Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging

Jianye Ching[1] and Yi-Chu Chen[2]

Abstract: This paper presents a newly developed simulation-based approach for Bayesian model updating, model class selection, and model averaging called the transitional Markov chain Monte Carlo (TMCMC) approach. The idea behind TMCMC is to avoid the problem of sampling from difficult target probability density functions (PDFs) but sampling from a series of intermediate PDFs that converge to the target PDF and are easier to sample. The TMCMC approach is motivated by the adaptive Metropolis–Hastings method developed by Beck and Au in 2002 and is based on Markov chain Monte Carlo. It is shown that TMCMC is able to draw samples from some difficult PDFs (e.g., multimodal PDFs, very peaked PDFs, and PDFs with flat manifold). The TMCMC approach can also estimate evidence of the chosen probabilistic model class conditioning on the measured data, a key component for Bayesian model class selection and model averaging. Three examples are used to demonstrate the effectiveness of the TMCMC approach in Bayesian model updating, model class selection, and model averaging.

DOI: 10.1061/(ASCE)0733-9399(2007)133:7(816)

CE Database subject headings: Bayesian analysis; Simulation; Markov chains; Monte Carlo method.

Introduction

Model updating refers to the methodology that determines the most plausible model for an instrumented civil engineering system given its measured response and, possibly, its excitation. In recent years, civil engineers have paid much attention to model updating techniques since they have broad applications in structural health monitoring (Natke and Yao 1988; Hjelmstad and Shin 1997; Sanayei et al. 1999; Chang 2001; Beck et al. 2001; Casciati 2002; Bernal et al. 2002). Among the model updating techniques, Bayesian model updating techniques (Beck and Katafygiotis 1998; Katafygiotis et al. 1998; Vanik et al. 2000; Kerschen et al. 2003; Yuen et al. 2004; Ching and Beck 2004) do not just find a single plausible model but a set of models whose predictions are weighted by the probabilities of these models conditional on the measured data. Due to their ability to consider more than one model, Bayesian model updating techniques are robust and suitable to characterize modeling uncertainties.

The problem of model updating is potentially ill-posed—that is, there may be more than one optimal solution (Katafygiotis and Beck 1998). The problem becomes even more challenging when only some of the degrees of freedom (DOF) of the model are measured and when modeling errors are explicitly acknowledged. A previous Bayesian model updating approach (Beck and Katafygiotis 1998) has been successful in resolving the aforementioned difficulties when the updated (posterior) probability density function (PDF) of the model parameters is very peaked (e.g., the amount of data is sufficiently large) so that an asymptotic approximation of the Bayesian predictive integrals is accurate. However, when the PDF is not very peaked (e.g., limited data or a flat region in the PDF), the validity of the asymptotic approximation is questionable.

On the other hand, stochastic simulation can handle more general cases than the asymptotic approximation approach. However, many stochastic simulation methods, including importance sampling (IS) and Metropolis–Hastings (MH) algorithm [a special case of Markov chain Monte Carlo (MCMC)], are often inefficient for practical Bayesian model updating problems: IS is not efficient for high-dimensional PDFs [Au and Beck (2003) have discussed this issue in the context of reliability evaluation]; the MH algorithm may be not efficient when the uncertain variables are highly correlated conditioning on the data and when the posterior PDF is very peaked (Beck and Au 2002). More seriously, most MCMC techniques are not applicable when the PDF is multimodal. It is highly desirable to have a stochastic simulation approach that is effective in all cases (e.g., high-dimensional PDFs, multimodal PDFs, very peaked PDFs, PDFs with flat regions, etc.), since the properties of the posterior PDF are unknown beforehand. However, such a solution is not yet available in the Bayesian model updating literature.

An attempt has been made by Beck and Au (2002), where they proposed an adaptive Metropolis–Hastings (AMH) method. The key idea behind their approach is to avoid directly sampling from the difficult target PDF but to sample from a series of simpler intermediate PDFs that converge to the target PDF. In principle,

[1]Assistant Professor, Dept. of Construction Engineering, National Taiwan Univ. of Science and Technology, Taiwan. E-mail: jyching@gmail.com

[2]Ph.D. Student, Dept. of Construction Engineering, National Taiwan Univ. of Science and Technology, Taiwan. E-mail: yiejue@gmail.com

Note. Associate Editor: Arvid Naess. Discussion open until December 1, 2007. 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 August 10, 2005; approved on December 14, 2006. This paper is part of the ***Journal of Engineering Mechanics***, Vol. 133, No. 7, July 1, 2007. ©ASCE, ISSN 0733-9399/2007/7-816–832/$25.00.

J. Eng. Mech., 2007, 133(7): 816-832

their approach is applicable to very peaked, flat, and multimodal posterior PDFs. However, a major limitation is that the approach is inefficient for high-dimensional problems because in the approach, kernel density estimation is required, which is inefficient for high-dimensional problems.

In this paper, it is shown that the key idea of AMH can be implemented without kernel density estimation. The resulting new approach is called the transitional Markov chain Monte Carlo (TMCMC) approach. It uses the same "transitional" intermediate PDF idea proposed by Beck and Au (2002) to avoid directly sampling from difficult PDFs, but when it applies MCMC, kernel density estimation is not necessary. This approach inherits the advantages of AMH (i.e., applicable to very peaked, flat, and multimodal PDFs), while it is more efficient for high-dimensional PDFs. Moreover, the TMCMC approach is capable of automatically selecting the intermediate PDFs.

TMCMC is able to evaluate the evidence for the assumed model class as a by-product, which is the key to Bayesian model class selection (Beck and Yuen 2004) and Bayesian model averaging (Hoeting et al. 1999). Bayesian model class selection and model averaging is one step ahead of Bayesian model updating: rather than assuming a single model class, several model classes are chosen, and their relative plausibilities are evaluated. Bayesian model class selection and model averaging answer an important research question that cannot be answered by common model updating methods: how do we choose the best model class among the chosen model classes and make predictions based on all the model classes?

The structure of this paper is as follows. First, the simulation-based Bayesian model updating, model class selection, and model averaging problems are defined, and the difficulties that are often encountered are addressed. Second, the TMCMC approach is introduced, and the statistical properties of the resulting estimators will be derived. Third, the effectiveness of the TMCMC approach on Bayesian model updating, model class selection, and model averaging is demonstrated through three examples: a mixture of Gaussians, a 2DOF shear building system, and a 4DOF shear building system. Finally, a discussion of the advantages and limitations of the new approach will be given.

Difficulties Encountered in Simulation-Based Bayesian Model Updating, Model Class Selection, and Model Averaging

As mentioned previously, many stochastic simulation methods cannot be efficiently applied to practical Bayesian model updating, model class selection, and model averaging problems. In this section, we discuss the difficulties and the possible solutions. Before the discussion, we first define the simulation-based Bayesian model updating, model class selection, and model averaging problems.

Problem Definition

Let M be the assumed probabilistic model class for the target system, θ be the uncertain model parameters, and D be the measured data from the system. The goal of simulation-based Bayesian model updating is to sample from the posterior PDF of θ conditioned on D

$$f(\theta|M,D)=\frac{f(D|M,\theta)\cdot f(\theta|M)}{f(D|M)}=\frac{f(D|M,\theta)\cdot f(\theta|M)}{\int f(D|M,\theta)\cdot f(\theta|M)\cdot d\theta} \quad (1)$$

where $f(\theta|M)$=prior PDF of θ; $f(D|M,\theta)$=likelihood; and $f(D|M)$ is called the evidence of M.

The goal of Bayesian model class selection is quite different. Given the chosen candidate probabilistic model classes $\{M^{(i)}: i=1,\ldots,N_{\text{class}}\}$, the goal of Bayesian model class selection is to calculate $P[M^{(i)}|D]$

$$P[M^{(i)}|D]=\frac{f[D|M^{(i)}]P[M^{(i)}]}{\sum_{i=1}^{N_{\text{class}}} f[D|M^{(i)}]P[M^{(i)}]} \quad (2)$$

where $P[M^{(i)}]$=prior probability of $M^{(i)}$. Notice that the calculation of $P[M^{(i)}|D]$ requires the determination of the evidence $\{f[D|M^{(i)}]: i=1,\ldots,N_{\text{class}}\}$. Once Bayesian model class selection is achieved, Bayesian model averaging is trivial because any quantity of interest g conditioning on the data and all the chosen model classes can be estimated using the following equation:

$$E(g|D)=\sum_{i=1}^{N_{\text{class}}} E[g|M^{(i)},D]\cdot P[M^{(i)}|D] \quad (3)$$

Difficulties for Simulation-Based Bayesian Model Updating

Simulation-based methods are valuable for Bayesian model updating because they are useful in obtaining samples from $f(\theta|M,D)$, using which we can estimate any quantity of interest $E(g|M,D)$ according to the Law of Large Number

$$E(g|M,D)\approx\frac{1}{N}\sum_{k=1}^{N} g(\theta_k) \quad (4)$$

where $\{\theta_k: k=1,\ldots,N\}$=samples from $f(\theta|M,D)$. However, many stochastic simulation methods are not efficient in drawing samples from $f(\theta|M,D)$, as discussed hereafter.

It is not trivial to apply IS to estimate $E(g|M,D)$ because its efficiency highly depends on the chosen important sampling PDF, denoted by IS PDF: the efficiency is high if the chosen IS PDF is roughly proportional to $g(\theta)\cdot f(\theta|M,D)$. This implies that the proper choice of the IS PDF relies on our knowledge on the $f(\theta|M,D)$ geometry, which is usually unknown beforehand. Furthermore, IS is not applicable to problems where the θ dimension is high.

It is, in general, not trivial to apply the MH algorithm to draw samples from $f(\theta|M,D)$ because its efficiency highly depends on the chosen proposal PDF, and the proper choice of the proposal PDF, again, relies on our understanding of the $f(\theta|M,D)$ geometry. This issue is aggravated when the uncertain parameters are highly correlated conditioning the data and when $f(\theta|M,D)$ is very peaked. For the latter circumstance, it is difficult to implement MH because a dilemma exists: if the proposal PDF is wide, it is likely that the very peaked region of $f(\theta|M,D)$ will never be reached; even if the very peaked region is reached, the sequential rejection rate will be high since it is difficult to find candidates better than the current sample. On the contrary, if the proposal PDF is narrow, the Markov chain will travel very slowly before reaching the very peaked region.

J. Eng. Mech., 2007, 133(7): 816-832

An even more difficult situation may arise for most MCMC methods when $f(\theta|M,D)$ is multimodal (i.e., it has more than one isolated peak). For most MCMC methods, sampling from a multimodal $f(\theta|M,D)$ may result in a nonergodic Markov chain (i.e., the Markov chain samples may be trapped in one of the peaks and never propagate to the others). For IS, it is highly nontrivial to find an IS PDF that is easy to sample and, at the same time, close to the multimodal target PDF.

Difficulties for Bayesian Model Class Selection

The estimation of $P[M^{(i)}|D]$ is also highly nontrivial. As seen in Eq. (2), the evaluation of $P[M^{(i)}|D]$ requires the knowledge of evidence if $\{f[D|M^{(i)}]: i=1,\ldots,N_{\text{class}}\}$, where

$$f[D|M^{(i)}] = \int f[D|M^{(i)},\theta^{(i)}] \cdot f[\theta^{(i)}|M^{(i)}] \cdot d\theta^{(i)} \tag{5}$$

where $\theta^{(i)}$ contains the uncertain model parameters in $M^{(i)}$. It seems easy to estimate $f[D|M^{(i)}]$ because if we obtain N_p samples of $\theta^{(i)}$ from the prior PDF $f(\theta^{(i)}|M^{(i)})$, according to the Law of Large Number

$$f[D|M^{(i)}] \approx \frac{1}{N_p}\sum_{k=1}^{N_p} f(D|M^{(i)},\theta_k^{(i)}) \tag{6}$$

where $\theta_k^{(i)}=k^{th}$ sample from $f[\theta^{(i)}|M^{(i)}]$. However, the main support region of $f[\theta^{(i)}|M^{(i)}]$ is usually very different from the high likelihood region of $f[D|M^{(i)},\theta^{(i)}]$. The consequence is that the variance of the estimator in Eq. (6) can be extremely large.

An approach proposed by Ching et al. (2005) avoids this issue by considering the following equality:

$$\ln\{f[D|M^{(i)}]\} = \int \ln\{f[D|M^{(i)},\theta]f[\theta^{(i)}|M^{(i)}]\}f[\theta^{(i)}|M^{(i)},D]d\theta^{(i)} + H\{f[\theta^{(i)}|M^{(i)},D]\} \tag{7}$$

where

$$H\{f[\theta^{(i)}|M^{(i)},D]\} = -\int \ln\{f[\theta^{(i)}|M^{(i)},D]\}f[\theta^{(i)}|M^{(i)},D]d\theta^{(i)} \tag{8}$$

is the differential entropy of $f[\theta^{(i)}|M^{(i)},D]$. Therefore

$$\ln\{f[D|M^{(i)}]\} \approx \frac{1}{N_p}\sum_{k=1}^{N_p} \ln\{f[D|M^{(i)},\theta_k^{(i)*}]f[\theta_k^{(i)*}|M^{(i)}]\} + \hat{H}\{f[\theta^{(i)}|M^{(i)},D]\} \tag{9}$$

where $\theta_k^{(i)*}=k$th sample from $f[\theta^{(i)}|M^{(i)},D]$; $\hat{H}\{f[\theta^{(i)}|M^{(i)},D]\}$ =estimated entropy of $f[\theta^{(i)}|M^{(i)},D]$. Note that the estimation of the first term at the right-hand side of the equation is usually robust since the main support region of $f[\theta^{(i)}|M^{(i)},D]$ is the same as the important region of $\ln\{f[D|M^{(i)},\theta^{(i)}]f[\theta^{(i)}|M^{(i)}]\}$. Therefore, this approach works reasonably well for most cases; however, it requires $f[\theta^{(i)}|M^{(i)},D]$ samples, which, in turn, requires an efficient Bayesian model updating algorithm. There are also asymptotic approaches of estimating $f[D|M^{(i)}]$ (Beck and Yuen 2004). These approaches are accurate when the amount of data is large but not accurate when the data are little.

Transitional Markov Chain Monte Carlo

In this paper, the novel method TMCMC is proposed to solve Bayesian model updating, model class selection, and model averaging problems. The new method has the following features: (1) it is based on MCMC, but unlike most MCMC methods, it is applicable to multimodal $f(\theta|M,D)$; (2) it is applicable to both very peaked and flat $f(\theta|M,D)$; (3) it is motivated by AMH proposed by Beck and Au (2002), but TMCMC is more efficient when θ dimension is higher; it can also automatically select the intermediate PDFs; and (3) it can estimate the evidence $f(D|M)$ as a by-product. The idea behind our new approach will be described in detail in this section. The entire algorithm will be formally presented in the next section.

Let us first consider the following equation:

$$f(\theta|M,D) \propto f(\theta|M) \cdot f(D|M,\theta) \tag{10}$$

As mentioned before, sampling from $f(\theta|M,D)$ using IS and MH can be difficult, mainly because the geometry of the likelihood $f(D|M,\theta)$ cannot be fully understood beforehand. The idea proposed by Beck and Au (2002) is to construct a series of intermediate PDFs that converge to the target PDF $f(\theta|M,D)$ from the prior PDF $f(\theta|M)$. In this paper, the same idea is implemented, but a universal way of constructing such intermediate PDFs is proposed as follows. Let us now consider a series of intermediate PDFs

$$f_j(\theta) \propto f(\theta|M) \cdot f(D|M,\theta)^{p_j}$$
$$j=0,\ldots,m \quad 0=p_0<p_1<\cdots<p_m=1 \tag{11}$$

where the index j denotes the stage number. As one can see, this construction fulfills the desirable properties: the series of intermediate PDFs starts from the prior PDF [note that $f_0(\theta)=f(\theta|M)$] and ends with $f(\theta|M,D)$ [note that $f_m(\theta)=f(\theta|M,D)$]. The idea is that although the geometry change from $f(\theta|M)$ to $f(\theta|M,D)$ can be dramatic (this dramatic change causes most of the difficulties), the change between two adjacent intermediate PDFs can be small. This small change makes it possible to efficiently obtain samples from $f_{j+1}(\theta)$ based on samples from $f_j(\theta)$.

This intermediate PDF idea was first proposed by Beck and Au (2002), where they proposed to use the $f_j(\theta)$ samples to estimate the PDF itself as a kernel density function (KDF), a mixture of weighted Gaussians centered at the samples. The resulting KDF is taken as the proposal PDF of the MH algorithm to draw samples from $f_{j+1}(\theta)$. Doing so sequentially will ultimately give us the $f(\theta|M,D)$ samples. This algorithm is the so-called AMH method. Unfortunately, since the employed proposal PDF is fixed (i.e., the KDF), rendering the MH algorithm is akin to importance sampling, and therefore not efficient in high dimension. Moreover, the KDF approach is not efficient for high-dimensional PDFs. Therefore, AMH can only effectively solve low-dimensional problems (from our experience, dimensions less than 5). Finally, in the AMH framework, it is not clear how to systematically select the intermediate PDFs.

The TMCMC algorithm takes a completely different strategy to obtain $f_{j+1}(\theta)$ samples based on $f_j(\theta)$ samples: a resampling approach is employed rather than the KDF approach. Since resampling is more robust against the θ dimension, the TMCMC can handle higher-dimensional problems. The TMCMC algorithm consists of a series of resampling stages, with each stage doing the following: given N_j samples from $f_j(\theta)$, denoted by $\{\theta_{j,k}: k=1,\ldots,N_j\}$, obtain samples from $f_{j+1}(\theta)$, denoted by $\{\theta_{j+1,k}: k=1,\ldots,N_{j+1}\}$. This can be done easily by the following:

J. Eng. Mech., 2007, 133(7): 816-832

with the samples $\{\theta_{j,k}:k=1,\ldots,N_j\}$ from $f_j(\theta)$, we can compute the "plausibility weights" of these samples with respect to $f_{j+1}(\theta)$

$$w(\theta_{j,k}) = \frac{f(\theta_{j,k}|M)f(D|M,\theta_{j,k})^{p_{j+1}}}{f(\theta_{j,k}|M)f(D|M,\theta_{j,k})^{p_j}} = f(D|M,\theta_{j,k})^{p_{j+1}-p_j}$$
$$k=1,\ldots,N_j \quad (12)$$

Now we resample the uncertain parameters according to the normalized weights, i.e., let

$$\theta_{j+1,k} = \theta_{j,l} \quad \text{w.p.} \quad \frac{w(\theta_{j,l})}{\sum_{l=1}^{N_j} w(\theta_{j,l})} \quad k=1,\ldots,N_{j+1} \quad (13)$$

where w.p. stands for "with probability" and l=dummy index. It can be shown that if N_j and N_{j+1} are large, $\{\theta_{j+1,k}: k=1,\ldots,N_{j+1}\}$ will be distributed as $f_{j+1}(\theta)$. Moreover, the expected value of $w(\theta_{j,k})$ is

$$E[w(\theta_{j,k})] = \int w(\theta)\cdot f_j(\theta)\cdot d\theta = \int f(D|M,\theta)^{p_{j+1}-p_j}\cdot f_j(\theta)\cdot d\theta$$
$$= \int f(D|M,\theta)^{p_{j+1}-p_j}\cdot \frac{f(\theta|M)f(D|M,\theta)^{p_j}}{\int f(\theta|M)f(D|M,\theta)^{p_j}d\theta}\cdot d\theta$$
$$= \frac{\int f(\theta|M)f(D|M,\theta)^{p_{j+1}}d\theta}{\int f(\theta|M)f(D|M,\theta)^{p_j}d\theta} \quad (14)$$

Therefore, $\sum_{k=1}^{N_j} w(\theta_{j,k})/N_j$=asymptotically unbiased estimator for $\int f(\theta|M)f(D|M,\theta)^{p_{j+1}}d\theta/\int f(\theta|M)f(D|M,\theta)^{p_j}d\theta$.

Based on the above results, the following algorithm can be applied to draw samples from $f(\theta|M,D)$ and to estimate $f(D|M)$.

Basic TMCMC algorithm:

1. Obtain samples $\{\theta_{0,k}:k=1,\ldots,N_0\}$ from the prior PDF $f_0(\theta)=f(\theta|M)$. Repeat the following 2–3 for $j=0,1,2,\ldots,m-1$.
2. Compute the plausibility weight $w(\theta_{j,k})=f(D|M,\theta_{j,k})^{p_{j+1}-p_j}$ for $k=1,\ldots,N_j$ and compute

$$S_j = \sum_{k=1}^{N_j} w(\theta_{j,k}) \Big/ N_j \quad (15)$$

   Now do the resampling step as follows to obtain the $f_{j+1}(\theta)$ samples:

$$\theta_{j+1,k} = \theta_{j,l} \quad \text{w.p.} \quad w(\theta_{j,l}) \Big/ \sum_{l=1}^{N_j} w(\theta_{j,l}) \quad k=1,\ldots,N_{j+1} \quad (16)$$

   $\{\theta_{j+1,k}:k=1,\ldots,N_{j+1}\}$ will be asymptotically distributed as $f_{j+1}(\theta)$.
3. At the end of the algorithm, $\{\theta_{m,k}:k=1,\ldots,N_m\}$ are asymptotically distributed as $f(\theta|M,D)$ and $\prod_{j=0}^{m-1}S_j$ is asymptotically unbiased for $f(D|M)$.

Remarks:

1. It is trivial to sample $f_0(\theta)$ because it is the prior PDF, and we always have the freedom to choose the prior PDF so that it is easy to sample.
2. $S=\prod_{j=0}^{m-1}S_j$ is asymptotically unbiased for $f(D|M)$ because S_j is asymptotically unbiased for $\int f(\theta|M)f(D|M,\theta)^{p_{j+1}}d\theta/\int f(\theta|M)f(D|M,\theta)^{p_j}d\theta$ and because $\int f_0(\theta)d\theta=1$ and $\int f(\theta|M)f(D|M,\theta)^{p_m}d\theta=\int f(\theta|M)f(D|M,\theta)d\theta=f(D|M)$.
3. As one can conceive from the algorithm, the number of distinct samples in $\{\theta_{j,k}:k=1,\ldots,N_j\}$ decreases as the stage number j increases because of the resampling step. At the end, only few samples in $\{\theta_{m,k}:k=1,\ldots,N_m\}$ are distinct, which is not desirable. A simple solution for this issue is to conduct MCMC for each resampled sample: apply MH with the stationary PDF equal to $f_{j+1}(\theta)$. More precisely, with probability $w(\theta_{j,k})/\sum_{l=1}^{N_j}w(\theta_{j,l})$, a Markov chain sample in the kth chain (the chain with $\theta_{j,k}$ as the leader) is generated using a Gaussian proposal PDF centered at the current sample of the kth chain with a covariance matrix equal to the scaled version of the estimated covariance matrix of $f_{j+1}(\theta)$

$$\Sigma_j = \beta^2 \sum_{k=1}^{N_j} w(\theta_{j,k})\left\{\theta_{j,k} - \left[\sum_{j=1}^{N_j} w(\theta_{j,l})\theta_{j,l} \Big/ \sum_{j=1}^{N_j} w(\theta_{j,l})\right]\right\}$$
$$\times\left\{\theta_{j,k} - \left[\sum_{j=1}^{N_j} w(\theta_{j,l})\theta_{j,l} \Big/ \sum_{j=1}^{N_j} w(\theta_{j,l})\right]\right\}^T \quad (17)$$

   where β=prescribed scaling factor. Note that Σ_j=product of β^2 and the estimated covariance of $f_{j+1}(\theta)$. The β value should be chosen to suppress the rejection rate and, at the same time, to make large MCMC jumps. It is found that 0.2 is a reasonable choice of the scaling parameter β. Do this N_{j+1} times to obtain $\{\theta_{j+1,k}:k=1,\ldots,N_{j+1}\}$. Note that we always generate new Markov chain samples at the tail of each chain. For example, if the kth chain is chosen for the third time, we generate the new sample using the second sample in that chain as the current sample, rather than the leader $\theta_{j,k}$.
4. There may be worries that TMCMC may not spread the samples widely in the θ space, and therefore the samples cannot represent the target PDF $f_{j+1}(\theta)$. This is usually an issue for most MCMC techniques. Fortunately, for TMCMC the θ samples in the first stage are obtained from direct Monte Carlo simulation of $f_0(\theta)$, so they spread widely in the θ space. Consequently, as long as the number of samples is sufficiently large, the θ samples in the later stages, including the samples in the last stage, are expected to also spread widely. Moreover, burn-in period is not an issue in the TMCMC algorithm, either, because all initial samples in the jth stage are already distributed as $f_j(\theta)$, meaning that all Markov chains reach stationary distribution from the very beginning.
5. The choice of $\{p_j:j=1,\ldots,m-1\}$ is essential. It is desirable to increase the p values slowly so that the transition between adjacent PDFs is smooth, but if the increase of the p values is too slow, the required number of intermediate stages (i.e., m value) will be huge. More intermediate stages mean more computational cost.

The following strategy is adopted to adaptively choose the p values. Notice that the degree of uniformity of the plausibility weights $\{w(\theta_{j,k}):k=1,\ldots,N_j\}$ is a good indicator of how close $f_{j+1}(\theta)$ is to $f_j(\theta)$, so p_{j+1} should be chosen so that the coefficient of variation (COV) of the plausibility weights is equal to a prescribed threshold. It is also convenient to force the thresholds of all stages to be the same. It is found that 100% is usually a reasonable choice of this threshold.

As a summary, the TMCMC algorithm is motivated by AMH developed in Beck and Au (2002), but it is more attractive because of the following benefits:

J. Eng. Mech., 2007, 133(7): 816-832

1. It is more efficient for higher-dimensional problems since resampling is employed to replace the kernel density function in AMH. Moreover, the resampling step in TMCMC requires very little computations, while the implementation of kernel density in AMH requires the evaluation of the kernel density function, which takes extra computations. This extra cost is not needed for resampling. Other than this, the total computation cost of TMCMC is the same as AMH as long as the sample number is taken to be the same.
2. It is able to estimate $f(D|M)$ (not addressed by AMH).
3. It provides a guideline of adaptively choosing the p values (not addressed by AMH).
4. There is no need to estimate the KDF, which is sometimes a tricky exercise.

The TMCMC algorithm still enjoys the benefits provided by AMH: it is applicable to difficult PDFs, such as multimodal PDFs, very peaked PDFs, and PDFs with flat regions.

TMCMC Algorithm

Modified according to the remarks in the previous section, the complete TMCMC algorithm is summarized as follows.

1. Obtain samples $\{\theta_{0,k}:k=1,\ldots,N_0\}$ from the prior PDF $f_0(\theta)$. Repeat 2–3 for $j=0,1,2,\ldots,m-1$.
2. Choose p_{j+1} such that the COV of $\{f(D|M,\theta_{j,k})^{p_{j+1}-p_j}: k=1,\ldots,N_j\}$ is equal to a prescribed threshold, then compute the plausibility weight $w(\theta_{j,k})=f(D|M,\theta_{j,k})^{p_{j+1}-p_j}$ for $k=1,\ldots,N_j$ and compute $S_j=\sum_{k=1}^{N_j}w(\theta_{j,k})/N_j$.
3. Initialize N_j Markov chain: the kth Markov chain starts its sample at $\theta_{j,k}^{\mathrm{now}}=\theta_{j,k}$, where $\theta_{j,k}^{\mathrm{now}}$ denotes the current sample in the kth Markov chain. Do the following for $k=1,\ldots,N_{j+1}$ to obtain $\{\theta_{j+1,k}:k=1,\ldots,N_{j+1}\}$: w.p. $w(\theta_{j,l})/\sum_{l=1}^{N_j}w(\theta_{j,l})$, θ^C (the superscript C denotes "candidate") is drawn from $N(\theta_{j,l}^{now},\Sigma_j)$, where Σ_j is defined in Eq. (28). Set $\theta_{j+1,k}=\theta^C$ and $\theta_{j,l}^{\mathrm{now}}=\theta^C$ w.p. $f_{j+1}(\theta^C)/f_{j+1}(\theta_{j,l}^{\mathrm{now}})$; otherwise, set $\theta_{j+1,k}=\theta_{j,l}^{now}$.

After the m stages, $\{\theta_{m,k}:k=1,\ldots,N_m\}$ samples are asymptotically distributed as $f(\theta|M,D)$. Moreover, $S\equiv\prod_{j=0}^{m-1}S_j$ is asymptotically unbiased for $f(D|M)$. According to the Law of Large Number, any quantity of interest $E(g|M,D)$ can be estimated as

$$E(g|M,D)\approx\frac{1}{N_m}\sum_{k=1}^{N_m}g(\theta_{m,k})\equiv g_{\mathrm{TMCMC}} \tag{18}$$

If the above procedure is repeated for all chosen model classes $\{M^{(i)}:i=1,\ldots,N_{\mathrm{class}}\}$ to obtain $\{(g_{\mathrm{TMCMC}}^{(i)},S^{(i)}):i=1,\ldots,N_{\mathrm{class}}\}$, Bayesian model class selection, and model averaging can be done as follows:

$$P[M^{(i)}|D]\approx S^{(i)}\cdot P(M^{(i)})\Bigg/\left[\sum_{l=1}^{N_{\mathrm{class}}}S^{(l)}\cdot P(M^{(l)})\right]\equiv P_{\mathrm{TMCMC}}^{(i)}$$

$$E[g|D]\approx\sum_{i=1}^{N_{\mathrm{class}}}g_{\mathrm{TMCMC}}^{(i)}\cdot P_{\mathrm{TMCMC}}^{(i)}\equiv g_{\mathrm{TMCMC}}^{\mathrm{average}} \tag{19}$$

Statistical Properties of the Estimators

In this section, we derive the statistical properties of the estimators g_{TMCMC}, S, and $g_{\mathrm{TMCMC}}^{\mathrm{average}}$. First of all, it is clear that g_{TMCMC} is an asymptotically unbiased estimator of $E(g|M,D)$; that is, $E(g_{\mathrm{TMCMC}})=E(g|M,D)$ when the number sample is large, because $\{\theta_{m,k}:k=1,\ldots,N_m\}$ are asymptotically distributed as $f(\theta|M,D)$. Moreover, g_{TMCMC} is asymptotically normally distributed due to the Central Limit Theorem. The variance of g_{TMCMC} is equal to

$$\mathrm{Var}(g_{\mathrm{TMCMC}})=\frac{1}{N_m^2}\left\{\sum_{k=1}^{N_m}\mathrm{Var}[g(\theta_{m,k})]+\sum_{k,l=1(k\neq l)}^{N_m}\mathrm{CO\text{–}V}[g(\theta_{m,k}),g(\theta_{m,l})]\right\} \tag{20}$$

where CO–V$[A,B]$ denotes the covariance between A and B. Note that the samples obtained in Step 3 in the algorithm are asymptotically from a stationary Markov chain; therefore

$$\mathrm{Var}(g_{\mathrm{TMCMC}})=\frac{1}{N_m^2}\left[\sum_{k=1}^{N_m}R(0)+\sum_{k,l=1(k\neq l)}^{N_m}R(k-l)\right] \tag{21}$$

where $R(q)=lag-q$ autocovariance of the sequence $\{g(\theta_{m,k}): k=1,\ldots,N_m\}$; and $R(0)$ is the same as $\mathrm{Var}[g(\theta_{m,k})]$. The following equation can be used to estimate $R(q)$ for $q=0,\ldots,N_m-1$:

$$R(q)\approx\frac{1}{N_m-q}\sum_{k=1}^{N_m-q}[g(\theta_{m,k})-g_{\mathrm{TMCMC}}][g(\theta_{m,k+q})-g_{\mathrm{TMCMC}}]\equiv r(q) \tag{22}$$

Using the fact that $R(q)=R(-q)$, $\mathrm{Var}(g_{\mathrm{TMCMC}})$ can be estimated as

$$\begin{aligned}\mathrm{Var}(g_{\mathrm{TMCMC}})&\approx\frac{1}{N_m^2}\left[\sum_{k=1}^{N_m}r(0)+2\sum_{k,l=1(k>l)}^{N_m}r(k-l)\right]\\&=\frac{1}{N_m^2}\left[N_m\cdot r(0)+2\sum_{q=1}^{N_m-1}(N_m-q)r(q)\right]\\&=\frac{r(0)}{N_m}(1+\gamma)\end{aligned} \tag{23}$$

where $\gamma=2\sum_{q=1}^{N_m-1}[1-q/N_m]\cdot r(q)/r(0)$. Consequently, the following 95.4% confidence interval for $E(g|M,D)$ can be built:

$$g_{\mathrm{TMCMC}}-2\sqrt{(1+\gamma)r(0)/N_m}\leqslant E(g|M,D)\leqslant g_{\mathrm{TMCMC}}+2\sqrt{(1+\gamma)r(0)/N_m} \tag{24}$$

and COV of g_{TMCMC} can be estimated as

$$\mathrm{COV}(g_{\mathrm{TMCMC}})\approx\frac{\sqrt{r(0)\cdot(1+\gamma)/N_m}}{g_{\mathrm{TMCMC}}}\equiv\delta \tag{25}$$

Similarly, $S_j=\sum_{k=1}^{N_j}w(\theta_{j,k})/N_j$=asymptotically unbiased estimator of $\int f(\theta|M)f(D|M,\theta)^{p_{j+1}}d\theta/\int f(\theta|M)f(D|M,\theta)^{p_j}d\theta$. Its variance can be estimated as

$$\mathrm{Var}[S_j]\approx\frac{r_j(0)}{N_j}(1+\lambda_j) \tag{26}$$

where

$$\lambda_j=2\sum_{q=1}^{N_j-1}(1-q/N_j)\cdot r_j(q)/r_j(0)$$

$$r_j(q)=\frac{1}{N_j-q}\sum_{k=1}^{N_j-q}[w(\theta_{j,k})-S_j][w(\theta_{j,k+q})-S_j] \tag{27}$$

J. Eng. Mech., 2007, 133(7): 816-832

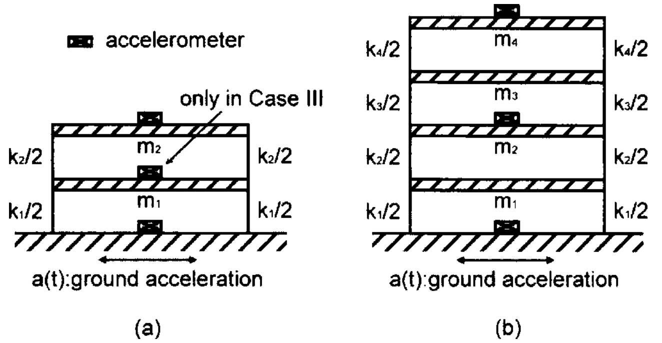

Fig. 1. 2DOF and 4DOF planar shear building in Example 1 and Example 3

As a consequence, $S \equiv \Pi_{j=0}^{m-1} S_j$=asymptotically unbiased estimator for $f(D|M)$. Under the assumption that $\{S_j : j=1,\ldots,m-1\}$ are independent, it can be shown that the COV of S can be estimated

$$\mathrm{COV}(S) \approx \sqrt{\prod_{j=0}^{m-1} [1 + r_j(0) \cdot (1+\lambda_j) \Big/ (N_j S_j^2) - 1} \quad (28)$$

Finally, since both g_{TMCMC} and S are asymptotically unbiased, $g_{\mathrm{TMCMC}}^{\mathrm{average}}$ is also an asymptotically unbiased estimator of $E(g|D)$. In the case that the nth model class dominates (i.e., $S^{(n)}$ is much larger than the rest), which is usually the case, we have

$$\mathrm{COV}(g_{\mathrm{TMCMC}}^{\mathrm{average}}) \approx \delta^{(n)} \quad (29)$$

Examples

Three examples are presented in this section to demonstrate the use of TMCMC. The goals of the demonstration include the following: (1) demonstrate that the new approach is able to effectively draw samples from $f(\theta|M,D)$ and to estimate $f(D|M)$; (2) compare the performance of the new approach with AMH and MH; (3) demonstrate that the new approach is applicable to difficult PDFs (e.g., multimodal, very peaked, and flat PDFs); and (4) demonstrate that the new approach works properly for higher-dimensional problems. For all case studies, the sample numbers of all stages are fixed (i.e., $N_0 = N_1 = \cdots = N_m = N$). To suppress the MCMC rejection rate in TMCMC, the scaling parameter β in Eq. (17) is chosen to be 0.2 for all examples. Except for the last stage $p_m = 1$, the $\{p_1, \ldots, p_{m=1}\}$ values are adaptively chosen in TMCMC so the COV of $\{w(\theta_{j,k}) : k=1,\ldots,1{,}000\}$ is equal to 100% for all stages in all examples. The same choice of the p values is implemented for AMH to ensure fair comparison between TMCMC and AMH. In AMH, the kernel density estimation approach recommended by Au and Beck (2003) is employed for most of the cases to estimate the KDF of each stage.

Example 1: Two Degrees-of-Freedom Structure

Consider a 2DOF linear structure shown in Fig. 1(a). The actual masses $m_1 = m_2 = 1$ and the actual stiffnesses are $k_1 = k_2 = 1{,}000$. The actual damping ratios for the two vibrational modes are identical $\xi_1 = \xi_2 = 3\%$. The goal is to draw samples from $f(\theta|M,D)$, the PDF of the uncertain stiffnesses and dampings (masses are assumed known) conditioning on the measured acceleration data from the structure. Three cases of various identifiability are considered:

1. Case I (flat PDF): only the roof and base accelerations are measured for a short period of time (1 s). The base acceleration is narrow banded so that only the first mode is excited.
2. Case II (bimodal PDF): the roof and base accelerations are measured for a short period of time (1 s). The base acceleration is wide banded so that both modes are excited. According to Beck and Katafygiotis (1998), the posterior PDF is bimodal.
3. Case III (PDF with a single sharp peak): the roof, base, and 2nd floor accelerations are measured for a longer period of time (10 s).

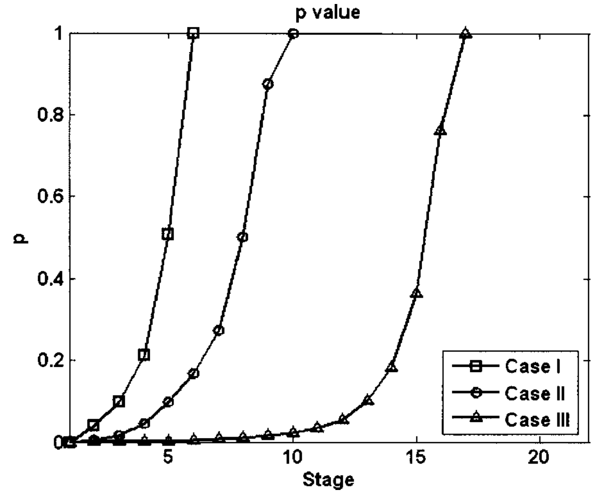

Fig. 2. Evolution of the p values with increasing stages for all cases in Example 1

For all cases, the accelerations are simulated by computers with sampling interval equal to 0.02 s; the roof and 2nd floor accelerations are contaminated by white Gaussian noises. The actual values of the variances of the noise are 0.2; however, they are treated as uncertain variables in the analysis.

The assignment of the prior PDF for θ is as follows: all stiffnesses are independently uniformly distributed over [0 3,000]; the damping ratio is uniformly distributed over [0.01 0.05]; and the variance parameter is uniformly distributed over [0 1]. For all cases, TMCMC and AMH with $N=1{,}000$ is implemented (i.e., 1,000 samples per stage). Therefore, in total, 1,000 m samples are needed for both TMCMC and AMH. It is found that the p values grow exponentially with increasing stages for all cases, as seen in Fig. 2. In AMH, it is found that the recommended choice of the bandwidth of the kernels in KDF is inefficient. Trial runs are taken to find the best bandwidth for KDF in AMH.

The samples obtained by TMCMC and AMH for some selected stages in all cases are plotted in Figs. 3–5 (only stiffness parameters are shown). In Fig. 3, the legend explains the significance of sizes of the dots in the figure. All the similar figures that follow will use the same notation. In general, the TMCMC samples spread more evenly according to the distribution with less repetitive samples, while the AMH samples finally cluster together despite the careful choice of the kernel bandwidth. For Case I, since only the first mode is excited, the posterior PDF is relatively flat (i.e., almost unidentifiable) (Beck and Katafygiotis 1998). This unidentifiability is clear for the TMCMC samples but not for the AMH samples. For Case II, the posterior PDF contains

J. Eng. Mech., 2007, 133(7): 816-832

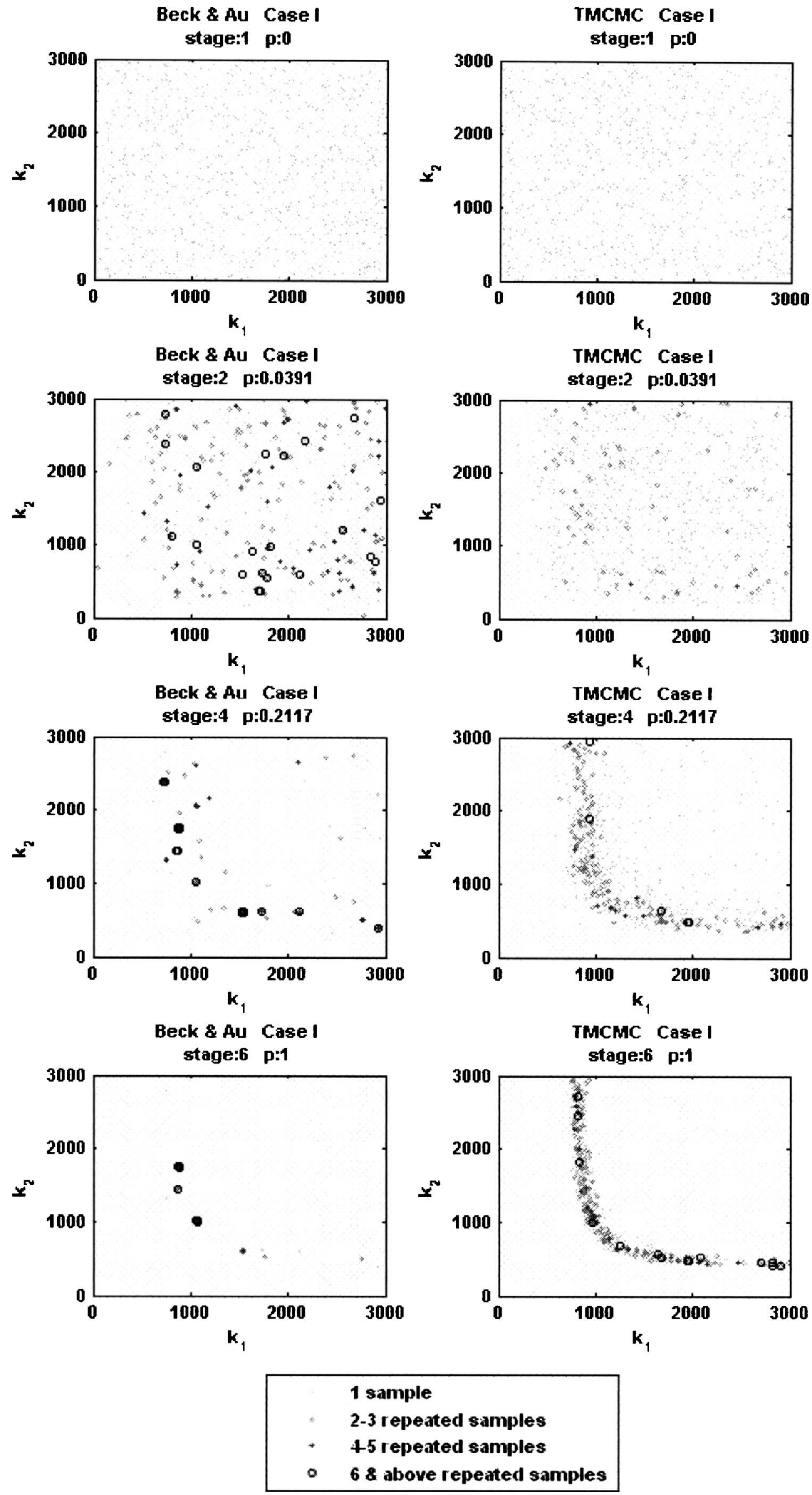

Fig. 3. Evolution of samples from AMH and TMCMC for selected stages in Case I

J. Eng. Mech., 2007, 133(7): 816-832

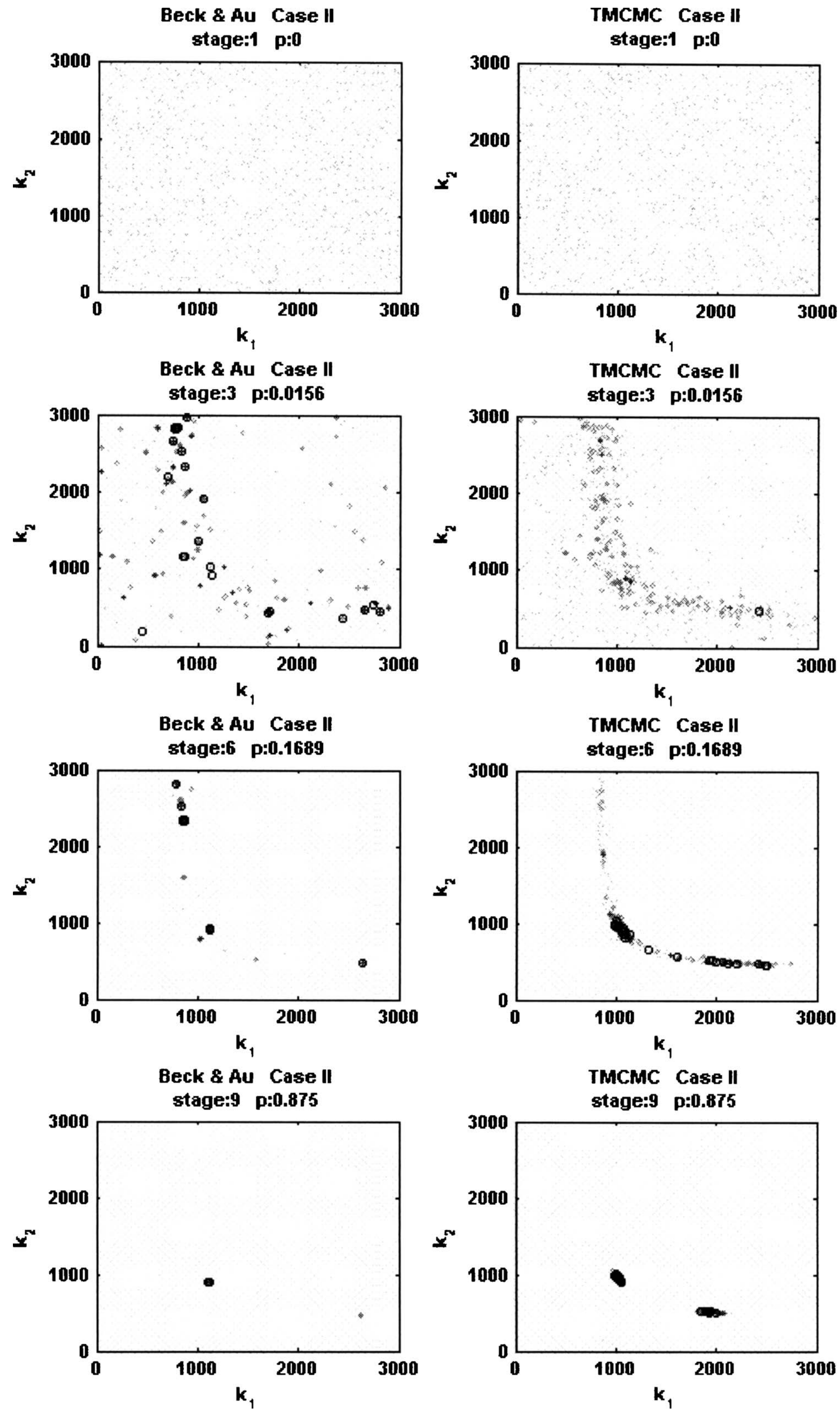

Fig. 4. Evolution of samples from AMH and TMCMC for selected stages in Case II

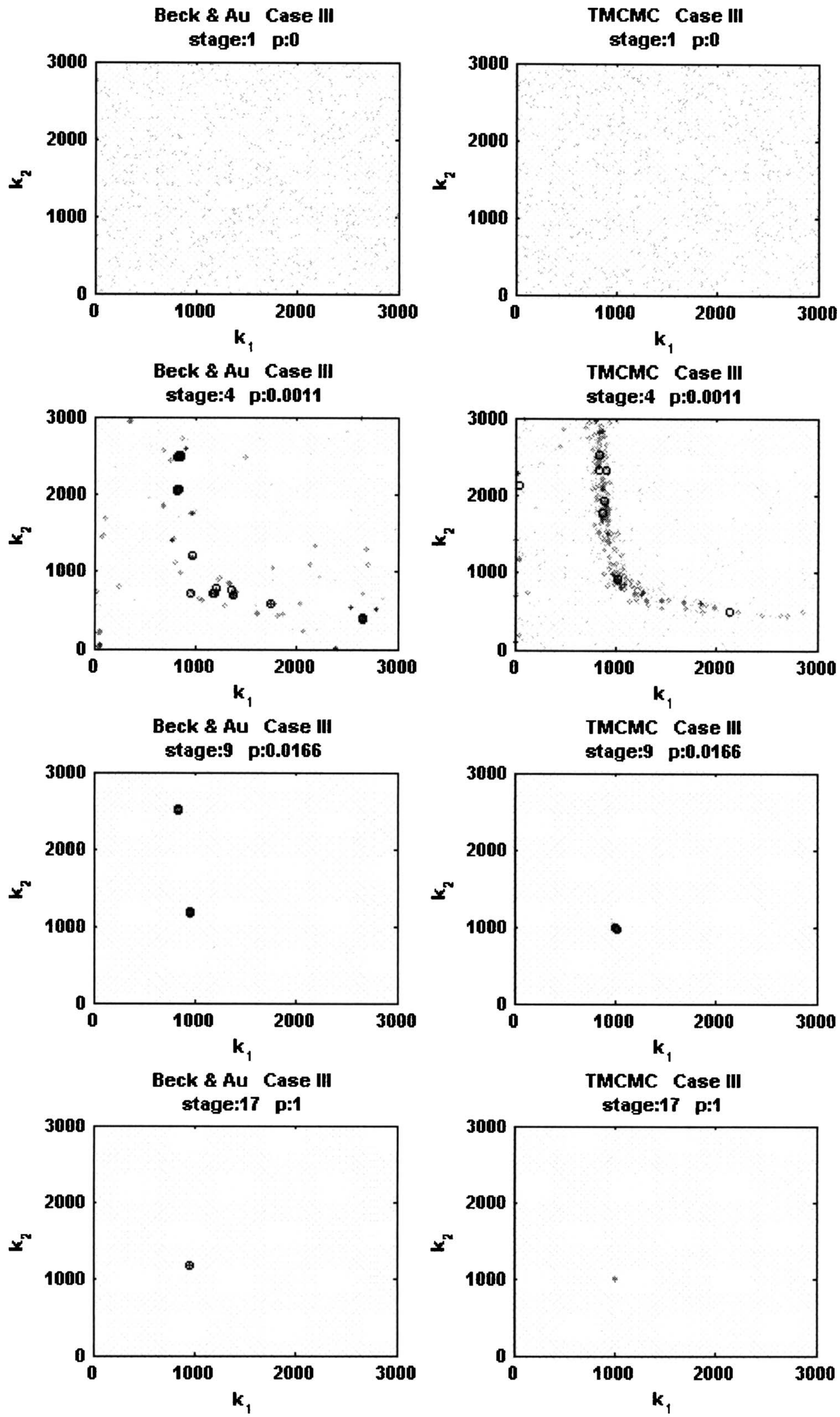

Fig. 5. Evolution of samples from AMH and TMCMC for selected stages in Case III

J. Eng. Mech., 2007, 133(7): 816-832

Table 1. Summary of the Analysis Results for Example 1

Case number	Data number	Measured channel	TMCMC $\log[f(D\|M)]$ (COV)	$E[\text{max drift}\|D]$ TMCMC (COV)	$E[\text{max drift}\|D]$ Beck and Au (2002) (COV)
I	50	Roof	8.93 (2.90%)	0.0047 (22.00%)	0.0037 (33.96%)
II	50	Roof	−7.62 (36.80%)	0.0039 (10.39%)	0.0059 (29.57%)
III	500	Roof and 2F	87.31 (53.30%)	0.0059 (0.46%)	0.0049 (29.12%)

two peaks (Beck and Katafygiotis 1998), as correctly indicated by the TMCMC samples. For Case III, the posterior PDF contains one sharp peak located at [1,000 1,000], as seen from the TMCMC samples. Note that the AMH samples converge to a location slightly different from [1,000 1,000].

The θ samples obtained from the TMCMC and AMH are also used to estimate the expected value of the maximum interstory drift of the building during the simulated excitation. The sample means and COVs of these estimates made by 50 independent TMCMC and AMH runs are listed in Table 1. It is clear that the TMCMC estimates are less variable. The sample mean and COV of the $\log[f(D|M)]$ estimates [i.e., $\log(S)$] from 50 independent TMCMC runs are also shown in Table 1. Note that the evaluation of $f(D|M)$ is not addressed by AMH.

Example 2: Mixture of Two Gaussians

In this example, the target PDF $f(\theta|M,D)$ is proportional to the product of the prior PDF $f(\theta|M)$ and the likelihood $f(D|M,\theta)$. To examine the efficiency of TMCMC for various dimensionality, three scenarios are studied: θ dimension is 2, 4, and 6. The prior PDF is uniform over the [−2 2] interval in all coordinate directions. The likelihood function is the mixture of two Gaussian functions centered at [0.5 … 0.5] and [−0.5 … −0.5]. The standard deviations of the two Gaussian functions are identical, while the weights of the main support regions of the two Gaussians can vary. It is clear that when the standard deviation of the Gaussians is small enough, the two Gaussian functions will become effectively disconnected; that is, $f(\theta|M,D)$ becomes bimodal. We implement MH, AMH, and TMCMC to the aforementioned cases summarized in Table 2. The purpose of this example is to examine the robustness of TMCMC against θ dimension and to verify its applicability on bimodal PDFs.

For all cases, TMCMC and AMH with $N=1{,}000$ is implemented (i.e., 1,000 samples per stage). Therefore, in total 1,000 m samples are needed for both TMCMC and AMH. It is found that the p values grow exponentially with increasing stages for all cases. For fair comparison, MH is implemented with 1,000 m samples, and the proposal PDF is carefully chosen to minimize the correlation between the Markov chain samples.

The $f(\theta|M,D)$ samples at the end of the final stage (stage m) for TMCMC in all cases are plotted in Fig. 6 for their first two coordinates (note that the dimensions for Cases III to VIII are more than 2). It is evident that the two Gaussian peaks are isolated for Cases II, IV, VII, and VIII. For all cases, TMCMC seems successful, indicating that TMCMC is able to draw samples from bimodal PDFs (i.e., Cases II, IV, VII, and VIII), flat PDFs (i.e., Cases I, III, and V), and higher-dimensional PDFs (i.e., Cases VII and VIII). As an example, the evolution of the TMCMC samples through stages for Case VIII is shown in Fig. 7. It is interesting to see that the samples gradually find the high probability region with increasing stages.

The results from AMH are shown in Fig. 8. It is found that AMH is also capable of drawing samples from these different PDFs; however, it is less efficient for higher-dimensional problems (i.e., Cases III and up), judging from the repetitive samples in the higher-dimensional problems. Even for the low-dimensional cases (i.e., Cases I and II), many samples are duplicated, indicating that the performance of AMH is poorer than TMCMC. The results from MH are shown in Fig. 9 for Cases I to IV (Cases V to VIII are omitted since the conclusion is similar). In general, the results are satisfactory (probably because the proposal PDF is carefully chosen with several trial runs), except when the two peaks are isolated, as the samples are trapped in one peak.

For more comparison, TMCMC, AMH, and MH are used to estimate $E[\max(\theta^l,\ldots,\theta^n)|M,D]$ (n=dimension of the problem) with the following estimator:

$$E[\max(\theta^l,\ldots,\theta^n)|M,D] \approx \frac{1}{\#\text{ of samples}}\sum_{k=1}^{\#\text{ of samples}} \max(\theta_k^l,\ldots,\theta_k^n) \qquad (30)$$

where $\theta_k^l = lth$ coordinate of the kth sample (for ordinary MCMC, burn-in-period samples are discarded; for TMCMC and AMH, only the samples in the final stage are used). The sample means and COVs of the $E[\max(\theta^l,\ldots,\theta^n)|M,D]$ estimates made by 50 independent TMCMC, AMH, and MH runs are listed in Table 3 (the "max" columns). The actual answers are also listed for reference. As one can see, except for Case VII with large variation, the $E[\max(\theta^l\ldots,\theta^n)|M,D]$ estimates made by TMCMC are satisfactory since they seem unbiased, although the coefficients of variation are sometimes large. The results from AMH seem biased although the coefficients of variation seem smaller. It is clear that the MH estimates for Cases II, IV, VII, and VIII are highly biased since the samples can be trapped in one peak.

The variation of the $E[\max(\theta^l,\ldots,\theta^n)|M,D]$ estimator made by the TMCMC approach can be estimated using Eq. (28) with a single TMCMC run. Table 3 shows the COV of the

Table 2. Considered Cases in Example 2

Case	θ dimension	Gaussian standard deviation	Fraction of first peak
I	2	0.5	0.5
II	2	0.1	0.9
III	4	0.5	0.5
IV	4	0.1	0.9
V	6	0.5	0.5
VI	6	0.3	0.5
VII	6	0.1	0.5
VIII	6	0.1	0.9

J. Eng. Mech., 2007, 133(7): 816-832

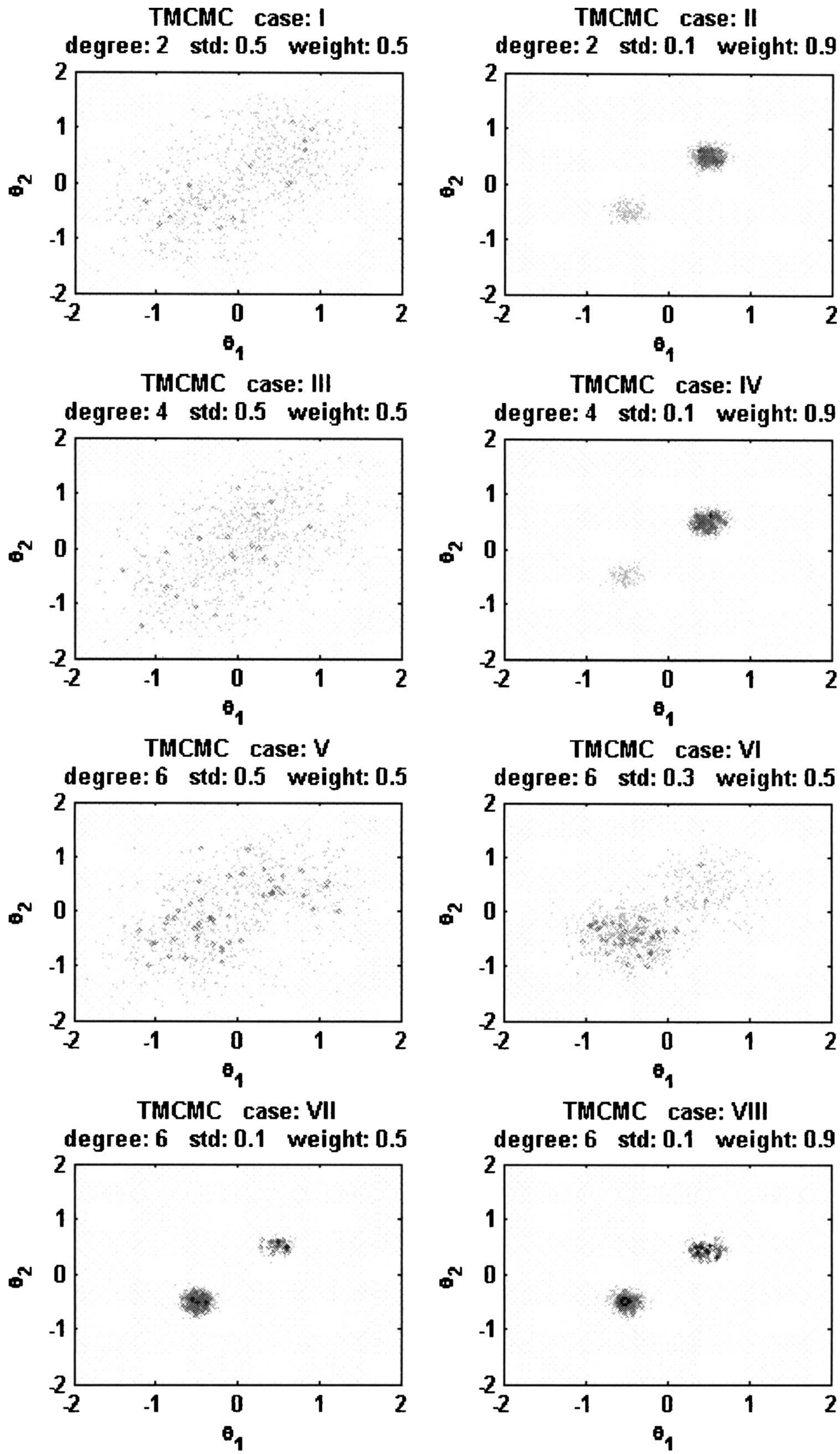

Fig. 6. TMCMC samples at the final stage for all cases in Example 2

J. Eng. Mech., 2007, 133(7): 816-832

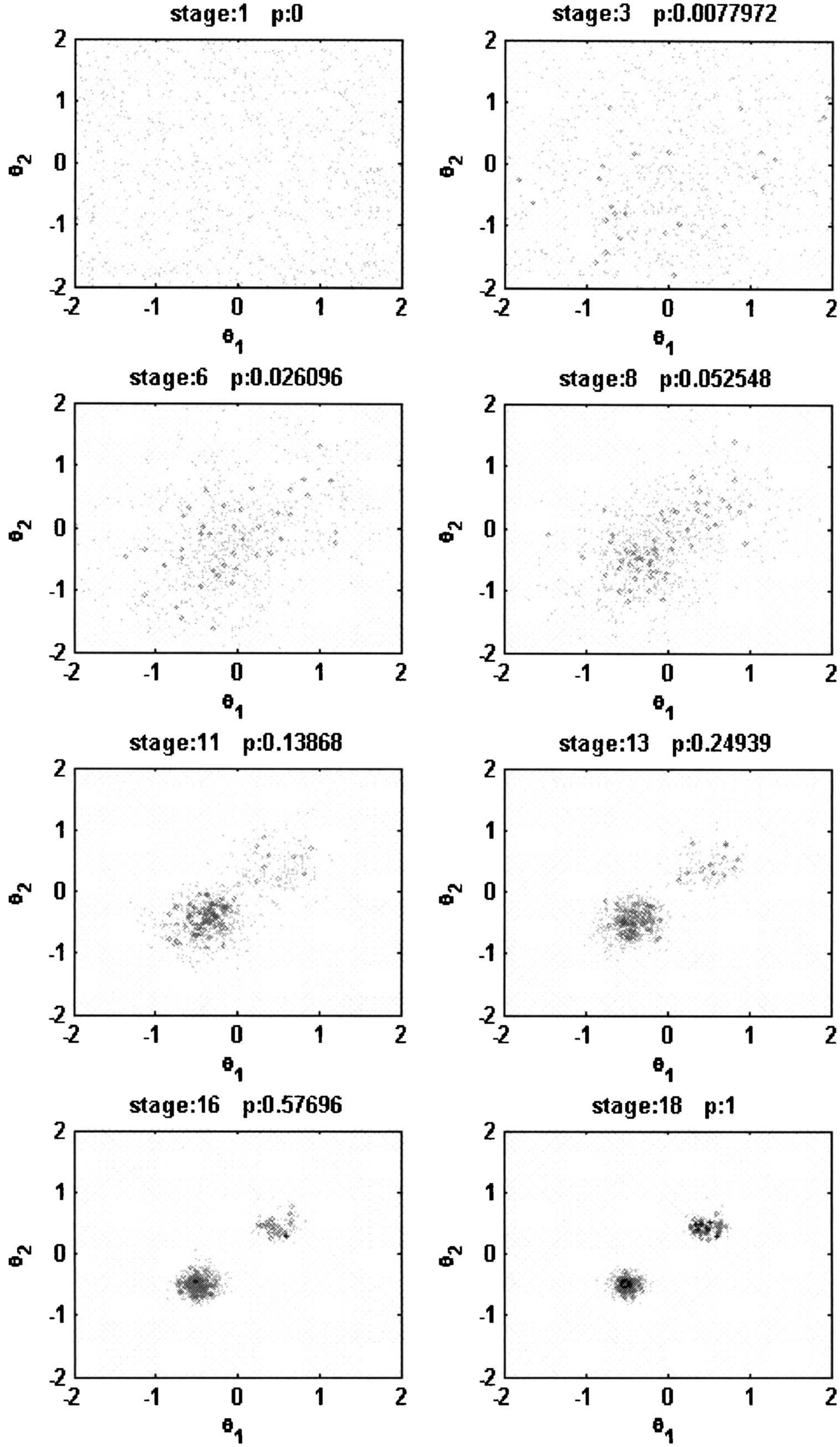

Fig. 7. Evolution of samples from TMCMC for selected stages in Case VIII

J. Eng. Mech., 2007, 133(7): 816-832

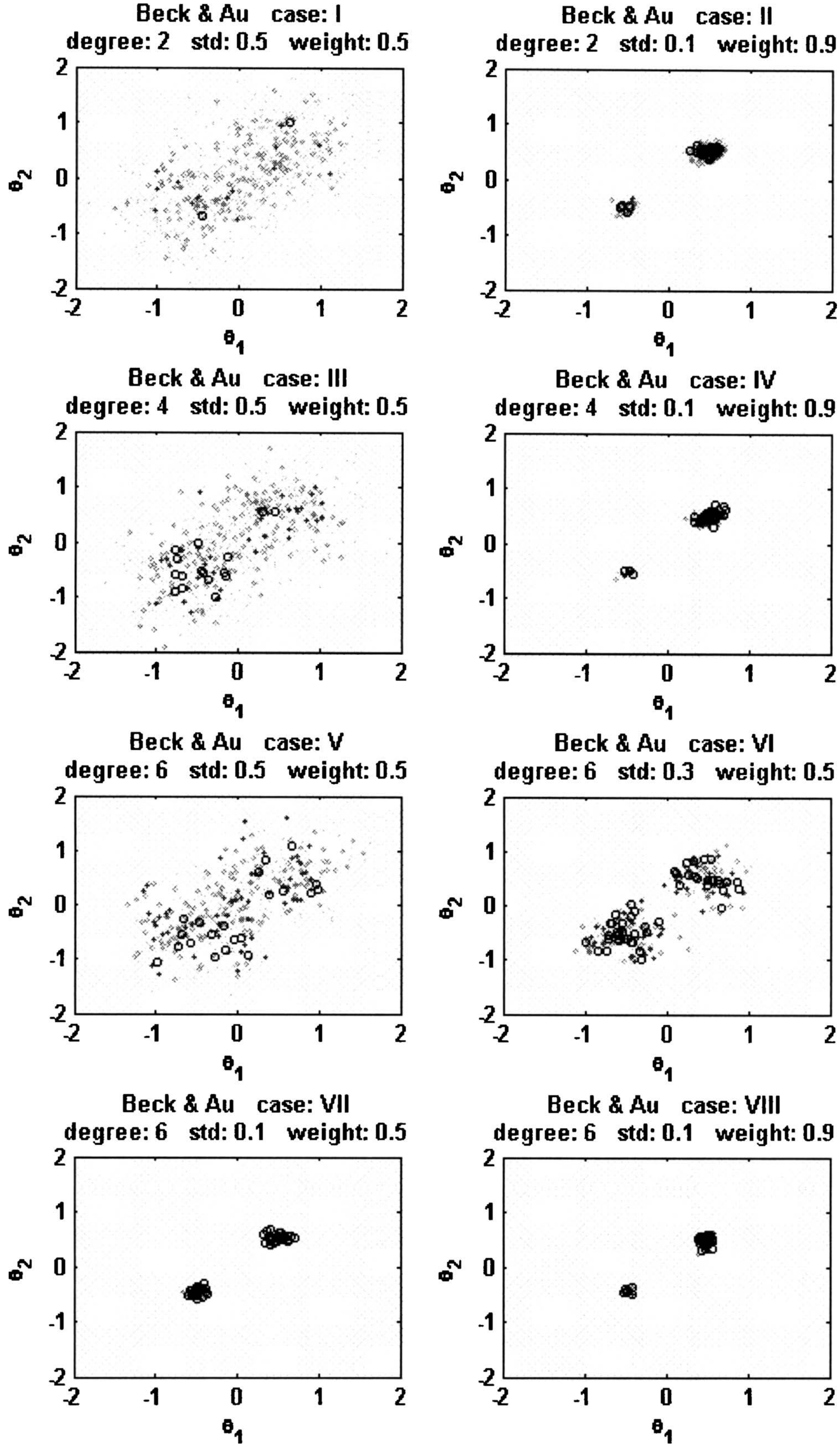

Fig. 8. AMH samples at the final stage for all cases in Example 2

J. Eng. Mech., 2007, 133(7): 816-832

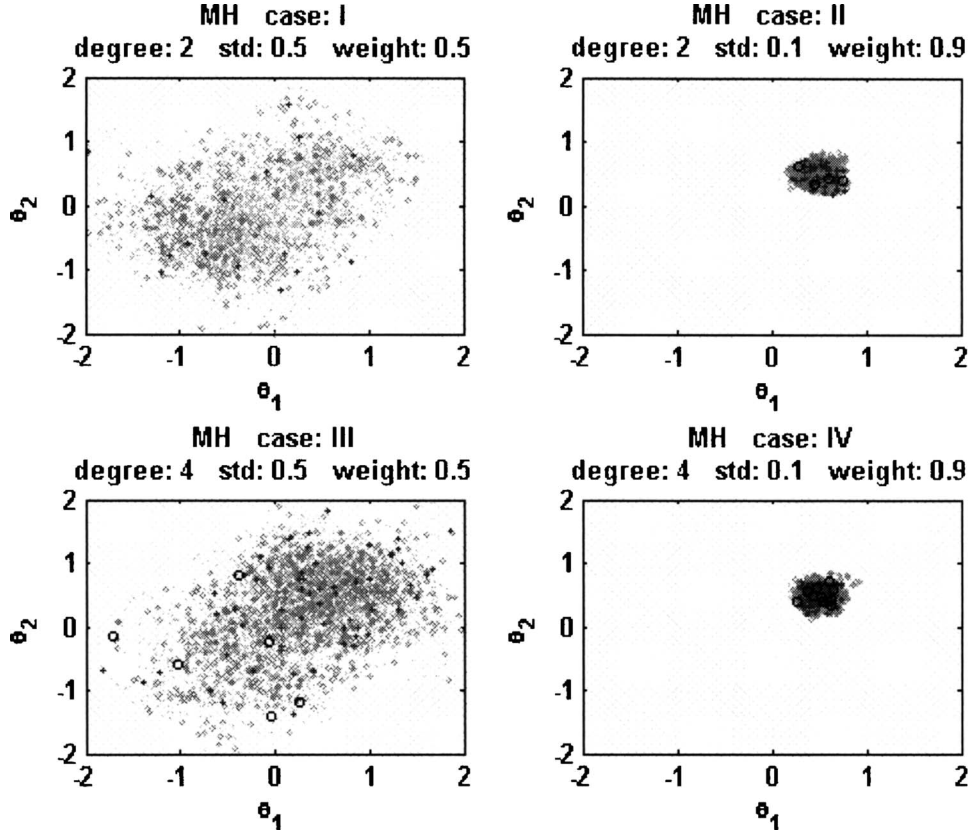

Fig. 9. MH samples for Cases I–IV in Example 2

$E[\max(\theta^l, \ldots, \theta^n) | M, D]$ estimator obtained from Eq. (28): they are in general consistent to the variation obtained from the 50 independent runs when the total number of stages is small. However, when the total number of stages is large, the COV seems underestimated.

In Table 3, it is seen that TMCMC can be more efficient than MH (e.g., in Case I the COV of the maximum value estimator from TMCMC is smaller). This deserves further discussion. For MH, samples are obtained by realizing a single Markov chain. In many times, the movement of the MH samples can be slow, especially for high-dimensional problems, even when the proposal PDF is carefully chosen. This means that the Markov chain samples from MH can be highly correlated. On the other hand, TMCMC uses independent parallel Markov chains to generate samples. More crucially, the TMCMC samples in the first stage are completely independent and spread widely, making it possible to let the samples in later stages also spread widely.

The fraction that the samples belong to the first peak can be estimated by AMH and TMCMC and is summarized in the table (the "first peak" columns). Surprisingly, the fraction estimators from AMH are more stable than those from TMCMC. This is due to the fact that the Gaussian proposal PDF in TMCMC is inefficient for multimodal PDFs, which is a drawback for TMCMC. To overcome this can be a future research topic. Finally, the log-scaled $f(D|M)$ estimates obtained by the TMCMC [i.e., $\log(S)$] for all cases are listed in Table 3. It is interesting to see that the log-scaled $f(D|M)$ estimates made by TMCMC are quite robust even when the fraction estimates are not.

Table 3. Summary of the Analysis Results for Example 2

	Exact		MH	AMH		TMCMC		
Scenario number	First peak	Max	Max (COV)	First peak (COV)	Max (COV)	First peak (COV)	Max (COV); (COV) from Eq. (28)	$\log[f(D\|M)]$ (COV)
I	0.5	0.29	0.28 (36.0%)	0.50 (5.3%)	0.25 (12.6%)	0.50 (5.6%)	0.28 (12.3%); (11.3%)	−2.34 (1.6%)
II	0.9	0.46	0.20 (244.6%)	0.91 (1.9%)	0.45 (4.0%)	0.90 (2.6%)	0.45 (5.2%); (5.1%)	−5.69 (3.0%)
III	0.5	0.51	0.52 (30.4%)	0.50 (6.4%)	0.42 (9.5%)	0.51 (9.4%)	0.54 (10.0%); (9.3%)	−4.72 (2.3%)
IV	0.9	0.50	0.30 (152.8%)	0.90 (3.9%)	0.48 (7.2%)	0.89 (5.9%)	0.50 (10.5%); (9.2%)	−11.36 (3.0%)
V	0.5	0.64	0.65 (28.1%)	0.50 (7.7%)	0.48 (10.4%)	0.50 (14.5%)	0.65 (15.7%); (10.1%)	−7.11 (2.9%)
VI	0.5	0.38	0.41 (117.1%)	0.50 (9.8%)	0.30 (18.5%)	0.48 (35.0%)	0.36 (49.0%); (20.2%)	−10.27 (3.8%)
VII	0.5	0.13	0.01 (7,190%)	0.47 (18.6%)	0.08 (112.4%)	0.49 (50.4%)	0.12 (201.1%); (45.6%)	−17.32 (3.0%)
VIII	0.9	0.53	0.13 (401.7%)	0.82 (10.4%)	0.43 (20.6%)	0.86 (17.1%)	0.49 (30.1%); (13.9%)	-17.27 (4.5%)

J. Eng. Mech., 2007, 133(7): 816-832

Table 4. Considered Cases in Example 3

Case number	Story number	Actual stiffness	Actual mass	Actual damping ratio	Actual standard deviation of measure error	Actual max-drift
I	1F	1,000	1	0.03	0.2	0.011278
	2F	1,000	1			
	3F	1,000	1			
	4F	1,000	1			
II	1F	1,000	1	0.03	0.2	0.009104
	2F	1,000	1			
	3F	800	1			
	4F	800	1			

Example 3: Instrumented Four-Story Linear Building

Consider a 4DOF four-story linear shear building instrumented at the ground floor, second floor, and roof [see Fig. 1(b)]. The actual masses are $m_1=m_2=m_3=m_4=1$, and two scenarios of the actual interstory stiffnesses are considered: (1) $k_1=k_2=k_3=k_4=1{,}000$; and (2) $k_1=k_2=1{,}000$ and $k_3=k_4=800$. The actual damping ratios for all the vibrational modes are equal to 0.03. The simulated 10-s earthquake ground-floor, second-floor, and roof acceleration data for Scenario 1 is shown in Fig. 10 (plots for Scenario 2 are skipped for brevity). For all cases, the accelerations are simulated by computers with sampling intervals equal to 0.02 s and are contaminated by white Gaussian noises. The actual values of the variances of the noise are 0.2; however, they are treated as uncertain variables in the analysis. Given the simulated acceleration data D, we would like to estimate the interstory masses, stiffnesses, and damping ratio [i.e., to sample from $f(\theta|M,D)$, where θ contains all uncertain parameters].

Three model classes $M^{(1)}$, $M^{(2)}$, and $M^{(3)}$ are chosen and described as follows: (1) M_1: a 4DOF four-story linear shear building model class where all four stiffnesses and masses can be different; (2) M_2: a 4DOF four-story linear shear building model class with the constraints $m_1=m_2$, $m_3=m_4$, $k_1=k_2$, and $k_3=k_4$; and (3) M_3: a 4DOF four-story linear shear building model class with the constraints $m_1=m_2=m_3=m_4$ and $k_1=k_2=k_3=k_4$. All model classes assume that all damping ratios for the four modes are uncertain but identical and that prediction errors are independent identically distributed Gaussian uncertainties, whose variance is uncertain. Among the three model classes, $M^{(3)}$ is highly biased for Scenario 2, while the other two model classes are both unbiased. All model classes are unbiased for Scenario 1. The assignment of the prior PDF for θ is as follows: all masses are independently uniformly distributed over the interval [0.95 1.05]; all stiffnesses are independently uniformly distributed over [500 1,200]; the damping ratio is uniformly distributed over [0.01 0.05]; and the prior PDF of the variance parameter is flat. The dimensions of the uncertain parameters are 10, 6, and 4 for $M^{(1)}$, $M^{(2)}$, and $M^{(3)}$, respectively.

For all three model classes, TMCMC with $N=1{,}000$ is implemented. It is found that the p values grow roughly exponentially with increasing stage for all cases. The statistics of the $f(\theta^{(i)}|M^{(i)},D)$ samples for the three model classes ($i=1,2,3$) for the two scenarios are summarized in Table 5, including the sample mean and COV of the mass, stiffness, and damping ratio parameter estimates from 50 independent TMCMC runs. It is clear that when the model class is unbiased [i.e., all model classes for Scenario 1 and $M^{(1)}$ and $M^{(2)}$ for Scenario 2], the estimated values of the parameters are close to their actual values.

The sample mean and COV of the $\log\{f[D|M^{(i)}]\}$ estimates, $\log[S^{(i)}]$, from 50 independent TMCMC runs are shown in the same table, which shows that the COVs of the estimates are quite small. It is evident that COVs of the $\log\{f[D|M^{(i)}]\}$ estimates grow with the θ dimension [i.e., the COV is the largest for $M^{(1)}$ and the smallest for $M^{(3)}$]. More interestingly, all model classes are unbiased for Scenario 1, but $S^{(3)}>S^{(2)}>S^{(1)}$. Similarly, both $M^{(1)}$ and $M^{(2)}$ are unbiased for Scenario 2, but $S^{(2)}>S^{(1)}$. The result fits in our intuition that simpler model classes should prevail if their explanatory power is the same as more complicated model classes (i.e., Occam's razor). Also, $S^{(3)}$ in Scenario 2 is significantly lower than the others, indicating that the wrong model class can be detected in a very clear way.

The θ samples obtained from the TMCMC are also used to estimate the expected value of the maximum interstory drift of the building during the simulated earthquake—that is, $E[\max \text{drift}|M^{(i)},D]$ for $i=1,2,3$. The sample means and COVs of these estimates made by 50 independent TMCMC runs are listed in Table 5 for all cases. The actual maximum interstory drift is 0.011278 for Scenario 1 and 0.009104 for Scenario 2. It is clear that when the model class is unbiased, the estimator is unbiased, while the estimator is biased when the model class is wrong. Although the COVs of the $E[\max \text{drift}|M^{(i)},D]$ ($i=1,2,3$) estimators are not large, they do increase with θ dimension.

If the following flat model class prior is chosen: $P[M^{(1)}]=P[M^{(2)}]=P[M^{(3)}]=1/3$, it is clear that the Bayesian model averaging result is

$$
\begin{aligned}
&E[\max \text{drift}|D] \\
&\quad\approx \sum_{i=1}^{3} E[\max \text{drift}|M^{(i)},D] \\
&\quad\quad\times \left\{ S^{(i)}\cdot P(M^{(i)}) \bigg/ \left[\sum_{l=1}^{3} S^{(l)}\cdot P(M^{(i)})\right]\right\} \\
&\quad\approx \begin{cases} E[\max \text{drift}|D,M^{(3)}]=0.011260 & \text{for Scenario 1} \\ E[\max \text{drift}|D,M^{(2)}]=0.009105 & \text{for Scenario 2} \end{cases}
\end{aligned}
\tag{31}
$$

which are both very close to the actual values. Note that although $M^{(3)}$ is a wrong model class for Scenario 2, the result is unaffected and still close to the actual value.

Discussion of the Examples

From the results of the examples, it is evident that the TMCMC algorithm is able to effectively draw samples from multimodal, very peaked, and higher-dimensional PDFs. However, we expect the TMCMC algorithm should start to show degradation when the dimension is extremely large, but for the examples in this section, the degradation is not severe. It is also seen that the TMCMC algorithm can effectively estimate the evidence. Moreover, the conclusion from the Bayesian model class selection is reasonable for example 3. The accuracy of the TMCMC estimators is found to decrease with increasing θ dimension, to improve what is left as a future research subject.

J. Eng. Mech., 2007, 133(7): 816-832

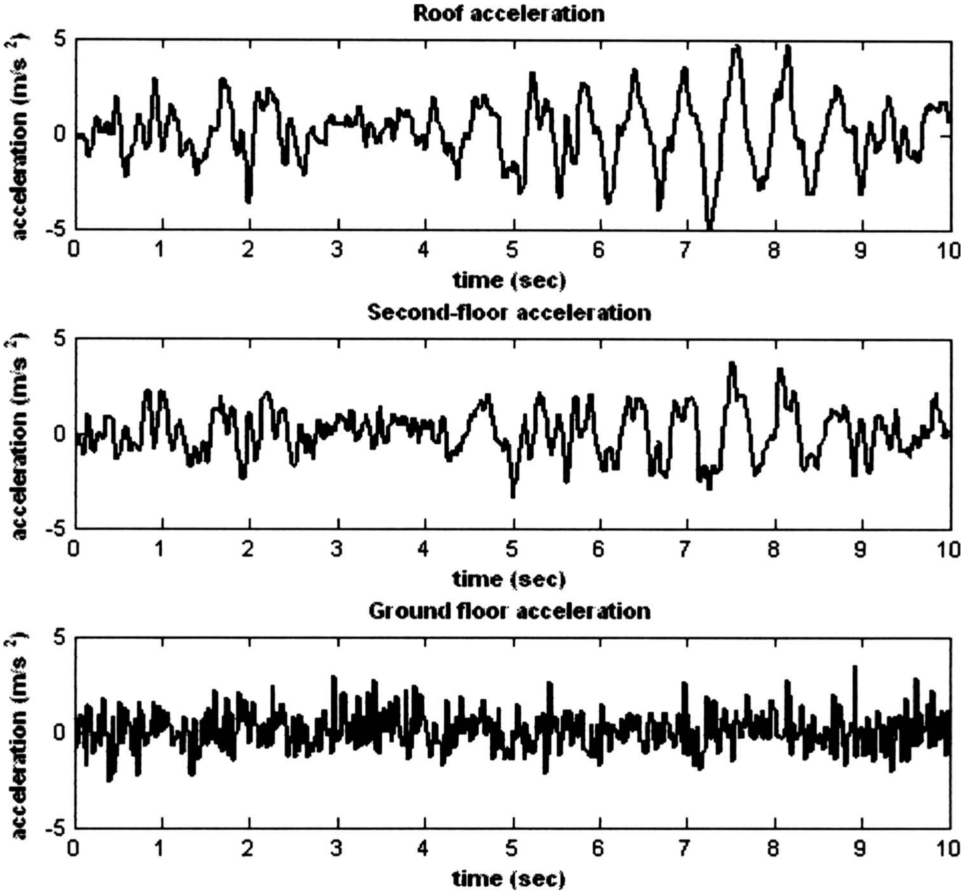

Fig. 10. Simulated acceleration data from the 4DOF shear building

Conclusion

A novel stochastic simulation approach, called the TMCMC, is developed for Bayesian model updating, model class selection, and model averaging problems. As discussed in the main texts, simulation-based Bayesian model updating and model class selection problems are, in general, quite nontrivial. The idea behind the TMCMC is to avoid the problem of sampling from difficult target PDFs but sampling from a series of PDFs that converge to the target PDF and that are easier to sample. The TMCMC approach is versatile: it can be used to sample from the posterior PDF of the uncertain parameters θ of the chosen model class M as well as estimate the evidence $f(D|M)$ for the chosen model class. The latter allows us to conduct Bayesian model class selection and model averaging. Three examples are used to demonstrate the use of the TMCMC, which show that the TMCMC is effective for

Table 5. Summary of the Analysis Results for Example 3

Scenario	Model number	Stiffness (COV)	Mass (COV)	Damping ratio (COV)	Standard deviation of measure error (COV)	$\log[f(D\|M)]$ (COV)	Max drift (COV)
1	1	1,003.39 (2.04%)	1.00 (2.14%)	0.03 (0.88%)	0.21 (4.46%)	155.50 (10.76%)	0.011300 (2.24%)
		1,003.14 (2.71%)	1.00 (2.50%)				
		1,006.10 (2.57%)	1.01 (2.43%)				
		1,005.91 (2.39%)	1.00(2.18%)				
	2	1,003.50 (2.22%)	1.00 (2.20%)	0.03 (0.33%)	0.20 (1.49%)	177.63 (3.08%)	0.011268 (0.46%)
		1,004.43 (2.20%)	1.01 (2.21%)				
	3	1,005.58 (1.75%)	1.01 (1.75%)	0.03 (0.24%)	0.20 (0.56%)	184.42 (0.48%)	0.011260 (0.41%)
2	1	1,004.45 (3.33%)	1.00 (2.67%)	0.03 (1.29%)	0.21 (6.98%)	129.05 (27.22%)	0.009098 (4.56%)
		993.20 (4.87%)	1.00 (2.30%)				
		804.64 (5.13%)	1.00 (2.78%)				
		795.86 (4.17%)	1.00 (2.60%)				
	2	1003.72 (2.27%)	1.00 (2.33%)	0.03 (0.40%)	0.20 (1.95%)	170.32 (3.64%)	0.009105 (0.58%)
		800.78 (2.33%)	1.00 (2.27%)				
	3	924.39 (1.17%)	1.00 (1.17%)	0.04 (0.39%)	0.57 (0.41%)	−868.10 (0.06%)	0.008263 (0.44%)

J. Eng. Mech., 2007, 133(7): 816-832

Bayesian model updating, model class selection, and model averaging. The results also show that the TMCMC is applicable when multimodal posterior PDFs are concerned and when θ dimension is higher. However, the accuracy of the TMCMC estimators decreases with increasing θ dimension, to improve what is left as a future research subject.

References

Au, S. K., and Beck, J. L. (2003). "Importance sampling in high dimensions." *Struct. Safety*, 25, 139–163.

Beck, J. L., and Au, S. K. (2002). "Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation." *J. Eng. Mech.*, 128(4), 380–391.

Beck, J. L., Au, S. K., and Vanik, M. W. (2001). "Monitoring structural health using a probabilistic measure." *Comput. Aided Civ. Infrastruct. Eng.*, 16, 1–11.

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.

Beck, J. L., and Yuen, K. V. (2004). "Model selection using response measurements: Bayesian probabilistic approach." *J. Eng. Mech.*, 130(2), 192–203.

Bernal, D., Dyke, S. J., Beck, J. L., and Lam, H. F. (2002). "Phase II of the ASCE benchmark study on structural health monitoring." *Proc., 15th Engineering Mechanics Division Conf. of the American Society of Civil Engineers*, ASCE, Reston, Va.

Casciati, F., ed. (2002). *Proc., 3rd World Conf. on Structural Control*, Wiley, New York.

Chang, F. K., ed. (2001). *Proc., 3rd Int. Workshop on Structural Health Monitoring*, Stanford Univ., Techomic, Penn.

Ching, J., and Beck, J. L. (2004). "Bayesian analysis of the Phase II IASC–ASCE structural health monitoring experimental benchmark data." *J. Eng. Mech.*, 130(10), 1233–1244.

Ching, J., Muto, M., and Beck, J. L. (2005). "Bayesian linear structural model updating using Gibbs sampler with modal data." *Proc., ICOSSAR 2005*, Millpress, Rotterdam, The Netherlands.

Hjelmstad, K. D., and Shin, S. (1997). "Damage detection and assessment of structures from static response." *J. Eng. Mech.*, 123(6), 568–576.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999). "Bayesian model averaging: A tutorial." *Stat. Sci.*, 14(4), 382–417.

Katafygiotis, L. S., and Beck, J. L. (1998). "Updating models and their uncertainties. Part II: 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, 495–507.

Kerschen, G., Golvinal, J. C., and Hemkez, F. M. (2003). "Bayesian model screening for the identification of nonlinear mechanical structures." *J. Vibr. Acoust.*, 125, 389–397.

Natke, H. G., and Yao, J. T. P., eds. (1988). *Proc., Workshop on Structural Safety Evaluation Based on System Identification Approaches*, Vieweg and Sons, Wiesbaden, Germany.

Sanayei, M., McClain, J. A. S., Wadia-Fascetti, S., and Santini, E. M. (1999). "Parameter estimation incorporating modal data and boundary conditions." *J. Struct. Eng.*, 125(9), 1048–1055.

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.

Yuen, K. V., Au, S. K., and Beck, J. L. (2004). "Two-stage structural health monitoring approach for Phase I benchmark studies." *J. Eng. Mech.*, 130(1), 16–33.

J. Eng. Mech., 2007, 133(7): 816-832