

Bayesian Updating of Structural Models and Reliability using Markov Chain Monte Carlo Simulation

James L. Beck¹ and Siu-Kui Au²

Abstract: In a full Bayesian probabilistic framework for “robust” system identification, structural response predictions and performance reliability are updated using structural test data \mathcal{D} by considering the predictions of a whole set of possible structural models that are weighted by their updated probability. This involves integrating $h(\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})$ over the whole parameter space, where $\boldsymbol{\theta}$ is a parameter vector defining each model within the set of possible models of the structure, $h(\boldsymbol{\theta})$ is a model prediction of a response quantity of interest, and $p(\boldsymbol{\theta}|\mathcal{D})$ is the updated probability density for $\boldsymbol{\theta}$, which provides a measure of how plausible each model is given the data \mathcal{D} . The evaluation of this integral is difficult because the dimension of the parameter space is usually too large for direct numerical integration and $p(\boldsymbol{\theta}|\mathcal{D})$ is concentrated in a small region in the parameter space and only known up to a scaling constant. An adaptive Markov chain Monte Carlo simulation approach is proposed to evaluate the desired integral that is based on the Metropolis-Hastings algorithm and a concept similar to simulated annealing. By carrying out a series of Markov chain simulations with limiting stationary distributions equal to a sequence of intermediate probability densities that converge on $p(\boldsymbol{\theta}|\mathcal{D})$, the region of concentration of $p(\boldsymbol{\theta}|\mathcal{D})$ is gradually portrayed. The Markov chain samples are used to estimate the desired integral by statistical averaging. The method is illustrated using simulated dynamic test data to update the robust response variance and reliability of a moment-resisting frame for two cases: one where the model is only locally identifiable based on the data and the other where it is unidentifiable.

DOI: 10.1061/(ASCE)0733-9399(2002)128:4(380)

CE Database keywords: Structural models; Markov chain; Reliability; Bayesian analysis; Monte Carlo method; Robust; Simulation.

Introduction

There has been increasing interest in using a Bayesian statistical framework (Box and Tiao 1992; Sivia 1996) to treat the challenging problem of updating a structural model based on test data from a structure (Beck and Katafygiotis 1998; Englund and Sorensen 1998; Geyskens et al. 1998; Katafygiotis and Beck 1998; Katafygiotis et al. 1998; Singhal and Kiremidjian 1998; Zheng and Ellingwood 1998; Enright and Frangopol 1999; Papadimitriou et al. 2000; Papadimitriou et al. 2001; Vanik et al. 2000). The need for model updating arises because there are always modeling errors associated with constructing a theoretical model of the behavior of a structure, and this leads to uncertain accuracy in the predicted response. There are many sources of modeling errors, such as inexact modeling of material constitutive behavior and boundary conditions (e.g., there are no perfectly pinned or fixed joints); unmodeled features such as in-plane diaphragm flexibility, neglected “nonstructural” components and foundation

flexibility; errors because of the spatial discretization of distributed structural systems and loads; variation of material properties during manufacture; and uncertainties introduced by the construction process. Because of these modeling errors and uncertainties, model updating is best tackled as a Bayesian statistical inference problem (Beck 1989).

The paper by Collins et al. (1974) is a pioneering effort in Bayesian structural model updating using identified modal parameters. A more rigorous and comprehensive Bayesian framework for model updating is described by Beck and Katafygiotis (1998). They define the concept of system identifiability and show how to treat the ill-conditioning and lack of identifiability that often characterizes the updating problem. Using this methodology, one does not identify a single “best” model but instead updates a probability distribution over a specified set of structural models and a specified set of prediction-error probability models for the uncertain error between the actual structural response and the corresponding model predictions.

Model updating is not an end in itself; it is usually motivated by the desire to improve the accuracy of predictions of the system response or its current condition or “health.” In this paper, we focus on updating the “robust” response moments and the “robust” reliability that a specified system performance is achieved under some set of possible excitations. As in robust structural control, the use of the term “robust” here implies that a set of possible structural models is used to represent the structure rather than a single nominal model (Papadimitriou et al. 2001).

The problem of interest is defined as follows. Let \mathcal{D} denote some test data from a structure and consider a set of possible structural models specified by \mathcal{M} that has been chosen to represent the behavior of the structure. Here, \mathcal{M} specifies the modeling assumptions used in the analysis (both structural and probabilis-

¹Professor of Applied Mechanics and Civil Engineering, Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125 (corresponding author).

²Assistant Professor, School of Civil and Environmental Engineering, Nanyang Technological Univ., Singapore 639798; formerly, PhD candidate, Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125.

Note. Associate Editor: Stein Sture. Discussion open until September 1, 2002. Separate discussions must be submitted for individual papers. To extend the closing date by one month, a written request must be filed with the ASCE Managing Editor. The manuscript for this paper was submitted for review and possible publication on October 13, 2000; approved on July 18, 2001. This paper is part of the *Journal of Engineering Mechanics*, Vol. 128, No. 4, April 1, 2002. ©ASCE, ISSN 0733-9399/2002/4-380-391/\$8.00+\$0.50 per page.

tic). The essence of Bayesian statistical system identification is that it gives a rigorous method of using \mathcal{D} to update an initial description of how plausible each model is as a representation of the structure, that is, the information in \mathcal{D} modifies the knowledge about the relative plausibilities of the different models specified by \mathcal{M} (Beck and Katafygiotis 1998). The plausibility of a model is quantified by a probability distribution over the model parameters $\boldsymbol{\theta}=[\theta_1, \dots, \theta_n]$ that define a model within the set of possible models (Cox 1961; Jaynes 1978). The updated probability distribution $p_{\mathcal{D}}(\boldsymbol{\theta})=p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M})$ is obtained using Bayes' theorem

$$p_{\mathcal{D}}(\boldsymbol{\theta})=p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})p_0(\boldsymbol{\theta}|\mathcal{M})/p(\mathcal{D}|\mathcal{M})=c p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})p_0(\boldsymbol{\theta}|\mathcal{M}) \quad (1)$$

where $p_0(\boldsymbol{\theta}|\mathcal{M})$ =initial ("prior") probability distribution specified by \mathcal{M} , which reflects the relative plausibility of each model before utilizing the data \mathcal{D} , and $c^{-1}=p(\mathcal{D}|\mathcal{M})=\int p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})p_0(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}$ =normalizing constant. The term $p(\mathcal{D}|\boldsymbol{\theta}, \mathcal{M})$ gives the probability of obtaining the data \mathcal{D} based on a model specified by the model parameters $\boldsymbol{\theta}$. It is formulated by using a probability model for the prediction error relating the "output" of the model specified by $\boldsymbol{\theta}$ to the actual measured structural output \mathcal{D} (Beck and Katafygiotis 1998). The conditioning on \mathcal{M} is included here to stress the fact that all probabilities involved in model updating are always conditional on the choice of the modeling assumptions.

Better response predictions can be made by utilizing the information in the data \mathcal{D} through the use of $p_{\mathcal{D}}$ given by Eq. (1). For example, if $h(\boldsymbol{\theta})$ is a structural response quantity of interest, then many useful performance measures can be formulated as

$$R_{\mathcal{D}}=\int h(\boldsymbol{\theta})p_{\mathcal{D}}(\boldsymbol{\theta})d\boldsymbol{\theta} \quad (2)$$

where the theorem of total probability is used. One example is $h(\boldsymbol{\theta})=\text{Var}[y|\boldsymbol{\theta}, \mathcal{M}]$, the conditional variance of a given structural response y computed with specified model parameters $\boldsymbol{\theta}$ when the structure is subjected to an uncertain excitation. In this case, $R_{\mathcal{D}}$ in Eq. (2) is the variance $\text{Var}[y|\mathcal{D}, \mathcal{M}]$ of the response given the set of structural models \mathcal{M} and the data \mathcal{D} , which addresses the uncertainty in which model best represents the structure while incorporating information from the data \mathcal{D} . Another example is $h(\boldsymbol{\theta})=P(F|\boldsymbol{\theta}, \mathcal{M})$, the probability of a failure event F of the structure, predicted using model parameters $\boldsymbol{\theta}$; then $R_{\mathcal{D}}$ is the robust failure probability $P(F|\mathcal{D}, \mathcal{M})$ which incorporates knowledge from \mathcal{M} and updated information from \mathcal{D} (Papadimitriou et al. 2001).

Suppose the model parameters $\boldsymbol{\theta}$ belong to a continuous parameter space of dimension 3 or more, then the evaluation of $R_{\mathcal{D}}$ by direct numerical integration is inefficient or infeasible. Also, the updated probability density function (PDF) $p_{\mathcal{D}}$ is known only up to a multiplicative constant because the normalizing constant c in Eq. (1) is given by an integral that is also difficult to evaluate. In fact, in terms of quantities whose values can be computed explicitly for a given $\boldsymbol{\theta}$, $R_{\mathcal{D}}$ can be expressed as (leaving dependence on \mathcal{M} implicit)

$$R_{\mathcal{D}}=\frac{\int h(\boldsymbol{\theta})p(\mathcal{D}|\boldsymbol{\theta})p_0(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int p(\mathcal{D}|\boldsymbol{\theta})p_0(\boldsymbol{\theta})d\boldsymbol{\theta}} \quad (3)$$

and so its evaluation involves two multidimensional integrals. Additional difficulties come from the nature of the updated PDF $p_{\mathcal{D}}$, which is usually concentrated in a small volume of the parameter space. In practical applications, the variation of $p_{\mathcal{D}}$ in the parameter space is more dominant than that of the response quantity h ,

and so methods for evaluating $R_{\mathcal{D}}$ are differentiated according to the shape of $p_{\mathcal{D}}$, which depends on the information that the available data \mathcal{D} produce about the model parameters $\boldsymbol{\theta}$.

The characterization of $p_{\mathcal{D}}$ has been studied for model identification of structures using as dynamic data the measured time histories of response and excitation (Beck and Katafygiotis 1998; Katafygiotis and Beck 1998; Katafygiotis et al. 1998), but the results can be applied in general. For a large amount of data (e.g., the number of data points M in the measured time histories is large), $p_{\mathcal{D}}$ is concentrated in the neighborhood \mathcal{N} of a lower-dimensional manifold \mathcal{S} in the parameter space on which $p(\mathcal{D}|\boldsymbol{\theta})$ is globally maximized. The thickness of \mathcal{N} around \mathcal{S} is of the order of $\epsilon=1/\sqrt{M}$. The characterization of $p_{\mathcal{D}}$ can be made according to the dimension of \mathcal{S} , which depends on the amount of data available compared to the information to be extracted from them. In the "identifiable" case where the number of model parameters is less than or equal to the number of "effective constraints" from the data \mathcal{D} , the dimension of \mathcal{S} is zero, that is, the updated PDF is concentrated in the close neighborhood \mathcal{N} of a finite number of isolated points, referred to as "optimal parameter" points. In this case, $p_{\mathcal{D}}$ can be well approximated by a weighted sum of Gaussian PDFs with spread of $O(\epsilon)$ centered at the optimal parameter points (Beck and Katafygiotis 1998). Consequently, an asymptotic approximation of $R_{\mathcal{D}}$ can be expressed as a weighted sum of values of h evaluated at the optimal parameter points, and the problem of calculating $R_{\mathcal{D}}$ in the identifiable case is reduced to finding the optimal parameter points and their associated probability weights. This leads to a nonconvex global optimization problem, which is not trivial to solve.

In the locally identifiable case (Katafygiotis and Beck 1998), there are multiple optimal parameter points and "global" optimization algorithms must be used to find them (Törn and Zilinskas 1989; Yang and Beck 1998). The computational effort is often much greater than is needed in a local optimization problem, especially when the dimension of the parameter space is not small and the number of optimal parameters is unknown, as is usually the case.

In the "unidentifiable" case where the number of model parameters is larger than the number of effective data constraints, the dimension of the manifold \mathcal{S} is greater than zero, and there exists a continuum of optimal parameter points lying on \mathcal{S} which give the same global maximum value of $p(\mathcal{D}|\boldsymbol{\theta})$. Deterministic search methods for computing an asymptotic approximation of $R_{\mathcal{D}}$ have been developed (Katafygiotis et al. 1998), which discretize the manifold \mathcal{S} using a finite number of representative points and then approximate $p_{\mathcal{D}}$ as a discrete probability mass distributed among the representative points. The main computational effort is spent on locating the representative points on \mathcal{S} , which requires a series of local optimizations in the parameter space. The representative points have to be located over the region of \mathcal{S} where the prior PDF p_0 is significant so that the contributions from different parts of \mathcal{S} , and hence \mathcal{N} , are accounted for. Consequently, the complexity and computational effort are expected to grow with the dimension of the manifold in a similar manner to that of direct numerical integration, making the method practical only when this dimension is no more than 2.

This paper presents a Markov chain simulation method to evaluate the integral in Eq. (2) for $R_{\mathcal{D}}$ without the need for optimization to find the manifold \mathcal{S} . It is based on the Metropolis-Hastings (MH) algorithm and a concept similar to simulated annealing (Fishman 1996) to gain information about the manifold in an iterative manner. By carrying out a series of Markov chain simulations with limiting stationary distributions equal to a se-

quence of intermediate PDFs that converge on $p_{\mathcal{D}}(\boldsymbol{\theta})$, the region \mathcal{N} of significant probability density of $p_{\mathcal{D}}$ is gradually portrayed. The Markov chain samples can be used to estimate $R_{\mathcal{D}}$ by statistical averaging.

The difficulties in applying simulation methods to evaluate the integral in Eq. (2) for $R_{\mathcal{D}}$ are discussed in the next section, followed by the proposed Markov chain simulation method. The statistical properties of the estimator for $R_{\mathcal{D}}$ are then studied. Finally, the method is illustrated with the problem of updating the robust response variance and reliability of a moment-resisting frame with measured modal properties.

Difficulties in Evaluating $R_{\mathcal{D}}$ by Simulation

Eqs. (2) and (3) suggest two ways of evaluating $R_{\mathcal{D}}$ by simulation. The former suggests estimating $R_{\mathcal{D}}$ as the average of h over samples simulated from $p_{\mathcal{D}}$, while the latter indicates that the integrals in the numerator and denominator could be estimated individually and then combined to give an estimate for $R_{\mathcal{D}}$. Since $p_{\mathcal{D}}$ is known only up to a multiplicative constant and in general a method for simulating independent samples from $p_{\mathcal{D}}$ is not available, the first option of using Eq. (2) is not feasible for existing methods such as Monte Carlo simulation (MCS) or importance sampling (Rubinstein 1981). Therefore, consider the application of these existing methods to evaluate $R_{\mathcal{D}}$ based on Eq. (3). Using MCS, the numerator is estimated as the average of $h(\boldsymbol{\theta})p(\mathcal{D}|\boldsymbol{\theta})$ over samples drawn from the prior PDF p_0 (assuming a method for simulating samples from p_0 is available). The resulting estimate, however, is very likely to be biased, since $p(\mathcal{D}|\boldsymbol{\theta})$ is concentrated in the small neighborhood \mathcal{N} of thickness $O(\epsilon)$ where ϵ is very small, and hence the chance of generating a sample from p_0 that lies in \mathcal{N} is extremely small. Similar difficulties will be encountered in evaluating the denominator. Using importance sampling, it is necessary to choose a sampling density that is concentrated in \mathcal{N} , otherwise similar problems as in a MCS will be encountered. However, this is extremely difficult since information about the manifold \mathcal{S} where the probability density is concentrated is not directly available.

Next, consider again evaluating $R_{\mathcal{D}}$ based on Eq. (2). Although a method for generating independent samples according to $p_{\mathcal{D}}$ is generally not available, it is noted that Markov chain Monte Carlo simulation, in particular, the MH algorithm, offers a feasible way to simulate samples according to an arbitrary distribution, at the expense of introducing dependence among the samples. However, direct application of the MH algorithm to simulate Markov chain samples according to $p_{\mathcal{D}}$ is not feasible due to the small region \mathcal{N} of probability concentration of $p_{\mathcal{D}}$. Nevertheless, as our proposed method is built on the MH algorithm, we first discuss its implementation.

Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is a simple procedure to simulate samples according to an arbitrary PDF where the target PDF need only be known up to a scaling constant. It was originally developed by Metropolis and his co-workers for computing canonical ensembles in statistical physics (Metropolis et al. 1953) and later generalized by Hastings in Bayesian statistics (Hastings 1970). Its potential use for solving reliability problems has been recently demonstrated by Au and Beck (1999).

In the MH method, samples are simulated as the states of a special Markov chain whose limiting stationary distribution is

equal to the target PDF. In other words, the PDF of the Markov chain sample $\boldsymbol{\theta}_k$ simulated at the k th Markov step tends to the target PDF as $k \rightarrow \infty$. The Markov chain samples, which are dependent in general, can be used for statistical averaging as if they were independent, although with some reduction of efficiency in the estimator.

Let $p^*(\boldsymbol{\xi}|\boldsymbol{\theta})$ be a chosen PDF, called the ‘‘proposal PDF,’’ which is a PDF for $\boldsymbol{\xi}$ that depends on $\boldsymbol{\theta}$. The role of p^* will become clear shortly. For convenience in notation, let $q(\boldsymbol{\theta}) = p(\mathcal{D}|\boldsymbol{\theta})p_0(\boldsymbol{\theta}) = c^{-1}p_{\mathcal{D}}(\boldsymbol{\theta})$. Note that the value of q can be computed readily for a given $\boldsymbol{\theta}$, while the same is not true for $p_{\mathcal{D}}$. The MH algorithm to simulate Markov chain samples $\{\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N\}$ with limiting stationary distribution equal to the target PDF $p_{\mathcal{D}}$ is described as follows. To start the Markov chain, let $\boldsymbol{\theta}_1$ be a point either chosen deterministically or simulated according to some PDF that approximates $p_{\mathcal{D}}$. In general, to simulate the next sample $\boldsymbol{\theta}_{k+1}$ from the current sample $\boldsymbol{\theta}_k$, $k=1, \dots, N-1$, first simulate a ‘‘candidate state’’ $\boldsymbol{\xi}$ from the proposal PDF $p^*(\boldsymbol{\xi}|\boldsymbol{\theta}_k)$. Compute the ratio

$$r = \frac{q(\boldsymbol{\xi})p^*(\boldsymbol{\theta}_k|\boldsymbol{\xi})}{q(\boldsymbol{\theta}_k)p^*(\boldsymbol{\xi}|\boldsymbol{\theta}_k)} \quad (4)$$

Then, accept the candidate state $\boldsymbol{\xi}$ with probability $\min\{1, r\}$ and reject with the remaining probability $1 - \min\{1, r\}$. If accepted, the candidate state will be taken as the next state of the Markov chain, i.e., $\boldsymbol{\theta}_{k+1} = \boldsymbol{\xi}$. Otherwise, the current state is taken as the next state, i.e., $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$. The process is repeated until N Markov chain samples have been simulated.

We now show that the next sample $\boldsymbol{\theta}_{k+1}$ will be distributed as $p_{\mathcal{D}}$ if the current sample $\boldsymbol{\theta}_k$ is. From the theorem of total probability, the PDF of the next sample is given by

$$p(\boldsymbol{\theta}_{k+1}) = \int p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k)d\boldsymbol{\theta}_k \quad (5)$$

where $p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)$ is the transition PDF governing the probabilistic properties of the Markov chain. From the MH algorithm, the transition PDF is given by, for $\boldsymbol{\theta}_{k+1} \neq \boldsymbol{\theta}_k$

$$p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k) = p^*(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)\min\{1, r(\boldsymbol{\theta}_{k+1}, \boldsymbol{\theta}_k)\} \quad (6)$$

Using Eqs. (4) and (6), along with the identity $\min\{1, a/b\}b = \min\{1, b/a\}a$ for any positive numbers a and b , and the fact that q differs from $p_{\mathcal{D}}$ only by a normalizing constant, one can readily show the following ‘‘reversibility’’ condition:

$$p(\boldsymbol{\theta}_{k+1}|\boldsymbol{\theta}_k)p_{\mathcal{D}}(\boldsymbol{\theta}_k) = p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})p_{\mathcal{D}}(\boldsymbol{\theta}_{k+1}) \quad (7)$$

which basically says that the transition rates between the two states $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_{k+1}$ are equal when $p(\boldsymbol{\theta}_k) = p_{\mathcal{D}}(\boldsymbol{\theta}_k)$ and $p(\boldsymbol{\theta}_{k+1}) = p_{\mathcal{D}}(\boldsymbol{\theta}_{k+1})$, that is, the Markov chain is stationary with stationary PDF $p_{\mathcal{D}}$. Note that Eq. (7) is trivial for $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k$. Assuming the current sample $\boldsymbol{\theta}_k$ is distributed as $p_{\mathcal{D}}$, i.e., $p(\boldsymbol{\theta}_k) = p_{\mathcal{D}}(\boldsymbol{\theta}_k)$, and using the reversibility condition in Eq. (7), $p(\boldsymbol{\theta}_{k+1})$ in Eq. (5) becomes

$$\begin{aligned} p(\boldsymbol{\theta}_{k+1}) &= \int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})p_{\mathcal{D}}(\boldsymbol{\theta}_{k+1})d\boldsymbol{\theta}_k \\ &= p_{\mathcal{D}}(\boldsymbol{\theta}_{k+1}) \int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})d\boldsymbol{\theta}_k = p_{\mathcal{D}}(\boldsymbol{\theta}_{k+1}) \end{aligned} \quad (8)$$

since $\int p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k+1})d\boldsymbol{\theta}_k = 1$. This means that if the current sample $\boldsymbol{\theta}_k$ is distributed as the target PDF $p_{\mathcal{D}}$, then so is the next sample $\boldsymbol{\theta}_{k+1}$, and hence $p_{\mathcal{D}}$ is the stationary PDF of the Markov chain.

In an actual implementation, the Markov chain is started with the initial state θ_1 simulated from a PDF different from p_D , so the Markov chain is in a transient state and its samples will not be distributed exactly as p_D . Under the assumption of ergodicity, however, the Markov chain will converge to the stationary state, and so the PDF of θ_k will tend to p_D as $k \rightarrow \infty$. Theoretically, the assumption of ergodicity is usually satisfied if the proposal PDF p^* has a positive density on the support of the target PDF. It is also usually satisfied even if p^* has a restricted support. With a finite sample size N used in an actual implementation, ergodicity often becomes an issue of whether the N Markov chain samples can populate sufficiently well the region of significant probability of the target PDF p_D . See Au and Beck (1999) for a more detailed discussion of ergodicity in applying the MH method to reliability problems.

Difficulties in Applying Metropolis-Hastings Algorithm to Reliability Updating

Using the Markov chain samples $\{\theta_1, \dots, \theta_N\}$ generated from the MH procedure, R_D is estimated as the average \tilde{R}_D of h over the samples, which is the same as the usual MCS estimator, except that the samples are simulated from a Markov chain instead of being independent and identically distributed (i.i.d.). Nevertheless, the estimator \tilde{R}_D has similar statistical properties to those of MCS estimators (see later). In order to reduce the initial transient effect of the Markov chain on the estimate, the first few samples (say 10) are often not used to compute the estimate \tilde{R}_D . In this work, the Markov chain samples $\{\theta_1, \dots, \theta_N\}$ used for computing the estimate \tilde{R}_D are those simulated after the initial transient stage.

The proposal PDF p^* in the MH algorithm affects the distribution of the candidate state ξ given the current state, and consequently the convergence rate of the estimator \tilde{R}_D to R_D . If the candidate state is rejected too often because small values of r in Eq. (4) are encountered during simulation, the Markov chain will consist of many repeated samples. As a result, the correlation among samples will be increased, slowing down the convergence of \tilde{R}_D . To understand how p^* affects the acceptance rate of the candidate state, first note that, if $p^*(\xi|\theta) = p_D(\xi)$, then according to Eq. (4), $r = 1$ and hence the candidate state is always accepted. In this case, the MH algorithm reduces to a standard Monte Carlo procedure with i.i.d. samples simulated from p_D . Although this choice is not possible, it indicates that, if p^* is chosen to be "nonadaptive," i.e., $p^*(\xi|\theta) = p^*(\xi)$, then the closer the p^* is to the target PDF, the better the acceptance rate of the candidate state, and the faster the convergence.

Choosing a nonadaptive proposal PDF p^* means that the information from the current sample is not used to explore the important region of significant probability density of the target PDF during simulation. When information about the important region is not available, constructing a nonadaptive p^* so that the candidate state simulated from it will lie in \mathcal{N} is similar to constructing an importance sampling density concentrated in \mathcal{N} , and is thus very difficult. It is therefore more desirable to choose an "adaptive" p^* which depends on the current sample. One popular choice is to have p^* localized and symmetric, i.e., $p^*(\xi|\theta) = p^*(\theta|\xi)$, which corresponds to the one used in the original Metropolis algorithm (Metropolis et al. 1953). In this case, p^* can be interpreted as a PDF localized at the current sample. The Markov chain simulation process can then be viewed as a "local random walk" in which the region of probability concentration of the target PDF is adaptively explored.

Direct application of the MH algorithm to simulate samples according to the target PDF p_D is not feasible, however, due to the problems arising from the small region \mathcal{N} of probability concentration of p_D , as described for the other methods discussed in the last section. In particular, it is difficult to choose the proposal PDF p^* so that the acceptance rate of the candidate state is not too small while at the same time the Markov chain samples effectively explore \mathcal{N} . To see this, first note that it is not possible to choose a nonadaptive p^* that can generate samples lying in \mathcal{N} , since the information about the manifold \mathcal{S} which localizes \mathcal{N} is not directly available. Thus, consider choosing an adaptive p^* , such as a symmetric one in the original Metropolis algorithm. If the Markov chain is started in a region not near \mathcal{N} , then the chance of generating a candidate state from p^* that visits \mathcal{N} is extremely small, and most of the candidate states will be rejected. As a result, most of the Markov chain samples are repeated, and clearly they cannot be used to estimate R_D . On the other hand, if the Markov chain is started in \mathcal{N} , then in order for the simulated candidate state to remain in \mathcal{N} and have a high probability of being accepted, the spread of p^* around the current sample has to be $O(\epsilon)$, because the direction along which \mathcal{N} extends around the current sample is not known. But this means the candidate state will be very close to the current sample, and as a result the Markov chain samples will not efficiently explore \mathcal{N} . In all these cases, the region visited by the Markov chain samples will be small compared to \mathcal{N} , leading to significant bias in the estimate for R_D .

Proposed Adaptive Simulation Method

The problems encountered in applying the simulation methods discussed in the previous sections, including the MH algorithm, arise from the fact that the updated PDF p_D is concentrated in a small neighborhood \mathcal{N} of the manifold \mathcal{S} containing the maxima of p_D . The process of adapting samples to \mathcal{N} in the MH algorithm is inhibited by the small scale ϵ of the thickness of \mathcal{N} compared to the size of the proposal PDF required to cover \mathcal{N} with an affordable number of samples. This suggests that direct adaptation using a proposal PDF which varies with a vastly different length scale from that of the target PDF will not be effective. In view of this, we introduce a sequence of intermediate PDFs which bridge the gap in length scale between the prior PDF p_0 and the target updated PDF p_D . By successively using the MH algorithm to utilize the information from the previous adapted intermediate PDFs, the region populated by the Markov chain samples gradually evolves into the desired region \mathcal{N} . Conceptually, this is similar to "simulated annealing" (Fishman 1996).

Let p_1, \dots, p_m be a chosen sequence of PDFs converging to $p_D (= p_m)$ so that their region of significant probability content gradually diminishes to that of p_D . For example, p_i may be chosen as the updated PDF from Bayes' theorem based on an increasing amount of data, i.e., $p_i = p_{D_i}$, where $\mathcal{D}_1 \subset \dots \subset \mathcal{D}_m = \mathcal{D}$. The proposed method is described as follows. Starting with the prior PDF p_0 as the proposal PDF, the MH algorithm is carried out to simulate samples $\{\theta_1^{(1)}, \dots, \theta_N^{(1)}\}$ with target PDF p_1 . A kernel sampling density κ_1 is constructed as a weighted sum of Gaussian PDFs centered among these samples to approximate p_1 (Silverman 1986; Ang et al. 1992; Au and Beck 1999).

$$\kappa_1(\theta) = \sum_{k=1}^N w_k \phi(\theta; \theta_k^{(1)}, C_k) \quad (9)$$

where $\phi(\theta; \theta_k^{(1)}, C_k) =$ multidimensional Gaussian PDF evaluated at θ with mean $\theta_k^{(1)}$ and covariance matrix C_k ; and

w_k =probability weights associated with the Gaussian PDFs, which satisfy $w_k \geq 0$ and $\sum_{k=1}^N w_k = 1$. The choice of w_k and C_k is completely defined by the samples $\{\theta_k^{(1)}\}$ (Au and Beck 1999), and therefore so is the choice of κ_1 . Since κ_1 is a weighted sum of Gaussian PDFs, the kernel marginal PDF for a particular component $\theta_j (j=1, \dots, n)$ of θ can be obtained by analytically integrating Eq. (9) with respect to all other components $\theta_l, l \neq j$

$$\kappa_1(\theta_j) = \sum_{k=1}^N w_k \phi(\theta_j; \theta_k^{(1)}(j), C_k(j, j)) \quad (10)$$

where $\phi(\theta_j; \theta_k^{(1)}(j), C_k(j, j))$ =one-dimensional Gaussian PDF evaluated at θ_j with mean $\theta_k^{(1)}(j)$ and variance $C_k(j, j)$; $\theta_k^{(1)}(j)$ = j th component of the vector $\theta_k^{(1)}$; and $C_k(j, j)$ = j th diagonal entry of the covariance matrix C_k . Similarly, the kernel marginal cumulative distribution function (CDF) of θ_j can be readily computed in terms of a weighted sum of Gaussian CDFs. More importantly, independent samples distributed as κ_1 can be readily simulated that are approximately distributed as p_1 and so lie in the region of significant probability of p_1 . In particular, to simulate a sample from κ_1 in Eq. (9), first draw an index l from the set $\{1, \dots, N\}$ with corresponding probabilities $\{w_1, \dots, w_N\}$, and then simulate the sample as a Gaussian vector with mean vector $\theta_l^{(1)}$ and covariance matrix C_l .

To proceed, κ_1 is used as the proposal PDF for simulating Markov chain samples $\{\theta_1^{(2)}, \dots, \theta_N^{(2)}\}$ with target PDF p_2 . These samples are then used to construct the kernel sampling density κ_2 as in Eq. (9), which gives an approximation to p_2 . In general, the kernel sampling density κ_i (which approximates p_i) is constructed using the Markov chain samples at the i th simulation level, which is then used as the proposal PDF for simulating Markov chain samples for the next level with target PDF p_{i+1} . This is continued until the m th simulation level, where Markov chain samples for the target updated PDF $p_D = p_m$ are simulated.

Let R_i be the expectation of the response quantity $h(\theta)$ when θ is distributed as p_i , that is

$$R_i = \int h(\theta) p_i(\theta) d\theta \quad (11)$$

Since p_i converges to p_D as the simulation level i increases to m , R_i converges to R_D defined by Eq. (2). Using the Markov chain samples $\{\theta_1^{(i)}, \dots, \theta_N^{(i)}\}$, R_i is estimated at the i th simulation level by \tilde{R}_i where

$$R_i \approx \tilde{R}_i = \frac{1}{N} \sum_{k=1}^N h(\theta_k^{(i)}) \quad (12)$$

The proposed strategy makes use of the fact that, if a nonadaptive proposed PDF is similar in shape to the target PDF, then the acceptance rate of the candidate state will not be small, and the MH algorithm will be effective in generating samples populating the important region of the target PDF. Thus, instead of using an adaptive proposal PDF within each simulation level, the adaptation is done from one simulation level to the next, where the latest kernel sampling density transfers the information about the important region from the current level to the next level. At each simulation level, since the candidate state is simulated independently of the current sample, successive Markov chain samples are correlated only if the candidate state is rejected. With a reasonable acceptance rate of the candidate states, this leads to a

global covering of the important region of the target PDF at each simulation level even with a relatively small number of samples, since the Markov chain develops globally rather than locally.

The success of the proposed strategy relies on effective application of the MH algorithm at each simulation level, which requires that p_{i-1} (which is approximated by k_{i-1}) varies with a similar length scale to p_i for $i=1, \dots, m$. The choice of the sequence $\{p_i\}$ (corresponding to the ‘‘annealing schedule’’ in the simulated annealing algorithm) is thus important to the success of the proposed method. The slower the ‘‘adaptation’’ schedule is, that is, the less p_{i+1} is changed compared to p_i , the higher the acceptance rate of the candidate state, and the better the result at each simulation level. However, a slow schedule requires more simulation levels to reach the final level for the updated PDF, and hence more Markov chain samples to be simulated. The prudent choice of an adaptation schedule is a trade-off between the efficiency of the MH algorithm at each simulation level and the total number of simulation levels required.

The choice of the adaptation schedule can be made by investigating the probabilistic properties of the ratio r in the MH algorithm appearing in Eq. (4), which governs the acceptance rate of the candidate states. It is found that a good choice for $\{p_i\}$ is $p_{i+1} \sim p_i^2$ (up to a normalizing constant). In this case, the size of the region of probability concentration for p_{i+1} is roughly $1/\sqrt{2}$ that of p_i . When the data \mathcal{D} consist of measured response time histories (Beck and Katafygiotis 1998), this schedule can be achieved by choosing $p_i = p_{\mathcal{D}_i}$, where the duration of time history data is doubled in successive simulation levels from \mathcal{D}_i to \mathcal{D}_{i+1} . On the other hand, when the updated PDF with data \mathcal{D} is of the form $p_{\mathcal{D}}(\theta) = c \exp[-J(\theta)/2\epsilon^2]$ (Vanik et al. 2000), where $J(\theta)$ is a measure-of-fit function between data and model, and ϵ is a measure of the size of the prediction error, then the sequence $\{p_i\}$ can be obtained by $p_i = c_i \exp[-J(\theta)/2\epsilon_i^2]$ where $\epsilon_i^2 = 2^{m-i}\epsilon^2$, $i = 1, \dots, m$, with $2^m \approx \epsilon^{-2}$, if the length scale of the prior PDF $p_0(\theta)$ is $O(1)$.

Statistical Properties of Estimators

The statistical properties of the estimator \tilde{R}_i in Eq. (12) are presented in this section, assuming the Markov chain generated according to the MH algorithm at each simulation level is ergodic.

In spite of the fact that \tilde{R}_i is computed using dependent samples from a Markov chain, it still has the usual properties of MCS estimators using i.i.d. samples (Doob 1953). For example, \tilde{R}_i converges to R_i with probability 1 as $N \rightarrow \infty$ (strong law of large numbers), and under similar conditions as those for Monte Carlo estimators, \tilde{R}_i is normally distributed as $N \rightarrow \infty$ (central limit theorem). If the Markov chain is started with the initial state $\theta_1^{(i)}$ distributed as the target PDF p_i , then the Markov chain is stationary, and \tilde{R}_i is unbiased, that is, $E[\tilde{R}_i] = R_i$. Otherwise, \tilde{R}_i is only asymptotically unbiased, although the bias decays exponentially with the number of Markov steps.

An expression for the coefficient of variation (COV) of \tilde{R}_i , defined as the ratio of the standard deviation to the mean of \tilde{R}_i , is next derived assuming that the Markov chain has settled into its stationary state and the proposal PDF for the i th simulation level is fixed. As a result of stationarity, $E[\tilde{R}_i] = R_i$. The variance of \tilde{R}_i is thus given by

$$\begin{aligned}
E[\tilde{R}_i - R_i]^2 &= E\left[\frac{1}{N} \sum_{k=1}^N [h(\boldsymbol{\theta}_k^{(i)}) - R_i]\right]^2 \\
&= \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N E\{(h(\boldsymbol{\theta}_j^{(i)}) - R_i)[h(\boldsymbol{\theta}_k^{(i)}) - R_i]\} \\
&= \frac{1}{N^2} \sum_{j=1}^N \sum_{k=1}^N c_i(j, k) \quad (13)
\end{aligned}$$

where $c_i(j, k)$ = covariance between $h(\boldsymbol{\theta}_j^{(i)})$ and $h(\boldsymbol{\theta}_k^{(i)})$. Since the Markov chain is stationary and $c_i(j, k)$ is symmetric in j and k , $c_i(j, k)$ depends only on the absolute difference between the indexes j and k , that is, $c_i(j, k) = c_i(|j - k|)$. Using this result and carrying out the double sum in Eq. (13) with respect to $(j - k)$, we have

$$\begin{aligned}
E[\tilde{R}_i - R_i]^2 &= \frac{1}{N^2} \left[Nc_i(0) + 2 \sum_{k=1}^{N-1} (N-k)c_i(k) \right] \\
&= \frac{c_i(0)}{N} \left[1 + 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \rho_i(k) \right] \quad (14)
\end{aligned}$$

where $\rho_i(k) = c_i(k)/c_i(0)$ = correlation coefficient at lag k , $k = 1, \dots, N-1$. The COV of \tilde{R}_i , denoted by δ_i , is then given by

$$\delta_i^2 = \frac{E[\tilde{R}_i - R_i]^2}{R_i^2} = \frac{\Delta_i^2}{N} (1 + \gamma_i) \quad (15)$$

where $\Delta_i = \sqrt{c_i(0)}/R_i$ is the COV of $h(\boldsymbol{\theta})$ when $\boldsymbol{\theta}$ is distributed as $p_i(\boldsymbol{\theta})$, and γ_i is a correlation factor

$$\gamma_i = 2 \sum_{k=1}^{N-1} \left(1 - \frac{k}{N}\right) \rho_i(k) \quad (16)$$

By estimating the covariance sequence $\{c_i(k): k=0, \dots, N-1\}$ from the Markov chain samples $\{\boldsymbol{\theta}_k^{(i)}: k=1, \dots, N\}$

$$c_i(k) \approx \frac{1}{N-k} \sum_{j=1}^{N-k} [h(\boldsymbol{\theta}_j^{(i)}) - \tilde{R}_i][h(\boldsymbol{\theta}_{j+k}^{(i)}) - \tilde{R}_i] \quad (17)$$

the COV Δ_i and the correlation sequence $\{\rho_i(k): k=1, \dots, N-1\}$ can be estimated. Consequently, γ_i in Eq. (16) and hence δ_i in Eq. (15) can be estimated, providing a means for assessing the variability of the estimate \tilde{R}_i using information from a single simulation run.

The term Δ_i^2/N in Eq. (15) is the familiar term for the square of the COV in MCS with N independent samples. The COV of \tilde{R}_i is thus equivalent to the COV in MCS with an effective number of independent samples $N/(1 + \gamma_i)$. The efficiency of the estimator using correlated samples of a Markov chain ($\gamma_i > 0$) is reduced compared to the case when the samples are uncorrelated ($\gamma_i = 0$), and smaller values of γ_i imply higher efficiency.

The result for δ_i in Eq. (15) is derived assuming that the proposal PDF is fixed in independent simulation runs. According to the proposed methodology, however, the proposal PDF is chosen as the kernel sampling density κ_i constructed using the Markov chain samples from the previous simulation level (except for the first simulation level where the prior PDF is used as the proposal PDF), and so κ_i is different for each independent simulation run. This gives rise to additional variability in \tilde{R}_i , and the actual COV of \tilde{R}_i will be greater than that given by Eq. (15). It can be argued

that this additional variability is of the order of the bias in \tilde{R}_i , and hence is often small. The numerical example shows that the COV predicted by Eq. (15) is quite close to the actual COV, and hence it can be used for assessing the variability of \tilde{R}_i .

Illustrative Example

The proposed methodology is applied to updating the response variance and reliability of a structure using its identified natural frequencies for two cases: locally identifiable and unidentifiable. The actual structure is a two-dimensional finite-element model of a two-story one-bay moment-resisting frame with a bay width of 8 m and equal story heights of 3 m. The columns have cross-sectional area of $18.8 \times 10^{-3} \text{ m}^2$ and moment of inertia $0.167 \times 10^{-3} \text{ m}^4$. The beams have cross-sectional area of $10.5 \times 10^{-3} \text{ m}^2$ and moment of inertia $0.562 \times 10^{-3} \text{ m}^4$. The elastic modulus and mass density of steel are assumed to be $2 \times 10^{11} \text{ N/m}^2$ and $7,850 \text{ kg/m}^3$, respectively. Lumped masses of $15 \times 10^3 \text{ kg}$ are put on the first and second floors. The natural frequencies of the first two modes are computed to be 3.16 and 9.67 Hz, respectively. Using noisy simulated response time histories, the identified natural frequencies are $\tilde{f}_1 = 3.13 \text{ Hz}$ and $\tilde{f}_2 = 9.83 \text{ Hz}$, which are used as the data \mathcal{D} in the updating.

Locally Identifiable Case: Uncertain Stiffness Parameters

A two-degree-of-freedom shear building model is used to model the moment-resisting frame structure in order to identify the interstory stiffnesses and to predict the response of the structure. The story masses are assumed to be 16.5×10^3 and $16.1 \times 10^3 \text{ kg}$ for the first and second stories, respectively, which are computed based on lumping half of the mass of the columns on the floors they are connected to. The interstory stiffnesses are parameterized as $k_1 = \bar{k}_1 \theta_1$ and $k_2 = \bar{k}_2 \theta_2$, where θ_1 and θ_2 are the stiffness parameters to be identified, and $\bar{k}_1 = \bar{k}_2 = 29.7 \times 10^6 \text{ N/m}$ are the nominal values for the interstory stiffnesses of the first and second stories, respectively. The nominal values are computed assuming the beams are rigid (shear building assumption). They do not correspond to the interstory stiffnesses of the actual structure, however, due to the flexibility of its beams.

The prior PDF p_0 for θ_1 and θ_2 is given by the product of two lognormal PDFs with most probable values (MPVs) 1.3 and 0.8 for θ_1 and θ_2 , respectively, and unit standard deviation for both θ_1 and θ_2 . These MPVs for the prior PDF reflect a 30% overestimation of the first-story stiffness and a 20% underestimation of the second-story stiffness relative to the nominal shear building model. The standard deviations for the prior PDF are chosen relatively large to express large initial uncertainty in the structural model.

Using the modal data \mathcal{D} , the updated PDF for the stiffness parameters $\boldsymbol{\theta} = [\theta_1, \theta_2]$ is formulated as (Vanik et al. 2000)

$$p_{\mathcal{D}}(\boldsymbol{\theta}) = c \exp[-J(\boldsymbol{\theta})/2\epsilon^2] p_0(\boldsymbol{\theta}) \quad (18)$$

where c = normalizing constant; and $J(\boldsymbol{\theta})$ = modal measure-of-fit function

$$J(\boldsymbol{\theta}) = \sum_{j=1}^2 \lambda_j^2 [f_j^2(\boldsymbol{\theta})/\tilde{f}_j^2 - 1]^2 \quad (19)$$

In Eqs. (18) and (19), ϵ/λ_j is the coefficient of variation of the prediction error between each \tilde{f}_j^2 and the corresponding model

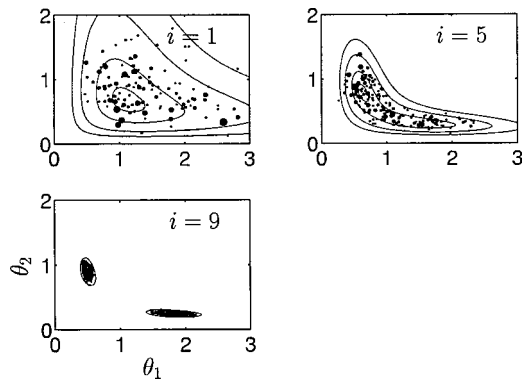


Fig. 1. Markov chain samples for θ_1 and θ_2 at simulation levels $i = 1, 5, 9$

squared frequency $f_j^2(\boldsymbol{\theta})$. The prediction error includes the uncertainty in the identified frequencies as well as the uncertainty in predicting the natural frequencies of the structure using the shear building model. We take $\lambda_1 = \lambda_2 = 1$ and explore the effect of different values for ϵ on the updated PDF for the model parameters and the corresponding updated robust reliability. Decreasing prediction error levels $\epsilon_i^2 = 1/2^{i-1}$ for successive simulation levels $i = 1, \dots, 9$ are investigated. The sequence of intermediate PDFs $\{p_i : i = 1, \dots, 9\}$ is constructed by successively substituting the sequence of values $\epsilon_i^2 = 1/2^{i-1}$ into Eq. (18).

Updated Joint Probability Density Function for Model Parameters θ_1 and θ_2

Fig. 1 shows the Markov chain samples at simulation levels $i = 1, 5, 9$, corresponding to $\epsilon = 1, 1/4, 1/16$ in Eq. (18). For each simulation level, after the first 10 Markov chain samples are ignored, $N = 300$ Markov chain samples are simulated, which are shown with dots in Fig. 1. Note that the Markov chain samples are not all distinct. To show the population of samples consistently, the area of the dots is shown proportional to the number of samples at each location. The numbers of distinct samples are 149, 187, and 142 for simulation levels $i = 1, 4$, and 9, respectively. The contour lines for each updated PDF are plotted at levels 0.01, 0.1, 0.5, and 0.9 relative to the peak value of the updated PDF in each figure. It can be seen that the area enclosed by the outermost contour (0.01), and hence the region of probability concentration of the updated PDF, diminishes as ϵ decreases. The samples populate well the important region of the updated PDFs. More importantly, the density of the population is consistent with the exact updated PDF.

The updated PDF bifurcates into two peaks as ϵ decreases because there are two “optimal” models that give the identified frequencies, showing that the class of model is only locally identifiable based on the modal data. In spite of this complexity, the samples still populate the important neighborhood of the two peaks, which are essentially disconnected from each other. This ability to populate samples in disconnected regions of significant probability density offers a robust and powerful way of locating the domain of influence of the optimal parameter points in Bayesian model updating problems, even when the model is not (globally) identifiable based on the available data, as in the present example. It should be noted that, if the Markov chain samples is generated using an adaptive proposal PDF, either one of the important regions at simulation level $i = 9$ is likely to be missed by the Markov chain samples, resulting in significant bias in the

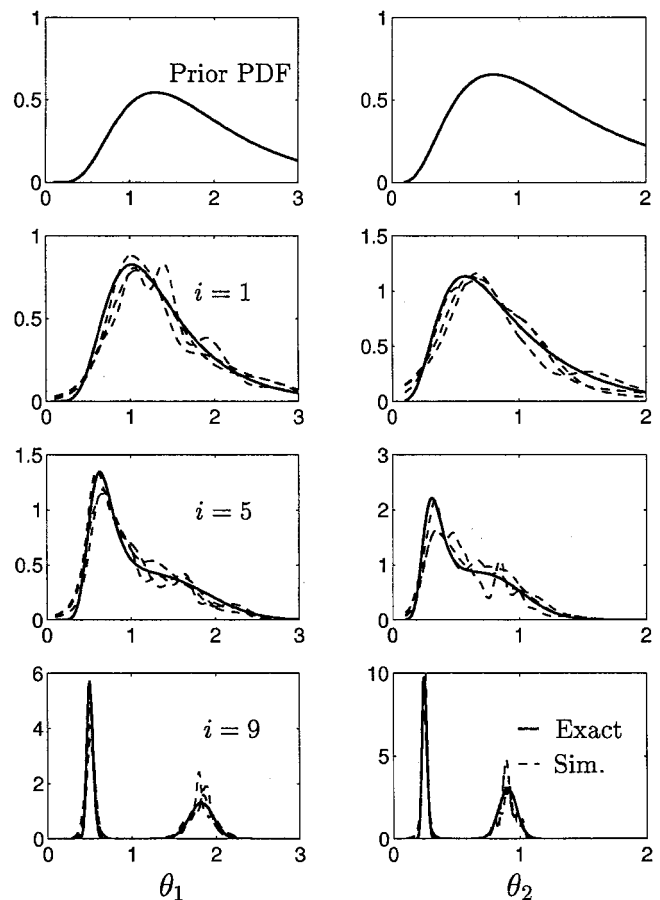


Fig. 2. Prior probability distribution function and updated marginal probability distribution functions for θ_1 and θ_2 at simulation levels $i = 1, 5, 9$

resulting estimators. The reason for this is that, in order to maintain a reasonable acceptance rate of the candidate state, the support size of the proposal PDF has to be similar to the size of the important region where the Markov chain is initiated. However, this results in a small chance of generating a sample lying in the other important region, since the spread of the proposal PDF is not large enough. In contrast, in the proposed strategy, the proposal PDF has significant probability content in the two important regions since it is constructed using the Markov chain samples from the previous simulation level ($i = 8$) which are clustered around the two regions. Consequently, the candidate states and hence the Markov chain samples at simulation level $i = 9$ are also clustered around the two important regions. Here, one sees the advantage of adapting from one simulation level to the next (using a kernel sampling density) rather than within each simulation level (using an adaptive proposal PDF) for updating robust reliability problems.

Updated Marginal Distributions

The marginal kernel PDFs for θ_1 and θ_2 constructed from the Markov chain samples are shown in Fig. 2. The corresponding marginal kernel CDFs are shown in Fig. 3. The prior PDFs and CDFs in Figs. 2 and 3, respectively, are included for reference. The results from three independent simulation runs are shown with dashed lines. Recall that the estimated marginal PDFs and CDFs are readily obtained from the joint kernel PDF without numerical integration. For comparison purposes, the exact mar-

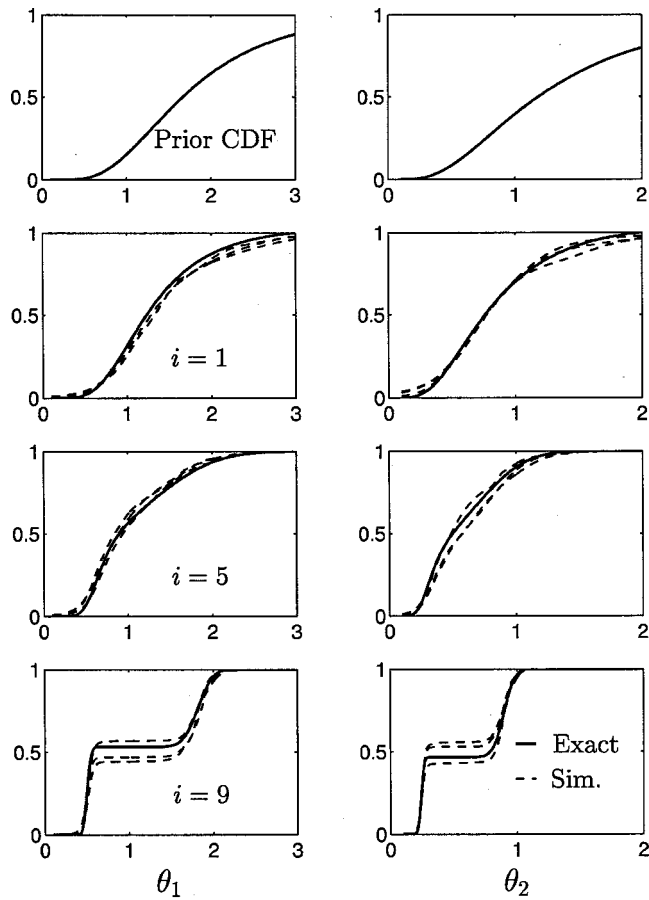


Fig. 3. Prior cumulative distribution function and updated marginal cumulative distribution functions for θ_1 and θ_2 at simulation levels $i=1, 5, 9$

ginal PDFs and CDFs obtained by numerical integration are also plotted with solid lines in Figs. 2 and 3, respectively. The results for the PDFs can be considered acceptable if fine detail is not required. The results for the CDFs shown in Fig. 3 exhibit a better match with the exact results, as the spurious noise in the PDFs is filtered out by integration in the CDFs. The errors in the marginal PDFs and CDFs come from two sources. The first source is that the finite number of Markov chain samples is not distributed exactly as the target PDF, which introduces bias in the estimates. The second is due to the approximate nature of the kernel PDF using a finite number of samples. Thus, even if the samples are exactly distributed like the target PDF, there will still be discrepancies between the kernel density estimates and the exact results. The dependent nature of the Markov chain samples in general does not bias the kernel PDF, but it does slow down the convergence of the kernel PDF compared to the case when the samples are independent.

Updated Robust Variance and Reliability

The response of the structure when it is subjected to stochastic excitations is predicted using the shear building model. Of interest are the interstory drift responses y_1 and y_2 of the first and second stories, respectively. The structure is assumed to be subjected to earthquake motion modeled by stationary Gaussian white noise with spectral intensity $S=1 \times 10^{-2} \text{ m}^2/\text{s}^3$. The stationary variances $\sigma_{y_1}^2(\boldsymbol{\theta})$ and $\sigma_{y_2}^2(\boldsymbol{\theta})$ of the interstory drifts for given model parameters $\boldsymbol{\theta}$ can be obtained by solving the

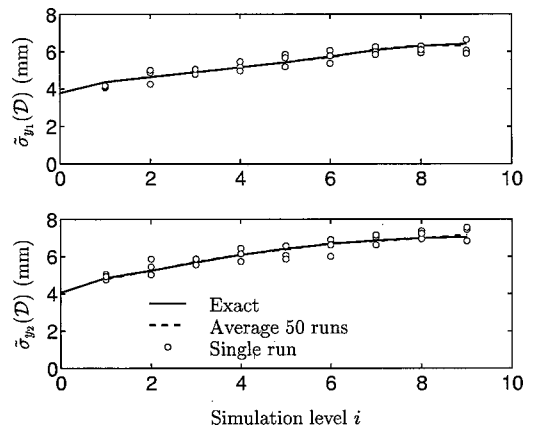


Fig. 4. Estimates $\tilde{\sigma}_{y_j}(\mathcal{D})$, $j=1, 2$, for updated robust drift standard deviations (uncertain stiffness)

Lyapunov equation (Soong and Grigoriu 1993). Using the updated PDF $p_{\mathcal{D}}$, the robust variances of the interstory drifts are given by Eq. (2) with $h(\boldsymbol{\theta}) = \sigma_{y_j}^2(\boldsymbol{\theta})$

$$\sigma_{y_j}^2(\mathcal{D}) = \int \sigma_{y_j}^2(\boldsymbol{\theta}) p_{\mathcal{D}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (20)$$

where $j=1, 2$ for the first and second stories, respectively.

The robust failure probability that the interstory drift of the j th story exceeds the threshold level $b=15 \text{ mm}$ within a duration of $T=10 \text{ s}$ is also computed

$$P(F_j|\mathcal{D}) = \int P(F_j|\boldsymbol{\theta}) p_{\mathcal{D}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \quad (21)$$

where $P(F_j|\boldsymbol{\theta})$ is the first excursion probability for given model parameters $\boldsymbol{\theta}$. Assuming the out-crossing events follow a Poisson process, $P(F_j|\boldsymbol{\theta})$ can be approximated by (Soong and Grigoriu 1993)

$$P(F_j|\mathcal{D}) = 1 - \exp[-2v_j(\boldsymbol{\theta})T] \quad (22)$$

where $v_j(\boldsymbol{\theta})$ is the up-crossing rate for given model parameters $\boldsymbol{\theta}$, given by Rice's formula

$$v_j(\boldsymbol{\theta}) = \frac{\sigma_{y_j}(\boldsymbol{\theta})}{2\pi\sigma_{y_j}(\boldsymbol{\theta})} \exp\left(-\frac{b^2}{2\sigma_{y_j}(\boldsymbol{\theta})^2}\right) \quad (23)$$

The estimates for updated robust standard deviations $\tilde{\sigma}_{y_j}(\mathcal{D})$ ($j=1,2$) are shown in Fig. 4 and the estimates for the updated robust failure probabilities $\tilde{P}(F_j|\mathcal{D})$ ($j=1,2$) are shown in Fig. 5. In these figures, three sample estimates, corresponding to the same simulation runs as in Figs. 2 and 3, are shown with circles. The exact results obtained by numerical integration are shown with solid lines. Simulation level $i=0$ refers to the case when no data are available and the robust response quantities are computed based on the prior PDF p_0 only, that is, by Eq. (2) with p_0 replacing $p_{\mathcal{D}}$. To investigate the bias of the simulation results, the average of the estimates over 50 independent simulation runs is computed and shown with dashed lines in the figures. In both Figs. 4 and 5, the results from averaging 50 simulation runs (dashed lines) almost overlap with the exact results (solid lines), showing that the bias from the initial transient stage of the simulated Markov chains is negligible, and the estimates are practically unbiased.

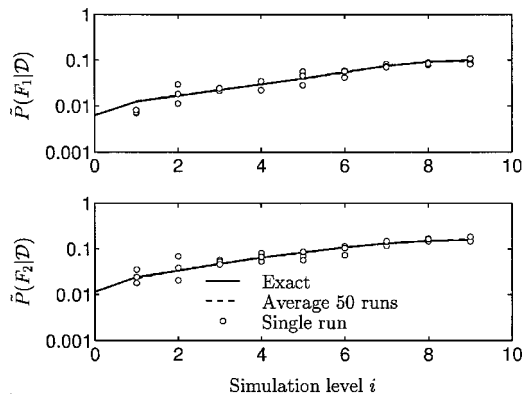


Fig. 5. Estimates $\tilde{P}(F_j|D)$, $j=1, 2$, for updated robust failure probabilities (uncertain stiffness)

Variability of Estimates

The sample COV of the updated robust drift variance and failure probability estimates computed from the 50 independent runs are shown in Figs. 6 and 7. Also, the COV estimates based on Eq. (15), which does not include the variability of the proposal PDF, are computed and shown in these figures: the COV estimates for the three simulation runs in Figs. 2 and 3 are shown with circles, while the averages of the COV estimates based on Eq. (15) over 50 simulation runs are shown with dashed lines. From Figs. 6 and 7, it can be seen that the actual COV of the robust drift variance and failure probability estimates (solid lines) are quite close to those predicted by Eq. (15) (circles and dashed lines), indicating that the additional variability due to the randomness of the proposal PDF is negligible and Eq. (15) is useful for assessing the variability of the updated robust estimates.

Unidentifiable Case: Uncertain Stiffness and Mass Parameters

Next, the story masses are also updated, in addition to the stiffnesses. The story masses are parametrized as $m_1 = \theta_3 \bar{m}_1$ and $m_2 = \theta_4 \bar{m}_2$, where θ_3 and θ_4 are the mass parameters to be updated, and $\bar{m}_1 = 16.5 \times 10^3$ kg and $\bar{m}_2 = 16.1 \times 10^3$ kg are the nominal values for the first- and second-story masses, respectively. There are thus four parameters (two stiffness parameters θ_1 and θ_2 , and two mass parameters θ_3 and θ_4) to be updated. In this example, the class of models is unidentifiable based on the data because

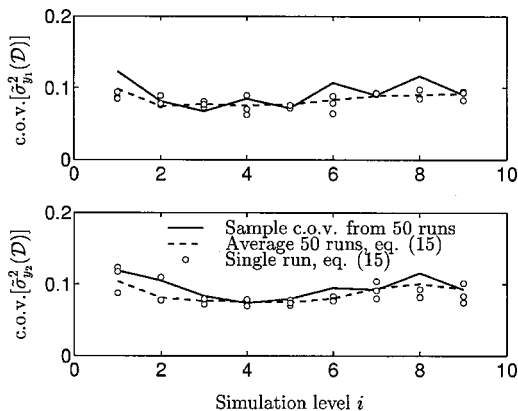


Fig. 6. Coefficient of variation of updated robust drift variance estimates (uncertain stiffness)

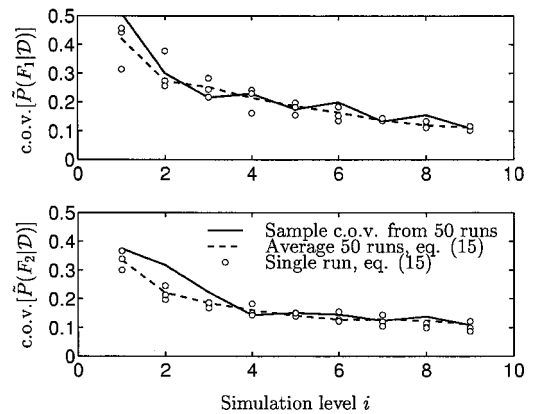


Fig. 7. Coefficient of variation of updated robust failure probability estimates (uncertain stiffness)

there is an infinite set of “optimal” mass and stiffness parameter values that give the two identified frequencies.

The prior PDF p_0 for θ_1 to θ_4 is given by the product of four lognormal PDFs with MPVs 1.3, 0.8, 0.95, 0.95, and standard deviations 1, 1, 0.1, 0.1, respectively. Relatively small values of the standard deviations are assumed for the mass parameters to reflect that, in practice, they can usually be determined more accurately from the structural drawings than the stiffness parameters.

Fig. 8 shows the Markov chain samples at simulation levels $i=1, 5, 9$ corresponding to $\epsilon=1, 1/4, 1/16$ in Eq. (18). At each level, $N=500$ Markov chain samples are simulated. The four components of the samples are shown in two groups: θ_2 versus θ_1 in the first column and θ_4 versus θ_3 in the second column of Fig. 8. The distribution of the samples for the stiffness parameters as shown in the first column of the figure is qualitatively similar to that shown in Fig. 1 but with more spread. There is no significant pattern in the distribution of the mass parameters as shown in the second column of Fig. 8, as a result of the small uncertainty assumed in their prior distribution. In fact, the prior distributions on the mass and stiffness parameters serve as a “regularizer” for this unidentifiable problem (in the sense of Tikhonov and Arsenin 1977).

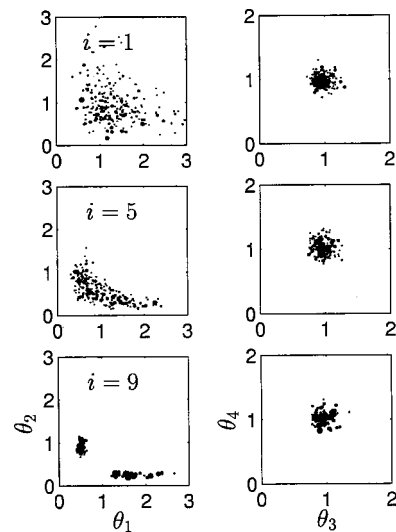


Fig. 8. Markov chain samples for θ_1 to θ_4 at simulation levels $i=1, 5, 9$

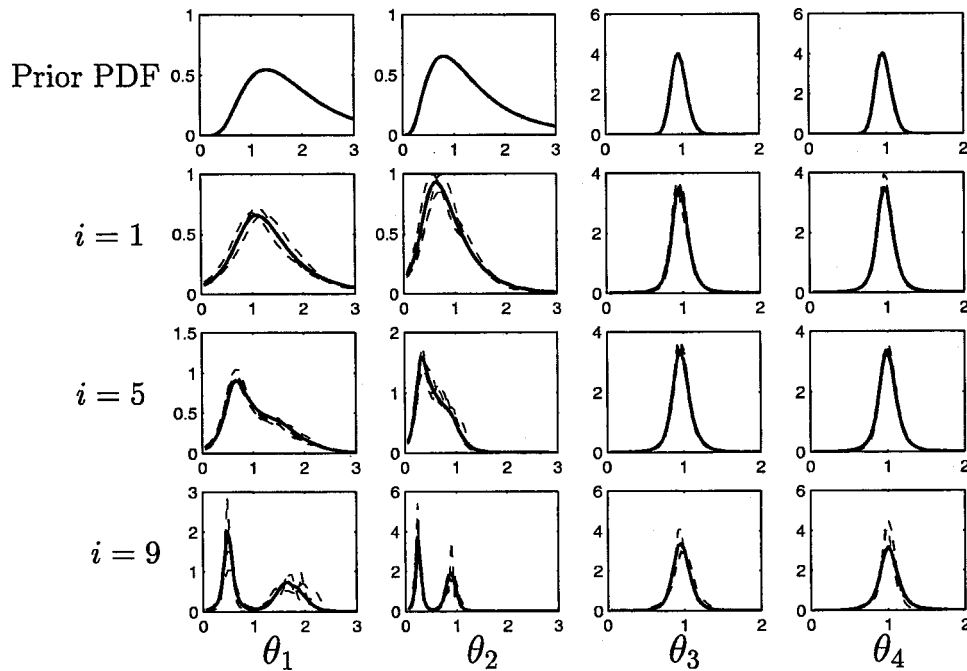


Fig. 9. Prior probability distribution function and updated marginal probability distribution functions for θ_1 to θ_4 at simulation levels $i=1, 5, 9$ (solid lines: average over 50 runs; dashed lines: three runs)

The marginal kernel PDFs for θ_1 to θ_4 constructed using the Markov chain samples are shown in Fig. 9. The corresponding marginal kernel CDFs are shown in Fig. 10. Calculation of the exact marginal PDFs and CDFs involves three-dimensional numerical integration, and was not attempted. Instead, the sample average over 50 simulation runs of the marginal kernel distributions are computed and are shown as solid lines in Figs. 9 and 10.

Finally, the estimates for the updated failure probabilities of the interstory drifts for the first and second stories are computed using the Markov chain samples. The results for the estimates are

shown in Fig. 11 and their COV is shown in Fig. 12. These results demonstrate the robustness of the proposed method to problems of higher dimensions.

Conclusions

The usual practice in system identification is to use system data to identify one model from a set of possible models and then to use this model for predicting system behavior. In contrast, the present

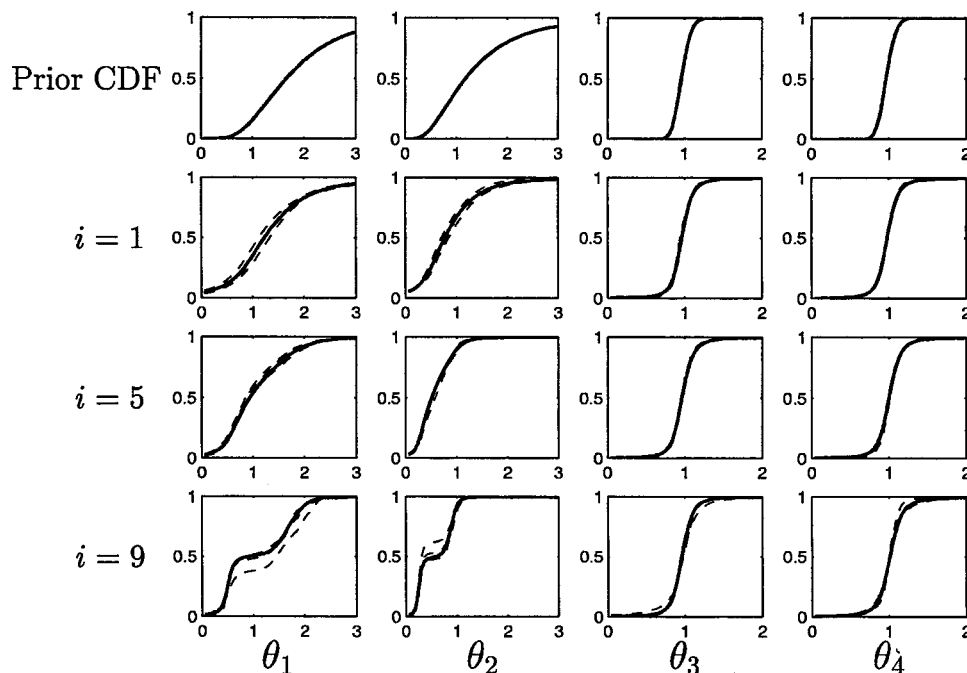


Fig. 10. Prior cumulative distribution function and updated marginal cumulative distribution functions for θ_1 to θ_4 at simulation levels $i=1, 5, 9$ (solid lines: average over 50 runs; dashed lines: three runs)

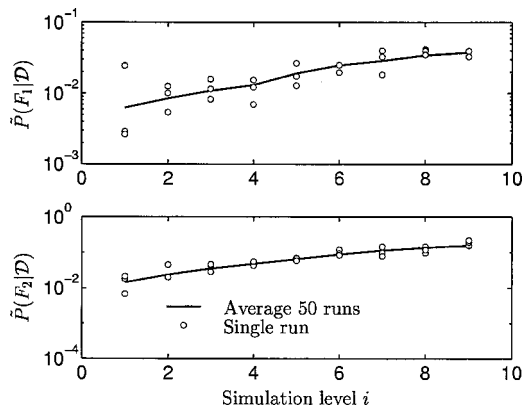


Fig. 11. Estimates $\tilde{P}(F_j|\mathcal{D})$, $j = 1, 2$, for updated robust failure probabilities (uncertain stiffness and mass)

robust predictive approach rigorously combines the predictions of all models, appropriately weighted by their updated probabilities based on the data. This approach works regardless of whether the class of models is identifiable for not.

The updated predictions can give useful information about the expected performance of a structure. However, the integrals involved are difficult to evaluate because they are generally multi-dimensional, and the updated PDF for the model parameters is known only up to a scaling constant and concentrated in a small region of the parameter space. Conventional simulation methods such as standard Monte Carlo or importance sampling simulation are found to be inapplicable. Markov chain simulation based on direct application of the Metropolis-Hastings algorithm is not applicable either, since it is difficult to choose a suitable proposal distribution in the absence of knowledge about the region of concentration of the updated PDF for the model parameters.

The proposed methodology combines the Metropolis-Hastings algorithm with a concept similar to that used in simulated annealing: it uses successive simulation levels to adapt to the region of probability concentration of the updated PDF. Theoretical and numerical investigations show that the Markov chain samples populate the region of concentration of the updated PDF in a probabilistically correct manner. The resulting robust reliability and response variance integrals estimated using the Markov chain

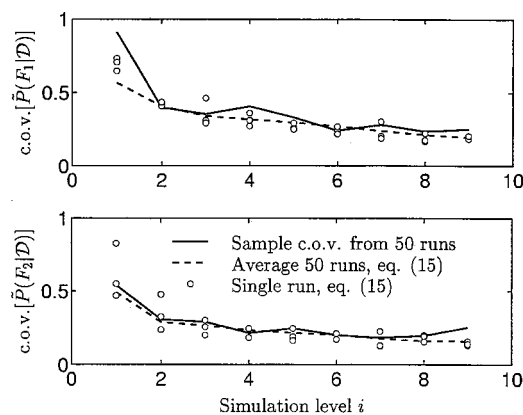


Fig. 12. Coefficient of variation of updated robust failure probability estimates (uncertain stiffness and mass)

samples are practically unbiased, and their variability can be consistently estimated by an available formula using information from a single simulation run.

The success of the proposed methodology relies on a good choice of the intermediate PDFs and their approximation by kernel PDFs. Future research may focus on developing better intermediate PDFs as well as enhancing the performance of the kernel PDF approximations.

Acknowledgments

This paper is based upon work partly supported by the Pacific Earthquake Engineering Research Center under National Science Foundation Cooperative Agreement No. CMS-9701568. This support is gratefully acknowledged.

References

- Ang, G. L., Ang, A. H.-S., and Tang, W. H. (1992). "Optimal importance sampling density estimator." *J. Eng. Mech.*, 118(6), 1146–1163.
- Au, S. K., and Beck, J. L. (1999). "A new adaptive importance sampling scheme." *Struct. Safety*, 21, 135–158.
- Beck, J. L. (1989). "Statistical system identification of structures." *Proc., 5th Int. Conf. Structural Safety and Reliability*, ASCE, New York, 1395–1402.
- Beck, J. L., and Katafygiotis, L. S. (1998). "Updating models and their uncertainties. Part I: Bayesian statistical framework." *J. Eng. Mech.*, 124(4), 455–461.
- Box, G. E. P., and Tiao, G. C. (1992). *Bayesian inference in statistical analysis*, Wiley, New York.
- Collins, J. D., Hart, G. C., Hasselman, T. K., and Kennedy, B. (1974). "Statistical identification of structures." *AIAA J.*, 12, 185–190.
- Cox, R. T. (1961). *The algebra of probable inference*, Johns Hopkins, Baltimore.
- Doob, J. L. (1953). *Stochastic processes*, Wiley, New York.
- Engelund, S., and Sorensen, J. D. (1998). "A probabilistic model for chloride-ingress and initiation of corrosion in reinforced concrete structures." *Struct. Safety*, 20(1), 69–89.
- Enright, M., and Frangopol, D. M. (1999). "Condition prediction of deteriorating concrete bridges." *J. Struct. Eng.*, 125(10), 1118–1125.
- Fishman, G. S. (1996). *Monte Carlo: Concepts, algorithms, and applications*, Springer-Verlag, New York.
- Geyskens, P., Der Kiureghian, A., and Monteiro, P. (1998). "Bayesian prediction of elastic modulus of concrete." *J. Struct. Eng.*, 124(1), 89–95.
- Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57, 97–109.
- Jaynes, E. T. (1978). "Where do we stand on maximum entropy?" *The maximum entropy formalism*, R. D. Levine and M. Tribus, eds., MIT, Cambridge, Mass.
- Katafygiotis, L. S., and Beck, J. L. (1998). "Updating models and their uncertainties—model identifiability." *J. Eng. Mech.*, 124(4), 463–467.
- Katafygiotis, L. S., Papadimitriou, C., and Lam, H. F. (1998). "A probabilistic approach to structural model updating." *Soil Dyn. Earthquake Eng.*, 17(7-8), 495–507.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., and Teller, A. H. (1953). "Equations of state calculations by fast computing machines." *J. Chem. Phys.*, 21(6), 1087–1092.
- Papadimitriou, C., Beck, J. L., and Au, S. K. (2000). "Entropy-based optimal sensor location for structural model updating." *J. Vib. Control*, 6(5), 781–800.

- Papadimitriou, C., Beck, J. L., and Katafygiotis, L. S. (2001). "Updating robust reliability using structural test data." *Probab. Eng. Mech.*, 16(2), 103–113.
- Rubinstein, R. Y. (1981). *Simulation and the Monte-Carlo method*, Wiley, New York.
- Silverman, B. W. (1986). *Density estimators*, Chapman and Hall, New York.
- Singhal, A., and Kiremidjian, A. S. (1998). "Bayesian updating of fragilities with application to RC frames." *J. Struct. Eng.*, 124(8), 922–929.
- Sivia, D. S. (1996). *Data analysis: A Bayesian tutorial*, Clarendon, Oxford.
- Soong, T. T., and Grigoriu, M. (1993). *Random vibration of mechanical and structural systems*, Prentice Hall, Englewood Cliffs, N.J.
- Tikhonov, A. N., and Arsenin, V. Y. (1977). *Solution of ill-posed problems*, Wiley, New York.
- Törn, A., and Zilinskas, A. (1989). *Global optimization, Vol. 350 of Lecture Notes in Computer Science*, Springer-Verlag, Berlin.
- Vanik, M. W., Beck, J. L., and Au, S. K. (2000). "Bayesian probabilistic approach to structural health monitoring." *J. Eng. Mech.*, 126(7), 738–745.
- Yang, C. M., and Beck, J. L. (1998). "Generalized trajectory methods for finding multiple extrema and roots of functions." *J. Optim. Theory Appl.*, 97(4), 211–227.
- Zheng, R. H., and Ellingwood, B. R. (1998). "Role of non-destructive evaluation in time-dependent reliability analysis." *Struct. Safety*, 20(4), 325–339.