame structure calculated using the posterior samples from $p(\theta|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(1)})$, j=2,3,4

5.2.2 Using data $\mathcal{D}_2$ from the validation experiment

Candidate model classes for the subsystem in the validation experiment are $\mathcal{M}_j^{(2)}$, j=1,2,3. The only difference between the subsystem here and that in the previous experiment is the longer beam length. The uncertain parameters $\boldsymbol{\theta}^{(2,j)}$ for $\mathcal{M}_j^{(2)}$ are the same as $\boldsymbol{\theta}^{(1,j+1)}$ for $\mathcal{M}_{j+1}^{(1)}$. The "prior" PDF $p(\boldsymbol{\theta}^{(2,j)}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$ for $\mathcal{M}_j^{(2)}$ is given by the "posterior" PDF $p(\boldsymbol{\theta}^{(1,j+1)}|\mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)})$ for $\mathcal{M}_{j+1}^{(1)}$. Data $\mathcal{D}_2^{(3)} = \{\delta L_v^{(i)}, i=1,2\ldots, N_v=10\}$ from the validation experiment are used to investigate the predictive performance, including the prediction consistency and accuracy of the model classes.

To evaluate *prediction accuracy*, we compute the probability that the response $\delta L_{v,p}$, which is the elongation of the bar in the validation experiment, predicted using the model classes updated by data from the previous experiment (i.e. data $\mathcal{D}_1^{(3)}$ from the calibration experiment), is within a certain b% (b= 5 and 10) of the measured quantity $\delta L_v^{(i)}$ in the validation experiment. This probability is given by the following updated robust predictive PDF conditioned on $\mathcal{D}_1^{(3)}$:

$$\begin{aligned} P(e_{v,p}{}^{(i)} \le b\% \,|\, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) &= \int P(e_{v,p}{}^{(i)} \le b\% \,|\, \boldsymbol{\theta}^{(2,j)}, \mathcal{M}_j^{(2)})\, p(\boldsymbol{\theta}^{(2,j)} \,|\, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) d\boldsymbol{\theta}^{(2,j)} \\ &= \int P(e_{v,p}{}^{(i)} \le b\% \,|\, \boldsymbol{\theta}^{(1,j+1)}, \mathcal{M}_j^{(2)})\, p(\boldsymbol{\theta}^{(1,j+1)} \,|\, \mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)}) d\boldsymbol{\theta}^{(1,j+1)} \end{aligned} \tag{5.24}$$

where

$$e_{v,p}{}^{(i)} = \left| \frac{\delta L_{v,p} - \delta L_v^{(i)}}{\delta L_v^{(i)}} \right| \tag{5.25}$$

For convenience, the superscripts in $\boldsymbol{\theta}^{(i,j)}$ will now be omitted. For the model class $\mathcal{M}_j^{(2)}$, j=1,2,3, given $\boldsymbol{\theta}$, it can be shown that the response $\delta L_{v,p}$ follows a Gaussian distribution with mean $\mu_v = K_v \mu_s$ and variance $\sigma_{v,j}{}^2 = \sigma_s{}^2 s_{v,j}(l_s, r)$ where $K_v = F_v L_v / A_v$ and $s_{v,j}$ are given by (5.16) with subscript 'c' replaced by 'v'. For j=1, r is equal to 1, and $s_{v,j}(l_s, r)$ is given by

(5.18) with subscript 'c' replaced by 'v' and for j=2, r is equal to 2, and $s_{v,j}(l_s,r)$ is given by (5.19) with subscript 'c' replaced by 'v'. Thus, the probability $P(e_{v,p}^{(i)} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$ in (5.24) becomes:

$$
\begin{aligned}
&P(e_{v,p}^{(i)} \le b\% \,|\, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) \\
&= \int [\Phi(\frac{(1+\frac{b}{100})\delta L_v^{(i)} - \mu_v(\boldsymbol{\theta})}{\sigma_{v,j}(\boldsymbol{\theta})}) - \Phi(\frac{(1-\frac{b}{100})\delta L_v^{(i)} - \mu_v(\boldsymbol{\theta})}{\sigma_{v,j}(\boldsymbol{\theta})})] p(\boldsymbol{\theta} \,|\, \mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)}) d\boldsymbol{\theta} \qquad (5.26) \\
&\approx \frac{1}{K}\sum_{k=1}^{K} [\Phi(\frac{(1+\frac{b}{100})\delta L_v^{(i)} - \mu_v(\boldsymbol{\theta}^{(k)})}{\sigma_{v,j}(\boldsymbol{\theta}^{(k)})}) - \Phi(\frac{(1-\frac{b}{100})\delta L_v^{(i)} - \mu_v(\boldsymbol{\theta}^{(k)})}{\sigma_{v,j}(\boldsymbol{\theta}^{(k)})})]
\end{aligned}
$$

where $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K, are posterior samples from $p(\boldsymbol{\theta}^{(1,j+1)}|\mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)})$. Similar to before, samples of $\delta L_{v,p}$ can be obtained as follows: For each $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K, which are the posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)})$, generate a sample $\delta L_{v,p}^{(k)}$ for $\delta L_{v,p}$ from a Gaussian distribution with mean $\mu_v(\boldsymbol{\theta})$ and variance $\sigma_{v,j}^{\,2}(\boldsymbol{\theta})$. These samples can also be used to find the above probability by approximating it as the proportion of samples that satisfies the condition $e_{v,p}^{(i)} \le b\%$ out of the K samples. It can be shown, however, that the estimator in (5.26) is always of a smaller c.o.v. and thus more accurate than the latter approximation.

The average prediction error probability, denoted $P(e_{v,p} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$, for a model class updated using data $\mathcal{D}_1^{(3)}$ can be obtained by taking the arithmetic mean of $P(e_{v,p}^{(i)} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$, i=1, 2…, N_v. Table 5.3 shows the results for $P(e_{v,p}^{(i)} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$ (the numbers outside the parenthesis) and their average $P(e_{v,p} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$ (the numbers inside the parenthesis) for j=1, 2, 3, and b=5 and 10. It can be seen from Table 5.3 that the model classes $\mathcal{M}_j^{(2)}$ (and so $\mathcal{M}_{j+1}^{(1)}$ updated using $\mathcal{D}_1^{(3)}$), for j=1, 2, 3, are sufficiently accurate. It is noted that the averages $P(e_{v,p} \le 5\% | \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$ for each j=1, 2, 3, are larger than 0.5 implying that it is more likely than not for the response prediction by the model classes to be accurate within 5% of the actual response. The averages $P(e_{v,p} \le 10\% | \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$ are all

very close to 1, showing that it is very probable that the prediction errors for each model class are less than 10%.

Table 5.3 Results of predicting δL_v using data $\mathcal{D}_1^{(3)}$ from the calibration experiment

	$\mathcal{M}_1^{(2)}$	$\mathcal{M}_2^{(2)}$	$\mathcal{M}_3^{(2)}$
$P(\lvert\delta L_{v,p}-\delta L_v^{(i)}\rvert/\lvert\delta L_v^{(i)}\rvert\le 5\%\lvert\mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)})$	0.325,0.732, 0.325,0.844, 0.579,0.325, 0.732,0.943, 0.149,0.844 (0.579)	0.368,0.774, 0.368,0.882, 0.624,0.368, 0.774,0.956, 0.160,0.882 (0.615)	0.327,0.730, 0.327,0.846, 0.579,0.327, 0.730,0.944, 0.137,0.846 (0.579)
$P(\lvert\delta L_{v,p}-\delta L_v^{(i)}\rvert/\lvert\delta L_v^{(i)}\rvert\le 10\%\lvert\mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)})$	0.940,0.994, 0.940,0.997, 0.984,0.940, 0.994,0.999, 0.815,0.997 (0.960)	0.956,0.997, 0.956,0.998, 0.988,0.956, 0.997,0.999, 0.854,0.998 (0.970)	0.943,0.993, 0.943,0.999, 0.984,0.943, 0.993,0.999, 0.817,0.999 (0.961)
$\frac{\delta L_v^{(i)} - E[\delta L_{v,p} \mid \mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)}]}{\sqrt{Var[\delta L_{v,p} \mid \mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)}]}}$	-2.40,-1.42 -2.40,-1.02 -1.81,-2.40 -1.42,-0.43 -2.99,-1.02	-2.40,-1.38 -2.40,-0.97 -1.79,-2.40 -1.38,-0.35 -3.01,-0.97	-2.41,-1.42 -2.40,-1.03 -1.82,-2.41 -1.42,-0.44 -3.00,-1.03

To evaluate *prediction consistency*, we calculate the difference of the measured quantity $\delta L_v^{(i)}$ and the posterior mean $E[\delta L_{v,p} \mid \mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)}]$ of the robust predicted response (measured in terms of the number of posterior standard deviations $\sqrt{Var[\delta L_{v,p} \mid \mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)}]}$) as follows:

$$c_{v,j}^{(i)} = \frac{\delta L_v^{(i)} - E[\delta L_{v,p} \mid \mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)}]}{\sqrt{Var[\delta L_{v,p} \mid \mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)}]}} \tag{5.27}$$

where

$$E[\delta L_{v,p} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}] = \int E[\delta L_{v,p} \mid \boldsymbol{\theta}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}] p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) d\boldsymbol{\theta}$$
$$= \int \mu_v(\boldsymbol{\theta}) p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)}) d\boldsymbol{\theta} = K_v \int \mu_s p(\mu_s \mid \mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)}) d\mu_s \approx \frac{K_v}{K} \sum_{k=1}^{K} \mu_s^{(k)} \tag{5.28}$$

where $\mu_s^{(k)}$ is the first component of $\boldsymbol{\theta}^{(k)}$, where $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K, are posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)})$. The variance in (5.27) is given by:

$$Var[\delta L_{v,p} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}] = E[\delta L_{v,p}{}^2 \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}] - E^2[\delta L_{v,p} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}] \tag{5.29}$$

where

$$E[\delta L_{v,p}{}^2 \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}] = \int E[\delta L_{v,p}{}^2 \mid \boldsymbol{\theta}, \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}] p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) d\boldsymbol{\theta}$$
$$= \int (\mu_v{}^2(\boldsymbol{\theta}) + \sigma_{v,j}{}^2(\boldsymbol{\theta})) p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)}) d\boldsymbol{\theta} \approx \frac{1}{K} \sum_{k=1}^{K} [\mu_v{}^2(\boldsymbol{\theta}^{(k)}) + \sigma_{v,j}{}^2(\boldsymbol{\theta}^{(k)})] \tag{5.30}$$

where $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K, are posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)})$. The last rows of Table 5.3 show the results for $c_{v,j}^{(i)}$, for j=1, 2, 3. It can be seen from these tables that the model classes $\mathcal{M}_j^{(2)}$ (and also $\mathcal{M}_{j+1}^{(1)}$) updated just using data $\mathcal{D}_1^{(3)}$, j=1, 2, 3, are sufficiently consistent since the results are all within about 3 standard deviations.

Using data $\mathcal{D}_2^{(3)}$, which is modeled as stochastically independent of $\mathcal{D}_1^{(3)}$ given $\boldsymbol{\theta}$, one can update uncertainties in $\boldsymbol{\theta}$ for all surviving model classes using Bayes' Theorem with $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)})$ as the prior (recall that in this case, $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) = p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{M}_{j+1}^{(1)})$):

$$p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)}) = c_2^{-1} p(\mathcal{D}_2^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_j^{(2)}) p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) \tag{5.31}$$

where the likelihood function is given by:

$$p(\mathcal{D}_2^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_j^{(2)}) = \frac{1}{(2\pi\sigma_{v,j}(\boldsymbol{\theta})^2)^{N_v/2}} \exp(-\frac{1}{2\sigma_{v,j}{}^2(\boldsymbol{\theta})} \sum_{i=1}^{N_v} (\delta L_v{}^{(i)} - \mu_v(\boldsymbol{\theta}))^2 \tag{5.32}$$

and the evidence $p(\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)}|\mathcal{M}_j^{(2)})$ for model class $\mathcal{M}_j^{(2)}$ provided by the data $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$ is given by:

$$p(\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)} \mid \mathcal{M}_j^{(2)}) = p(\mathcal{D}_1^{(3)} \mid \mathcal{M}_j^{(2)})\, p(\mathcal{D}_2^{(3)} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) \tag{5.33}$$

where $p(\mathcal{D}_1^{(3)}|\mathcal{M}_j^{(2)})$ is equal to $p(\mathcal{D}_1^{(3)}|\mathcal{M}_{j+1}^{(1)})$ which has already been determined from previous analyses, while $p(\mathcal{D}_2^{(3)}|\mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)})$ is given by:

$$p(\mathcal{D}_2^{(3)} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) = \int p(\mathcal{D}_2^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_j^{(2)})\, p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{M}_j^{(2)}) d\boldsymbol{\theta} \tag{5.34}$$

which is determined using the stochastic simulation method in Appendix 5A as before. The samples from the prior $p(\boldsymbol{\theta}|\mathcal{D}_1^{(l)}, \mathcal{M}_j^{(2)})$ (calibration test posterior $p(\boldsymbol{\theta}|\mathcal{D}_1^{(l)}, \mathcal{M}_{j+1}^{(1)})$), obtained from the previous analyses, are used.

Table 5.4 shows the statistical results using data $\mathcal{D}_2^{(3)}$ in addition to $\mathcal{D}_1^{(3)}$. Compared to Table 5.2, it can be seen that the posterior c.o.v. of the parameters updated using additional data $\mathcal{D}_2^{(l)}$ is reduced somewhat for σ_s^2, l_s. The posterior means of the parameters σ_s^2 and l_s using $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$ are significantly higher than the means using only $\mathcal{D}_1^{(3)}$. There are several possible reasons: 1) additional information is provided by the additional data $\mathcal{D}_2^{(3)}$; and 2) uncertainties of the estimators due to a finite number of samples used in the stochastic simulation. Similar to before, it can be seen from the posterior correlation coefficient matrix that there is only weak correlation between most pairs of parameters. The posterior mean of r in $\mathcal{M}_3^{(2)}$ is 1.79 but the uncertainty in r is significant (40% c.o.v.). The results show that given both $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$, $\mathcal{M}_1^{(2)}$, $\mathcal{M}_2^{(2)}$ and $\mathcal{M}_3^{(2)}$ are significantly probable. Thus, based on the calibration data and validation data, all the model classes $\mathcal{M}_1^{(2)}$, $\mathcal{M}_2^{(2)}$ and $\mathcal{M}_3^{(2)}$ are considered in subsequent analyses.

It can also be seen that the predicted robust failure probability $P(F|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_2^{(2)})$ of the target frame structure using model class $\mathcal{M}_2^{(2)}$ is smaller than that using model classes $\mathcal{M}_1^{(2)}$ and $\mathcal{M}_3^{(2)}$. The predicted hyper-robust failure probability $P(F|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},M_2)$ is 1.25×10^{-5}.

Table 5.4 Statistical results using data $\mathcal{D}_2^{(3)}$ from the validation experiment in addition to $\mathcal{D}_1^{(3)}$

		$\mathcal{M}_1^{(2)}$	$\mathcal{M}_2^{(2)}$	$\mathcal{M}_3^{(2)}$
Statistics of parameters	μ_s (Pa^{-1})	8.70×10^{-11};0.62%	8.68×10^{-11};0.63%	8.68×10^{-11};0.6%
	σ_s^2(Pa^{-2})	6.00×10^{-23};17.2%	5.80×10^{-23};17.8%	5.70×10^{-23};20.1%
	l_s(m)	0.0383;25.4%	0.0384;19.0%	0.0398;25.6%
	r			1.79;39.6%
	R	$\begin{bmatrix} 1 & 0.02 & 0.08 \\ & 1 & -0.14 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.04 & 0.10 \\ & 1 & -0.10 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.02 & 0.10 & -0.13 \\ & 1 & -0.28 & 0.26 \\ & & 1 & -0.41 \\ & & & 1 \end{bmatrix}$
Log evidence		1174.56	1173.82	1173.83
$E[\ln p(\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)}\|\boldsymbol{\theta},\mathcal{M}_j^{(2)})]$		1182.70	1182.83	1182.72
Expected Information gain		8.14	9.01	8.90
$P(\mathcal{M}_j^{(2)}\|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},M_2)$		0.510	0.244	0.246
$P(F\|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)})$		1.32×10^{-5}(20.6%)	3.43×10^{-6}(32.1%)	1.99×10^{-5}(22.2%)
$P(F\|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},M_2)$		1.25×10^{-5}		

Table 5.5 Consistency assessment of model classes in predicting δL_v using data $\mathcal{D}_2^{(3)}$ from the validation experiment in addition to $\mathcal{D}_1^{(3)}$ from the calibration experiment

	$\mathcal{M}_1^{(2)}$	$\mathcal{M}_2^{(2)}$	$\mathcal{M}_3^{(2)}$
$\dfrac{\delta L_v^{(i)} - E[\delta L_{v,p}\mid\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)}]}{\sqrt{Var[\delta L_{v,p}\mid\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)}]}}$	-1.30,-0.43, -1.30,-0.08, -0.78,-1.30, -0.43,0.44, -1.83,-0.08	-1.34,-0.42, -1.34,-0.05, -0.79,-1.34, -0.42,0.50, -1.90,-0.05	-1.33,-0.43, -1.33,-0.07, -0.79,-1.33, -0.43,0.47, -1.87,-0.07

Table 5.5 shows the results for checking, using the following index, the consistency of the model classes $\mathcal{M}_j^{(2)}$, j =1, 2, 3, in predicting the response δL_v using data $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$:

$$\frac{\delta L_v^{(i)} - E[\delta L_{v,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)}]}{\sqrt{Var[\delta L_{v,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)}]}} \tag{5.35}$$

where $E[\delta L_{v,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)}]$ and $Var[\delta L_{v,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)}]$ can be determined using (5.28), (5.29) and (5.30) except that the samples from the most recently updated posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)})$ are used instead of $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{M}_j^{(2)})$. By comparing Table 5.3 and Table 5.5, it can be seen that the consistency of the model classes improves over the case without data $\mathcal{D}_2^{(3)}$, with the ratios in (5.35) all being less than 2 standard deviations.

The accuracy of the model classes $\mathcal{M}_j^{(2)}$, j =1, 2, 3, in predicting δL_v using data $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$ can be assessed, similar to the case without data $\mathcal{D}_2^{(3)}$, by evaluating i) $P(e_{v,p}^{(i)} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)})$, i=1, 2…, N_v, which can be determined using (5.26) except that the samples from the most recently updated posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)})$ are used instead, and ii) the average prediction error probability $P(e_{v,p} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)})$ of a model class updated using data $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$, which can be obtained by taking the arithmetic mean of $P(e_{v,p}^{(i)} \le b\% | \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)})$, i=1, 2…, N_v. The corresponding results are not shown here for brevity but they show high probability that the prediction errors for each model class will be less than 5%, with even higher probabilities for 10% (see Cheung and Beck (2008b) for details).

5.2.3 Using data $\mathcal{D}_3$ from the accreditation experiment

Candidate model classes for the subsystem in the accreditation experiment are $\mathcal{M}_j^{(3)}$, j=1,2,3. The uncertain parameters $\boldsymbol{\theta}^{(3,j)}$ for $\mathcal{M}_j^{(3)}$ are the same as $\boldsymbol{\theta}^{(2,j)}$ for $\mathcal{M}_j^{(2)}$. The "prior" PDF $p(\boldsymbol{\theta}^{(3,j)}|\mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)})$ for $\mathcal{M}_j^{(3)}$ is given by the "posterior" PDF $p(\boldsymbol{\theta}^{(2,j)}|\mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)})$ for $\mathcal{M}_j^{(2)}$. Similar analyses to the above are carried out as follows. Data $\mathcal{D}_3^{(3)} = \{w_a^{(i)}, i=1, 2\}$

from the accreditation experiment are used to investigate the predictive performance of the model classes. The probability that the response $w_{a,p}$ (the vertical displacement of point Q of the frame structure in the accreditation experiment) predicted using the model classes updated by data from the previous two experiments is within a certain b% of the measured quantity $w_a^{(i)}$ is given by the following updated robust predictive PDF conditioned on $\mathcal{D}_1^{(3)}$and $\mathcal{D}_2^{(3)}$:

$$\begin{aligned} P(e_{a,p}{}^{(i)} \le b\% \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}) &= \int P(e_{a,p}{}^{(i)} \le b\% \mid \boldsymbol{\theta}, \mathcal{M}_j^{(3)})\, p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}) d\boldsymbol{\theta} \\ &= \int P(e_{a,p}{}^{(i)} \le b\% \mid \boldsymbol{\theta}, \mathcal{M}_j^{(3)})\, p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)}) d\boldsymbol{\theta} \end{aligned} \tag{5.36}$$

where

$$e_{a,p}{}^{(i)} = \left| \frac{w_{a,p} - w_a^{(i)}}{w_a^{(i)}} \right| \tag{5.37}$$

For the model class $\mathcal{M}_j^{(3)}$, j=1, 2, 3, given $\boldsymbol{\theta}$, it can be shown that the response $w_{a,p}$ follows a Gaussian distribution with mean $\mu_a = K_a \mu_s$ and variance $\sigma_{a,j}{}^2 = \sigma_s{}^2 s_{a,j}(l_s, r)$ where K_a is given as follows:

$$K_a = \frac{1}{2}\left[\frac{F_1 L_1}{A_1} - \sqrt{2}\left(\frac{F_2 L_2}{A_2} + \frac{F_4 L_4}{A_4}\right)\right] - \frac{F_a L_1{}^3}{48 I} \tag{5.38}$$

The expression for $s_{a,j}$ is given in Appendix III in Cheung and Beck (2008b). Thus,

$$P(e_{a,p}^{(i)} \le b\% \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)})$$
$$= \mathrm{sgn}(w_a^{(i)}) \int [\Phi(\frac{(1+\frac{b}{100})w_a^{(i)} - \mu_a(\boldsymbol{\theta})}{\sigma_{a,j}(\boldsymbol{\theta})}) - \Phi(\frac{(1-\frac{b}{100})w_a^{(i)} - \mu_a(\boldsymbol{\theta})}{\sigma_{a,j}(\boldsymbol{\theta})})] p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(2)}) d\boldsymbol{\theta}$$
$$\approx \frac{\mathrm{sgn}(w_a^{(i)})}{K} \sum_{k=1}^{K} \Phi(\frac{(1+\frac{b}{100})w_a^{(i)} - \mu_a(\boldsymbol{\theta}^{(k)})}{\sigma_{a,j}(\boldsymbol{\theta}^{(k)})}) - \Phi(\frac{(1-\frac{b}{100})w_a^{(i)} - \mu_a(\boldsymbol{\theta}^{(k)})}{\sigma_{a,j}(\boldsymbol{\theta}^{(k)})}) \tag{5.39}$$

where $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K, are posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)})$.

Table 5.6 shows the results for $P(e_{a,p}^{(i)} \le b\% | \mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$ (the numbers outside the parenthesis) and the average prediction error probability $P(e_{a,p} \le b\% | \mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$ (the numbers inside the parenthesis), for j=1, 2, 3, and b=5 and 10 using $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$. It can be seen that, the model classes $\mathcal{M}_j^{(3)}$ (and so $\mathcal{M}_j^{(2)}$), j=1, 2, 3, updated using $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$, are sufficiently accurate. It is noted that all $P(e_{a,p} \le 5\% | \mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$ are larger than 0.84 implying that there is a high probability for the response prediction by the model classes to be within 5% of the actual response measurements.

Table 5.6 Results of predicting w_a using data $\mathcal{D}_2^{(3)}$ from the validation experiment in addition to $\mathcal{D}_1^{(3)}$ from the calibration experiment

	$\mathcal{M}_1^{(3)}$	$\mathcal{M}_2^{(3)}$	$\mathcal{M}_3^{(3)}$
$P(\lvert w_{a,p} - w_a^{(i)} \rvert / \lvert w_a^{(i)} \rvert \le 5\% \mid \mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$	0.896,0.788 (0.842)	0.907,0.782 (0.844)	0.902,0.795 (0.848)
$P(\lvert w_{a,p} - w_a^{(i)} \rvert / \lvert w_a^{(i)} \rvert \le 10\% \mid \mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$	0.997,0.992 (0.994)	0.999,0.995, (0.997)	0.9995,0.994 (0.997)
$\frac{w_a^{(i)} - E[w_{a,p} \mid \mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)}]}{\sqrt{Var[w_{a,p} \mid \mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)}]}}$	0.26,-0.89	0.24,-0.96	0.26,-0.94

The difference between the measured quantity $w_a^{(i)}$ and the posterior mean $E[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]$ of the robust predicted response (measured in terms of the number of posterior standard deviations $\sqrt{Var[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]}$) is given by:

$$c_{a,j}^{(i)} = \frac{w_a^{(i)} - E[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]}{\sqrt{Var[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]}} \tag{5.40}$$

where $E[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]$ and $Var[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]$ can be calculated using (5.28), (5.29) and (5.30) with $\delta L_v^{(i)}$ replaced by $w_a^{(i)}$, $\delta L_{v,p}$ by $w_{a,p}$, $\mathcal{D}_1^{(3)}$ by $\mathcal{D}_1^{(3)}$, $\mathcal{D}_2^{(3)}$, $\mathcal{M}_j^{(2)}$ by $\mathcal{M}_j^{(3)}$, the subscript "v" replaced by "a" and where $\mu_s^{(k)}$ is the first component of $\boldsymbol{\theta}^{(k)}$, where $\boldsymbol{\theta}^{(k)}$, k=1,2,...,K, are posterior samples from $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)})$. The last row of Table 5.6 shows the results for $c_{a,j}^{(i)}$, for j=1, 2, 3. It can be seen from this table that the model classes $\mathcal{M}_j^{(3)}$, j=1, 2,3, (and so $\mathcal{M}_j^{(2)}$) updated using $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$ are sufficiently consistent since the results are all within a standard deviation.

Using data $\mathcal{D}_3^{(3)}$, which is modelled as stochastically independent of $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$ given $\boldsymbol{\theta}$, one can update the uncertainties in $\boldsymbol{\theta}$ for all the model classes using Bayes' Theorem with the previous posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(2)})$ as the prior $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$:

$$p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)}) = c_3^{-1} p(\mathcal{D}_3^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_j^{(3)}) p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}) \tag{5.41}$$

where the likelihood function is given by (with N_a=2):

$$p(\mathcal{D}_3^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_j^{(3)}) = \frac{1}{(2\pi\sigma_{a,j}(\boldsymbol{\theta})^2)^{N_a/2}} \exp(-\frac{1}{2\sigma_{a,j}^2(\boldsymbol{\theta})} \sum_{i=1}^{N_a} (w_a^{(i)} - \mu_a(\boldsymbol{\theta}))^2 \tag{5.42}$$

The evidence $p(\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)}|\mathcal{M}_j^{(3)})$ for model class $\mathcal{M}_j^{(3)}$ that is provided by the data $\mathcal{D}_1^{(3)}$, $\mathcal{D}_2^{(3)}$ and $\mathcal{D}_3^{(3)}$ is given by:

$$p(\mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)} \mid \mathcal{M}_j^{(3)}) = p(\mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)} \mid \mathcal{M}_j^{(3)})\, p(\mathcal{D}_3^{(3)} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}) \quad (5.43)$$

where $p(\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)}|\mathcal{M}_j^{(3)})$ has already been determined and $p(\mathcal{D}_3^{(3)}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$ is given by:

$$p(\mathcal{D}_3^{(3)} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}) = \int p(\mathcal{D}_3^{(3)} \mid \boldsymbol{\theta}, \mathcal{M}_j^{(3)})\, p(\boldsymbol{\theta} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}) d\boldsymbol{\theta} \quad (5.44)$$

which is determined using the same stochastic simulation method as before. The samples from the prior $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{M}_j^{(3)})$ obtained from the previous analyses are used.

Table 5.7 Statistical results using data $\mathcal{D}_3^{(3)}$ from the accreditation experiment in addition to $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$

		$\mathcal{M}_1^{(3)}$	$\mathcal{M}_2^{(3)}$	$\mathcal{M}_3^{(3)}$
Statistics of parameters	μ_s (Pa^{-1})	8.69×10^{-11};0.57%	8.69×10^{-11};0.59%	8.69×10^{-11};0.6%
	σ_s^2(Pa^{-2})	5.88×10^{-23};18.0%	5.75×10^{-23};17.5%	5.61×10^{-23};20.0%
	l_s(m)	0.0374;25.5%	0.0378;18.9%	0.0392;26.5%
	r			1.81;40.4%
	R	$\begin{bmatrix} 1 & -0.04 & 0.06 \\ & 1 & -0.17 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.06 & 0.09 \\ & 1 & -0.19 \\ & & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & -0.06 & 0.16 & -0.18 \\ & 1 & -0.30 & 0.21 \\ & & 1 & -0.36 \\ & & & 1 \end{bmatrix}$
Log evidence		1193.94	1193.21	1193.21
$P(\mathcal{M}_j^{(3)}\|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},M_3)$		0.510	0.245	0.245
$P(F\|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},\mathcal{M}_j^{(3)})$		8.98×10^{-6}(11.8%)	1.29×10^{-6}(16.6%)	2.68×10^{-5}(20.0%)
$P(F\|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},M_3)$		1.14×10^{-5}		

The system involved in this accreditation experiment is a lot more complicated than the one in the validation experiment. In practice, one may want to introduce additional parameters to take into account the additional uncertainties involved. Nonetheless, for illustration, we

have kept the same number of uncertain parameters as before, which is consistent with the statement of the validation challenge problem, and used data $\mathcal{D}_3^{(3)}$ to update the uncertainties in the parameters. Table 5.7 shows the statistical results using data $\mathcal{D}_3^{(3)}$ in addition to the data $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$ from the previous experiments. Compared to Tables 5.2 and 5.4, some of the differences observed in the posterior mean, c.o.v. and correlation coefficient of parameters are due to: 1) additional information provided by the additional data $\mathcal{D}_3^{(3)}$; and 2) uncertainties of the estimators due to a finite number of samples used in stochastic simulation. Similar to before, it can be seen from the posterior correlation coefficient matrix that there is only weak correlation between most pairs of parameters. The posterior means of r in $\mathcal{M}_3^{(3)}$ is 1.81 but the uncertainty in r is still significant since $\mathcal{D}_3^{(3)}$ provides only 2 additional data. The results show that given $\mathcal{D}_1^{(3)}$, $\mathcal{D}_2^{(3)}$ and $\mathcal{D}_3^{(3)}$, $\mathcal{M}_1^{(3)}$, $\mathcal{M}_2^{(3)}$ and $\mathcal{M}_3^{(3)}$ are significantly probable and the posterior probabilities are essentially unchanged from Table 5.4. Thus, all of the model classes $\mathcal{M}_1^{(3)}$, $\mathcal{M}_2^{(3)}$ and $\mathcal{M}_3^{(3)}$ are utilized to make robust predictions.

It can also be seen from Table 5.7 that the predicted robust failure probability $P(F|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},\mathcal{M}_2^{(3)})$ of the target frame structure using model class $\mathcal{M}_2^{(3)}$ is again smaller than that using model classes $\mathcal{M}_1^{(3)}$ and $\mathcal{M}_3^{(3)}$. The predicted hyper-robust failure probability $P(F|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},M_3)$ is 1.14×10^{-5}. By comparing Table 5.4 and Table 5.7, it can be seen that the predicted hyper-robust failure probability changes little compared to that based on only data $\mathcal{D}_1^{(3)}$ and $\mathcal{D}_2^{(3)}$. $P(F|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},\mathcal{M}_2^{(3)})P(\mathcal{M}_2^{(3)}|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},M_3)$ is small compared to $P(F|\mathcal{D}_1^{(3)},\mathcal{D}_2^{(3)},\mathcal{D}_3^{(3)},M_3)$ and thus the contribution of $\mathcal{M}_2^{(3)}$ to the prediction quantity of interest is small.

Table 5.8 shows the results for checking the consistency of the model classes $\mathcal{M}_j^{(3)}$, j =1, 2, 3, in predicting the response w_a using data $\mathcal{D}_1^{(3)}$, $\mathcal{D}_2^{(3)}$ and $\mathcal{D}_3^{(3)}$:

$$\frac{w_a^{(i)} - E[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)}]}{\sqrt{Var[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)} \mathcal{M}_j^{(3)}]}} \tag{5.45}$$

where $E[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)}]$ and $Var[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)}]$ can be determined by using the equations for calculating $E[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]$ and $Var[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{M}_j^{(3)}]$ except that the samples from the most recently updated posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)})$ are used instead. By comparing Table 5.6 and Table 5.8, it can be seen that the consistency of the model classes is similar to the case without data $\mathcal{D}_3^{(3)}$ since $\mathcal{D}_3^{(3)}$ provides only two additional data.

Table 5.8 Consistency assessment of model classes in predicting w_a using data $\mathcal{D}_3^{(3)}$ from the accreditation experiment in addition to $\mathcal{D}_1^{(3)}$ from the calibration experiment and $\mathcal{D}_2^{(3)}$ from the validation experiment

	$\mathcal{M}_1^{(3)}$	$\mathcal{M}_2^{(3)}$	$\mathcal{M}_3^{(3)}$
$\frac{w_a^{(i)} - E[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)}]}{\sqrt{Var[w_{a,p} \mid \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)}]}}$	0.30,-0.88	0.28,-0.94	0.28, -0.92

The accuracy of the model classes $\mathcal{M}_j^{(3)}$, j =1, 2, 3, in predicting w_a using data $\mathcal{D}_1^{(3)}$, $\mathcal{D}_2^{(3)}$ and $\mathcal{D}_3^{(3)}$ can be assessed, similar to the case without data $\mathcal{D}_3^{(3)}$, by evaluating i) $P(e_{a,p}^{(i)} \leq b\% | \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)})$, i=1, 2, which can be determined using (5.39) except that the samples from the most recently updated posterior PDF $p(\boldsymbol{\theta}|\mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)})$ are used instead, and ii) the average prediction error probability $P(e_{a,p} \leq b\% | \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)})$ of a model class updated using data $\mathcal{D}_1^{(3)}$, $\mathcal{D}_2^{(3)}$ and $\mathcal{D}_3^{(3)}$, which can be obtained by taking the arithmetic mean of $P(e_{a,p}^{(i)} \leq b\% | \mathcal{D}_1^{(3)}, \mathcal{D}_2^{(3)}, \mathcal{D}_3^{(3)}, \mathcal{M}_j^{(3)})$, i=1, 2. The corresponding results are not shown here for brevity but they show high prediction

accuracy (high probability for prediction errors less than 5%, with even higher probabilities for 10%) (see Cheung and Beck (2008b) for details).

5.3 Concluding remarks

A novel methodology based on Bayesian updating of hierarchical stochastic system model classes is proposed for uncertainty quantification, model updating, model selection, model validation and robust prediction of the response of a system for which some subsystems have been separately tested. It uses full Bayesian updating of the model classes, along with model class comparison and prediction consistency and accuracy assessment. In the proposed methodology, all the results are rigorously derived from the probability axioms and all the information in the available data are considered to make predictions. The concepts and computational tools of the proposed methodology are illustrated with a previously-studied validation challenge problem, although the methodology can handle a more general process of hierarchical subsystem testing.

As shown by the illustrative example, within a model class, there are many plausible models and the predictions of response and failure probability of the final system can often vary greatly from one model to another, showing that the consequences of the uncertainties in the parameters are significant. Ignoring the uncertainty in the modeling parameters and solely relying on the MAP model (corresponding to the maximum of the posterior PDF) or the MLE model (corresponding to the maximum likelihood parameter value) for predictions can be dangerous and misleading since such predictions can greatly underestimate the failure probability and the uncertainty in the response. It is shown how more robust predictions by a model class can be obtained by taking into account the predictions from all the plausible models in the model class where the plausibilities are quantified by their respective posterior PDF values.

Multiple model classes are investigated for the illustrative example. The response and failure probability prediction vary greatly from one model class to another. Hyper-robust

predictions of response and failure probability are also obtained by a weighted average of the robust predictions given by each model class where the weight is given by the posterior probability of the model class. The posterior probability of one of the candidate model classes is so small based on the calibration data that its contribution to the prediction is negligible, so it is discarded from further predictive analysis after the calibration tests.

The computational problems resulting from full Bayesian updating of hierarchical model classes, as well as model class comparison, can be challenging, especially for problems with many uncertain parameters. A number of powerful computational tools based on stochastic simulation are used to solve efficiently the computational problems involved; in particular, for the illustrative example studied, the Hybrid Gibbs TMCMC algorithm worked well.

If a model class performs well in predicting the response for the subsystems involved in all of the experiments, one can gain more confidence in its predictive performance for the final constructed system. However, it should be stressed that 1) whether the predictive performance of the model classes is acceptable or not depends on which criteria the decision maker thinks are critical, and 2) there is no guarantee that a model class which performs well enough to satisfy the selected criteria in predicting the response of the subsystems in these experiments will always predict the response of the final system well, especially in the case where some of the uncertainties in the final system which are critical to the prediction are not present in the subsystem tests (for example, there can be uncertainties in support or joint conditions in the final system, and uncertainties in input loadings, such as stronger amplitude inputs which may be experienced by the final system that cause it to behave very differently than the subsystems during their tests).

Although it did not occur in the illustrative example, in the case where all candidate model classes give poor performance in predicting the response for subsystems involved in an experiment, one should check whether some of the uncertainties have not been adequately

modeled in the failing subsystem tests and, if so, modify the candidate model classes to properly take into account these uncertainties.

To test the performance of the proposed methodology, future work should use data collected from real systems, preferably with a larger degree of complexity than the one considered in the illustrative example of this paper.

Appendix 5A: Hybrid Gibbs TMCMC algorithm for posterior sampling

Part of our methodology involves a sequential update of the posterior PDF given the data from the experiments collected from the subsystems. The following algorithm is proposed for this purpose. At the end of the experiment where data are collected from the i-th subsystem, we need to characterize $p(\boldsymbol{\theta}|D_i,\mathcal{M}_j^{(i)})$ given the data $\mathcal{D}_i$ collected from the most current subsystem experiment and all the data $D_{i\text{-}1} =\{\mathcal{D}_1,\ldots, \mathcal{D}_{i\text{-}1}\}$ collected from the previous subsystem experiments, where $D_i = D_{i\text{-}1} \cup \mathcal{D}_i$. The prior PDF corresponding to this posterior PDF is $p(\boldsymbol{\theta}|D_{i\text{-}1},\mathcal{M}_j^{(i)})$ from which samples have been previously generated and the evidences $p(D_{i\text{-}1}|\mathcal{M}_j^{(i)})$ for each model class $\mathcal{M}_j^{(i)}$ which have been obtained. Note that in the analysis below, we use the conventions $p(\boldsymbol{\theta}|D_0,\mathcal{M}_j^{(i)}) = p(\boldsymbol{\theta}|\mathcal{M}_j^{(i)})$ and $p(D_0|\mathcal{M}_j^{(i)})=1$.

For a given $\boldsymbol{\theta}$, $\mathcal{D}_1,\ldots, \mathcal{D}_i$ are modeled as stochastically independent. We propose a hybrid approach making use of the TMCMC method (Ching and Chen 2007), Metropolis Hastings algorithm and Gibbs sampling to generate samples from the posterior PDF $\pi(\boldsymbol{\theta})=p(\boldsymbol{\theta}|D_i,\mathcal{M}_j^{(i)})= p(\mathcal{D}_i|\boldsymbol{\theta},\mathcal{M}_j^{(i)})p(\boldsymbol{\theta}|D_{i\text{-}1},\mathcal{M}_j^{(i)})/p(\mathcal{D}_i|D_{i\text{-}1},\mathcal{M}_j^{(i)})$ and to calculate the evidence $p(\mathcal{D}_i|D_{i\text{-}1},\mathcal{M}_j^{(i)})$.

Consider a sequence of intermediate PDFs $\pi_l(\boldsymbol{\theta})$ for l=0,1,…, L, such that the first and last PDFs, $\pi_0(\boldsymbol{\theta})$ and $\pi_L(\boldsymbol{\theta}) = \pi(\boldsymbol{\theta})$, in the sequence are the prior $p(\boldsymbol{\theta}|D_{i\text{-}1},\mathcal{M}_j^{(i)})$ and posterior $p(\boldsymbol{\theta}|D_i,\mathcal{M}_j^{(i)})$, respectively:

$$\pi_l(\boldsymbol{\theta}) \propto p^{\tau_l}(\mathcal{D}_i \mid \boldsymbol{\theta}, \mathcal{M}_j^{(i)}) p(\boldsymbol{\theta} \mid D_{i-1}, \mathcal{M}_j^{(i)}) \tag{A5.1}$$

where $0=\tau_0<\tau_1<\ldots<\tau_L=1$. Divide $\boldsymbol{\theta}$ into B groups of components. Denote the b-th component group of $\boldsymbol{\theta}$ as $\boldsymbol{\theta}_b$.

First, N_0 samples are generated from the prior $p(\boldsymbol{\theta}|D_{i-1},\mathcal{M}_j^{(i)})$. Then do the following procedures for $l=1,\ldots,L$. At the beginning of the l-th level, we have the samples $\boldsymbol{\theta}_{l-1}^{(m)}$, $m=1,2,\ldots,N_{l-1}$, from $\pi_{l-1}(\boldsymbol{\theta})$. First, select τ_l such that the effective sample size $1/\sum_{s=1}^{N_{l-1}} \tilde{w}_s^{\,2} =$ some threshold (e.g., 0.9 N_{l-1}) (Cheung and Beck 2008c; Chapter 2 in this thesis), where $\tilde{w}_s = w_s / \sum_{s=1}^{N_{l-1}} w_s$ and $w_s = p^{\tau_l - \tau_{l-1}}(\mathcal{D}_i \mid \boldsymbol{\theta}_{l-1}^{(s)}, \mathcal{M}_j)$, $s=1,2,\ldots,N_{l-1}$. If $\tau_l>1$, then set $L=l$ and $\tau_l=1$, then recompute w_s and $\tilde{w}_s$. Compute an estimate for the sample covariance matrix for $\pi_l(\boldsymbol{\theta})$ as follows:

$$\Sigma = \sum_{m=1}^{N_{l-1}} \tilde{w}_m (\boldsymbol{\theta}_{l-1}^{(m)} - \overline{\boldsymbol{\theta}})(\boldsymbol{\theta}_{l-1}^{(m)} - \overline{\boldsymbol{\theta}})^T, \ \overline{\boldsymbol{\theta}} = \sum_{m=1}^{N_{l-1}} \tilde{w}_m \boldsymbol{\theta}_{l-1}^{(m)} \tag{A5.2}$$

Set $E_l = \sum_{s=1}^{N_{l-1}} w_s / N_{l-1}$. Then the N_l samples $\boldsymbol{\theta}_l^{(n)}$ from $\pi_l(\boldsymbol{\theta})$ are generated by doing the following for $n=1,2,\ldots,N_l$:

1. Draw a number s' from a discrete distribution $p(S=s)=\tilde{w}_s$, $s=1,2,\ldots,N_{l-1}$.
2. Fixing the last component group of $\boldsymbol{\theta}$ at the values of $\boldsymbol{\theta}_{l-1,B}^{(s')}$, draw the samples $\boldsymbol{\theta}_{l,1}^{(n)}$, …, $\boldsymbol{\theta}_{l,B-1}^{(n)}$ for the first B-1 component groups of $\boldsymbol{\theta}$, one after another, using Gibbs sampling as described later. Set $\boldsymbol{\theta}_{l-1,b}^{(s')} = \boldsymbol{\theta}_{l,b}^{(n)}$ for $b=1,\ldots,B$-1.

3. Fixing the first B-1 component groups at the values of $\boldsymbol{\theta}_{l,1}^{(n)}$, …, $\boldsymbol{\theta}_{l,B-1}^{(n)}$, generate a sample $\boldsymbol{\theta}_{l,B}^{(n)}$ for the last component group of $\boldsymbol{\theta}$ by the Metropolis-Hastings algorithm: Generate $\boldsymbol{\theta}^*$ from a Gaussian PDF with mean $\boldsymbol{\theta}_{l-1,B}^{(s')}$ and covariance matrix $\eta\Sigma_B$ where Σ_B is the submatrix that corresponds to the last component group (i.e., the B-th component group) in the covariance matrix Σ. Compute the acceptance probability $r''=\min\{r',1\}$ where r' is given by:

$$r' = \frac{p^{\tau_i}(\mathcal{D}_i \mid \boldsymbol{\theta}_{l,1}^{(n)},...,\boldsymbol{\theta}_{l,B-1}^{(n)},\boldsymbol{\theta}^*,\mathcal{M}_j^{(i)})}{p^{\tau_i}(\mathcal{D}_i \mid \boldsymbol{\theta}_{l,1}^{(n)},...,\boldsymbol{\theta}_{l,B-1}^{(n)},\boldsymbol{\theta}_{l-1,B}^{(s')},\mathcal{M}_j^{(i)})} \times \frac{[\prod_{t=1}^{i-1} p(\mathcal{D}_t \mid \boldsymbol{\theta}_{l,1}^{(n)},...,\boldsymbol{\theta}_{l,B-1}^{(n)},\boldsymbol{\theta}^*,\mathcal{M}_j^{(i)})]p(\boldsymbol{\theta}_{l,1}^{(n)},...,\boldsymbol{\theta}_{l,B-1}^{(n)},\boldsymbol{\theta}^* \mid \mathcal{M}_j^{(i)})}{[\prod_{t=1}^{i-1} p(\mathcal{D}_t \mid \boldsymbol{\theta}_{l,1}^{(n)},...,\boldsymbol{\theta}_{l,B-1}^{(n)},\boldsymbol{\theta}_{l-1,B}^{(s')},\mathcal{M}_j^{(i)})]p(\boldsymbol{\theta}_{l,1}^{(n)},...,\boldsymbol{\theta}_{l,B-1}^{(n)},\boldsymbol{\theta}_{l-1,B}^{(s')} \mid \mathcal{M}_j^{(i)})} \tag{A5.3}$$

If $r''>U(0,1)$ where $U(0,1)$ is a uniformly distributed number between 0 and 1, $\boldsymbol{\theta}_{l,B}^{(n)}=\boldsymbol{\theta}^*$, $\boldsymbol{\theta}_{l-1,B}^{(s')}=\boldsymbol{\theta}^*$. Otherwise, $\boldsymbol{\theta}_{l,B}^{(n)}=\boldsymbol{\theta}_{l-1,B}^{(s')}$.

Thus, the n-th sample for $\boldsymbol{\theta}$ with the target PDF $\pi_l(\boldsymbol{\theta})$ is given by $\boldsymbol{\theta}_l^{(n)}=[\boldsymbol{\theta}_{l,1}^{(n)}\ \boldsymbol{\theta}_{l,2}^{(n)}\ ....\boldsymbol{\theta}_{l,B}^{(n)}]$.

In step 3, η (e.g., 0.2^2) is chosen such that the average acceptance probability is larger than some threshold (e.g., 0.7). Other MCMC algorithms such as Hybrid Monte Carlo methods (Cheung and Beck 2007, 2008a; Chapter 2 in this thesis) can also be used in place of the Metropolis-Hastings algorithm in step 3 for more effective sampling, as is done in Cheung and Beck (2008e, f; Chapter 3 in this thesis). The evidence $p(\mathcal{D}_i|D_{i-1},\mathcal{M}_j^{(i)})$ for $\mathcal{M}_j^{(i)}$ given by data $\mathcal{D}_i$ can be estimated as follows:

$$p(\mathcal{D}_i \mid D_{i-1}, \mathcal{M}_j^{(i)}) \approx \prod_{l=1}^{L} E_l \tag{A5.4}$$

Gibbs sampling for the posterior PDF in the illustrative example with data D_1 (i=1)

Now we describe how Gibbs sampling can be performed for the posterior PDF in the illustrative example with data D_1 (i=1). For $\mathcal{M}_1^{(1)}$ (i=1, j=1), $\boldsymbol{\theta}$ is divided into 2 component groups: $\boldsymbol{\theta}_1 = \mu_s$, $\boldsymbol{\theta}_2 = [\sigma_s^2 \ \sigma_\varepsilon^2]$. Gibbs sampling in step 2 of the above algorithm is performed on the first component group as follows: draw $\boldsymbol{\theta}_{l,1}^{(n)}$ from a truncated Gaussian PDF (constrained to be positive) which is proportional to a Gaussian distribution with mean μ and variance σ^2 given below:

$$\mu = \frac{H_{11}\frac{F_c L_c}{A_c}\sum_{i=1}^{N_c}\delta L_c^{(i)} + H_{12}\left(\frac{F_c L_c}{A_c}\sum_{k=1}^{N_c} S_c^{(k)}(L_c/2) + \sum_{k=1}^{N_c}\delta L_c^{(k)}\right) + H_{22}\sum_{k=1}^{N_c} S_c^{(k)}(L_c/2) + \frac{\mu_0}{\tau_l \sigma_0^{\ 2}}}{N_c\left(H_{11}\left(\frac{F_c L_c}{A_c}\right)^2 + 2H_{12}\frac{F_c L_c}{A_c} + H_{22}\right) + \frac{1}{\tau_l \sigma_0^{\ 2}}} \tag{A5.5}$$

$$\sigma^2 = \frac{1}{\tau_l\left[N_c\left(H_{11}\left(\frac{F_c L_c}{A_c}\right)^2 + 2H_{12}\frac{F_c L_c}{A_c} + H_{22}\right) + \frac{1}{\tau_l \sigma_0^{\ 2}}\right]} \tag{A5.6}$$

where H_{11}, H_{12} and H_{22} are the (1,1), (1,2) and (2,2) entries of the inverse of $\mathbf{C}(\sigma_s^{\ 2}, \sigma_\varepsilon^{\ 2})$ in equation (5.13) with $[\sigma_s^2 \ \sigma_\varepsilon^2] = \boldsymbol{\theta}_{l-1,2}^{(s')}$; μ_0 and $\sigma_0^{\ 2}$ are the mean and variance of the prior PDF $p(\mu_s|\mathcal{M}_j^{(1)})$ of μ_s respectively

For $\mathcal{M}_4^{(1)}$ (i=1, j=4), $\boldsymbol{\theta}$ is divided into 3 component groups: $\boldsymbol{\theta}_1 = \mu_s$, $\boldsymbol{\theta}_2 = \sigma_s^2$, $\boldsymbol{\theta}_3 = [l_s^2 \ r]$. Gibbs sampling in step 2 of the proposed algorithm is performed on the first two component groups as follows: draw $\boldsymbol{\theta}_{l,1}^{(n)}$ from a truncated Gaussian PDF (constrained to be

positive) which is proportional to a Gaussian distribution with mean μ' and variance σ'^2 given below:

$$\mu' = \frac{H_{11}\frac{F_c L_c}{A_c}\sum_{k=1}^{N_c}\delta L_c^{(k)} + H_{12}(\frac{F_c L_c}{A_c}\sum_{k=1}^{N_c}S_c^{(k)}(L_c/2) + \sum_{k=1}^{N_c}\delta L_c^{(k)}) + H_{22}\sum_{k=1}^{N_c}S_c^{(k)}(L_c/2) + \frac{\sigma_s^{\,2}\mu_0}{\tau_l\sigma_0^{\,2}}}{N_c(H_{11}(\frac{F_c L_c}{A_c})^2 + 2H_{12}\frac{F_c L_c}{A_c} + H_{22}) + \frac{\sigma_s^{\,2}}{\tau_l\sigma_0^{\,2}}} \tag{A5.7}$$

$$\sigma'^2 = \frac{\sigma_s^{\,2}}{\tau_l[N_c(H_{11}(\frac{F_c L_c}{A_c})^2 + 2H_{12}\frac{F_c L_c}{A_c} + H_{22}) + \frac{\sigma_s^{\,2}}{\tau_l\sigma_0^{\,2}}]} \tag{A5.8}$$

In the above equations, $\sigma_s^2 = \boldsymbol{\theta}_{l-1,2}^{(s')}$ and H_{11}, H_{12} and H_{22} are the (1,1), (1,2) and (2,2) entries of the inverse of $\mathbf{C}(l_s, r)$ in equation (5.15) with $[l_s\ r] = \boldsymbol{\theta}_{l-1,3}^{(s')}$. Then draw $\boldsymbol{\theta}_{l,2}^{(n)}$ from an inverse gamma distribution with PDF proportional to $(\theta_2')^{-\alpha'-1}\exp(-\beta'/\theta_2')$ where $\alpha'=\alpha+\tau_l N_c$ and β' is given by:

$$\beta' = \beta + \frac{\tau_l}{2}\sum_{k=1}^{N_c}[\mathbf{y}^{(k)} - \boldsymbol{\mu}(\mu_s)]^T\mathbf{C}^{-1}(l_s, r)[\mathbf{y}^{(k)} - \boldsymbol{\mu}(\mu_s)] \tag{A5.9}$$

where α and β are the parameters for the prior PDF $p(\sigma_s^2|\mathcal{M}_j^{(1)})$ of σ_s^2, the terms in the above are given by (5.11), (5.12) and (5.15) with $\mu_s = \boldsymbol{\theta}_{l,1}^{(n)}$, $[l_s\ r] = \boldsymbol{\theta}_{l-1,3}^{(s')}$. For $\mathcal{M}_2^{(1)}(i=1, j=2)$ and $\mathcal{M}_3^{(1)}(i=1, j=3)$, everything is the same as for $\mathcal{M}_4^{(1)}(i=1, j=4)$ except that r is fixed at 1 and 2 respectively.

Gibbs sampling for the posterior PDF in the illustrative example with data D_2 (i=2)

Now we describe how Gibbs sampling can be performed for the posterior PDF in the illustrative example with data $D_2=\{\mathcal{D}_1, \mathcal{D}_2\}$ (i=2), for $\mathcal{M}_3^{(2)}(i=2, j=3)$, $\boldsymbol{\theta}$ is divided into 3

component groups: $\boldsymbol{\theta}_1 = \mu_s$, $\boldsymbol{\theta}_2 = \sigma_s^2$, $\boldsymbol{\theta}_3 = [l_s^2\ r]$. Gibbs sampling in step 2 of the proposed stochastic simulation algorithm is performed on the first two component groups as follows: draw $\boldsymbol{\theta}_{l,1}^{(n)}$ from a truncated Gaussian PDF (constrained to be positive) which is proportional to a Gaussian distribution with mean μ'' and variance σ''^2 given below:

$$\mu'' = \sigma''^2 \left(\frac{\mu'}{\sigma'^2} + \frac{\tau_l K_v \sum_{k=1}^{N_v} \delta L_v^{(k)}}{\sigma_{v,j}(\sigma_s^2, l_s, r)^2}\right) \tag{A5.10}$$

$$\sigma''^2 = \frac{1}{\frac{1}{\sigma'^2} + \frac{N_v K_v^2 \tau_l}{\sigma_{v,j}(\sigma_s^2, l_s, r)^2}} \tag{A5.11}$$

$$\mu' = \frac{H_{11}\frac{F_c L_c}{A_c}\sum_{k=1}^{N_c}\delta L_c^{(k)} + H_{12}\left(\frac{F_c L_c}{A_c}\sum_{k=1}^{N_c} S_c^{(k)}(L_c/2) - \sum_{k=1}^{N_c}\delta L_c^{(k)}\right) + H_{22}\sum_{k=1}^{N_c} S_c^{(k)}(L_c/2) + \frac{\sigma_s^2 \mu_0}{\sigma_0^2}}{N_c\left(H_{11}\left(\frac{F_c L_c}{A_c}\right)^2 + 2H_{12}\frac{F_c L_c}{A_c} + H_{22}\right) + \frac{\sigma_s^2}{\sigma_0^2}} \tag{A5.12}$$

$$\sigma'^2 = \frac{\sigma_s^2}{N_c\left(H_{11}\left(\frac{F_c L_c}{A_c}\right)^2 + 2H_{12}\frac{F_c L_c}{A_c} + H_{22}\right) + \frac{\sigma_s^2}{\sigma_0^2}} \tag{A5.13}$$

In the above equations, $\sigma_s^2 = \boldsymbol{\theta}_{l-1,2}^{(s')}$, $[l_s\ r] = \boldsymbol{\theta}_{l-1,3}^{(s')}$; H_{11}, H_{12} and H_{22} are the (1,1), (1,2) and (2,2) entries of the inverse of $\mathbf{C}(l_s, r)$ in (5.15); K_v is given in section 5.2; $\sigma_{v,j}(\sigma_s^2, l_s, r)^2 = \sigma_s^2 s_{v,j}(l_s, r)$ where $s_{v,j}(l_s, r)$ is given in section 5.2. Then draw $\boldsymbol{\theta}_{l,2}^{(n)}$ from an inverse gamma distribution with PDF proportional to $(\theta_2'')^{-\alpha''-1}\exp(-\beta''/\theta_2'')$ where $\alpha'' = \alpha + N_c + \tau_l N_v/2$ and β'' is given by:

$$\beta'' = \beta + \frac{1}{2}\sum_{k=1}^{N_c}[\mathbf{y}^{(k)} - \boldsymbol{\mu}(\mu_s)]^T \mathbf{C}^{-1}(l_s, r)[\mathbf{y}^{(k)} - \boldsymbol{\mu}(\mu_s)] + \frac{\tau_l}{2s_v(l_s, r)}\sum_{k=1}^{N_v}(\delta L_v^{(k)} - K_v\mu_s)^2 \quad \text{(A5.14)}$$

where α and β are the parameters for the PDF $p(\sigma_s^2|\mathcal{M}_{j+1}^{(1)})$ of σ_s^2, the terms in the above are given by (5.11), (5.12) and (5.15) with $\mu_s = \boldsymbol{\theta}_{l,1}^{(n)}$, $[l_s\ r] = \boldsymbol{\theta}_{l-1,3}^{(s')}$. For $\mathcal{M}_1^{(2)}(i=2, j=1)$ and $\mathcal{M}_2^{(2)}$ $(i=2, j=2)$, everything is the same as for $\mathcal{M}_3^{(2)}(i=2, j=3)$ except that r is fixed at 1 and 2 respectively.

Gibbs sampling for the posterior PDF in the illustrative example with data D_3 (i=3)

Now we describe how Gibbs sampling can be performed for the posterior PDF in the illustrative example with data $D_3=\{\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3\}$ (i=3), for $\mathcal{M}_3^{(3)}$ $(i=3, j=3)$, $\boldsymbol{\theta}$ is divided into 3 component groups: $\boldsymbol{\theta}_1= \mu_s$, $\boldsymbol{\theta}_2=\sigma_s^2$, $\boldsymbol{\theta}_3=[l_s^2\ r]$. Gibbs sampling in step 2 of the proposed stochastic simulation algorithm is performed on the first two component groups as follows: draw $\boldsymbol{\theta}_{l,1}^{(n)}$ from a truncated Gaussian PDF (constrained to be positive) which is proportional to a Gaussian distribution with mean μ''' and variance σ'''^2 given below:

$$\mu''' = \sigma'''^2\left(\frac{\mu'}{\sigma'^2} + \frac{K_v\sum_{k=1}^{N_v}\delta L_v^{(k)}}{\sigma_{v,j}(\sigma_s^2, l_s, r)^2} + \frac{\tau_l K_a\sum_{k=1}^{N_a} w_a^{(k)}}{\sigma_{a,j}(\sigma_s^2, l_s, r)^2}\right) \quad \text{(A5.15)}$$

$$\sigma'''^2 = \frac{1}{\frac{1}{\sigma'^2} + \frac{N_v K_v^2}{\sigma_{v,j}(\sigma_s^2, l_s, r)^2} + \frac{N_a K_a^2 \tau_l}{\sigma_{a,j}(\sigma_s^2, l_s, r)^2}} \quad \text{(A5.16)}$$

In the above equations, $\sigma_s^2 = \boldsymbol{\theta}_{l-1,2}^{(s')}$, $[l_s\ r] = \boldsymbol{\theta}_{l-1,3}^{(s')}$; $\sigma_{a,j}(\sigma_s^2, l_s, r)^2 = \sigma_s^2\, s_{a,j}(l_s, r)$ where $s_{a,j}(l_s, r)$ is given in Appendix III in Cheung and Beck (2008b). Then draw $\boldsymbol{\theta}_{l,2}^{(n)}$ from an inverse gamma distribution with PDF proportional to $(\theta_2''')^{-\alpha'''-1}\exp(-\beta'''/\theta_2''')$ where $\alpha'''=\alpha+N_c+N_v/2+\tau_l N_a/2$ and β''' is given by:

$$\beta''' = \beta + \frac{1}{2}\sum_{k=1}^{N_c}[\mathbf{y}^{(k)} - \boldsymbol{\mu}(\mu_s)]^T \mathbf{C}^{-1}(l_s, r)[\mathbf{y}^{(k)} - \boldsymbol{\mu}(\mu_s)] + \frac{\sum_{k=1}^{N_v}(\delta L_v{}^{(k)} - K_v\mu_s)^2}{2s_v(l_s, r)} + \frac{\tau_l\sum_{k=1}^{N_a}(w_a{}^{(k)} - K_a\mu_s)^2}{2s_a(l_s, r)} \quad \text{(A5.17)}$$

where $\mu_s = \boldsymbol{\theta}_{l,1}^{(n)}$, $[l_s\ r] = \boldsymbol{\theta}_{l-1,3}^{(s')}$. For $\mathcal{M}_1^{(3)}(i=3, j=1)$ and $\mathcal{M}_2^{(3)}(i=3, j=2)$, everything is the same as for $\mathcal{M}_3^{(3)}(i=3, j=3)$ except that r is fixed at 1 and 2 respectively.

Gibbs sampling in step 3 of the hybrid Gibbs TMCMC algorithm exploits the form of $p(\boldsymbol{\theta}|D_i, \mathcal{M}_j^{(i)})$ which allows direct sampling from the conditional PDF for some groups. In the case where the form of $p(\boldsymbol{\theta}|D_i, \mathcal{M}_j^{(i)})$ cannot be exploited to carry out Gibbs sampling, step 2 is skipped and $\boldsymbol{\theta}$ has only one component group which includes all the parameters and so the algorithm reduces to the original TMCMC algorithm.

Appendix 5B: Analytical integration of part of integrals

Consider the following multi-dimensional integral:

$$E[g(\boldsymbol{\xi})] = \int g(\boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi} \quad \text{(B5.1)}$$

The above is the expectation of $g(\boldsymbol{\xi})$ with respect to a PDF $f(\boldsymbol{\xi})$. Recall that by MCS, the above integral can be estimated as follows using iid samples $\boldsymbol{\xi}_k$, $k=1,2,\ldots,K$ from $f(\boldsymbol{\xi})$ as follows:

$$E[g(\boldsymbol{\xi})] \approx \frac{1}{K}\sum_{k=1}^{K} g(\boldsymbol{\xi}_k) \triangleq \overline{g}_{MCS,K} \quad \text{(B5.2)}$$

For $E_f[g(\boldsymbol{\xi})] \neq 0$, the c.o.v. $\delta_{MCS,K}$ of the MCS estimator using iid samples $\boldsymbol{\xi}_k$, $k=1,2,\ldots,K$ from $f(\boldsymbol{\xi})$ is given by:

$$\delta_{MCS,K} = \frac{\Delta_{MCS}}{\sqrt{K}} \tag{B5.3}$$

where the unit c.o.v. Δ_{MCS} is given by:

$$\Delta_{MCS} = \sqrt{Var[g(\boldsymbol{\xi})]} / E[g(\boldsymbol{\xi})] \tag{B5.4}$$

Assume $\boldsymbol{\xi}$ can be splitted into two groups, say $\boldsymbol{\xi}=\left[\boldsymbol{\xi}_1^T \ \boldsymbol{\xi}_2^T\right]^T$, such that $g(\boldsymbol{\xi})$ can be integrated analytically with respect to $f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)= f(\boldsymbol{\xi})/f(\boldsymbol{\xi}_2)$. $E[g(\boldsymbol{\xi})]$ can be calculated as follows:

$$\begin{aligned} &E[g(\boldsymbol{\xi})] = \int g(\boldsymbol{\xi}) f(\boldsymbol{\xi}) d\boldsymbol{\xi} = \iint g(\boldsymbol{\xi}) f(\boldsymbol{\xi}_1 \mid \boldsymbol{\xi}_2) f(\boldsymbol{\xi}_2) d\boldsymbol{\xi}_1 d\boldsymbol{\xi}_2 \\ &= \int E_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \mid \boldsymbol{\xi}_2] f(\boldsymbol{\xi}_2) d\boldsymbol{\xi}_2 \\ &\approx \frac{1}{K}\sum_{k=1}^{K} \tilde{g}(\boldsymbol{\xi}_2^{(k)}) \triangleq \overline{g}_{AI,K}, \ \text{ where } \tilde{g}(\boldsymbol{\xi}_2) = E_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \mid \boldsymbol{\xi}_2] \end{aligned} \tag{B5.5}$$

where $\boldsymbol{\xi}_2^{(k)}$, k=1,…,K are independently identically distributed samples from $f(\boldsymbol{\xi}_2)$. The above estimator has the mean equal to $E[g(\boldsymbol{\xi})]$ and always has a smaller variance and thus c.o.v. than the MCS estimator $\overline{g}_{MCS,K}$ for a given sample size K.

By Law of Total Variance,

$$\begin{aligned} &Var_{f(\boldsymbol{\xi})}[g(\boldsymbol{\xi})] = E_{f(\boldsymbol{\xi}_2)}[Var_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_2]] + Var_{f(\boldsymbol{\xi}_2)}[E_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_2]] \\ &\geq Var_{f(\boldsymbol{\xi}_2)}[E_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_2]] (\because Var_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_2] \geq 0 \Rightarrow E_{f(\boldsymbol{\xi}_2)}[Var_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_2]] \geq 0) \end{aligned}$$

The sampling efficiency is given by:

$$\frac{K_{AI}}{K_{MCS}} = \frac{Var_{f(\boldsymbol{\xi}_2)}[\tilde{g}(\boldsymbol{\xi}_2)]}{Var_{f(\boldsymbol{\xi})}[g(\boldsymbol{\xi})]} = 1 - \frac{E_{f(\boldsymbol{\xi}_2)}[Var_{f(\boldsymbol{\xi}_1|\boldsymbol{\xi}_2)}[g(\boldsymbol{\xi}) \mid \boldsymbol{\xi}_2]]}{Var_{f(\boldsymbol{\xi})}[g(\boldsymbol{\xi})]} \leq 1$$

where K_{AI} and K_{MCS} are the minimum number of samples required to achieve the same c.o.v. in the estimator $\overline{g}_{AI,K}$ and the MCS estimator $\overline{g}_{MCS,K}$ respectively. The above result implies that one should always carry out analytical integration of the integrals as far as possible which agrees with intuition. The above proof provides a general proof the case which allows an analytical integration of part of the integrals during the calculation of the failure probability $P(F)$ (where $g(\xi)$ is an indicator function equal to 1 if ξ belongs to F and 0 if otherwise) which always leads to an estimator with a smaller c.o.v.

The following provides the proof of Law of Total Variance:

$$
\begin{aligned}
&Var_{f(\xi)}[g(\xi)] = E_{f(\xi)}[g^2(\xi)] - E_{f(\xi)}{}^2[g(\xi)] \\
&= E_{f(\xi_2)}[E_{f(\xi_1|\xi_2)}[g^2(\xi)\,|\,\xi_2]] - (E_{f(\xi_2)}[E_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2]])^2 \\
&= E_{f(\xi_2)}[Var_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2] + (E_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2])^2] - (E_{f(\xi_2)}[E_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2]])^2 \\
&= E_{f(\xi_2)}[Var_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2]] + E_{f(\xi_2)}[(E_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2])^2] - (E_{f(\xi_2)}[E_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2]])^2 \\
&= E_{f(\xi_2)}[Var_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2]] + Var_{f(\xi_2)}[E_{f(\xi_1|\xi_2)}[g(\xi)\,|\,\xi_2]]
\end{aligned}
$$

In our case, $\xi_2^{(k)}$, k=1,…,K are dependent samples. The above proof can be modified using the same idea as in Appendix 2C to handle this case.

CHAPTER 6

New stochastic simulation method for updating robust reliability of dynamic systems

6.1 Introduction

Before presenting the proposed method, it is instuctive to go over and review the commonly used importance sampling for evaluating multi-dimensional integrals as follows:

$$E_f[g(\xi)] = \int g(\xi) f(\xi) d\xi \tag{6.1}$$

Importance sampling is a stochastic simulation technique that makes use of samples drawn from another PDF $q(\xi)$, referred to as the importance sampling density (ISD) as follows:

$$E_f[g(\xi)] = \int g(\xi) \frac{f(\xi)}{q(\xi)} q(\xi) d\xi = E_q[g(\xi) \frac{f(\xi)}{q(\xi)}] \approx \frac{1}{K} \sum_{k=1}^{K} g(\xi_k) \frac{f(\xi_k)}{q(\xi_k)} \triangleq \overline{g}_{IS,K} \tag{6.2}$$

where $\xi^{(k)}$, k=1,2,…,K are samples drawn from $q(\xi)$. Here to ensure the above estimator has finite variance, we require supp $q \supset$ supp f. With finite variance, the Central Limit Theorem is applicable to the IS estimator, just like the MCS estimator $\overline{g}_{MCS,K}$.

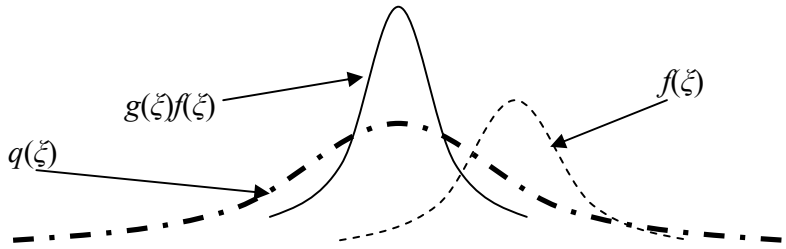

Figure 6.1: Schematic plot of importance sampling density

This method is often used:

1. to simulate more samples in the region which give significant contributions to the integral rather than wasting too much effort sampling in the region which contributes little. This often leads to an estimator with a smaller variance.
2. when drawing samples from $f(\xi)$ is not trivial or easy.

The variance of the IS estimator is given by:

$$Var[\overline{g}_{IS,K}] = \frac{1}{K} Var_q[\frac{g(\xi)f(\xi)}{q(\xi)}] \tag{6.3}$$

where

$$Var_q[\frac{g(\xi)f(\xi)}{q(\xi)}] = E_q[\frac{g^2(\xi)f^2(\xi)}{q^2(\xi)}] - E_q{}^2[\frac{g(\xi)f(\xi)}{q(\xi)}] \tag{6.4}$$

If $E_f[g(\xi)] \neq 0$, the c.o.v. $\delta_{IS,K}$ of the IS estimator using identitically and independently distributed (iid) samples ξ_k, k=1,2,…,K from $q(\xi)$ is given by:

$$\delta_{IS,K} = \frac{\Delta_{IS}}{\sqrt{K}} \tag{6.5}$$

where the unit c.o.v. Δ_{IS} is given by:

$$\Delta_{IS} = \sqrt{Var_q[\frac{g(\xi)f(\xi)}{q(\xi)}]} / E_q[\frac{g(\xi)f(\xi)}{q(\xi)}] \tag{6.6}$$

$$\begin{aligned} E_q[\frac{g(\xi)f(\xi)}{q(\xi)}] &\approx \frac{1}{K}\sum_{k=1}^{K}\frac{g(\xi_k)f(\xi_k)}{q(\xi_k)} \\ E_q[\frac{g^2(\xi)f^2(\xi)}{q^2(\xi)}] &\approx \frac{1}{K}\sum_{k=1}^{K}(\frac{g(\xi_k)f(\xi_k)}{q(\xi_k)})^2 \\ Var_q[\frac{g(\xi)f(\xi)}{q(\xi)}] &= E_q[\frac{g^2(\xi)f^2(\xi)}{q^2(\xi)}] - E_q{}^2[\frac{g(\xi)f(\xi)}{q(\xi)}] \end{aligned} \tag{6.7}$$

To exploit the advantage of the IS, an ISD $q(\xi)$ should be chosen such that $Var_q[\frac{g(\xi)f(\xi)}{q(\xi)}]$ is as small as possible. Let's manipulate Equation (6.4) further as follows:

$$\begin{aligned} Var_q[\frac{g(\xi)f(\xi)}{q(\xi)}] &= E_q[\frac{g^2(\xi)f^2(\xi)}{q^2(\xi)}] - E_q{}^2[\frac{g(\xi)f(\xi)}{q(\xi)}] \\ &= \int \frac{g^2(\xi)f^2(\xi)}{q^2(\xi)} q(\xi)d\xi - E_f{}^2[g(\xi)] \\ &= \int \frac{g^2(\xi)f^2(\xi)}{q(\xi)} d\xi - E_f{}^2[g(\xi)] \end{aligned} \tag{6.8}$$

It can be seen that the second term in the last expression in the above equation is independent of $q(\xi)$. For a given K, the variance of the IS estimator is minimized if the ISD $q(\xi)$ is chosen to be the optimal ISD $q^*(\xi)$ that minimizes the first integral in the last expression in (6.8). It can be shown that $q^*(\xi)$ is given by:

$$q^*(\xi) = \frac{|g(\xi)|f(\xi)}{\int |g(\xi)|f(\xi)d\xi} \tag{6.9}$$

The above is proved in Appendix 6A.

In practice, it is often not straightforward to simulate from $q^*(\xi)$ (note that the normalizing constant $\int |g(\xi)| f(\xi) d\xi$ in Equation (6.9) is often not known analytically and, in fact, is the original integral of interest in (6.1) if $g(\xi)>0$ on its support). However, one can expect a reduction in the variance of the IS estimator if $q(\xi)$ is constructed to be close enough to $q^*(\xi)$ while still ensuring that samples of $q(\xi)$ can be readily obtained. There are at least two methods of constructing such ISD $q(\xi)$:

1. Find all the local maxima of $|g(\xi)|f(\xi)$ and construct ISD $q(\xi)$ so that one can sample in the neighborhood of these maxima, by e.g., Laplace's asymptotic approximation; see, for example, Au et al.(1999) and Papadimitriou et al. (2001).
2. Generate some presamples from $q^*(\xi)$ and construct ISD $q(\xi)$ using these samples, e.g., by constructing a kernel sampling density (a common choice is a PDF which is a weighted sum of Gaussian PDFs) to approximate $q^*(\xi)$; see, for example, Ang et al. (1992) and Au and Beck (1999).

For problems with multiple maxima of $|g(\xi)|f(\xi)$, being unable to simulate in the neighborhood of some of the maxima (especially those whose contribution to the integral are not negligible) can lead to a bias in the IS estimate for finite sample sizes. The c.o.v. estimated by IS samples from only one simulation (using Equations (6.6) and (6.7)) can then be misleading because, for instance, the estimated c.o.v. of the IS estimator can be small while the actual c.o.v. can be very large. If the sample size is sufficiently large, a small number of points in the neighborhood of omitted maxima can lead to occasional sudden jumps in the estimate.

It is in general inefficient to use IS if ξ has high dimensions except for the special case where a lot of information regarding the underlying problem can be exploited (Au and Beck 2001a). For high-dimensional ξ, it is computationally expensive or prohibitive to find all the 'significant' local maxima of $|g(\xi)|f(\xi)$ as required in Method 1 above. Method 2 is shown to be in general inapplicable in high dimensions (Au and Beck 2003) which is the case of interest in this thesis.

To assess the system performance subjected to dynamic excitation, a stochastic system analysis considering all the uncertainties involved has to be performed. In engineering, evaluating the robust failure probability (or its complement, reliability) of the system is a very important part of such stochastic system analysis.

During the design stage, the prior robust failure probability can be employed to evaluate the system performance. Such probability takes into account the prior knowledge of the stochastic system model based on engineering judgment and experience. Efficient stochastic simulation algorithms such as Subset Simulation (Au and Beck 2001b) can be used to calculate such failure probabilities when they are very small (in which case ordinary Monte Carlo simulation is very inefficient). The proof for stationarity of the Markov chain in the original presentation of Subset Simulation by Au and Beck (2001b) is not exactly correct. The corrected proof is presented in Appendix 6B.

After, or while, the system is constructed, there is the opportunity to measure system input and output and then use these data to obtain a more accurate evaluation of the system performance by updating the robust failure probability for the system. During system operation, the behavior, and thus the robust failure probability of the system, can change from time to time due to deterioration or damage. For example, for structures, deterioration can be due to corrosion or fatigue, and damage can also result after the structure is subjected to severe loading from explosions, strong winds or earthquakes. The consequences of such changes in the system behavior can be assessed quantitatively by monitoring the dynamic response of the system and using it to update the robust failure probability of the system.

Let $\boldsymbol{\theta}$ be the vector consisting of the uncertain parameters for a model class $\mathcal{M}$ which are to be updated by data $\mathcal{D}$ from the system (for example, structural parameters and parameters related to prediction errors as in previous chapters). Let $\mathbf{U}_n=[\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_n]$ denote the input at different times, which in turn is specified by a stochastic input model class $\mathcal{U}$ with model

parameters $\boldsymbol{\theta}_U$. $\boldsymbol{\theta}_U$ can comprise of model parameters 1) $\boldsymbol{\theta}_u$ (with uncertainty quantified by $p(\boldsymbol{\theta}_u|\mathcal{U})$) which is not part of $\boldsymbol{\theta}$ and not updated by $\mathcal{D}$, and 2) $\boldsymbol{\theta}_p$ which are some components of $\boldsymbol{\theta}$ for $\mathcal{M}$ (with uncertainty quantified by $p(\boldsymbol{\theta}_p|\mathcal{D},\mathcal{M})$ which is a marginal PDF of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ corresponding to some components of $\boldsymbol{\theta}$), i.e. $\boldsymbol{\theta}_U = [\boldsymbol{\theta}_u^T \; \boldsymbol{\theta}_p^T]^T$. The uncertainty in $\boldsymbol{\theta}_U$ is quantified by $p(\boldsymbol{\theta}_U|\mathcal{D},\mathcal{U})$ given as follows:

$$p(\boldsymbol{\theta}_U \mid \mathcal{D},\mathcal{U}) = p(\boldsymbol{\theta}_u \mid \mathcal{U})\, p(\boldsymbol{\theta}_p \mid \mathcal{D},\mathcal{M}) \tag{6.10}$$

This model class can be viewed as a special case of hierarchical model classes presented in Chapter 5. The uncertainty in $\mathbf{U}$ is thus quantified by $p(\mathbf{U}|\mathcal{D},\mathcal{U})$. Here we are interested in the failure F which corresponds to the event(s) where the system performs unsatisfactorily when subjected to future excitations/inputs modeled by $\mathcal{U}$. Let $\mathcal{D}$ denote the dynamic data from the system, which can include output response data and possibly input data. The updated (posterior) robust failure probability given $\mathcal{D}$ based on $\mathcal{M}$ and $\mathcal{U}$ is given by:

$$P(F \mid \mathcal{D},\mathcal{M},\mathcal{U}) = \int P(F \mid \boldsymbol{\theta}, \mathbf{U}_n, \mathcal{D},\mathcal{M},\mathcal{U}) p(\boldsymbol{\theta} \mid \mathcal{D},\mathcal{M})\, p(\mathbf{U}_n \mid \mathcal{D},\mathcal{U}) d\mathbf{U}_n d\boldsymbol{\theta} \tag{6.11}$$

Often the performance measures defining the failure are functions of $\boldsymbol{\theta}$, $\mathbf{U}_n$ and some uncertain variables $\mathbf{Z}$ (for example, those related to prediction errors like $\mathbf{W}$ and $\mathbf{V}$ in (4.30)), then:

$$\begin{aligned} &P(F \mid \mathcal{D},\mathcal{M},\mathcal{U}) \\ &= \int I_F(\boldsymbol{\theta}, \mathbf{U}_n, \mathbf{Z}) p(\mathbf{U}_n \mid \boldsymbol{\theta}_u, \boldsymbol{\theta}_p, \mathcal{U})\, p(\boldsymbol{\theta}_u \mid \mathcal{U})\, p(\mathbf{Z} \mid \boldsymbol{\theta},\mathcal{M})\, p(\boldsymbol{\theta} \mid \mathcal{D},\mathcal{M}) d\mathbf{Z} d\mathbf{U}_n d\boldsymbol{\theta} \end{aligned} \tag{6.12}$$

The plausibility of each model within a class $\mathcal{M}$ of models for a system, based on data $\mathcal{D}$, is quantified by the updated joint probability density function $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ (posterior PDF). By Bayes' Theorem, the posterior PDF of $\boldsymbol{\theta}$ is given by $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M}) = c^{-1} p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M}) p(\boldsymbol{\theta}|\mathcal{M})$ where $c = p(\mathcal{D}|\mathcal{M})$ is the normalizing constant (also called the evidence) which makes the probability volume under the posterior PDF equal to unity; $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})$ is the likelihood function based on the predictive PDF for the response given by model class $\mathcal{M}$; $p(\boldsymbol{\theta}|\mathcal{M})$ is the prior PDF for

the model class $\mathcal{M}$ in which one can incorporate engineering judgment through experience or previous analysis to quantify the initial plausibility of each predictive model defined by the value of the parameters $\boldsymbol{\theta}$.

For simplicity in presentation, the conditioning on $\mathcal{M}$ and $\mathcal{U}$ will be left implicit in the rest of this chapter.

Very few publications have appeared that tackle the problem of updating the robust failure probability of a system given dynamic data since it is computationally very challenging. In Papadimitriou et al. (2001), Laplace's method of asymptotic approximation was adopted to calculate the updated robust reliability with an illustration based on linear dynamics. However, the accuracy of such an approximation is questionable when (i) the amount of data is not sufficiently large or (ii) the chosen class of models turns out to be unidentifiable based on the available data. Also, such an approximation requires a non-convex optimization in what is usually a high-dimensional parameter space, which is computationally challenging, especially when the model class is not globally identifiable. It is shown in Cheung and Beck (2008b,g) that the robust failure probability may require information of the posterior PDF in regions of the uncertain parameter space, that are not in the high probability region of the posterior PDF. The asymptotic approximation will usually not give a good approximation in the region of the uncertain parameter space that lies outside the high probability content of the posterior PDF, leading to a poor estimate of the robust failure probability. Beck and Au (2002) proposed to update the system reliability using a level-adaptive Metropolis algorithm (like simulated annealing) with global proposal PDFs. However, their approach can only be applied for the case where the dimension of the modeling parameters is quite small because of the kernel densities used as the global proposal PDFs. Ching and Beck (2007) proposed a method to update the reliability based on combining a Kalman filter and smoother and modifying the algorithm ISEE (Au and Beck 2001a). Such an approach is only applicable to linear systems with no uncertainties in model parameters. Ching and Hsieh (2006) proposed a method based on analytical

approximation of some of the required PDFs by maximum entropy PDFs. The method is applicable regardless of the dimension of $\boldsymbol{\theta}$ but can only be applied to very low dimensional system output data $\mathcal{D}$. In practice, dynamic data is of very high dimension (say of the order of hundreds or thousands). In this chapter, a new method for calculating the updated robust failure probability of a dynamic system for a model class subjected to future stochastic excitation is proposed. Part of the materials in this chapter is presented in Cheung and Beck (2007b). If there are multiple model classes, as in Chapter 4 and 5, the proposed method in this chapter can be combined with Bayesian model averaging procedures to obtain hyper robust failure probabilities.

6.2 The proposed method

6.2.1 Theory and formulation

By Bayes' Theorem, the updated probability of failure conditional on data $\mathcal{D}$ (and implicitly, the model classes $\mathcal{M}$ and $\mathcal{U}$), $P(F|\mathcal{D})$ is given by:

$$P(F \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid F)P(F)}{p(\mathcal{D} \mid F)p(F) + p(\mathcal{D} \mid \sim F)(1 - P(F))} = \frac{1}{1 + \dfrac{p(\mathcal{D} \mid \sim F)}{p(\mathcal{D} \mid F)}(P^{-1}(F) - 1)} \tag{6.13}$$

where $P(F)$ is the prior probability of failure and $\sim F$ denotes non-failure, so $P(\sim F)$= 1- $P(F)$. The new idea here is to compute $p(\mathcal{D}|F)$ and $p(\mathcal{D}|\sim F)$ by expressing each of them as a product of factors and calculating each of the factors one by one as follows:

$$p(\mathcal{D} \mid F) = \prod_{i=0}^{l} \alpha_i, \; p(\mathcal{D} \mid \sim F) = \prod_{i=0}^{l} \beta_i \tag{6.14}$$

where

$$\alpha_i = \frac{p(\mathcal{D} \mid F, t_{i+1})}{p(\mathcal{D} \mid F, t_i)},\ \beta_i = \frac{p(\mathcal{D} \mid \sim F, t_{i+1})}{p(\mathcal{D} \mid \sim F, t_i)} \tag{6.15}$$

and where $0 = t_0 < t_1 < \ldots < t_{l+1} = 1$ and $p(\mathcal{D} \mid F,t)$ is given by:

$$p(\mathcal{D} \mid F, t) = \int p(\mathcal{D} \mid \boldsymbol{\theta}, F, t)\, p(\boldsymbol{\theta} \mid F)\, d\boldsymbol{\theta} \tag{6.16}$$

The likelihood $p(\mathcal{D} \mid \boldsymbol{\theta}, t)$ for the model class defined by $\mathcal{M}$ and t is given by:

$$p(\mathcal{D} \mid \boldsymbol{\theta}, t) = p^t(\mathcal{D} \mid \boldsymbol{\theta}) = p(\mathcal{D} \mid \boldsymbol{\theta}, F, t) = p(\mathcal{D} \mid \boldsymbol{\theta}, \sim F, t) \tag{6.17}$$

If there is a time period between the time when the data is collected and the time of interest in the future, one can assume that given $\boldsymbol{\theta}$, the failure or non-failure in the future does not affect the PDFs of data collected in the present or in the past, so (6.17) is valid. Thus, $p(\mathcal{D} \mid F,t)$ is given by:

$$p(\mathcal{D} \mid F, t) = \int p(\mathcal{D} \mid \boldsymbol{\theta}, t)\, p(\boldsymbol{\theta} \mid F)\, d\boldsymbol{\theta} = \int p(\mathcal{D} \mid \boldsymbol{\theta}, t) \frac{P(F \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta})}{P(F)}\, d\boldsymbol{\theta} \tag{6.18}$$

Similarly, $p(\mathcal{D} \mid \sim F,t)$ is given by (6.18) with F replaced by $\sim F$. Obviously $p(\mathcal{D} \mid F,t_0) = p(\mathcal{D} \mid \sim F,t_0) = 1$. Now define the PDF $p(\boldsymbol{\theta} \mid F, \mathcal{D}, t)$ as follows:

$$p(\boldsymbol{\theta} \mid F, \mathcal{D}, t) = \frac{p(\mathcal{D} \mid \boldsymbol{\theta}, t)\, p(\boldsymbol{\theta} \mid F)}{p(\mathcal{D} \mid F, t)} \propto p^t(\mathcal{D} \mid \boldsymbol{\theta})\, p(\boldsymbol{\theta} \mid F) \tag{6.19}$$

Similarly, $p(\boldsymbol{\theta} \mid \sim F, \mathcal{D}, t)$ is given by (6.19) with F replaced by $\sim F$. With this, it can be shown that α_i and β_i can be estimated by stochastic simulation using the following (shown in Appendix 6C):

$$\alpha_i = \frac{p(\mathcal{D} \mid F, t_{i+1})}{p(\mathcal{D} \mid F, t_i)} \approx \frac{1}{N} \sum_{k=1}^{N} p^{t_{i+1} - t_i}(D \mid \boldsymbol{\theta}^{(k)}) \tag{6.20}$$

$$\beta_i = \frac{p(\mathcal{D}|\sim F, t_{i+1})}{p(\mathcal{D}|\sim F, t_i)} \approx \frac{1}{N'}\sum_{m=1}^{N'} p^{t_{i+1}-t_i}(D|\tilde{\boldsymbol{\theta}}^{(m)}) \tag{6.21}$$

where $\boldsymbol{\theta}^{(k)}$, k=1, 2,…, N, are samples from $p(\boldsymbol{\theta}|F,\mathcal{D}, t_i)$ and $\tilde{\boldsymbol{\theta}}^{(m)}$, m=1, 2,…, N', are drawn from $p(\boldsymbol{\theta}|\sim F, \mathcal{D}, t_i)$.

6.2.2 Algorithm of proposed method

Let $\mathbf{Z}$ denote the vector consisting of the uncertain parameters, which are not to be updated by the data (for example, those used to model the uncertain input excitation $\mathbf{U_n}$). The proposed method is summarized as follows:

1. Set t_0=0. Using efficient procedures such as Subset Simulation given by Au and Beck (2001b) for the parameter space of $\boldsymbol{\theta}$, $\boldsymbol{\theta}_u$, $\mathbf{U}_n$ and $\mathbf{Z}$, calculate the prior robust failure probability $P(F)$ given by (6.12) with the conditioning on $\mathcal{D}$ removed and obtain the samples from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F)= p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F,\mathcal{D},t_0)$ and $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|\sim F)= p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|\sim F,\mathcal{D},t_0)$. Take the $\boldsymbol{\theta}$ part of these samples to give samples from $p(\boldsymbol{\theta}|F)=p(\boldsymbol{\theta}|F,\mathcal{D}, t_0)$ and $p(\boldsymbol{\theta}|F)=p(\boldsymbol{\theta}|\sim F,\mathcal{D}, t_0)$.
2. Repeat the following for i=0,1,2,…,l:

   (a) Let $\boldsymbol{\theta}^{(k)}$, k=1, 2,…, N, be samples from $p(\boldsymbol{\theta}|F,\mathcal{D}, t_i)$ and $\tilde{\boldsymbol{\theta}}^{(m)}$, m=1, 2,…, N', be samples from $p(\boldsymbol{\theta}|\sim F,\mathcal{D},t_i)$. Select $\tilde{t}_{i+1}$ such that the effective sample size $1/\sum_{s=1}^{N}\tilde{w}_s^{\,2}$ is equal to some threshold (Cheung and Beck 2008c; Chapter 2 in this thesis) (e.g., 0.9N) where $\tilde{w}_k = w_k / \sum_{k=1}^{N} w_k$ and $w_k = p^{t_{i+1}-t_i}(\mathcal{D}|\boldsymbol{\theta}^{(k)})$. Select $\hat{t}_{i+1}$ such that the effective sample size $1/\sum_{m=1}^{N'}\tilde{w}_m^{\,2}$ is equal some threshold (e.g., 0.9 N') where $\tilde{w}_m = w_m / \sum_{s=1}^{N'} w_m$ and $w_m = p^{t_{i+1}-t_i}(\mathcal{D}|\tilde{\boldsymbol{\theta}}^{(m)})$. Set t_{l+1}=min$\{\tilde{t}_{i+1}, \hat{t}_{i+1}\}$. If $t_{l+1}\geq 1$, set t_{l+1}=1;

   (b) Obtain an estimate for α_i and β_i using (6.20)-(6.21) and go to step 3 if t_{l+1}=1;

   (c) Using samples from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F,\mathcal{D},t_i)$ as starting points, simulate samples from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F,\mathcal{D},t_{i+1})$. Similarly, using samples from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|\sim F,\mathcal{D},t_i)$ as starting points, simulate samples from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|\sim F,\mathcal{D},t_{i+1})$. The detailed

procedures are described in the next section. Take the $\boldsymbol{\theta}$ part of these samples to give samples from $p(\boldsymbol{\theta}|F,\mathcal{D},t_{i+1})$ and $p(\boldsymbol{\theta}|\sim F,\mathcal{D},t_{i+1})$ for use in (6.20) and (6.21).

3. Compute the estimate $p(\mathcal{D}|F)$ and $p(\mathcal{D}|\sim F)$ by substituting α_i's and β_i's found above into (6.14). Based on (6.13), the estimate for $P(F|\mathcal{D})$ is then given by:

$$P(F\mid\mathcal{D})\approx\frac{1}{1+(\prod_{i=0}^{l}\frac{\beta_i}{\alpha_i})(P^{-1}(F)-1)} \tag{6.22}$$

It is interesting to note that the ratio R of the updated robust reliability and prior robust reliability is approximately equal to the following for sufficiently small $P(F)$:

$$R=\frac{P(F\mid\mathcal{D})}{P(F)}\approx\prod_{i=0}^{l}\frac{\alpha_i}{\beta_i}\qquad\text{if } P(F)<<\prod_{i=0}^{l}\frac{\beta_i}{\alpha_i} \tag{6.23}$$

6.2.3 Simulations of samples from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F,\mathcal{D},t_{i+1})$

In the i-th step of the algorithm, we have the samples $\boldsymbol{\theta}^{(k)}$, $\boldsymbol{\theta}_u^{(k)}$, $\mathbf{U}_n^{(k)}$, $\mathbf{Z}^{(k)}$, k=1, 2,…, N, from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F,\mathcal{D},t_i)$. We need to simulate samples from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F,\mathcal{D},t_{i+1})$ to move on to the next level. Here we propose the following algorithm to simulate these samples:

1. Define the probability p_k as follows:

$$p_k=\frac{p^{t_{i+1}-t_i}(\mathcal{D}\mid\boldsymbol{\theta}^{(k)})}{\sum_{k=1}^{N}p^{t_{i+1}-t_i}(\mathcal{D}\mid\boldsymbol{\theta}^{(k)})} \tag{6.24}$$

2. Repeat the following to simulate samples $(\widehat{\boldsymbol{\theta}}^{(j)},\widehat{\boldsymbol{\theta}}_u^{(j)},\widehat{\mathbf{U}}_n^{(j)},\widehat{\mathbf{Z}}^{(j)})$ from $p(\boldsymbol{\theta},\boldsymbol{\theta}_u,\mathbf{U}_n,\mathbf{Z}|F,\mathcal{D},t_{i+1})$ for j=1, 2, …N:

2.1. Draw a point $(\hat{\boldsymbol{\theta}}^{(j)}, \hat{\boldsymbol{\theta}}_u^{(j)}, \hat{\mathbf{U}}_n^{(j)}, \hat{\mathbf{Z}}^{(j)}) = (\boldsymbol{\theta}^{(k)}, \boldsymbol{\theta}_u^{(k)}, \mathbf{U}_n^{(k)}, \mathbf{Z}^{(k)})$ with probability p_k. Starting with $\hat{\boldsymbol{\theta}}^{(j)}$, perform a 1-step MCMC procedure such as those presented in Chapter 2 (for example, multiple-group MCMC in TMCMC) to obtain the candidate $\boldsymbol{\theta}_c^{(j)}$ for $\widehat{\boldsymbol{\theta}}^{(j)}$. Similarly, starting with $\hat{\boldsymbol{\theta}}_u^{(j)}$, $\hat{\mathbf{U}}_n^{(j)}$, $\hat{\mathbf{Z}}^{(j)}$, perform multigroup MCMC procedure (using a procedure similar to modified Metropolis-Hastings algorithm in Subset Simulation) to obtain the candidate $\boldsymbol{\theta}_{u,c}^{(j)}$, $\mathbf{U}_{n,c}^{(j)}, \mathbf{Z}_c^{(j)}$ for $\widehat{\boldsymbol{\theta}}_u^{(j)}, \widehat{\mathbf{U}}_n^{(j)}, \widehat{\mathbf{Z}}^{(j)}$, respectively.

2.2. If $(\boldsymbol{\theta}_c^{(j)}, \mathbf{Z}_c^{(j)})$ leads to failure, $(\widehat{\boldsymbol{\theta}}^{(j)}, \widehat{\mathbf{Z}}^{(j)}) = (\boldsymbol{\theta}_c^{(j)}, \mathbf{Z}_c^{(j)})$, $(\boldsymbol{\theta}^{(k)}, \mathbf{Z}^{(k)}) = (\boldsymbol{\theta}_c^{(j)}, \mathbf{Z}_c^{(j)})$. Otherwise, $(\widehat{\boldsymbol{\theta}}^{(j)}, \widehat{\mathbf{Z}}^{(j)}) = (\boldsymbol{\theta}^{(k)}, \mathbf{Z}^{(k)})$.

Samples from $p(\boldsymbol{\theta}, \boldsymbol{\theta}_u, \mathbf{U}_n, \mathbf{Z} | \sim F, \mathcal{D}, t_{i+1})$ can be generated using the same procedures as the above with F replaced by $\sim F$.

6.2 Illustrative example

For illustration of the proposed method, consider a 4-story building modeled as an inelastic shear building with the hysteretic restoring force model shown in Figure 3.4 and Rayleigh damping. The simulated noisy accelerometer data $\mathcal{D}$ consist of 10s (with a sample interval Δt of 0.01s) of the total acceleration at the base and at all the floors. The simulated Gaussian white noise has a noise-to-signal ratio of 10% rms of the roof acceleration. The data $\mathcal{D}$ are generated from a shear building model with Rayleigh damping and hysteretic bilinear interstory restoring forces, a similar system as used earlier in Chapter 3.

The lumped masses m_i, i=1, 2, 3, 4, on each floor are assumed fixed at 2×10^4kg for all floors. The vector $\boldsymbol{\theta}$ to be updated by the dynamic data $\mathcal{D}$ consists of D=15 parameters with

the first component θ_1 equal to the prediction error variance σ^2 and for s=2,…,D, θ_s = $\log(\varphi_{s\text{-}1}/l_{s\text{-}1})$ where $\varphi_{s\text{-}1}$'s are comprised of the following 16 structural parameters: for i=1,2,3,4, the initial stiffness k_i, post-yield stiffness reduction factor r_i, yield displacement u_i and the damping coefficient c_i of the viscous damper of the i-th floor and the $l_{s\text{-}1}$'s are the corresponding nominal values given later. Let $q_i(n;\theta_2,...,\theta_D)$ denote the output at time t_n= $n\Delta t$ (Δt=0.01s) at the i-th observed degree of freedom predicted by the proposed structural model and $y_i(n)$ denote the corresponding measured output. The combined prediction and measurement errors $\varepsilon_i(n) = y_i(n) - q_i(n;\boldsymbol{\theta})$ for n=1,…, N_T =1000 and i=1,…,N_o = 4 are modeled as independently and identically distributed Gaussian variables with mean zero and some unknown prediction-error variance σ^2. Thus the likelihood function $p(\mathcal{D}|\boldsymbol{\theta},\mathcal{M})$ is given by:

$$p(\mathcal{D}\,|\,\boldsymbol{\theta},\mathcal{M}) = \frac{1}{(2\pi\sigma^2)^{N_o N_T/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N_o}\sum_{n=1}^{N_T}[y_i(n) - q_i(n;\theta_2,...,\theta_D)]^2\right) \qquad (6.25)$$

The prior PDF for $\boldsymbol{\theta}$ is chosen as the product of independent distributions: the structural parameters $\varphi_{s\text{-}1}$ including k_i, r_i, u_i, ρ and γ follow a lognormal distribution with median equal to the corresponding nominal values $l_{s\text{-}1}$ and the corresponding log standard deviations equal to 0.6 and thus the θ_s, for s=2,…,D, follow a Gaussian distribution with zero mean and standard deviation of 0.6; $\theta_1=\sigma^2$ follows an inverse gamma distribution with mean μ equal to its nominal value and c.o.v. δ =1.0, i.e., $p(\sigma^2) \propto (\sigma^2)^{-\alpha-1}\exp(-\beta/\sigma^2)$ where $\alpha=\delta^{-2}+2, \beta=\mu(\alpha-1)$. The nominal values for the structural parameters k_1, k_2, k_3, k_4 are 2.2, 2.0, 1.7, 1.45 (10^7Nm^{-1}) respectively; the nominal values for r_i are 0.1 for all i; the nominal values for u_i are 8mm for i=1,2 and 7mm for i=3,4;. The nominal values for ρ, γ are 0.7959 and 2.50×10^{-3} so that the corresponding nominal modal damping ratios for the first 2 modes are 5%. The nominal value for σ^2 is the square of 10% of the maximum of the r.m.s of the total accelerations measured at each of the 4 floors. $q_i(n;\boldsymbol{\theta})$ is the i-th component at time t_n of $\mathbf{q}(t_n)$ which satisfies the following equation of motion:

$$\mathbf{M_s}\ddot{\mathbf{q}}(t)+\mathbf{C_s}\dot{\mathbf{q}}(t)+\mathbf{F}(\mathbf{Q}(t),\dot{\mathbf{Q}}(t))=-\mathbf{M}_s\begin{bmatrix}1\\ \vdots\\ 1\end{bmatrix}a_g(t) \tag{6.26}$$

where the mass matrix $\mathbf{M_s}$, is a diagonal matrix diag(m_1, m_2, m_3, m_4); damping matrix $\mathbf{C_s}$ is equal to $\rho\mathbf{M_s}+\gamma\mathbf{K_s}$ where $\mathbf{M_s}$ and $\mathbf{K_s}$ are the mass and stiffness matrix of the shear building model in $\mathcal{M}$, respectively, and ρ, γ are some uncertain positive scalars (such that a higher mode has the same or larger modal damping ratio than a lower mode). The hysteretic restoring force $\mathbf{F}(\mathbf{Q}(t),\dot{\mathbf{Q}}(t))$, which depends on the whole time history $[\mathbf{Q}(t),\dot{\mathbf{Q}}(t)]$ of responses from time=0 up to time τ, i.e., $\mathbf{q}(\tau)$ and $\dot{\mathbf{q}}(\tau)$ for all $\tau\in[0,t]$, is modeled by a hysteretic bilinear restoring force model as mentioned above. This model class contains the system used to generate the simulated noisy data $\mathcal{D}$. For this case, the uncertain parameter vector $\boldsymbol{\theta}$ to be updated by the dynamic data $\mathcal{D}$ consists of D=15 parameters.

The goal here is to calculate the updated robust failure probability of the building for future ground shaking from earthquakes. The model class $\mathcal{U}$ for modeling the future horizontal acceleration $\ddot{a}$ of the base of the building is given in the illustrative example in Chapter 4. The updated robust failure probability will be compared with the nominal failure probability (failure probability using the nominal structural model) and prior robust failure probability.

For the purpose of illustration, first consider failure F defined as the exceedance over some threshold of the interstory drift of any one of the stories at any time within the 10s of ground shaking:

$$\begin{aligned}F&=\bigcup_{n=0}^{1000}\bigcup_{l=1}^{4}\{|x_l(t_n)-x_{l-1}(t_n)|\geq b_l\cup|x_1(t_n)|\geq b_1\}\\&=\max_{\substack{n\in\{0,1,\ldots,1000\}\\ l\in\{1,\ldots,4\}}}\{\frac{|x_l(t_n)-x_{l-1}(t_n)|}{b_l},\frac{|x_1(t_n)|}{b_1}\}\geq 1\end{aligned} \tag{6.27}$$

where the threshold b_l for all the stories is the same, i.e., $b_l=b$; ; $x_l(t)$ denotes the l-th story displacement relative to the ground at time t. Figure 6.1 shows the posterior robust failure probability (solid curve) of the structure, prior robust failure probability (dashed curve) and the nominal failure probability (dot-dashed curve) for different threshold levels of maximum interstory drift. It can be seen that the posterior robust failure probability is quite different from the other failure probabilities due to different levels of model uncertainties, confirming the importance of using data to update the failure probability.

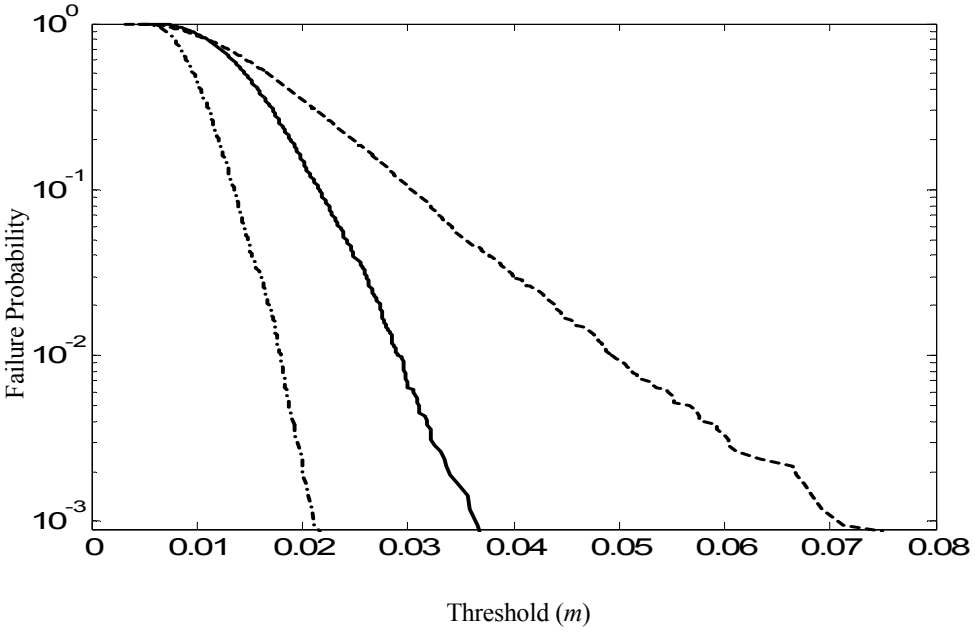

Figure 6.1: Posterior robust (solid curve), prior robust (dashed) and nominal (dot-dashed) failure probabilities plotted against the threshold of maximum interstory drift of all floors

Next, consider failure F defined as the exceedance over some threshold of the displacement of any one of the stories relative to the ground at any time within the 10s of ground shaking:

$$F = \bigcup_{n=0}^{1000} \bigcup_{l=1}^{4} \{ | x_l(t_n) | \geq b_l \} = \max_{\substack{n \in \{0,1,\ldots,1000\} \\ l \in \{1,\ldots,4\}}} \frac{| x_l(t_n) |}{b_l} \geq 1 \tag{6.28}$$

where the threshold b_l for all the stories is the same, i.e., $b_l=b$. Figure 6.2 shows the posterior robust failure probability (solid curve) of the structure, prior robust failure probability (dashed curve) and the nominal failure probability (dot-dashed curve) for different threshold levels of maximum displacement relative to the ground. Once again, it can be seen that the posterior robust failure probability is quite different from the nominal and the prior robust failure probability.

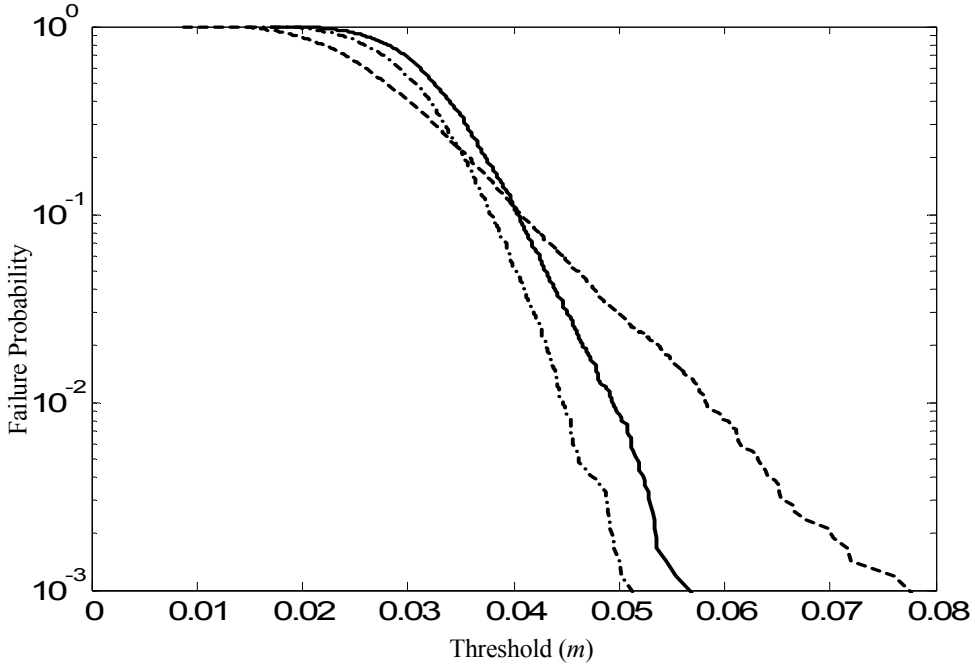

Figure 6.2: Posterior robust (solid curve), prior robust (dashed) and nominal (dot-dashed) failure probabilities plotted against the threshold of maximum displacements of all floors relative to the ground

Finally, consider failure F defined as the exceedance over some threshold of the absolute acceleration of any one of the stories at any time within the 10s of ground shaking:

$$F = \bigcup_{n=0}^{1000} \bigcup_{l=1}^{4} \{ | a_l(t_n) | \geq b_l \} = \max_{\substack{n \in \{0,1,\ldots,1000\} \\ l \in \{1,\ldots,4\}}} \frac{| a_l(t_n) |}{b_l} \geq 1 \tag{6.29}$$

where the threshold b_l for all the stories is the same, i.e., $b_l=b$; $a_l(t)$ denotes the l-th story absolute acceleration at time t. Figure 6.3 shows the posterior robust failure probability (solid curve) of the structure, prior robust failure probability (dashed curve) and the nominal failure probability (dot-dashed curve) for different threshold levels of maximum absolute acceleration. Similar observation can be seen once again as in the above two cases of failure as shown in Figures 6.1 and 6.2.

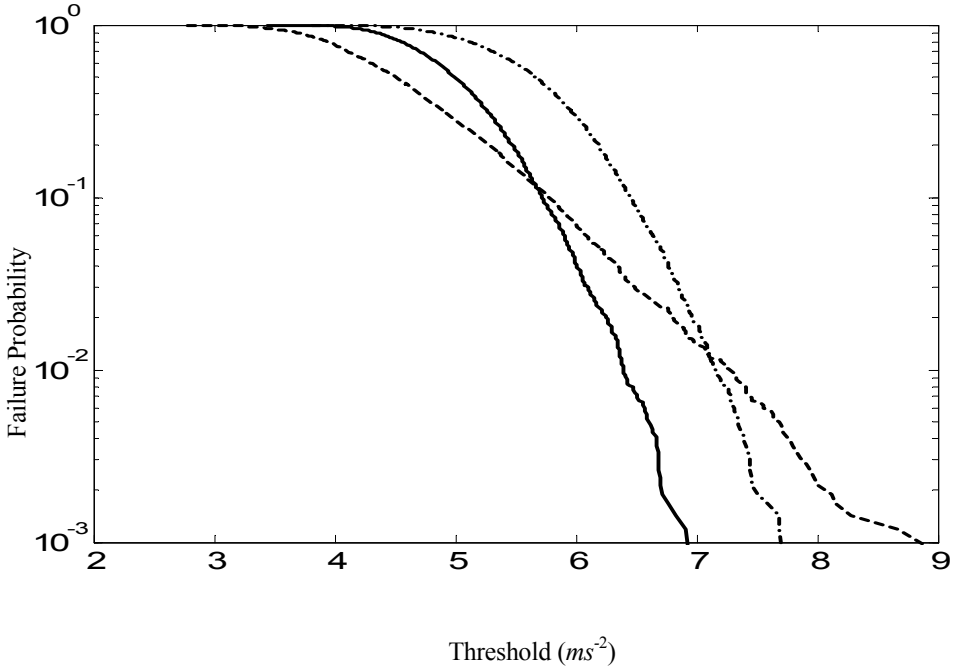

Figure 6.3: Posterior robust (solid curve), prior robust (dashed) and nominal (dot-dashed) failure probability against the threshold of maximum absolute acceleration of all floors

Appendix 6A

Let $L(q) = \frac{g^2(\xi) f^2(\xi)}{q(\xi)}$. $q^*(\xi)$ is the solution of the following constrained optimization problem:

$$q^*(\xi) = \arg\min_q \int L(q) d\xi$$
$$\text{s.t.}$$
$$\int q(\xi) d\xi = 1, q(\xi) \geq 0$$

By Calculus of Variation, it can be shown that $q^*(\xi)$ is the solution of the Euler-Lagrange Equation:

$$\frac{\partial (L(q) + \lambda q)}{\partial q} |_{q^*} = 0$$
$$\text{s.t.} \qquad \text{(A6.1)}$$
$$\int q^*(\xi) d\xi = 1, q^*(\xi) \geq 0$$

$$\frac{\partial (L(q) + \lambda q)}{\partial q} |_{q^*} = 0 \Rightarrow q^*(\xi) = \frac{|g(\xi)| f(\xi)}{\sqrt{\lambda}} (\because q^*(\xi) \geq 0) \text{ where } \lambda > 0$$

Substitute this to $\int q^*(\xi) d\xi = 1$, we obtain

$$\lambda = (\int |g(\xi)| f(\xi) d\xi)^2 \text{ and thus } q^*(\xi) = \frac{|g(\xi)| f(\xi)}{\int |g(\xi)| f(\xi) d\xi}$$

The corresponding minimum variance $Var_{q^*}[\frac{g(\xi) f(\xi)}{q^*(\xi)}]$ is given by:

$$Var_{q^*}[\frac{g(\xi) f(\xi)}{q^*(\xi)}] = (\int |g(\xi)| f(\xi) d\xi)^2 - E_f{}^2[g(\xi)] \qquad \text{(A6.2)}$$
$$= (\int |g(\xi)| f(\xi) d\xi)^2 - (\int g(\xi) f(\xi) d\xi)^2$$

If $g(\boldsymbol{\xi})\geq 0$, it can be seen that the variance of the IS estimator using the optimal ISD $q^*(\boldsymbol{\xi})$ will be zero.

Appendix 6B

The transition PDF of modified Metropolis-Hastings algorithm used in Subset Simulation is given by the following:

$$
\begin{aligned}
&K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}) \\
&= I(\boldsymbol{\theta}^*)\prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j) + (1-a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_j)] \\
&+[1-\int I(\tilde{\boldsymbol{\theta}})\prod_{j=1}^{G}[T_j(\tilde{\boldsymbol{\theta}}_j \mid \boldsymbol{\theta}_j) + (1-a(\boldsymbol{\theta}_j))\delta(\widetilde{\boldsymbol{\theta}_j} - \boldsymbol{\theta}_j)]d\tilde{\boldsymbol{\theta}}]\delta(\boldsymbol{\theta}^* - \boldsymbol{\theta}\ )
\end{aligned}
$$

where we have the following:

$$T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j) = T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_j^*)\pi(\boldsymbol{\theta}_j^*) \tag{B6.1}$$

From Appendix 2F, it can be seen that in general the above transition PDF will not satisfy the reversibility condition. To prove the validity of modified Metropolis-Hastings algorithms in Subset Simulation directly, we need to prove the above transition PDF satisfies the stationarity condition:

$$p(\boldsymbol{\theta}^*) = \int K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid F)d\boldsymbol{\theta} = \pi(\boldsymbol{\theta}^* \mid F) \tag{B6.2}$$

Our trick here is to expand $\prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_j^*) + (1-a(\boldsymbol{\theta}_j^*))\delta(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)]$ into the sum of terms (here there will be 2^G) since the integration will depend on the number of delta functions involved in the term. It can be seen that the number of terms which involves the product of k delta functions and G-k transition functions is equal to $C_k^G = G!/[(G\text{-}k)!k!]$.

$$
\begin{aligned}
&\prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_j^*) + (1-a(\boldsymbol{\theta}_j^*))\delta(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*)] \\
&= \prod_{j=1}^{G}[T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}_j^*)] + \prod_{j=1}^{G}(1-a(\boldsymbol{\theta}_j^*))\delta(\boldsymbol{\theta}_j - \boldsymbol{\theta}_j^*) \\
&+ \sum_{k=1}^{G-1}\sum_{m=1}^{C_k^n}[(1-a(\boldsymbol{\theta}_{n_1^m}^*))\delta(\boldsymbol{\theta}_{n_1^m} - \boldsymbol{\theta}_{n_1^m}^*)]...[(1-a(\boldsymbol{\theta}_{n_k^m}^*))\delta(\boldsymbol{\theta}_{n_k^m} - \boldsymbol{\theta}_{n_k^m}^*)] \\
&T_{i_1^m}(\boldsymbol{\theta}_{i_1^m} \mid \boldsymbol{\theta}_{i_1^m}^*)T_{i_2^m}(\boldsymbol{\theta}_{i_2^m} \mid \boldsymbol{\theta}_{i_2^m}^*)...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m} \mid \boldsymbol{\theta}_{i_{G-k}^m}^*)
\end{aligned}
\tag{B6.3}
$$

$$
\begin{aligned}
p(\boldsymbol{\theta}^*) &= \int K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid F)d\boldsymbol{\theta} = I_1 + I_2 + I_3 + \pi(\boldsymbol{\theta}^* \mid F) - J \\
&= I_1 + I_2 + I_3 + \pi(\boldsymbol{\theta}^* \mid F) - (J_1 + J_2 + J_3)
\end{aligned}
\tag{B6.4}
$$

where I_1, I_2, I_3, J are as follows:

$$
\begin{aligned}
I_1 &= \int I(\boldsymbol{\theta}^*)[\prod_{j=1}^{G} T_j(\boldsymbol{\theta}_j^* \mid \boldsymbol{\theta}_j)]\pi(\boldsymbol{\theta} \mid F)d\boldsymbol{\theta} \\
I_2 &= \int I(\boldsymbol{\theta}^*)[\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}_j^* - \boldsymbol{\theta}_j)]\pi(\boldsymbol{\theta} \mid F)d\boldsymbol{\theta} \\
I_3 &= \sum_{k=1}^{G-1}\sum_{m=1}^{C_k^n} I_{k,m} \\
I_{k,m} &= \int \frac{I(\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}))\delta(\boldsymbol{\theta}_{n_1^m}^* - \boldsymbol{\theta}_{n_1^m})]...[(1-a(\boldsymbol{\theta}_{n_k^m}))\delta(\boldsymbol{\theta}_{n_k^m}^* - \boldsymbol{\theta}_{n_k^m})]}{T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}^* \mid \boldsymbol{\theta}_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}_{i_2^m}^* \mid \boldsymbol{\theta}_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}^* \mid \boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta} \mid F)d\boldsymbol{\theta}} \\
J &= \int[\int I(\tilde{\boldsymbol{\theta}})\prod_{j=1}^{G}[T_j(\tilde{\boldsymbol{\theta}}_j \mid \boldsymbol{\theta}_j) + (1-a(\boldsymbol{\theta}_j))\delta(\widetilde{\boldsymbol{\theta}_j} - \boldsymbol{\theta}_j)]d\tilde{\boldsymbol{\theta}}]\delta(\boldsymbol{\theta}^* - \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid F)d\boldsymbol{\theta} \\
&= \pi(\boldsymbol{\theta}^* \mid F)\int I(\tilde{\boldsymbol{\theta}})\prod_{j=1}^{G}[T_j(\tilde{\boldsymbol{\theta}}_j \mid \boldsymbol{\theta}_j^*) + (1-a(\boldsymbol{\theta}_j^*))\delta(\widetilde{\boldsymbol{\theta}_j} - \boldsymbol{\theta}_j^*)]d\tilde{\boldsymbol{\theta}}] \\
&= J_1 + J_2 + J_3 \\
J_1 &= \pi(\boldsymbol{\theta}^* \mid F)\int I(\tilde{\boldsymbol{\theta}})\prod_{j=1}^{G} T_j(\tilde{\boldsymbol{\theta}}_j \mid \boldsymbol{\theta}_j^*)d\tilde{\boldsymbol{\theta}}
\end{aligned}
\tag{B6.5}
$$

$$
\begin{aligned}
J_2 &= \pi(\boldsymbol{\theta}^* \mid F)\int I(\tilde{\boldsymbol{\theta}})\prod_{j=1}^{G}[(1-a(\boldsymbol{\theta}^*_j))\delta(\widetilde{\boldsymbol{\theta}_j}-\boldsymbol{\theta}^*_j)]d\tilde{\boldsymbol{\theta}} \\
&= \pi(\boldsymbol{\theta}^* \mid F)I(\boldsymbol{\theta}^*)\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}^*_j)) \\
&= \pi(\boldsymbol{\theta}^* \mid F)\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}^*_j)) \\
J_3 &= \sum_{k=1}^{G-1}\sum_{m=1}^{C_k^n} J_{k,m} \\
J_{k,m} &= \pi(\boldsymbol{\theta}^* \mid F)\times \\
&\int \frac{I(\tilde{\boldsymbol{\theta}})[(1-a(\boldsymbol{\theta}^*_{n_1^m}))\delta(\tilde{\boldsymbol{\theta}}_{n_1^m}-\boldsymbol{\theta}^*_{n_1^m})]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))\delta(\tilde{\boldsymbol{\theta}}_{n_k^m}-\boldsymbol{\theta}^*_{n_k^m})]}{T_{i_1^m}(\tilde{\boldsymbol{\theta}}_{i_1^m} \mid \boldsymbol{\theta}^*_{i_1^m})T_{i_2^m}(\tilde{\boldsymbol{\theta}}_{i_2^m} \mid \boldsymbol{\theta}^*_{i_2^m})...T_{i_{G-k}^m}(\tilde{\boldsymbol{\theta}}_{i_{G-k}^m} \mid \boldsymbol{\theta}^*_{i_{G-k}^m})}d\tilde{\boldsymbol{\theta}} \\
&= \pi(\boldsymbol{\theta}^* \mid F)(1-a(\boldsymbol{\theta}^*_{n_1^m}))...(1-a(\boldsymbol{\theta}^*_{n_k^m})) \\
&\int \frac{I(\boldsymbol{\theta}^*_{n_1^m},...,\boldsymbol{\theta}^*_{n_k^m},\tilde{\boldsymbol{\theta}}_{i_1^m},...,\tilde{\boldsymbol{\theta}}_{i_{G-k}^m})T_{i_1^m}(\tilde{\boldsymbol{\theta}}_{i_1^m} \mid \boldsymbol{\theta}^*_{i_1^m})T_{i_2^m}(\tilde{\boldsymbol{\theta}}_{i_2^m} \mid \boldsymbol{\theta}^*_{i_2^m})...T_{i_{G-k}^m}(\tilde{\boldsymbol{\theta}}_{i_{G-k}^m} \mid \boldsymbol{\theta}^*_{i_{G-k}^m})}{d\tilde{\boldsymbol{\theta}}_{i_1^m}d\tilde{\boldsymbol{\theta}}_{i_2^m}...d\tilde{\boldsymbol{\theta}}_{i_{G-k}^m}}
\end{aligned}
\tag{B6.6}
$$

Now let's evaluate $I_1, I_2, I_3, J_1, J_2, J_3$

$$
\begin{aligned}
I_1 &= \int I(\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}^*_j \mid \boldsymbol{\theta}_j)]\pi(\boldsymbol{\theta} \mid F)d\boldsymbol{\theta} = \int I(\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}^*_j \mid \boldsymbol{\theta}_j)\pi(\boldsymbol{\theta}_j)]I(\boldsymbol{\theta})/P_F d\boldsymbol{\theta} \\
&= \int I(\boldsymbol{\theta}^*)[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}^*_j)\pi(\boldsymbol{\theta}^*_j)]I(\boldsymbol{\theta})/P_F d\boldsymbol{\theta} \quad \text{(by B6.1)} \\
&= \int I(\boldsymbol{\theta})[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}^*_j)]\pi(\boldsymbol{\theta}^*)I(\boldsymbol{\theta}^*)/P_F d\boldsymbol{\theta} = \int I(\boldsymbol{\theta})[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}^*_j)]\pi(\boldsymbol{\theta}^* \mid F)d\boldsymbol{\theta} \\
&= \pi(\boldsymbol{\theta}^* \mid F)[\int I(\boldsymbol{\theta})[\prod_{j=1}^{G}T_j(\boldsymbol{\theta}_j \mid \boldsymbol{\theta}^*_j)]d\boldsymbol{\theta}] = J_1
\end{aligned}
\tag{B6.7}
$$

$$I_2 = \int I(\boldsymbol{\theta}^*)[\prod_{j=1}^{G}(1-a(\boldsymbol{\theta}_j))\delta(\boldsymbol{\theta}_j^*-\boldsymbol{\theta}_j)]\pi(\boldsymbol{\theta}\mid F)d\boldsymbol{\theta}$$

$$= I(\boldsymbol{\theta}^*)\prod_{j=1}^{G}\{[1-a(\boldsymbol{\theta}_j^*)]\}\pi(\boldsymbol{\theta}^*\mid F)$$

$$= \pi(\boldsymbol{\theta}^*\mid F)\prod_{j=1}^{G}\{[1-a(\boldsymbol{\theta}_j^*)]\} \qquad \text{(B6.8)}$$

$$(\because I(\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*\mid F) = I(\boldsymbol{\theta}^*)I(\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)/P_F = I(\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)/P_F)$$

$$= J_2$$

$I_{k,m}$ is given by:

$$I_{k,m} = \int \frac{I(\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}))\delta(\boldsymbol{\theta}_{n_1^m}^*-\boldsymbol{\theta}_{n_1^m})]...[(1-a(\boldsymbol{\theta}_{n_k^m}))\delta(\boldsymbol{\theta}_{n_k^m}^*-\boldsymbol{\theta}_{n_k^m})]}{T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}^*\mid\boldsymbol{\theta}_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}_{i_2^m}^*\mid\boldsymbol{\theta}_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}^*\mid\boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta}\mid F)d\boldsymbol{\theta}}$$

$$= \int \frac{I(\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}))\delta(\boldsymbol{\theta}_{n_1^m}^*-\boldsymbol{\theta}_{n_1^m})]...[(1-a(\boldsymbol{\theta}_{n_k^m}))\delta(\boldsymbol{\theta}_{n_k^m}^*-\boldsymbol{\theta}_{n_k^m})]}{T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}^*\mid\boldsymbol{\theta}_{i_1^m})T_{i_2^m}(\boldsymbol{\theta}_{i_2^m}^*\mid\boldsymbol{\theta}_{i_2^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}^*\mid\boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta}_{n_1^m})...\pi(\boldsymbol{\theta}_{n_k^m})\pi(\boldsymbol{\theta}_{i_1^m})...\pi(\boldsymbol{\theta}_{i_{G-k}^m})I(\boldsymbol{\theta})/P_F d\boldsymbol{\theta}}$$

$$= \int \frac{I(\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}^*))]...[(1-a(\boldsymbol{\theta}_{n_k^m}^*))]}{[T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}^*\mid\boldsymbol{\theta}_{i_1^m})\pi(\boldsymbol{\theta}_{i_1^m})]...[T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}^*\mid\boldsymbol{\theta}_{i_{G-k}^m})\pi(\boldsymbol{\theta}_{i_{G-k}^m})]\pi(\boldsymbol{\theta}_{n_1^m}^*)...\pi(\boldsymbol{\theta}_{n_k^m}^*)I(\boldsymbol{\theta}_{n_1^m}^*,...,\boldsymbol{\theta}_{n_k^m}^*,\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m})/P_F d\boldsymbol{\theta}_{i_1^m}d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m}}$$

$$= \int \frac{I(\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}^*))]...[(1-a(\boldsymbol{\theta}_{n_k^m}^*))]}{[T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}\mid\boldsymbol{\theta}_{i_1^m}^*)\pi(\boldsymbol{\theta}_{i_1^m}^*)]...[T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}\mid\boldsymbol{\theta}_{i_{G-k}^m}^*)\pi(\boldsymbol{\theta}_{i_{G-k}^m}^*)]\pi(\boldsymbol{\theta}_{n_1^m}^*)...\pi(\boldsymbol{\theta}_{n_k^m}^*)I(\boldsymbol{\theta}_{n_1^m}^*,...,\boldsymbol{\theta}_{n_k^m}^*,\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m})/P_F d\boldsymbol{\theta}_{i_1^m}d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m}}$$

$$= \int \frac{I(\boldsymbol{\theta}^*)[(1-a(\boldsymbol{\theta}_{n_1^m}^*))]...[(1-a(\boldsymbol{\theta}_{n_k^m}^*))]}{T_{i_1^m}(\boldsymbol{\theta}_{i_1^m}\mid\boldsymbol{\theta}_{i_1^m}^*)...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m}\mid\boldsymbol{\theta}_{i_{G-k}^m}^*)\pi(\boldsymbol{\theta}^*)I(\boldsymbol{\theta}_{n_1^m}^*,...,\boldsymbol{\theta}_{n_k^m}^*,\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m})/P_F d\boldsymbol{\theta}_{i_1^m}d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m}}$$

$$= \pi(\boldsymbol{\theta}^* \mid F)[(1-a(\boldsymbol{\theta}^*_{n_1^m}))]...[(1-a(\boldsymbol{\theta}^*_{n_k^m}))]$$

$$\int T_{i_1^m}(\boldsymbol{\theta}_{i_1^m} \mid \boldsymbol{\theta}^*_{i_1^m})...T_{i_{G-k}^m}(\boldsymbol{\theta}_{i_{G-k}^m} \mid \boldsymbol{\theta}^*_{i_{G-k}^m}) I(\boldsymbol{\theta}^*_{n_1^m},...,\boldsymbol{\theta}^*_{n_k^m},\boldsymbol{\theta}_{i_1^m},\boldsymbol{\theta}_{i_2^m},...,\boldsymbol{\theta}_{i_{G-k}^m}) d\boldsymbol{\theta}_{i_1^m} d\boldsymbol{\theta}_{i_2^m}...d\boldsymbol{\theta}_{i_{G-k}^m} \quad \text{(B6.9)}$$

$$= J_{k,m}$$

Combining (B6.5)-(B6.9), we have:

$$I_1 + I_2 + I_3 = J_1 + J_2 + J_3$$

Thus by this and (B6.4), given $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta} \mid F)$, we have:

$$p(\boldsymbol{\theta}^*) = \int K(\boldsymbol{\theta}^* \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid F) d\boldsymbol{\theta} = \pi(\boldsymbol{\theta}^* \mid F)$$

Appendix 6C

With (6.15)-(6.19), we can then derive (6.20) as follows:

$$\alpha_i = \frac{p(\mathcal{D} \mid F, t_{i+1})}{p(\mathcal{D} \mid F, t_i)} = \frac{\int p^{t_{i+1}}(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid F) d\boldsymbol{\theta}}{p(\mathcal{D} \mid F, t_i)}$$

$$= \int \frac{p^{t_{i+1}}(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid F)}{p(\boldsymbol{\theta} \mid F, \mathcal{D}, t_i)} \frac{p(\boldsymbol{\theta} \mid F, \mathcal{D}, t_i)}{p(\mathcal{D} \mid F, t_i)} d\boldsymbol{\theta}$$

$$= \int \frac{p^{t_{i+1}}(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid F)}{p(\boldsymbol{\theta} \mid F, \mathcal{D}, t_i) p(\mathcal{D} \mid F, t_i)} p(\boldsymbol{\theta} \mid F, \mathcal{D}, t_i) d\boldsymbol{\theta}$$

$$= \int \frac{p^{t_{i+1}}(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid F)}{p^{t_i}(\mathcal{D} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta} \mid F)} p(\boldsymbol{\theta} \mid F, \mathcal{D}, t_i) d\boldsymbol{\theta}$$

$$= \int \frac{p^{t_{i+1}}(\mathcal{D} \mid \boldsymbol{\theta})}{p^{t_i}(\mathcal{D} \mid \boldsymbol{\theta})} p(\boldsymbol{\theta} \mid F, \mathcal{D}, t_i) d\boldsymbol{\theta}$$

$$\approx \frac{1}{N} \sum_{k=1}^{N} \frac{p^{t_{i+1}}(\mathcal{D} \mid \boldsymbol{\theta}^{(k)})}{p^{t_i}(\mathcal{D} \mid \boldsymbol{\theta}^{(k)})} = \frac{1}{N} \sum_{k=1}^{N} p^{t_{i+1}-t_i}(\mathcal{D} \mid \boldsymbol{\theta}^{(k)})$$

where $\boldsymbol{\theta}^{(k)}$, k=1, 2,…, N follows $p(\boldsymbol{\theta}|F,\mathcal{D},t_i)$. Similar to the above, it can be shown that the following is true by repeating the above proof by replacing F by $\sim F$:

$$\beta_i = \frac{p(\mathcal{D}|\sim F, t_{i+1})}{p(\mathcal{D}|\sim F, t_i)} = \int \frac{p^{t_{i+1}}(\mathcal{D}|\boldsymbol{\theta})}{p^{t_i}(\mathcal{D}|\boldsymbol{\theta})} p(\boldsymbol{\theta}|\sim F, \mathcal{D}, t_i) d\boldsymbol{\theta}$$
$$\approx \frac{1}{N'} \sum_{k=1}^{N'} \frac{p^{t_{i+1}}(\mathcal{D}|\tilde{\boldsymbol{\theta}}^{(m)})}{p^{t_i}(\mathcal{D}|\tilde{\boldsymbol{\theta}}^{(m)})} = \frac{1}{N'} \sum_{m=1}^{N'} p^{t_{i+1}-t_i}(\mathcal{D}|\tilde{\boldsymbol{\theta}}^{(m)})$$

where $\tilde{\boldsymbol{\theta}}^{(m)}$, m=1, 2,…, N', follows $p(\boldsymbol{\theta}|\sim F,\mathcal{D},t_i)$.

CHAPTER 7

Updating reliability of nonlinear dynamic systems using near real-time data

Using real-time data to assess the uncertain system performance and to evaluate various failure probabilities when the system is subjected to severe dynamic excitations, such as explosions, strong winds or earthquakes, is a very challenging problem. There are two possible important problems to consider. The first problem is to use the data from the monitored system to update its reliability against future excitations, which has been considered in Chapter 6. The second one is to use the data to update the reliability for unobserved quantities during recent excitation. It is often of interest to the owners, design engineers, or insurance companies to know, immediately after a severe dynamic event, the performance of the structure during the event. In this chapter, our focus will be on this aforementioned second problem. Data from an instrumented structure are often incomplete and sparse and the corresponding input or excitation may or may not be measured. Ching and Beck (2007) proposed a method to update the reliability using real-time dynamic data for linear dynamic systems with no uncertainties in the model parameters. Here we tackle the problem of calculating the probability that any unobserved system response of interest exceeds its threshold during the time when the system is subjected to dynamic excitation, based on real-time measurements of some output and possibly input from the system. A novel stochastic simulation method is used that updates in near real-time the reliability of

this system. Part of the material presented in this chapter is presented in Cheung and Beck (2008d).

7.1 Proposed stochastic simulation method

Failure F is defined as the event that the system performs unsatisfactorily. One common type of failure of interest is the event that any unobserved response of interest of the system exceeds some specified threshold over any time duration of interest when the system is subjected to dynamic excitation. Such unobserved response of interest is a function of the unobserved state vector $X_N=[x_0,x_1,\ldots,x_N]$ at different discrete times where $x_n \in \mathbb{R}^{N_s}$. Now suppose that during some event, measurements are made of the system output (response) $Y_N= [y_1,y_2,\ldots,y_N]$ where $y_n \in \mathbb{R}^{N_o}$ and its input (excitation) $U_N= [u_1,u_2,\ldots,u_N]$ where $u_n \in \mathbb{R}^{N_i}$. The updated robust failure probability given these data and a class $\mathcal{M}$ of models for the system is given by:

$$P(F \mid Y_N, U_N, \mathcal{M}) = \int P(F \mid X_N, Y_N, U_N, \mathcal{M}) p(X_N \mid Y_N, U_N, \mathcal{M}) dX_N \tag{7.1}$$

For simplicity in presentation, the conditioning on $\mathcal{M}$ and U_N will be left implicit.

Evaluation of $P(F|Y_N)$ is computationally very challenging. First, one needs to obtain the probabilistic information $p(X_N|Y_N)$ through Bayes' Theorem. However, for nonlinear systems, regardless of whether there are uncertainties in the model parameters or uncertainties in the excitation, an analytical form of $p(X_N|Y_N)$ is generally not available. Second, the integral in (7.1) involves an integration in a very high dimensional space which cannot be evaluated analytically or by straightforward numerical quadrature. To solve the first difficulty, a stochastic simulation method is proposed which generates samples from $p(X_N|Y_N)$ that provide a characterization of the probabilistic information in the PDF. An appropriate stochastic simulation method is then used to solve the second difficulty by using the samples from $p(X_N|Y_N)$.

7.1.1 Simulation of samples from $p(X_N|Y_N)$ for the calculation of $P(F|Y_N)$

We consider the following general stochastic discrete-time state-space model $\mathcal{M}$ of a dynamical system:

$$\begin{aligned} x_n &= f_{n-1}(x_{n-1}, u_{n-1}, \gamma_{n-1}) \overset{\gamma_n \text{ PDF}}{\Rightarrow} p(x_n \mid x_{n-1}, u_{n-1}) \text{ [state transition]} \\ y_n &= h_n(x_n, u_n, v_n) \overset{v_n \text{ PDF}}{\Rightarrow} p(y_n \mid x_n, u_n) \text{ [observation output equation]} \end{aligned} \tag{7.2}$$

where $x_n \in \mathbb{R}^{N_s}$ denotes the model state, $u_n \in \mathbb{R}^{N_i}$ denotes the system input, $y_n \in \mathbb{R}^{N_o}$ denotes the observed system output, $\gamma_n \in \mathbb{R}^{N_l}$ denotes the uncertain disturbances, and $v_n \in \mathbb{R}^{N_r}$ denotes the prediction errors, all being at discrete time n. The probability models for the γ_n's and v_n's are prescribed; as usual, they are usually taken as Gaussian which is justified by the Principle of Maximum Information Entropy (Jaynes 2003). The two fundamental system probability models given in (7.2), along with the specification of the PDF $p(x_0)$ for the initial states, completely define the stochastic dynamics of the system. Any unknown model parameters can be augmented into the model state. Therefore, whether the system inputs or excitations u_n's are uncertain or known, they will be left implicit in the notation.

Let $\hat{Y}_n = [\hat{y}_1, \hat{y}_2, ..., \hat{y}_n]$ be the measured system output up to the current time n. The first step of our proposed method requires performing the following Particle Filtering (PF) algorithm to generate samples from $p(x_n \mid \hat{Y}_n)$ (Doucet et al. 2000, Ching et al. 2006b) (an overview of Particle Filtering is given in Appendix 7A-7D):

1. Draw K samples $x_0^{(k)}$ from $q(x_0) = p(x_0)$ and initialize the importance weights $\beta_{0,k} = 1/K$ for k=1,2,...,K.
2. Repeat the following for time n = 1,2,…, N:

2.1. Draw K candidate samples $\tilde{x}_n^{(k)}$ from a proposal PDF $q(x_n \mid x_{n-1}^{(k)}, \hat{y}_n)$ and update the importance weights as follows for k=1,2,...,K:

$$\beta_{n,k} = \beta_{n-1,k} \frac{p(\tilde{x}_n^{(k)} \mid x_{n-1}^{(k)}) p(\hat{y}_n \mid x_n^{(k)})}{q(\tilde{x}_n^{(k)} \mid x_{n-1}^{(k)}, \hat{y}_n)} \tag{7.3}$$

2.2. Compute the normalized weight $w_{n,k} = \beta_{n,k} / \sum_{j=1}^{K} \beta_{n,j}$.

2.3. Calculate the effective number of samples:

$$N_e = 1 / \sum_{k=1}^{K} w_{n,k}^2 \tag{7.4}$$

2.4. Set $\hat{x}_{n-1}^{(j)} = \hat{\hat{x}}_{n-1}^{(j)}$ for j=1,2,…,K. If $N_e \geq K_o$, a prescribed threshold, set $x_n^{(k)} = \tilde{x}_n^{(k)}$, $\beta_{n,k} = w_{n,k}$ for k=1,2,...,K. Otherwise, do the resampling as follows for j=1,2,…,K:

$$x_n^{(j)} = \tilde{x}_n^{(k)} \quad \text{with probability} \quad w_{n,k} \tag{7.5}$$

$$\hat{\hat{x}}_{n-1}^{(j)} = \hat{x}_{n-1}^{(k)} \tag{7.6}$$

and set $w_{n,k} = \beta_{n,k} = 1/K$ for k=1,2,...,K.

2.5. If resampling is implemented, repeat the following for M times, for each k=1,2,…,K: A candidate sample x_{cand} is drawn from a proposal PDF $q_{MH}(x_{\text{cand}} \mid x_n^{(k)})$. Compute the acceptance probability r :

$$r = \frac{p(x_n = x_{\text{cand}} \mid x_{n-1} = \hat{\hat{x}}_{n-1}^{(k)}) p(\hat{y}_n \mid x_n = x_{cand}) q_{MH}(x_n^{(k)} \mid x_{\text{cand}})}{p(x_n = x_n^{(k)} \mid x_{n-1} = \hat{\hat{x}}_{n-1}^{(k)}) p(\hat{y}_n \mid x_n = x_n^{(k)}) q_{MH}(x_{\text{cand}} \mid x_n^{(k)})} \tag{7.7}$$

If $r>u$ where u~Uniform(0,1), $x_n^{(k)} = x_{\text{cand}}$. Otherwise, $x_n^{(k)}$ remains unchanged.

2.6. $\hat{x}_n^{(k)} = x_n^{(k)}$, k=1,2,…,K.

N_e is the effective number of samples due to the non-uniformity of weights, $\beta_{n,k}$ k=1,…,K; K_o in Step 2.4 is the threshold prescribed to decide whether resampling should be carried out; it is chosen to be a certain fraction of K, e.g. 0.5K is used in our example later. Ching et al. (2006bc) apply the PF algorithm to Bayesian state estimation of uncertain dynamical systems but use the size of the coefficient of variation of $\beta_{n,k}$, k=1,…,K, to decide whether resampling should be performed, which is equivalent to our choice.

The PF algorithm obtains probabilistic information for $p(x_n \mid \hat{Y}_n)$ but recall from (7.1) that to update the reliability, it is critical to obtain samples from $p(X_N \mid \hat{Y}_N)$ where X_N is the whole time history of responses instead of just at a particular time. The samples $x_n^{(k)}$ obtained from the above procedure lie in the high-probability region of $p(x_n \mid \hat{Y}_n)$ which can be approximated as follows using $w_{n,k}$ and $x_n^{(k)}$:

$$p(x_n \mid \hat{Y}_n) \approx \sum_{k=1}^{K} w_{n,k} \delta(x_n - x_n^{(k)}) \tag{7.8}$$

However, the $[x_0^{(k)}\ x_1^{(k)} \ldots x_N^{(k)}]$ do not necessarily lie in the high-probability region of $p(X_N \mid \hat{Y}_N)$. To generate samples in this region, the following steps can be added to the PF algorithm. At the beginning of Step 2.4, let $\hat{x}_p^{(j)} = \hat{x}_p^{(j)}$ for p=0,1,…,n-1 and j=1,2,…,K, then after (7.5), add:

$$\hat{x}_p^{(j)} = \hat{x}_p^{(k)} \text{ for } p=0,1,\ldots,n\text{-}1;\ \hat{x}_n^{(j)} = x_n^{(j)} \tag{7.9}$$

After Step 2.4, add an additional step, set $\hat{x}_n^{(k)} = x_n^{(k)}$, k=1,2,.., K. Let $\hat{X}_n^{(k)} = [\hat{x}_0^{(k)} \hat{x}_1^{(k)} ... \hat{x}_n^{(k)}]$ for k=1,2,…, K. It can be shown that $\hat{X}_N^{(k)}$, k=1,2,…,K, can be used to approximate $p(X_N \mid \hat{Y}_N)$ as follows:

$$p(X_N \mid \hat{Y}_N) \approx \sum_{k=1}^{K} w_{N,k} \delta(X_N - \hat{X}_N^{(k)}) \tag{7.10}$$

Theoretically, the samples from $p(X_N \mid \hat{Y}_N)$ can be obtained using resampling from $\hat{X}_N^{(k)}$'s with weights $w_{N,k}$'s. However, it is expected that samples simulated relying on $\hat{X}_N^{(k)}$'s will give a poor representation of $p(X_N \mid \hat{Y}_N)$ because a lot of $\hat{x}_p^{(k)}$'s at time step p<N (especially for p not close to N) are repeated due to the resampling step in the PF algorithm. This point is confirmed by the example considered later. Thus, we present a way to alleviate this problem. Note that $p(X_N \mid \hat{Y}_N)$ can be expressed as:

$$p(X_N \mid \hat{Y}_N) = p(x_N \mid \hat{Y}_N) \prod_{n=0}^{N-1} p(x_n \mid x_{n+1}, ..., x_N, \hat{Y}_N) = p(x_N \mid \hat{Y}_N) \prod_{n=0}^{N-1} p(x_n \mid x_{n+1}, \hat{Y}_n) \tag{7.11}$$

Thus, samples $X_N^{*(j)} = [x_0^{*(j)} x_1^{*(j)} ... x_N^{*(j)}]$, j=1, 2,…,J, from $p(X_N \mid \hat{Y}_N)$ can be simulated by the following algorithm:

3.1. After completing Step 2 of the PF algorithm, $x_N^{(j)}$ should be in the high-probability region of $p(x_N \mid \hat{Y}_N)$. If $w_{N,k} \neq 1/K$ for all k, do the resampling step as in (7.5) and Step 2.5 to obtain better samples $x_N^{(j)}$, j=1,2,…,K, from $p(x_N \mid \hat{Y}_N)$. Thus, $x_N^{*(j)} = x_N^{(j)}$, j=1,…,K.
3.2. For n=N-1,N-2, …, 0, given $x_{n+1}^{*(j)}$, simulate $x_n^{*(j)}$ from $p(x_n \mid x_{n+1}, \hat{Y}_n)$ as follows. By noting $p(x_n \mid x_{n+1}, \hat{Y}_n)$ may be expressed in terms of $p(x_n \mid \hat{Y}_n)$ using Bayes' Theorem:

$$p(x_n \mid x_{n+1}, \hat{Y}_n) = \frac{p(x_n \mid \hat{Y}_n) p(x_{n+1} \mid x_n, \hat{Y}_n)}{p(x_{n+1} \mid \hat{Y}_n)} \propto p(x_n \mid \hat{Y}_n) p(x_{n+1} \mid x_n), \tag{7.12}$$

simulate samples $x_n^{*(j)}$ from $p(x_n \mid x_{n+1}, \hat{Y}_n)$, given $x_{n+1}^{*(j)}$, as follows: For each j, $x_n^{*(j)} = x_n^{(k)}$ with probability $w_{n,k}$* given by:

$$w_{n,k}^* = w_{n,k}\, p(x_{n+1}^{*(j)} \mid x_n^{(k)}) / \sum_{k=1}^{K} w_{n,k}\, p(x_{n+1}^{*(j)} \mid x_n^{(k)}) \tag{7.13}$$

The $x_n^{(k)}$'s from the PF algorithm lie in the high-probability region of $p(x_n \mid \hat{Y}_n)$ and not necessarily that of $p(x_n \mid x_{n+1}, \hat{Y}_n)$. Thus, the weight of each sample needs to be adjusted for correct resampling as in Step 3.1.

7.1.2 Calculation of $P(F \mid \hat{Y}_N)$

For simplicity, consider the case of (1) where $P(F \mid \hat{Y}_N) = I(X_N \in F)$ (=1 if $X_N \in F$ and 0 otherwise), is estimated using Monte Carlo simulation as follows:

$$P(F \mid \hat{Y}_N) \approx \frac{1}{K} \sum_{k=1}^{K} I(X_N^{*(k)} \in F) \tag{7.14}$$

However, for the case where the updated failure probability $P(F \mid \hat{Y}_N)$ is small (e.g. <0.1), a large number of samples (and thus number of dynamic analyses) is required to obtain a reasonably accurate estimate of $P(F \mid \hat{Y}_N)$. For increased computational efficiency, a novel stochastic simulation method incorporating Subset Simulation (Au and Beck 2001b) has recently been developed as given as follows.

7.1.2.1 Subset Simulation with a novel hybrid Gibbs-MCMC conditional-on-failure algorithm

$P(F|\mathcal{D})$ can be calculated using the framework of Subset Simulation (SS) (Au and Beck 2001b) as follows:

$$P(F \mid \mathcal{D}) = P(F_1 \mid \mathcal{D}) \prod_{m=1}^{L-1} P(F_{m+1} \mid F_m, \mathcal{D}) \tag{15}$$

where $F_1 \supset F_2 \supset ... \supset F_L = F$ and $P(F_1 \mid \mathcal{D})$ can be estimated using Monte Carlo simulation as follows:

$$P(F_1 \mid \mathcal{D}) \approx \frac{1}{K} \sum_{k=1}^{K} I(X_N^{*(k)} \in F_1) \tag{16}$$

where $I(X_N^{*(k)} \in F_1)$ equals 1 if $X_N^{*(k)} \in F_1$ and 0 otherwise. For m>0, $P(F_{m+1} \mid F_m, \mathcal{D})$ can be estimated as follows:

$$P(F_{m+1} \mid F_m, \mathcal{D}) \approx \frac{1}{K} \sum_{k=1}^{K} I(X_{N,m}^{*(k)} \in F_{m+1}) \tag{17}$$

where $X_{N,m}^{*(k)}$ are the samples from $p(X_N \mid F_m, \hat{Y}_N)$; $I(X_{N,m}^{*(k)} \in F_{m+1})$ equals 1 if $X_{N,m}^{*(k)} \in F_{m+1}$ and 0 otherwise. $F_1, F_2 ..., F_{L-1}$ are selected such that $P(F_{m+1} \mid F_m, \mathcal{D})$, m=1,..,L and $P(F_1 \mid \mathcal{D})$ are not smaller than some pre-specified threshold p_0, e.g., 0.1. In level m, after estimating $P(F_m \mid F_{m-1}, \mathcal{D})$, we have some failure samples (about p_0K samples) from $p(X_N \mid F_m, \hat{Y}_N)$ and more samples (about (1- p_0)K more) from $p(X_N \mid F_m, \hat{Y}_n)$ are required to estimate $P(F_{m+1} \mid F_m, \mathcal{D})$. Due to the structure of $p(X_N \mid \hat{Y}_N)$, it is very challenging in real practice to simulate samples from $p(X_N \mid F_m, \hat{Y}_N)$ even we already have some samples from $p(X_N \mid F_m, \hat{Y}_N)$ and the modified Metropolis Hastings algorithm proposed in Au and Beck (2001b) is not applicable here. The detailed explanation for this will be presented in a future publication. Here a novel Hybrid Gibbs-MCMC conditional-on-failure algorithm is proposed for simulating samples from $p(X_N \mid F_m, \hat{Y}_N)$ based on some samples from $p(X_N \mid F_m, \hat{Y}_N)$ as follows: About p_0K Markov chains of samples are generated in parallel

using each of the p_0K available samples from $p(X_N \mid F_m, \hat{Y}_N)$ as the starting point for each Markov chain. Along a chain, a new sample from $p(X_N \mid F_m, \hat{Y}_N)$ is generated based on a previous sample on the chain. This is repeated until about $(1-p_0)/p_0$ more samples are generated on each chain. Given a sample $X_N^* = [\, x_0^* \; x_1^* \ldots x_N^* \,]$ from $p(X_N \mid F_m, \hat{Y}_N)$ on a chain, a new sample X_N^{**} from $p(X_N \mid F_m, \hat{Y}_N)$ can be simulated using the following procedures:

1. Let $X_{\text{cand}} = [x_{0,\text{cand}} x_{1,\text{cand}} \ldots x_{N,\text{cand}}]$ be the candidate sample. Simulate $x_{0,\text{cand}}$* from $q(x_{0,\text{cand}}* \mid x_0^*, \ldots, x_N^*)$. Compute the acceptance probability r:

$$r = \frac{p(x_{0,\text{cand}}*)\, p(x_1 = x_1^* \mid x_0 = x_{0,\text{cand}}*)\, q(x_0^*, \ldots, x_N^* \mid x_{0,\text{cand}}*)}{p(x_0^*)\, p(x_1 = x_1^* \mid x_0 = x_0^*)\, q(x_{0,\text{cand}}* \mid x_0^*, \ldots, x_N^*)} \tag{18}$$

If $r>u$ where u~Uniform(0,1), $x_{0,\text{cand}} = x_{0,\text{cand}}$*. Otherwise, $x_{0,\text{cand}} = x_0^*$.

2. For n= 1, 2, …, N, simulate $x_{n,\text{cand}}$* from $q(x_{n,\text{cand}}* \mid x_{0,\text{cand}}, \ldots, x_{n\text{-}1,\text{cand}}, x_n^*, \ldots, x_N^*)$. Compute the acceptance probability r:

$$\begin{aligned} r = {} & \frac{p(x_{n+1} = x_{n+1}^* \mid x_n = x_{n,\text{cand}}*)\, p(\hat{y}_n \mid x_n = x_{n,\text{cand}}*)\, p(x_n = x_{n,\text{cand}}* \mid x_{n-1} = x_{n-1,\text{cand}})}{p(x_{n+1} = x_{n+1}^* \mid x_n = x_n^*)\, p(\hat{y}_n \mid x_n = x_n^*)\, p(x_n = x_n^* \mid x_{n-1} = x_{n-1,\text{cand}})} \\ & \times \frac{q(x_{0,\text{cand}}, \ldots, x_{n\text{-}1,\text{cand}}, x_n^*, \ldots, x_N^* \mid x_{n,\text{cand}}*)}{q(x_{n,\text{cand}}* \mid x_{0,\text{cand}}, \ldots, x_{n\text{-}1,\text{cand}}, x_n^*, \ldots, x_N^*)} \end{aligned} \tag{19}$$

If $r>u$ where u~Uniform(0,1), $x_{n,\text{cand}} = x_{n,\text{cand}}$*. Otherwise, $x_{n,\text{cand}} = x_n^*$.

3. If $X_{\text{cand}} \in F_m$, $X_N^{**} = X_{\text{cand}}$; otherwise, $X_N^{**} = X_N^*$.

The details for the choice of $q(x_{0,\text{cand}}* \mid x_0^*, \ldots, x_N^*)$, $q(x_{n,\text{cand}}* \mid x_{0,\text{cand}}, \ldots, x_{n\text{-}1,\text{cand}}, x_n^*, \ldots, x_N^*)$ are not discussed here for brevity and are presented in a future publication.

7.2 Illustrative example with real seismic data from a seven-story hotel

In this example, a seven-story hotel located in Van Nuys in the San Fernando Valley of Los Angeles County is considered. It is a reinforced-concrete moment-frame building. It was subjected to severe damage during the 1994 Northridge earthquake. We are interested in using accelerometer data collected during this earthquake to do post-earthquake assessment of the "failure" probability of the building during the event. The data is available online from the CSMIP program of the California Geological Survey (http://db.cosmos-eq.org).

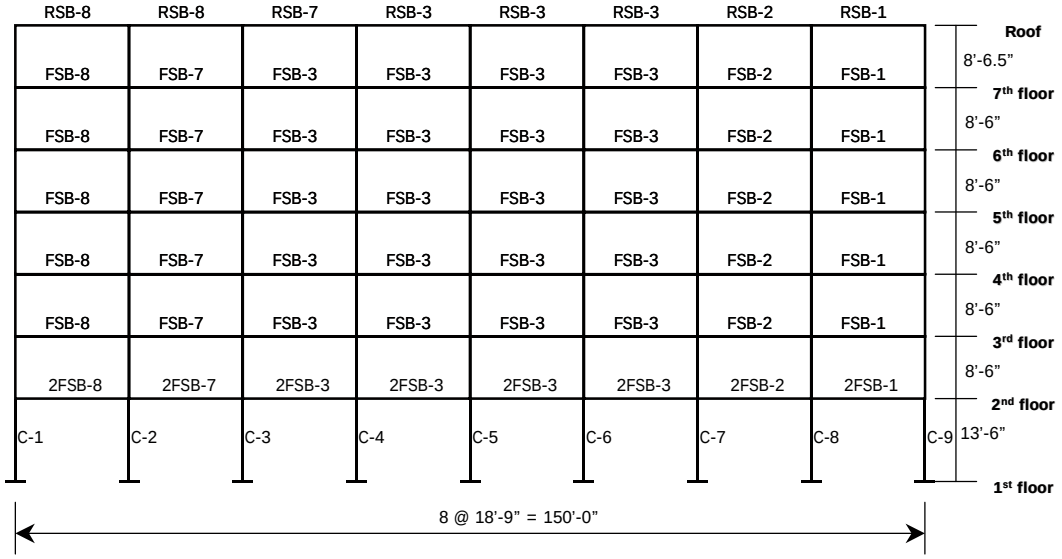

Figure 7.1: South frame elevation (Ching et al. 2006c)

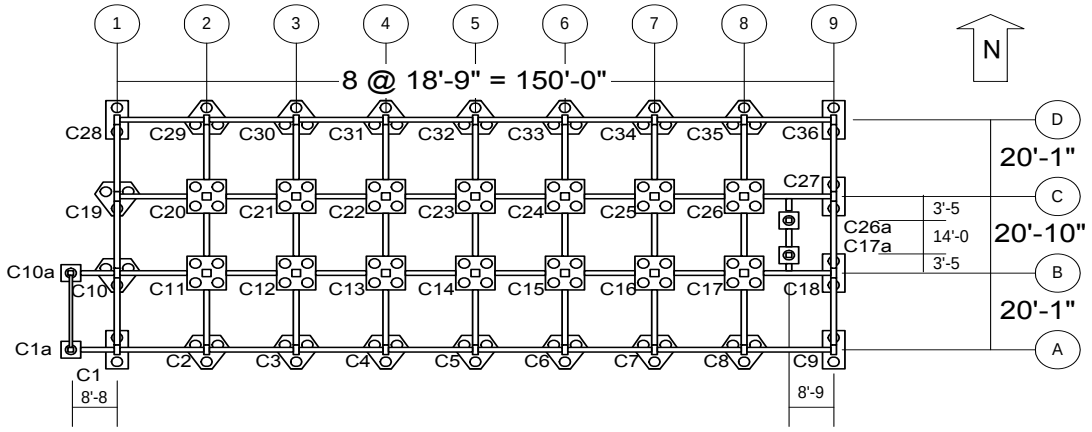

Figure 7.2: Hotel column plan (Ching et al. 2006c)

The E-W acceleration data of the ground floor, the second floor, third floor, sixth floor and the roof of the hotel during the earthquake are used. The south frame elevation with column and beam numbering is shown in Figure 7.1 and the column plan is shown in Figure 7.2. The following seven-DOF deteriorating shear-building model developed by Ching et al. (2006c) is used as the stochastic identification model (which is highly nonlinear) in this example where $x(t)=[x_7(t)\ \ldots x_1(t)]^T$ denotes the displacement relative to the ground and the mass matrix is a diagonal matrix $diag(\{m_7, \ldots, m_1\})$.

$$\frac{d}{dt}\begin{bmatrix} x(t) \\ \dot{x}(t) \\ \theta(t) \end{bmatrix} = \begin{bmatrix} \dot{x}(t) \\ -[\mathbf{M}^{-1}\mathbf{K}(t) \quad \mathbf{M}^{-1}\mathbf{C}(t)]\begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \mathbf{M}^{-1}\mathbf{F} \\ 0 \end{bmatrix} \cdot \ddot{u}_g(t) + \begin{bmatrix} 0 \\ 0 \\ \mathbf{G} \end{bmatrix} \cdot w(t) \tag{7.20}$$

$$y(t) = -\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} [\mathbf{M}^{-1}\mathbf{K}(t) \quad \mathbf{M}^{-1}\mathbf{C}(t)] \begin{bmatrix} x(t) \\ \dot{x}(t) \end{bmatrix} + \mathbf{H}(\theta(t)) \cdot v(t) \tag{7.21}$$

$$F = [-m_4 \quad -m_3 \quad -m_2 \quad -m_1]^T \tag{7.22}$$

$$\mathbf{K}(t)=\begin{bmatrix} k_7(t) & -k_7(t) & 0 & 0 \\ -k_7(t) & k_7(t)+k_6(t) & \ddots & 0 \\ 0 & \ddots & \ddots & -k_2(t) \\ 0 & 0 & -k_2(t) & k_2(t)+k_1(t) \end{bmatrix} \tag{7.23}$$

$$\mathbf{C}(t)=\begin{bmatrix} c_7(t) & -c_7(t) & 0 & 0 \\ -c_7(t) & c_7(t)+c_6(t) & \ddots & 0 \\ 0 & \ddots & \ddots & -c_2(t) \\ 0 & 0 & -c_2(t) & c_2(t)+c_1(t) \end{bmatrix} \tag{7.24}$$

where

$$k_i(t)=k_i(0)\cdot e^{-\gamma\cdot\mu_i(t)^{\rho}}$$

$$c_i(t)=\delta+\lambda\cdot\mu_i(t)$$

$$\mu_1(t)=\max_{0\le k\le t}\{|x_1(k)|/h_1\}$$
$$\mu_i(t)=\max_{0\le k\le t}|x_i(k)-x_{i-1}(k)|/h_i \qquad i=2,3,...7$$

$\mathbf{H}$=diag(H_1,H_2,H_3,H_4)

$$\theta(t)=[\gamma(t)\quad \rho(t)\quad \delta(t)\quad \lambda(t)\quad \alpha(t)\quad H_1(t)\quad H_2(t)\quad H_3(t)\quad H_4(t)]^T$$

and h_i is the story height of the i-th story.

There are 23 components in the uncertain state vector; the first seven are the relative displacements of each floor; the 8th to 14th are relative velocities of each floor, the last nine are related to nonlinear stiffness and damping parameters, and the prediction error variances. For the choice of prior PDFs for the uncertain parameters, one can refer to Ching et al. (2006c).

For the purpose of illustration, consider failure F defined as the exceedance over some threshold of the interstory displacement of any one of the stories at any time within the 40s of ground shaking (time interval of 0.04s):

$$\begin{aligned} F &= \bigcup_{n=0}^{1000} \bigcup_{l=1}^{7} \{ | x_l(t_n) - x_{l-1}(t_n) | \geq b_l \cup | x_1(t_n) | \geq b_1 \} \\ &= \max_{\substack{n \in \{0,1,\ldots,1000\} \\ l \in \{1,\ldots,7\}}} \{ \frac{| x_l(t_n) - x_{l-1}(t_n) |}{b_l}, \frac{| x_1(t_n) |}{b_1} \} \geq 1 \end{aligned} \tag{7.25}$$

where the threshold b_l for all the stories is the same, i.e., $b_l=b$; and $a_l(t)$ denotes the l-th story absolute acceleration at time t.

Figure 7.3 shows for different thresholds b, the interstory exceedance probability conditional on the aforementioned data (solid curve) and that conditional on only the earthquake input record and thus the uncertainties in the states are not updated (dashed curve). These two curves give the complementary cumulative distribution function (CDF) of the peak interstory drift during the earthquake based on incomplete data, or equivalently, the solid curve is the updated robust failure probability. The solid curve drops off a lot more sharply than the dashed curve, showing that the incomplete floor acceleration data greatly reduce the uncertainty in the predicted peak interstory drift. From Figure 7.4, it can be seen that the predicted mean interstory displacement of the first story obtained using samples from $p(X_N \mid \hat{Y}_N)$ (dashed curve) captures quite well the evolution of the inferred one obtained by filtering numerically double-integrated versions of the adjacent acceleration records (solid curve). Although not shown here, all the measured responses lie within the 5 percentile and 95-percentile of the predicted response at most times. Also our results show that the predicted interstory drift corresponding to the fourth story (where no measurements were made) is the largest. This is consistent with the observation that the most severe damage occurred at the fourth story during the Northridge earthquake. All of the results are obtained using 2000 samples from $p(X_N \mid \hat{Y}_N)$.

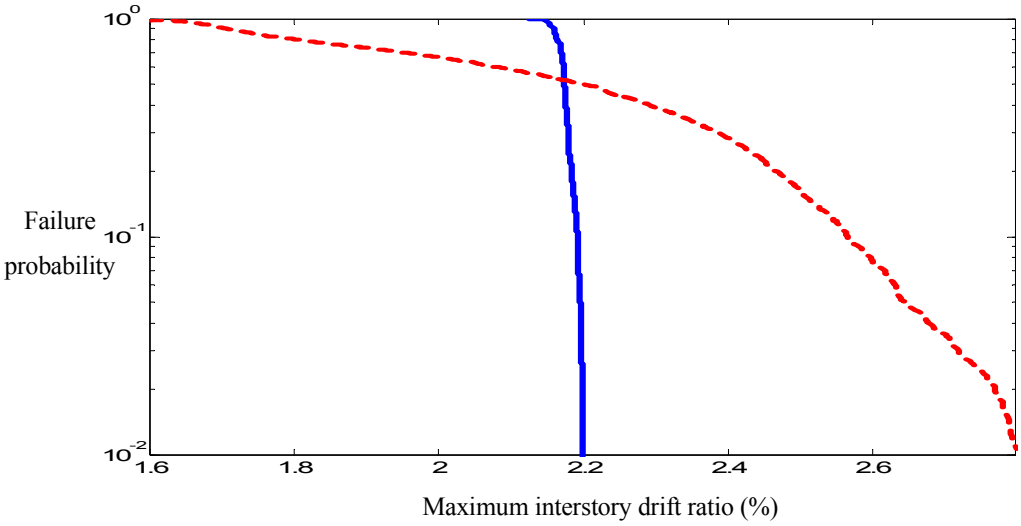

Figure 7.3: Exceedance probability for maximum interstory drift

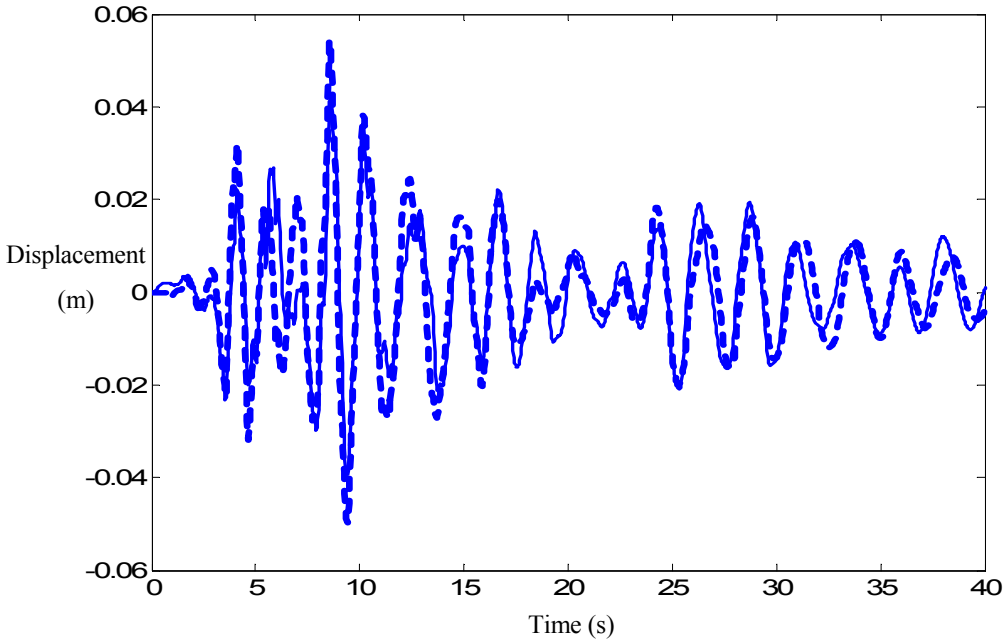

Figure 7.4: Predicted time history of interstory displacement of the first story (dashed) vs the measured interstory displacement (solid)

Appendix 7A

Assume we are interested in estimating the expectation $E_\pi[h(\boldsymbol{\theta})]$ of $h(\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ follows a certain target PDF $\pi(\boldsymbol{\theta})$, i.e., $\boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$:

$$E_\pi[h(\boldsymbol{\theta})] = \int h(\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{A7.1}$$

By MCS, $E_\pi[h(\boldsymbol{\theta})] \approx \frac{1}{K}\sum_{k=1}^{K} h(\boldsymbol{\theta}^{(k)})$ where $\boldsymbol{\theta}^{(k)}$ are samples drawn from $\pi(\boldsymbol{\theta})$.

Importance sampling is a variance reduction technique, which makes use of samples drawn from another PDF $q(\boldsymbol{\theta})$, referred to as the importance sampling density, which is often chosen to simulate more samples in the region which give significant contributions to the integral thus often leading to an estimator with a smaller variance:

$$E_\pi[h(\boldsymbol{\theta})] = \int h(\boldsymbol{\theta})\frac{\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}q(\boldsymbol{\theta})d\boldsymbol{\theta} = E_q[h(\boldsymbol{\theta})\frac{\pi(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}] \approx \frac{1}{K}\sum_{k=1}^{K}\frac{\pi(\boldsymbol{\theta}^{(k)})}{q(\boldsymbol{\theta}^{(k)})}h(\boldsymbol{\theta}^{(k)}) \tag{A7.2}$$

where $\boldsymbol{\theta}^{(k)}$ are samples drawn from $q(\boldsymbol{\theta})$. Here to ensure the above estimator has finite variance, we require supp $q \supset$ supp π. With this, by the Strong Law of Large Numbers, the estimator in (A7.2) converges to $E_\pi[h(\boldsymbol{\theta})]$ as $K \to \infty$.

If $\pi(\boldsymbol{\theta}) = cf(\boldsymbol{\theta})$ and $q(\boldsymbol{\theta}) = dg(\boldsymbol{\theta})$ where normalizing constants c and d need not be known a priori. An alternative estimator can be obtained by the following weighted average:

$$E_\pi[h(\boldsymbol{\theta})] \approx \sum_{k=1}^{K} w_k h(\boldsymbol{\theta}^{(k)}) \tag{A7.3}$$

where the normalized weight w_k corresponding to each sample is given by:

$$w_k = \frac{\pi(\boldsymbol{\theta}^{(k)})/q(\boldsymbol{\theta}^{(k)})}{\sum_{j=1}^{K}\pi(\boldsymbol{\theta}^{(j)})/q(\boldsymbol{\theta}^{(j)})} = \frac{f(\boldsymbol{\theta}^{(k)})/g(\boldsymbol{\theta}^{(k)})}{\sum_{j=1}^{K}f(\boldsymbol{\theta}^{(j)})/g(\boldsymbol{\theta}^{(j)})} \tag{A7.4}$$

where $\boldsymbol{\theta}^{(k)}$ are samples drawn from $q(\boldsymbol{\theta})$. By the Strong Law of Large Numbers, $\frac{1}{K}\sum_{j=1}^{K}\pi(\boldsymbol{\theta}^{(j)})/q(\boldsymbol{\theta}^{(j)})$ converges to 1 and $\frac{1}{K}\sum_{k=1}^{K}\frac{\pi(\boldsymbol{\theta}^{(k)})}{q(\boldsymbol{\theta}^{(k)})}h(\boldsymbol{\theta}^{(k)})$ converges to $E_\pi[h(\boldsymbol{\theta})]$ as $K\to\infty$ and thus the estimator in (A7.3) converges to $E_\pi[h(\boldsymbol{\theta})]$ as $K\to\infty$.

Sampling Importance Resampling (SIR)

Let $\pi(\boldsymbol{\theta})$ be the target PDF we want to draw samples from. SIR draws samples $\hat{\boldsymbol{\theta}}^{(k)}$, k=1, 2,…, K for $\pi(\boldsymbol{\theta})$ by first drawing samples from an importance sampling density $q(\boldsymbol{\theta})$ using the following procedure (Assume supp $q \supset$ supp π):

1. Draw K samples $\boldsymbol{\theta}^{(k)}$, k=1, 2,…, K, from $q(\boldsymbol{\theta})$.
2. Evaluate the weight w_k corresponding to each $\boldsymbol{\theta}^{(k)}$ using (A7.4).
3. For j=1,2,…,K, $\hat{\boldsymbol{\theta}}^{(j)} = \boldsymbol{\theta}^{(k)}$ with probability w_k (i.e., the index k is randomly drawn from the set {1,2,3,…,K} with $P(k=m) = w_m$)

One simple way to do step 3 is as follows:

1. Calculate the cumulative distribution function $F(k)$ (CDF) for the discrete distribution for the index k as $F(k) = \sum_{j=1}^{k} w_j$ for k=1,2,…,K
2. Draw a number u from Uniform(0,1). $k = m$ such that m satisfies $\sum_{j=1}^{m-1} w_j < u \le \sum_{j=1}^{m} w_j$ ($\sum_{j=1}^{m-1} w_j$ =0 for m=1).

The expectation $E_\pi[h(\boldsymbol{\theta})]$ can be estimated using the samples from the resampling step as follows:

$$E_\pi[h(\boldsymbol{\theta})] \approx \frac{1}{K}\sum_{k=1}^{K} h(\hat{\boldsymbol{\theta}}^{(k)}) \tag{A7.5}$$

The following proves that this estimator converges to $E_\pi[h(\boldsymbol{\theta})]$ as $K \to \infty$.

$$\begin{aligned}
E[h(\hat{\boldsymbol{\theta}}^{(k)})] &= \int h(\hat{\boldsymbol{\theta}}^{(k)})\, p(\hat{\boldsymbol{\theta}}^{(k)})\, d\hat{\boldsymbol{\theta}}^{(k)} \\
&= \int h(\hat{\boldsymbol{\theta}}^{(k)})\, p(\hat{\boldsymbol{\theta}}^{(k)} \mid \boldsymbol{\theta}^{(1)},...,\boldsymbol{\theta}^{(K)})\, p(\boldsymbol{\theta}^{(1)},...,\boldsymbol{\theta}^{(K)})\, d\boldsymbol{\theta}^{(1)}...d\boldsymbol{\theta}^{(K)} d\hat{\boldsymbol{\theta}}^{(k)} \\
&= \int h(\hat{\boldsymbol{\theta}}^{(k)})[\sum_{j=1}^{K} w_j \delta(\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}^{(j)})] \prod_{l=1}^{K} q(\boldsymbol{\theta}^{(l)})\, d\boldsymbol{\theta}^{(1)}...d\boldsymbol{\theta}^{(K)} d\hat{\boldsymbol{\theta}}^{(k)} \\
&= \sum_{j=1}^{K} \int\int h(\hat{\boldsymbol{\theta}}^{(k)})\, \delta(\hat{\boldsymbol{\theta}}^{(k)} - \boldsymbol{\theta}^{(j)})\, d\hat{\boldsymbol{\theta}}^{(k)}\, w_j(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)},..., \boldsymbol{\theta}^{(K)}) \prod_{l=1}^{K} q(\boldsymbol{\theta}^{(l)})\, d\boldsymbol{\theta}^{(1)}...d\boldsymbol{\theta}^{(K)} \\
&= \sum_{j=1}^{K} \int h(\boldsymbol{\theta}^{(j)})\, w_j(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)},..., \boldsymbol{\theta}^{(K)}) \prod_{l=1}^{K} q(\boldsymbol{\theta}^{(l)})\, d\boldsymbol{\theta}^{(1)}...d\boldsymbol{\theta}^{(K)}, \text{ for all } k
\end{aligned}$$

$$\begin{aligned}
E[\sum_{k=1}^{K} w_k h(\boldsymbol{\theta}^{(k)})] &= \sum_{k=1}^{K} E[w_k h(\boldsymbol{\theta}^{(k)})] \\
&= \sum_{k=1}^{K} \int w_k(\boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(2)},..., \boldsymbol{\theta}^{(K)}) h(\boldsymbol{\theta}^{(k)}) \prod_{l=1}^{K} q(\boldsymbol{\theta}^{(l)})\, d\boldsymbol{\theta}^{(1)}...d\boldsymbol{\theta}^{(K)}
\end{aligned}$$

$$\Rightarrow E[h(\hat{\boldsymbol{\theta}}^{(k)})] = E[\sum_{k=1}^{K} w_k h(\boldsymbol{\theta}^{(k)})]. \text{ (Thus } E[\frac{1}{K}\sum_{k=}^{K} h(\hat{\boldsymbol{\theta}}^{(k)})] = E[h(\hat{\boldsymbol{\theta}}^{(k)})] = E[\sum_{k=1}^{K} w_k h(\boldsymbol{\theta}^{(k)})])$$

Thus $E[\frac{1}{K}\sum_{k=1}^{K} h(\hat{\boldsymbol{\theta}}^{(k)})] = E[h(\hat{\boldsymbol{\theta}}^{(k)})]$ converges to $E_\pi[h(\boldsymbol{\theta})]$ as $K \to \infty$ since $\sum_{k=1}^{K} w_k h(\boldsymbol{\theta}^{(k)})$ converges to $E_\pi[h(\boldsymbol{\theta})]$ as $K \to \infty$ according to p.1. With this and the Strong Law of Large Numbers, we can conclude that $\frac{1}{K}\sum_{k=1}^{K} h(\hat{\boldsymbol{\theta}}^{(k)})$ converges to $E_\pi[h(\boldsymbol{\theta})]$ as $K \to \infty$.

Note: $\boldsymbol{\theta}^{(k)}$'s with larger weights are duplicated many times while the one with very small weight are eliminated. After doing the resampling, the weight corresponding to each $\hat{\boldsymbol{\theta}}^{(j)}$ becomes uniform: 1/*K*. We will discuss this in more detail later. Asymptotically, these samples from resampling are distributed according to $\pi(\boldsymbol{\theta})$.

Appendix 7B: Particle Filter (PF)

Consider the stochastic discrete-time state-space model $\mathcal{M}$ of a dynamical system as in (7.2).

Denote $X_n=[x_0\ \ x_1 \ldots.\ x_n]$, $Y_n=[y_1\ \ y_1 \ldots.\ y_n]$ and $U_n=[u_1\ \ u_2 \ldots.\ u_n]$. Our objective here is to evaluate sequentially the PDF $p(x_n \mid \hat{Y}_n)$ for the state at every time n as the measured system input $\hat{U}_n$ and output $\hat{Y}_n$ are collected, i.e., to perform a sequential update of the conditional PDF using the new measured system input $\hat{u}_n$ and output $\hat{y}_n$ to update $p(x_{n-1} \mid \hat{Y}_{n-1})$. For convenience, the conditioning of the PDF on $\hat{U}_n$ and the model class $\mathcal{M}$ is left out.

From the Theorem of Total Probability, we can get a predictor equation:

$$p(x_n \mid \hat{Y}_{n-1}) = \int p(x_n \mid x_{n-1}, \hat{Y}_{n-1}) p(x_{n-1} \mid \hat{Y}_{n-1}) dx_{n-1} = \int p(x_n \mid x_{n-1}) p(x_{n-1} \mid \hat{Y}_{n-1}) dx_{n-1} \qquad \text{(B7.1)}$$

By Bayes' Theorem, we get the updater equation:

$$p(x_n \mid \hat{Y}_n) = \frac{p(\hat{y}_n \mid x_n, \hat{Y}_{n-1}) p(\hat{Y}_{n-1} \mid x_n) p(x_n)}{p(\hat{Y}_n)} = \frac{p(\hat{y}_n \mid x_n) p(x_n \mid \hat{Y}_{n-1}) p(\hat{Y}_{n-1})}{p(\hat{Y}_n)}$$

$$= \frac{p(\hat{y}_n \mid x_n) p(x_n \mid \hat{Y}_{n-1})}{p(\hat{y}_n \mid \hat{Y}_{n-1})} = \frac{p(\hat{y}_n \mid x_n) p(x_n \mid \hat{Y}_{n-1})}{\int p(\hat{y}_n \mid x_n) p(x_n \mid \hat{Y}_{n-1}) dx_n} \qquad \text{(B7.2)}$$

When f_n and h_n are linear in x_n, u_n, γ_{n-1} and x_n, u_n, v_n respectively and x_0, γ_n's and v_n's are Gaussian, $p(x_n \mid \hat{Y}_n)$ is Gaussian with a certain mean and covariance matrix that can be found analytically from the above equations. This leads to the mean and covariance matrix being updated sequentially using the Kalman Filter (KF), i.e. KF is Bayesian sequential updating of the state. In the case of a nonlinear model, the Extended Kalman Filter (EKF)

provides an approximate filter by linearizing the state space model. It can be applied to slightly nonlinear systems but its performance is very poor for highly nonlinear systems. The Particle Filter is a sequential stochastic simulation method that can deal with any nonlinear model, even if the uncertainties are not modeled as Gaussian.

Notice that $p(x_n \mid \hat{Y}_n)$ is just the marginal PDF of $p(X_n \mid \hat{Y}_n)$, the joint PDF of the state history up to time *n*. It is useful to consider $p(X_n \mid \hat{Y}_n)$; it will become clear why in the coming section. By Bayes' Theorem,

$$
\begin{aligned}
p(X_n \mid \hat{Y}_n) &= \frac{p(X_n \mid \hat{Y}_{n-1}) p(\hat{y}_n \mid X_n, \hat{Y}_{n-1})}{p(\hat{y}_n \mid \hat{Y}_{n-1})} \\
&= \frac{p(X_{n-1} \mid \hat{Y}_{n-1}) p(x_n \mid X_{n-1}, \hat{Y}_{n-1}) p(\hat{y}_n \mid X_n, \hat{Y}_{n-1})}{p(\hat{y}_n \mid \hat{Y}_{n-1})} \\
&= p(X_{n-1} \mid \hat{Y}_{n-1}) \frac{p(x_n \mid x_{n-1}) p(\hat{y}_n \mid x_n)}{p(\hat{y}_n \mid \hat{Y}_{n-1})} \\
&= p(x_0) \prod_{m=1}^{n} \frac{p(x_m \mid x_{m-1}) p(\hat{y}_m \mid x_m)}{p(\hat{y}_m \mid \hat{Y}_{m-1})} \\
&= \frac{p(x_0)}{p(\hat{Y}_n)} \prod_{m=1}^{n} p(x_m \mid x_{m-1}) p(\hat{y}_m \mid x_m)
\end{aligned}
$$

One can estimate the expectation of any function $\tilde{h}(X_n)$ of X_n given the data $\hat{Y}_n$, using importance sampling:

$$
E[\tilde{h}(X_n) \mid \hat{Y}_n] \approx \sum_{k=1}^{K} w_{n,k} \tilde{h}(X_n^{(k)}) \tag{B7.3}
$$

where $w_{n,k} = \dfrac{\beta_{n,k}}{\sum_{j=1}^{K} \beta_{n,j}}$ and $\beta_{n,k}$ is given by:

$$\beta_{n,k} = \frac{p(x_0^{(k)}) \prod_{m=1}^{n} p(x_m^{(k)} \mid x_{m-1}^{(k)}) p(\hat{y}_m \mid x_m^{(k)})}{q(X_n^{(k)} \mid \hat{Y}_n)} \tag{B7.4}$$

where $X_n^{(k)}$ are samples (particle trajectories) drawn from an importance sampling density $q(X_n \mid \hat{Y}_n)$ which can readily be sampled. The essence of PF is the smart choice of this q. The expectation of any function $h(x_n)$ of the state x_n given the data $\hat{Y}_n$ can be estimated readily using importance sampling:

$$E[h(x_n) \mid \hat{Y}_n] \approx \sum_{k=1}^{K} w_{n,k} h(x_n^{(k)}) \tag{B7.5}$$

To allow the sequential update in time, the following form of $q(X_n \mid \hat{Y}_n)$ is adopted:

$$q(X_n \mid \hat{Y}_n) = q_0(x_0) \prod_{m=1}^{n} q_m(x_m \mid X_{m-1}, \hat{Y}_m) \tag{B7.6}$$

With this choice, the weight $\beta_{n,k}$ can be evaluated sequentially:

$$\beta_{n,k} = \beta_{n-1,k} \frac{p(x_n^{(k)} \mid x_{n-1}^{(k)}) p(\hat{y}_n \mid x_n^{(k)})}{q(x_n^{(k)} \mid X_{n-1}^{(k)}, \hat{Y}_n)} \tag{B7.7}$$

PF algorithm 1

1. Draw K samples $x_0^{(k)}$ from $q(x_0) = p(x_0)$ where $\beta_{0,k} = 1/K$ for k=1,2,...,K (gives initial position of K particles).
2. Repeat the following for time n = 1,2,…, N (generates the K particle trajectories):
   2.1. Draw K samples $x_n^{(k)}$ from $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n)$ and update the importance weight as follows for k=1,2,...,K:

$$\beta_{n,k} = \beta_{n-1,k} \frac{p(x_n^{(k)} \mid x_{n-1}^{(k)}) p(\hat{y}_n \mid x_n^{(k)})}{q(x_n^{(k)} \mid X_{n-1}^{(k)}, \hat{Y}_n)} \tag{B7.8}$$

2.2. $w_{n,k} = \dfrac{\beta_{n,k}}{\sum_{j=1}^{K} \beta_{n,j}}$ and $E[h(x_n) \mid \hat{Y}_n]$ can be estimated by:

$$E[h(x_n) \mid \hat{Y}_n] \approx \sum_{k=1}^{K} w_{n,k} h(x_n^{(k)}) \tag{B7.9}$$

Note: All the K particles evolve through time independently. As n increases, the importance weights become far from uniform. This leads to downgrading of some particles because eventually only a few particles will have weights much larger than the others and only these few particles will contribute to (B7.9). It was shown in Kong et al. (1994) that the variance of the importance weights conditioned on $\hat{Y}_n$ increases with time.

PF algorithm 2 (with resampling)

1. Draw K samples $x_0^{(k)}$ from $q(x_0) = p(x_0)$ where $\beta_{0,k} = 1/K$ for k=1,2,...,K.
2. Repeat the following for time n = 1,2,…, N:
   2.1. Draw K candidate samples $\tilde{x}_n^{(k)}$ from $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n)$ and update the importance weight as follows for k=1,2,...,K:

$$\beta_{n,k} = \beta_{n-1,k} \frac{p(\tilde{x}_n^{(k)} \mid x_{n-1}^{(k)}) p(\hat{y}_n \mid x_n^{(k)})}{q(\tilde{x}_n^{(k)} \mid X_{n-1}^{(k)}, \hat{Y}_n)} \tag{B7.10}$$

a) Calculate the coefficient of variation δ_n (c.o.v.) of $\{\beta_{n,1}, \beta_{n,2}, ..., \beta_{n,K}\}$:

$$\overline{\beta}_n = \frac{1}{K}\sum_{k=1}^{K}\beta_{n,k},\ \delta_n = \frac{\sqrt{\frac{1}{K-1}\sum_{k=1}^{K}(\beta_{n,k}-\overline{\beta}_n)^2}}{\overline{\beta}_n} \tag{B7.11}$$

b) Compute the normalized weight $w_{n,k} = \dfrac{\beta_{n,k}}{\sum_{j=1}^{K}\beta_{n,j}}$.

c) If $\delta_n \le \delta_{th}$, set $x_n^{(k)} = \tilde{x}_n^{(k)}$, $\beta_{n,k} = w_{n,k}$ for k=1,2,...,K. Otherwise, do the resampling (SIR) as follows for j=1,2,…,K:

$$x_n^{(j)} = \tilde{x}_n^{(k)} \quad \text{with probability} \quad w_{n,k} \tag{B7.12}$$

and then set $w_{n,k} = \beta_{n,k} = 1/K$ for k=1,2,...,K. (particle cloning and elimination).

d) $E[h(x_n)\,|\,\hat{Y}_n]$ can be estimated by:

$$E[h(x_n)\,|\,\hat{Y}_n] \approx \sum_{k=1}^{K} w_{n,k}\, h(x_n^{(k)}) \tag{B7.13}$$

Note:

1. It is desirable to have the importance weights as uniform as possible so that all samples contribute to the estimation. After the resampling, the importance weights become uniform. As mentioned before, resampling duplicates particles (cloning) with larger weights and eliminates particles with smaller weights. This puts the computational effort into particles that will explore the high probability content region of $p(x_n\,|\,\hat{Y}_n)$. However, the samples become increasingly dependent and so the effective number of distinct particles to explore the state space decreases. Therefore, the resampling step should only be carried out if the importance weights are highly non-uniform (one way is to do the resampling step only when the c.o.v. exceeds some threshold as in step 2.4).

2. To reduce the dependency introduced by the resampling step, one way is to perform independent PF algorithms in parallel. Another way is to use instead the following algorithm.

PF algorithm 3 (with resampling and MCMC)

This is the same as PF algorithm 2 with an additional MCMC step(s) whenever the resampling is performed. Here I illustrate the idea with MH algorithm as an example:

For each k=1,2,…,K, repeat the following for M times:
After the resampling step in 2.4, a candidate sample x_{cand} is drawn from a proposal PDF $q_{MH}(x_{\text{cand}} \mid x_n^{(k)})$. Compute the acceptance probability r:

$$r = \frac{p(x_n = x_{\text{cand}} \mid x_{n-1}^{(k)}) p(\hat{y}_n \mid x_n = x_{cand}) q_{MH}(x_n^{(k)} \mid x_{\text{cand}})}{p(x_n = x_n^{(k)} \mid x_{n-1}^{(k)}) p(\hat{y}_n \mid x_n = x_n^{(k)}) q_{MH}(x_{\text{cand}} \mid x_n^{(k)})} \tag{B7.14}$$

If $r>u$ where u~Uniform(0,1), $x_n^{(j)} = x_{\text{cand}}$. Otherwise, $x_n^{(j)}$ remains unchanged.

Note:

This procedure allows the duplicate particles to move to new positions, thus improving the exploration of the state space at the expense of additional computational effort.

Appendix 7C: Choice of $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n)$:

1. $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n) = p(x_n \mid x_{n-1}^{(k)})$: For this case, the importance weight is updated as follows:

$$\beta_{n,k} = \beta_{n-1,k}\, p(\hat{y}_n \mid x_n^{(k)}) \tag{C7.1}$$

where $p(\hat{y}_n \mid x_n^{(k)})$ is one of the fundamental PDFs of $\mathcal{M}$. The advantage of this choice is that sampling from $p(x_n \mid x_{n-1}^{(k)})$ can be done readily since the PDF of γ_n is prescribed in such a way as to be readily sampled (eg. Gaussian). However, one drawback is that the exploration of the state space can be very ineffective since the new measured data $\hat{y}_n$ is not used.

2. $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n) = p(x_n \mid x_{n-1}^{(k)}, \hat{y}_n)$**:** For this case, the importance weight is updated as follows:

$$\beta_{n,k} = \beta_{n-1,k}\, \frac{p(x_n^{(k)} \mid x_{n-1}^{(k)})\, p(\hat{y}_n \mid x_n^{(k)})}{p(x_n^{(k)} \mid x_{n-1}^{(k)}, \hat{y}_n)} = \beta_{n-1,k}\, p(\hat{y}_n \mid x_{n-1}^{(k)}) \tag{C7.2}$$

Doucet et al. (2000) shows that this choice of $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n)$ is optimal in the sense that it minimizes the variance of the importance weights $\beta_{n,k}$ conditioned on $X_{n-1}^{(k)}$ and $\hat{Y}_n$. This choice of $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n)$ has two drawbacks: 1) it requires the ability to draw samples from $p(x_n \mid x_{n-1}^{(k)}, \hat{y}_n)$**,** which is generally non-Gaussian, and 2) $p(\hat{y}_n \mid x_{n-1}^{(k)})$ in general is not known analytically because:

$$p(\hat{y}_n \mid x_{n-1}^{(k)}) = \int p(\hat{y}_n \mid x_n)\, p(x_n \mid x_{n-1}^{(k)})\, dx_n \tag{C7.3}$$

One way to get around this is to use a Gaussian PDF obtained by local linearization of the state space model (as is done in the EKF algorithm). As an alternative, we can impose some special structure on $f_{n-1}(x_{n-1}, u_{n-1}, \gamma_n)$ and $h_n(x_n, u_n, v_n)$ in (7.2) and prescribe probability models for the γ_n's and v_n's ; for example:

$$
\begin{aligned}
x_n &= f_{n-1}(x_{n-1}, u_{n-1}) + B_n \gamma_n \\
y_n &= C_n x_n + D_n u_n + E_n v_n
\end{aligned}
\tag{C7.4}
$$

where $\gamma_n \sim N(0, \sum_\gamma)$ and $v_n \sim N(0, \sum_\nu)$ are independent and $f_{n-1}(x_{n-1}, u_{n-1})$ can be nonlinear. Many common models in use belong to this class. For this class, it can be shown (shown in Appendix 7D) that to construct the optimal $q(x_n \mid X_{n-1}^{(k)}, \hat{Y}_n)$, 1) there is no need to linearize $f_{n-1}(x_{n-1}, u_{n-1})$ even if the state space model is nonlinear; 2) $p(x_n \mid x_{n-1}^{(k)}, \hat{y}_n)$ is a multivariate Gaussian which allows direct simulation; and 3) $p(\hat{y}_n \mid x_{n-1}^{(k)})$ is known analytically:

$$
\begin{aligned}
& y_n \mid x_{n-1} \sim N(\mu_{y,n} = C_n f_{n-1}(x_{n-1}, u_{n-1}) + D_n u_n, \sum_{y,n} = C_n B_n \sum_\gamma B_n^T C_n^T + E_n \sum_\nu E_n^T) \\
& \Rightarrow p(\hat{y}_n \mid x_{n-1}) \propto \exp(-\frac{1}{2}(\hat{y}_n - \mu_{y,n})^T \sum_{y,n}^{-1} (\hat{y}_n - \mu_{y,n}))
\end{aligned}
\tag{C7.5}
$$

$p(x_n \mid x_{n-1}^{(k)}, \hat{y}_n) \sim N(\mu_{x,n}, \sum_{x,n})$, that is a multivariate Gaussian with mean $\mu_{x,n}$ and covariance matrix $\sum_{x,n}$ where

$$
\begin{aligned}
\mu_{x,n} &= \sum_{x,n}^{-1} [(B_n \sum_\gamma B_n^T)^{-1} f_{n-1}(x_{n-1}^{(k)}, u_{n-1}) + C_n^T (E_n \sum_\nu E_n^T)^{-1} (\hat{y}_n - D_n u_n)] \\
\sum_{x,n}^{-1} &= C_n^T (E_n \sum_\nu E_n^T)^{-1} C_n + (B_n \sum_\gamma B_n^T)^{-1}
\end{aligned}
\tag{C7.6}
$$

Appendix 7D

By substituting the state equation of (C7.4) into the observation equation, we obtain:

$$
y_n = C_n f_{n-1}(x_{n-1}, u_{n-1}) + C_n B_n \gamma_n + D_n u_n + E_n v_n
$$

$$
\begin{aligned}
& y_n \mid x_{n-1} \sim N(\mu_{y,n} = C_n f_{n-1}(x_{n-1}, u_{n-1}) + D_n u_n, \sum_{y,n} = C_n B_n \sum_\gamma B_n^T C_n^T + E_n \sum_\nu E_n^T) \\
& \Rightarrow p(\hat{y}_n \mid x_{n-1}) \propto \exp(-\frac{1}{2}(\hat{y}_n - \mu_{y,n})^T \sum_{y,n}^{-1} (\hat{y}_n - \mu_{y,n}))
\end{aligned}
$$

$$p(x_n \mid x_{n-1}^{(k)}, \hat{y}_n) = \frac{p(\hat{y}_n \mid x_n, x_{n-1}^{(k)}) p(x_n \mid x_{n-1}^{(k)})}{p(\hat{y}_n \mid x_{n-1}^{(k)})} = \frac{p(\hat{y}_n \mid x_n) p(x_n \mid x_{n-1}^{(k)})}{p(\hat{y}_n \mid x_{n-1}^{(k)})}$$

$$\propto p(\hat{y}_n \mid x_n) p(x_n \mid x_{n-1}^{(k)})$$

$$\propto \exp(-\frac{1}{2}(\hat{y}_n - C_n x_n - D_n u_n)^T (E_n \sum_v E_n^T)^{-1} (\hat{y}_n - C_n x_n - D_n u_n)) \times$$

$$\exp(-\frac{1}{2}(x_n - f_{n-1}(x_{n-1}^{(k)}, u_{n-1}))^T (B_n \sum_\gamma B_n^T)^{-1} (x_n - f_{n-1}(x_{n-1}^{(k)}, u_{n-1})))$$

$$\propto \exp(-\frac{1}{2}[x_n^T (C_n^T (E_n \sum_v E_n^T)^{-1} C_n + (B_n \sum_\gamma B_n^T)^{-1}) x_n$$
$$- x_n^T [(B_n \sum_\gamma B_n^T)^{-1} f_{n-1}(x_{n-1}^{(k)}, u_{n-1}) + C_n^T (E_n \sum_v E_n^T)^{-1} (\hat{y}_n - D_n u_n)]$$
$$- [x_n^T [(B_n \sum_\gamma B_n^T)^{-1} f_{n-1}(x_{n-1}^{(k)}, u_{n-1}) + C_n^T (E_n \sum_v E_n^T)^{-1} (\hat{y}_n - D_n u_n)]]^T)$$

$$\propto \exp(-\frac{1}{2}(x_n - \mu_{x,n})^T \sum_{x,n} (x_n - \mu_{x,n}))$$

where

$$\sum_{x,n} \mu_{x,n} = (B_n \sum_\gamma B_n^T)^{-1} f_{n-1}(x_{n-1}^{(k)}, u_{n-1}) + C_n^T (E_n \sum_v E_n^T)^{-1} (\hat{y}_n - D_n u_n)$$

$$\Rightarrow \mu_{x,n} = \sum_{x,n}^{-1} [(B_n \sum_\gamma B_n^T)^{-1} f_{n-1}(x_{n-1}^{(k)}, u_{n-1}) + C_n^T (E_n \sum_v E_n^T)^{-1} (\hat{y}_n - D_n u_n)]$$

$$\sum_{x,n}^{-1} = C_n^T (E_n \sum_v E_n^T)^{-1} C_n + (B_n \sum_\gamma B_n^T)^{-1}$$

Thus $p(x_n \mid x_{n-1}^{(k)}, \hat{y}_n) \sim N(\mu_{x,n}, \sum_{x,n})$, that is a multivariate Gaussian with mean $\mu_{x,n}$ and covariance matrix $\sum_{x,n}$.

CHAPTER 8

Conclusions

This thesis addresses the problem of stochastic system analysis, model and reliability updating of complex systems with special attention to complex dynamic systems and high-dimensional uncertainties. For stochastic system analysis, special attention is paid to evaluating robust failure probability. Full Bayesian model updating approach is adopted to provide a robust and rigorous framework to characterize modeling uncertainties associated with the underlying system and its environment. The following summarizes the conclusions for all the chapters in this thesis.

8.1.1 Conclusions to Chapter 2

The proposed algorithms presented in Chapter 2 provide powerful and effective computational tools for solving model updating problems in higher-dimensional parameter spaces, even unidentifiable ones, which are well known to present a challenging computational problem. Any type of model can be used: physics-based or blackbox, linear or nonlinear, without restriction on the type of data. Although the focus of application is on system identification and model updating of dynamic systems, there are other possible areas of potential application such as Bayesian regression and classification problems (e.g. Oh et al., 2008).

Advanced Monte Carlo algorithms are presented and their features are discussed and reviewed in detail. Improvements are proposed to make the algorithms more effective and efficient for solving higher-dimensional model updating problems for dynamic systems. New formulae for Markov Chain convergence assessment are also derived. The illustrative numerical example shows that based on acceleration data from the structure, the proposed fully probabilistic Bayesian model updating approach is able to characterize modeling uncertainties associated with the underlying structural system and can provide robust estimation even when the model class is unidentifiable based on the recorded response.

8.1.2 Conclusions to Chapter 3

Bayesian model class comparison based on the evidence for each model class provided by the data is very general and can deal with any type of model: physically-based or blackbox, parametric or nonparametric, linear or nonlinear, deterministic or probabilistic, without restriction on the type of data. A computational method is proposed for calculating the evidence for each candidate model class provided by the data, and so for calculating the posterior probability of each model class, by using its posterior samples generated using a Markov Chain Monte Carlo algorithm. In addition, this method allows for an efficient calculation of the information entropy and information (entropy) gain about each model classes given the data. This method can be applied in general to efficiently solve problems involving many uncertain parameters, especially where the previously-published Laplace asymptotic approximation (Beck and Yuen 2004) for the evidence does not perform well (e.g. unidentifiable model classes) or is computationally prohibitive because of the inherent optimization problem in high-dimensional spaces. Besides calculating the evidence required in Bayesian model class comparison, the proposed method can be used to calculate integrals with non-negative integrands in higher-dimensions by simulating samples from the PDF proportional to the integrand. Examples of potential application include system identification, regression, classification, and calculating reliability and expectations of functions of uncertain parameters. The presented examples show that among a set of candidate model classes, the most plausible model class based on the data is

identified and the plausibility of each model class is quantified based on its posterior probability.

8.1.3 Conclusions to Chapter 4

Past applications of the framework for model updating of dynamic systems focus on model classes which consider an uncertain prediction error as the difference between the real system output and the model output and model it probabilistically using Jaynes' Principle of Maximum Information Entropy. In this paper, an extension of such model classes is considered to allow more flexibility in modeling uncertainties for updating of state space models and for making robust predictions by introducing prediction errors in the state vector equation in addition to those in system output vector equation. State-of-the-art algorithms are used to solve the computational problems resulting from these extended model classes. For the illustrative example which involves a benchmark structure from the IASC-ASCE Structural Health Monitoring Task Group, it is shown by Bayesian model class selection that the posterior probability of the extended model class is significantly larger than the original model class. The posterior robust failure probability of the benchmark structure subjected to a future earthquake for these model classes are calculated for different threshold levels. The results show that the posterior failure probability for these model classes can be quite different from each other even though they have the same type of underlying deterministic state-space model. Thus, the posterior robust failure probability is sensitive to the choice of model classes and hence to the way that model uncertainties are treated. This confirms the importance of implementation of model class comparison and averaging when predicting the system response, especially when calculating the robust failure probability.

8.1.4 Conclusions to Chapter 5

A novel methodology based on Bayesian updating of hierarchical stochastic system model classes is proposed for uncertainty quantification, model updating, model selection, model

validation and robust prediction of the response of a system for which some subsystems have been separately tested. It uses full Bayesian updating of the model classes, along with model class comparison and prediction consistency and accuracy assessment. In the proposed methodology, all the results are rigorously derived from the probability axioms and all the information in the available data are considered to make predictions. The concepts and computational tools of the proposed methodology are illustrated with a previously-studied validation challenge problem, although the methodology can handle a more general process of hierarchical subsystem testing.

As shown by the illustrative example, within a model class, there are many plausible models and the predictions of response and failure probability of the final system can often vary greatly from one model to another, showing that the consequences of the uncertainties in the parameters are significant. Ignoring the uncertainty in the modeling parameters and solely relying on the MAP model (corresponding to the maximum of the posterior PDF) or the MLE model (corresponding to the maximum likelihood parameter value) for predictions can be dangerous and misleading since such predictions can greatly underestimate the failure probability and the uncertainty in the response. It is shown how more robust predictions by a model class can be obtained by taking into account the predictions from all the plausible models in the model class where the plausibilities are quantified by their respective posterior PDF values.

Multiple model classes are investigated for the illustrative example. The response and failure probability prediction vary greatly from one model class to another. Hyper-robust predictions of response and failure probability are also obtained by a weighted average of the robust predictions given by each model class where the weight is given by the posterior probability of the model class. The posterior probability of one of the candidate model classes is so small based on the calibration data that its contribution to the prediction is negligible, so it is discarded from further predictive analysis after the calibration tests.

The computational problems resulting from full Bayesian updating of hierarchical model classes, as well as model class comparison, can be challenging, especially for problems with many uncertain parameters. A number of powerful computational tools based on stochastic simulation are used to solve efficiently the computational problems involved; in particular, for the illustrative example studied, the Hybrid Gibbs TMCMC algorithm worked well.

If a model class performs well in predicting the response for the subsystems involved in all of the experiments, one can gain more confidence in its predictive performance for the final constructed system. However, it should be stressed that 1) whether the predictive performance of the model classes is acceptable or not depends on which criteria the decision maker thinks are critical, and 2) there is no guarantee that a model class which performs well enough to satisfy the selected criteria in predicting the response of the subsystems in these experiments will always predict the response of the final system well, especially in the case where some of the uncertainties in the final system which are critical to the prediction are not present in the subsystem tests (for example, there can be uncertainties in support or joint conditions in the final system, and uncertainties in input loadings, such as stronger amplitude inputs which may be experienced by the final system that cause it to behave very differently than the subsystems during their tests).

Although it did not occur in the illustrative example, in the case where all candidate model classes give poor performance in predicting the response for subsystems involved in an experiment, one should check whether some of the uncertainties have not been adequately modeled in the failing subsystem tests and, if so, modify the candidate model classes to properly take into account these uncertainties.

8.1.5 Conclusions to Chapter 6

All types of uncertainties, including those from dynamic system modeling and/or the modeling of the uncertain excitation, are considered during the computation of the robust

reliability of a dynamic system subjected to future uncertain excitation. The prior robust reliability can be updated by using system data. This updating problem has been rarely tackled in the past due to the fact that it involves high-dimensional integrations of complicated integrands with respect to the uncertain parameters, leading to a computationally very challenging problem. A new approach is presented that is based on a stochastic simulation method and the availaibility of partial output data from the dynamic system. The proposed method is illustrated by a numerical example involving an inelastic hysteretic four-story building.

8.1.6 Conclusions to Chapter 7

A novel stochastic simulation method is proposed for updating in near real time the robust reliability of a dynamic system. The performance of the method is illustrated by an example which updates the failure probability using a nonlinear dynamic model of a seven-story reinforced-concrete hotel based on incomplete floor acceleration data obtained during the 1994 Northridge earthquake. Using the observed response greatly reduces the uncertainty in the predicted peak interstory drift which characterizes the reliability of the system. In addition, the proposed method gives an updated probabilistic description of the entire time history of the complete response, conditional on the observed response.

8.1.7 Conclusions for the whole thesis

This thesis addresses the problem of stochastic system analysis, model and reliability updating of complex systems (with special attention to complex dynamic systems and high-dimensional uncertainties) and applications to structural dynamics problems. For stochastic system analysis, special attention is paid to evaluating robust failure probability. This thesis contributes to both methodological and algorithmic developments for the problem. It is shown that the proposed methods which are based on probability logic and full Bayesian model updating provide a robust, rigorous and powerful framework to tackle the problem of stochastic system analysis, model and reliability updating of complex systems. The

proposed computational tools in this thesis is efficient, effective and very general (regardless of whether the system behaves linearly or nonlinearly, whether the system has a large number of uncertain modeling parameters or not). As confirmed by many illustrative examples, the proposed methods can tackle the case involving complex systems very well (e.g., dynamic systems with a large number of uncertain parameters).

8.1.8 Future Works

One of the plans is to investigate the application of the methodologies and computational methods presented in this thesis for more model and reliability updating problems with real data cases. Data were collected from real systems including 1) four-story ASCE-IASC benchmark structure; 2) six-story full-scale steel-frame structure tested pseudo-dynamically at BRI, Tsukuba, Japan in mid 1980s and 3) Milikan library during 1994 Northridge Earthquake. Applications to systems involving multi-physics interaction, for example, systems involving fluid-structure interaction such as offshore platforms will also be considered.

Hamiltonian Markov Chain Method and the Multi-level multiple-group MCMC algorithm presented in Chapter 2 will be further improved so that they can be more efficient for the case with a huge number of uncertain parameters.

Algorithms for calculating complicated likelihood functions resulting from different complexity in the stochastic model classes and for generating posterior samples from such model classes will be further improved. More studies will be carried out to study the effects and contributions of different stochastic model classes with embedded stochastic nonlinear dynamic models on the prediction of failure probability updated by the data. Similar to what was done in Chapter 4, multiple stochastic model class comparison and robust system reliability predictions will be implemented using modal data collected from four-story ASCE-IASC benchmark structure and the same set of seismic data (as in Chapter 7) collected from the seven-story hotel located in Van Nuys.

The new Bayesian model validation methodology presented in Chapter 5 will be applied to a problem involving a real, complicated system where data from the corresponding hierarchical subsystem tests can be obtained. More theorems and results will be presented in future publications about the new algorithms presented in Chapters 6 and 7 for updating robust future reliability and updating robust near real-time reliability of dynamic systems. Data from real structures will be used in these studies.

References

Au, S.K. and Beck, J.L. 1999. A new adaptive importance sampling scheme for reliability calculations. *Structural Safety*. 21(2): 135-158.

Au, S.K. and Beck, J.L. 2001a. First excursion probabilities for linear systems by very efficient importance sampling. *Probabilistic engineering mechanics*. 16(3): 193-207.

Au, S.K. and Beck, J.L. 2001b. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic engineering mechanics*. 16(4): 263-277.

Au, S.K. and Beck, J.L. 2003. Importance sampling in high dimension. *Structural Safety*. 25(2): 139-163.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*. 19 (6): 716-723.

Babuška, I. and Oden, J.T. 2004. Verification and validation in computational engineering and science: basic concepts. *Computer Methods in Applied Mechanics and Engineering*. 193(36-38): 4057-4066.

Babuška, I., Nobile, F. and Tempone, R. 2006. Reliability of computational science. *Numerical methods for partial differential equations*. 23(4):753-784

Babuška, I., Nobile, F. and Tempone, R. 2008. Formulation of static frame problem. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32): 2496-2499

Babuška, I., Nobile, F. and Tempone, R. 2008. A systematic approach to model validation based on Bayesian updates and prediction related rejection criteria. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32): 2517-2539

Beal, M.J. 2003. Variational algorithms for approximate Bayesian inference. PhD. Thesis, Gatsby Computational Neuroscience Unit, University College London.

Beck, J.L. and Katafygiotis, L.S. 1991. Updating of a model and its uncertainties utilizing dynamic test data. *Proc., First International Conference on Computational Stochastic Mechanics*, Computational Mechanics Publications, Boston, 125-136.

Beck, J.L. and Katafygiotis, L.S. 1998. Updating models and their uncertainties: Bayesian statistical framework. *ASCE Journal of Engineering Mechanics*, 124(4): 455-461.

Beck, J.L. and Au, S.K. 2000. Updating robust reliability using Markov Chain simulation. *Proc., International Conference on Monte Carlo Simulation*, Monte Carlo, Monaco, June 2000.

Beck, J.L. and Au, S.K. 2002. Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation. *ASCE Journal of Engineering Mechanics*, 128(2): 380-391.

Beck, J.L., Au, S.K. and Vanik, M.W. 2001. Monitoring structural health using a probabilistic measure. *Computer-Aided Civil and Infrastructure Engineering* 16: 1-11.

Beck, J.L. and Yuen, K.V. 2004. Model selection using response measurements: A Bayesian probabilistic approach. *ASCE Journal of Engineering Mechanics*, 130(2):192-203.

Beck, J.L. and Cheung, S.H. 2009. Probability logic, model uncertainty and robust predictive system analysis. *Proc. 10th International Conference on Structural Safety and Reliability*, Osaka, Japan, September 13-17, 2009.

Berger, J. and Delampady, M. 1987. Testing precise hypotheses. *Statistical Science*. 3:317-352.

Berger, J. and Pericchi, L. 1996. The intrinsic Bayes factor for model selection and prediction. *Journal of American Statististical Association*. 91:109-122.

Bishop, C.M. 2006. Pattern recognition and machine learning. Springer.

Cheung, S.H. and Beck, J.L. 2007a. New stochastic simulation method for updating robust reliability of dynamic systems, *18th Engineering Mechanics Division Conference for the American Society of Civil Engineers* (*EMD2007*), Blacksburg, Virginia, USA, June 3-6, 2007.

Cheung, S.H. and Beck, J.L. 2007b. Bayesian model class selection of higher-dimensional dynamic systems using posterior samples, *18th Engineering Mechanics Division Conference for the American Society of Civil Engineers* (*EMD2007*), Blacksburg, Virginia, USA, June 3-6, 2007.

Cheung, S.H. and Beck, J.L. 2007c. Bayesian model updating of higher-dimensional dynamic systems. *10th International Conference on Applications of Statistics and Probability in Civil Engineering (ICASP10),* the University of Tokyo, Tokyo, Japan, July 31-August 3, 2007.

Cheung, S.H. and Beck, J.L. 2007d. Algorithms for Bayesian model class selection of higher-dimensional dynamic systems. *ASME 2007 Intl Design Engineering & Computers and Information in Engineering Conferences*, Las Vegas, Nevada, USA, September 4-7, 2007.

Cheung, S.H. and Beck, J.L. 2008a. Bayesian model updating using Hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters, *ASCE Journal of Engineering Mechanics*, In print in April 2009.

Cheung, S.H. and Beck, J.L. 2008b. New Bayesian updating methodology for model validation and robust predictions based on data from hierarchical subsystem tests. EERL Report No. 2008-04, California Institute of Technology.

Cheung, S.H. and Beck, J.L. 2008c. Near real-time loss estimation of structures subjected to strong seismic excitation. *Inaugural International Conference of the Engineering Mechanics Institute (EM08)*, University of Minnesota, Minneapolis, Minnesota, USA, May 18-21, 2008.

Cheung, S.H. and Beck, J.L. 2008d. Updating reliability of monitored nonlinear structural dynamic systems using real-time data. *Proc. Inaugural International Conference of the Engineering Mechanics Institute* (EM08), University of Minnesota, Minneapolis, Minnesota, USA, May 18-21, 2008.

Cheung, S.H. and Beck, J.L. 2008e. On using posterior samples for model selection for structural identification" *Asian-Pacific Symposium on Structural Reliability and its Applications 2008 (APSSRA'08)*, Hong Kong University of Science and Technology, Hong Kong, China, June 18-20, 2008.

Cheung, S.H. and Beck, J.L. 2008f. Calculation of the posterior probability for Bayesian model class selection and averaging from posterior samples based on dynamic system data. *Computer-Aided Civil and Infrastructure Engineering*, Accepted for Publication.

Cheung, S.H. and Beck, J.L. 2008g. New Bayesian updating methodology for model validation and robust predictions of a target system based on hierarchical subsystem tests. *Computer Methods in Applied Mechanics and Engineering*, Accepted for Publication.

Cheung, S.H. and Beck, J.L. 2009a. Model class comparison for Bayesian updating and robust system reliability predictions using stochastic nonlinear dynamic system models". *Workshop on Statistical Methods for Dynamic Systems*, Vancouver, Canada, June 4-6, 2009.

Cheung, S.H. and Beck, J.L. 2009b. Comparison of different model classes for Bayesian updating and robust predictions using stochastic state-space system models. *Proc. 10th International Conference on Structural Safety and Reliability* (ICOSSAR09), Osaka, Japan, September 13-17, 2009.

Chib, S. 1995, Marginal likelihood from the Gibbs output. *Journal of American Statististical Association*. 90:1313-1321.

Chib, S. and Jeliazkov I. 2001. Marginal likelihood from the Metropolis-Hastings output. *Journal of American Statististical Association*. 96:270-281.

Chleboun, J. An approach to the Sandia workshop static frame challenge problem: A combination of elementary probabilistic, fuzzy set, and worst scenario tools. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32): 2500-2516

Ching, J., Muto, M. and Beck, J.L. 2005, Bayesian linear structural model updating using Gibbs sampler with modal data, *Proc. International Conference on Structural Safety and Reliability*, Rome, Italy, June 2005.

Ching, J. and Hsieh, Y.H. 2006. Updating reliability of instrumented geotechnical systems via simple Monte Carlo simulation. *Journal of GeoEngineering*. 1(2): 71-78.

Ching, J., Muto, M. and Beck, J.L. 2006a. Structural model updating and health monitoring with incomplete modal data using Gibbs Sampler. *Computer-Aided Civil and Infrastructure Engineering*. 21(4): 242-257.

Ching, J., Beck, J.L. and Porter, K.A. 2006b. Bayesian state and parameter estimation of uncertain dynamical systems. *Probabilistic Engineering Mechanics*, 21, 81-96.

Ching, J., Beck, J.L., Porter, K.A. and Shaikhutdinov, R. 2006c. Bayesian state estimation method for nonlinear systems and its application to recorded seismic response. *Journal of Engineering Mechanics*, 132, 396-410.

Ching, J. and Chen, Y.J. 2007. Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. *ASCE Journal of Engineering Mechanics*, 133(7): 816-832.

Ching, J. and Beck, J.L. 2007. Real-time reliability estimation for serviceability limit states in structures using only incomplete output data. *Probabilistic engineering mechanics*, 22(1): 50-62.

Cover, T.M. and Thomas, J.A. 2001. Elements of information theory. Wiley Series in Telecommunications, John Wiley & Sons, Inc.

Cox, R.T. 1946. Probability, Frequency, and reasonable expectation. *American Journal of Physics*, 14: 1–13.

Cox, R.T. 1961. The algebra of probable inference. Johns Hopkins University Press, Baltimore, MD.

Doucet, A., Freitas, J.F.G. de. and Gordon, N. 2000. Introduction to sequential Monte Carlo methods. *Sequential Monte Carlo Methods in Practice*, A. Doucet, J.F.G. de Freitas and N.J. Gordan, eds., Springer-Verlag, Berlin.

Dowding, K.J., Pilch, M. and Hills, R.G. 2008. Formulation of the thermal problem. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32): 2385-2389

Duane, S., Kennedy, A.D., Pendleton, B.J., and Roweth, D. 1987. Hybrid Monte Carlo. *Physics Letter B*, 195(2): 216-222.

Forest, E. and Ruth, R.D. 1990. Fourth-order symplectic integration. *Physica D*, 43(1): 105-117.

Gelfand, A.E. and Dey, D.K. 1994. Bayesian model choice: asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B*. 56: 501-514

Geman, S. and Geman, D. 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6): 194-207.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B. 1995. Bayesian data analysis. *Chapman and Hall*, London

Griewank, A. 1989. On automatic differentiation. mathematical programming: recent developments and applications. *Kluwer Academic Publishers*, M. Iri and K. Tanabe, eds., 83-108.

Grigoriu, M.D. and Field Jr., R.V. A solution to the static frame validation challenge problem using Bayesian model selection. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32): 2540-2549

Gull, S.F. 1988. Bayesian inductive inference and maximum entropy. *Maximum-Entropy and Bayesian Methods in Science and Engineering*. 1: 53-74

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

Hills, R.G., Pilch, M., Dowding, K.J., Red-Horse, J., Paez, T.L., Babuška, I., and Tempone, R. 2008. Validation challenge workshop. Computer Methods in Applied Mechanics and Engineering. 197(29-32): 2375-2380

Hoeting, J.A., Madigan, D., Raftery, A.E., and Volinsky, C.T. 1999. Bayesian model averaging: a tutorial (with discussion). *Statistical Science*. 14(4):382-417

Hoeting, J.A. Methodology for Bayesian model averaging: An update. 2002. *Proceedings - Manuscripts of invited paper presentations, International Biometric Conference*, *2002*, *Freiburg*, *Germany*, 231-240.

Hurvich, C.M. and C.L. Tsai. 1989. Regression and time series model selection in small samples. *Biometrika*. 76:297-307.

Jaynes, E.T. 2003. Probability theory: The logic of science. Cambridge University Press, London

Jeffreys, H. 1939(1st Edn), 1961(2nd Edn). Theory of probability. Oxford University Press, London.

Johnson, E.A., Lam, H.F., Katafygiotis, L.S. and Beck, J.L. 2004. Phase I IASC-ASCE Structural health monitoring benchmark problem using simulated data. *Journal of Engineering Mechanics* 130(1): 3-15.

Kagiwada, H., Kalaba, R., Rosakhoo, N.and Spingarn, K. 1986. Numerical derivatives and nonlinear analysis. *Mathematical Concepts and Methods in Science and Engineering 31*, Plenum Press, New York and London

Kass, R.E. and Raftery, A.E. 1993. Bayes factors and model uncertainty. Technical report 254, University of Washington.

Katafygiotis, LS. and Beck, JL. 1998. Updating models and their uncertainties: model identifiability. *ASCE Journal of Engineering Mechanics*, 124(4): 463-467.

Katafygiotis, L.S. and Lam, H.F. 2002. Tangential-projection algorithm for manifold representation in unidentifiable model updating problems. *Earthquake Engineering and Structural Dynamics*, 31(4): 791-812.

Katafygiotis. L.S. and Cheung, S.H. (2004). "Wedge simulation method for calculating the reliability of linear dynamical systems." *Journal of Probabilistic Engineering Mechanics*, 19(3), 229-238.

Katafygiotis. L.S. and Cheung, S.H. (2005). "A Two-Stage subset-Simulation-based approach for calculating the reliability of inelastic structural systems subjected to Gaussian random excitations." *Computer Methods in Applied Mechanics and Engineering*, 194(1), 1581-1595.

Katafygiotis. L.S. and Cheung, S.H. (2006). "Domain decomposition method for calculating the failure probability of linear dynamic systems subjected to Gaussian stochastic loads." *Journal of Engineering Mechanics*, 132(5), 475-486.

Katafygiotis, L.S., Moan, T., and Cheung, S.H. (2007). "Auxiliary domain method for solving multi-objective dynamic reliability problems for nonlinear structures" *International Journal of Structural Engineering and Mechanics*, 347(2), 25-33.

Katafygiotis, L.S., Cheung, S.H. (2007). "Application of Spherical Subset Simulation method and Auxiliary Domain Method on a benchmark reliability study." *Structural Safety*, 29(3), 194-207.

Katafygiotis, L.S., Cheung, S.H., and Yuen, K.V. (2008). "Spherical Subset Simulation (S^3) for solving nonlinear dynamical reliability problems." *International Journal of Reliability and Safety*. Accepted for publication.

Kleinman, N., Spall, JC. and Naiman, DQ. 1999. Simulation-based optimization using stochastic approximation using Common Random Numbers. *Management Science*, 45(11): 1570-1578.

Kullback, S., and Leibler, R.A. 1951. On information and sufficiency. *Annals of Mathematical Statistics*. 22: 79-86

Lam, H.F., Katafygiotis, L.S. and Mickleborough, N.C. 2004, Application of a statistical Model Updating Approach on Phase I of the IASC-ASCE SHM Benchmark Study. *ASCE Journal of Engineering Mechanics*, 130(1): 34-48.

Lindley, D.V. 1957. A Statistical Paradox. *Biometrika*. 44: 187-192.

Lindley, D.V. 1980. L.J. Savage-His work in probability and statistics. *Annals of Statistics*. 8: 1-14.

Mackay, D.J.C. 1992. Bayesian methods for adaptive methods. PhD. Thesis, California Institute of Technology, Computation and Neural Systems.

Mackay, D.J.C. 1993. Bayesian nonlinear modeling for the energy prediction competition. *ASHARE Transactions*, 100: 1053-1062.

Mackenzie, P. B. 1989. An improved hybrid Monte Carlo method. *Physics Letters B*, 226(3): 369-371.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21(6):1087-1092

Meng, X.L and Wong, W.H. 1996. Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*. 6: 831-860

Muto, M. and Beck, J.L. 2008. Bayesian updating of hysteretic structural models using stochastic simulation. *Journal of Vibration and Control*. 14(1-2):7-34.

Neal, R.M. 1994. An improved acceptance procedure for the Hybrid Monte Carlo Algorithm. *Journal of Computational Physics*, 111(1): 194-203.

Neal, R.M. 2001. Annealed importance sampling. *Statistics and Computing*. 11:125-139

Neal, R.M. 2005. Estimating ratios of normalizing constants using Linked Importance Sampling. Technical Report No. 0511, Dept. of Statistics, University of Toronto

Newton, M.A. and Raftery, A.E. 1994. Approximate Bayesian inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society, Series B*. 56: 3-48.

Oberkampf, W.L., Helton, J.C., Joslyn, C.A., Wojtkiewicz, S.F. and Ferson, S. 2004. Challenge problems: uncertainty in system response given uncertain parameters. *Reliability Engineering and System Safety*. 85(1-3): 11-19

Oh, C.K., Beck, J.L. and Yamada, M. 2008. Bayesian Learning using Automatic Relevance Determination Prior with an Application to Earthquake Early Warning. *Journal of Engineering Mechanics*. 134(12): 1013-1020

Papadimitriou C, Beck JL and Katafygiotis LS. 2001 Updating robust reliability using structural test data. *Probabilistic engineering mechanics*, 16(2): 103-113.

Pradlwarter, H.J. and Schuëller, G.I. The use of kernel densities and confidence intervals to cope with insufficient data in validation experiments. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32): 2550-2560

Rall, LB. 1981. Automatic Differentiation-Techniques and Applications. *Lecture Notes in Computer Science*, Vol .120, Springer, Berlin

Raftery, A.E., Madigan, D., and Hoeting, J.A. 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*. 92:179-191.

Rebba, R. and Cafeo, J. 2008. Probabilistic analysis of a static frame model. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32): 2561-2571.

Red-Horse, J.R. and Paez, T.L. 2008. Sandia National Laboratories Validation Workshop: Structural dynamics application. *Computer Methods in Applied Mechanics and Engineering*. 197(29-32):2578-2584

Robert, C.P. and Casella, G. 1999, 2004. Monte Carlo statistical methods. *Springer-Verlag*, New York.

Sadegh, P. and Spall, JC. 1998. Optimal random perturbations for multivariate stochastic approximation using a simultaneous perturbation gradient approximation. *IEEE Transactions on Automatic Control*, 43(10): 1480-1484

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics*. 6(2):461-464.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Technical J*. 27, 379-423 and 623-656.

Silverman, B.W. 1986. Density estimation for statistics and data analysis. London: Chapman and Hall.

Spall, J.C. 1997. Accelerated second-order stochastic optimization using only function measurements. *Proc., the 36th IEEE Conference on Decision and Control*, 1417-1424

Spall, J.C. 1998a. An overview of the simultaneous perturbation method for efficient optimization. *John Hopkins APL Technical Digest*, 19(4): 482-492.

Spall, J.C. 1998b. Implementation of the simultaneous perturbation algorithm for stochastic optimization. *IEEE Transactions on Aerospace and Electronic Systems*, 34(3): 817-823

Spiegelhalter, D.J., Nicola G.B., Bradley P. C. and Angelika V.D.L. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*. 64 (4): 583–639.

Schueller, G. and Pradlwarter, H.J. 2007. Benchmark study on reliability estimation in higher dimensions of structural systems-An overview. *Structural Safety* 29:167-182.

Tierney, L. 1994. Markov Chain for exploring posterior distributions. *Annals of Statistics*, 22(4): 1701-1762.

Wolfe, P. 1982. Checking the calculation of gradients. *ACM TOMS*, 6(4): 337-343

Yuen, K.V., Beck, J.L. and Au, S.K. (2004). Structural damage detection and assessment by adaptive Markov chain Monte Carlo simulation. *Structural Control and Health Monitoring*.11: 327-347.

Yuen, K.V. and Lam, H.F. 2006 On the complexity of artificial neural networks for Smart Structures Monitoring. *Journal of Engineering Structures* 28(7): 977-984.