Bayesian Updating and Model Class Selection for Hysteretic Structural Models Using Stochastic Simulation

MATTHEW MUTO
JAMES L. BECK
Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125 USA (jimbeck@its.caltech.edu)

Abstract: System identification of structures using their measured earthquake response can play a key role in structural health monitoring, structural control and improving performance-based design. Implementation using data from strong seismic shaking is complicated by the nonlinear hysteretic response of structures. Furthermore, this inverse problem is ill-conditioned; for example, even if some components in the structure show substantial yielding, others will exhibit nearly elastic response, producing no information about their yielding behavior. Classical least-squares or maximum likelihood estimation will not work with a realistic class of hysteretic models because it will be unidentifiable based on the data. It is shown here that Bayesian updating and model class selection provide a powerful and rigorous approach to tackle this problem when implemented using a recently developed stochastic simulation algorithm called Transitional Markov Chain Monte Carlo. The updating and model class selection is performed on a previously-developed class of Masing hysteretic structural models that are relatively simple yet can give realistic responses to seismic loading. The theory for the Masing hysteretic models, and the theory used to perform the updating and model class selection, are presented and discussed. An illustrative example is given that uses simulated dynamic response data and shows the ability of the algorithm to identify hysteretic systems even when the class of models is unidentifiable based on the data.

1. INTRODUCTION

Current methods for developing finite-element models can produce structural responses that are consistent qualitatively with behavior observed during strong earthquake shaking, but there has long been an interest in using system identification methods for quantitative assessment of structural models using recorded seismic response (see Beck, 1978 for early work and Beck, 1996 for a review). The objective may be to improve the predictive capabilities of structural models for dynamic design or for the design of structural control systems, or to implement structural health monitoring. System identification based on updating of finite-element models using measured seismic response is challenging, however, because the large number of uncertain parameters associated with these models makes the inverse problem extremely ill-conditioned.

Simplified models can be used in the identification procedure but the selection of an appropriate class of models to employ is complicated by the nonlinear response of structures under strong seismic loading; in particular, the structural restoring forces are hysteretic, depending on the previous time history of the structural response rather than on an instantaneous finite-dimensional state. Although some research into the identification of hysteretic systems has been carried out (Jayakumar, 1987; Jayakumar and Beck, 1988; Cifuentes and Iwan, 1989; Benedettini et al., 1995, Ashrafi and Smyth, 2005), this previous work did not quantify the modeling uncertainties and did not properly deal with the ill-conditioning inherent in this inverse problem. However, the uncertainty associated with structural model predictions can have a significant impact on the decision-making process in structural design, control and health monitoring. Furthermore, classical estimation techniques such as least-squares and maximum likelihood do not usually work properly when applied to hysteretic model classes because they are nearly always unidentifiable based on the available data.

The Bayesian updating approach treats the probability of all models within a set of candidate models for a system, and consequently has the advantage of being able to quantify all of the uncertainties associated with modeling of a system and to handle ill-conditioned identification problems. Note that the probability of a model will not make sense if one interprets probability as a long-run frequency of an event, but it does when probability is interpreted as a multi-valued logic that expresses the degree of plausibility of a proposition conditioned on the given information (an interpretation given a rigorous foundation by Cox, 1961; see also Jaynes, 2003). Although Bayesian methods are widely used in many fields, their application to identification of dynamic hysteretic models seems to be very limited.

Beck (1989) and Beck and Katafygiotis (1991, 1998) presented a Bayesian statistical framework for model updating and predictions for linear or nonlinear dynamic systems that explicitly treats prediction-error and model uncertainties. This earlier work utilized Laplace's method for asymptotic approximation to evaluate the Bayesian predictive integrals. An introduction to this theory is given in Papadimitriou and Katafygiotis (2005). A basic concept is that *any* set of possible deterministic dynamic models for a system can be embedded in a set of predictive probability models for the system by specifying a probability distribution (usually Gaussian) for the uncertain prediction error, which is the difference between the actual system output and the deterministic model output. Each predictive probability model is assumed to be uniquely specified by assigning a value to a model parameter vector. Therefore, a probability distribution over the set of possible predictive models that specifies the plausibility of each such model is equivalent to a probability distribution over a corresponding set of possible values for the model parameter vector. When dynamic data is available from the system, a chosen initial (*prior*) probability distribution over the parameters can be updated using Bayes' Theorem to give a *posterior* probability distribution.

It is useful to characterize the topology of this posterior as a function of the model parameter vector by whether it has a global maximum at a single most probable parameter value, at a finite number of them, or at a continuum of most probable parameter values lying on some manifold in the parameter vector space. These three cases may be described as *globally identifiable*, *locally identifiable*, and *unidentifiable* model classes based on given dynamic data from the system. The Laplace asymptotic approximation is most useful when there is a large amount of data and the model class is globally identifiable, although it can

be applied in the locally identifiable case (Beck and Katafygiotis, 1991, 1998) and even unidentifiable cases when the manifold of most probable values of the parameter vector is of very low dimension (Papadimitriou et al., 2001; Katafygiotis and Lam, 2002). However, it requires a non-convex high-dimensional optimization to find the most probable parameter vectors, which can be computationally challenging. To avoid these difficult optimizations and to more readily treat cases where the model class is not globally identifiable (which will occur in finite-element model updating because of the large number of uncertain parameters that occur in realistic structural models), in recent years attention has been focused on stochastic simulation methods for Bayesian updating and prediction, especially Markov Chain Monte Carlo methods, such as the Metropolis–Hastings, Gibbs Sampler and Hybrid Monte Carlo algorithms. The emergence of these stochastic simulation methods has led to a renaissance in Bayesian methods across all disciplines in science and engineering because the high-dimensional integrations that are involved can now be readily evaluated.

The goal of the stochastic simulation methods is to generate samples which are distributed according to the posterior probability density function (PDF). The posterior PDF gives the plausibility of each of the candidate models in the model class (specified by a corresponding vector of model parameters), based on the data. While a variety of stochastic simulation methods are available, many of them are not useful for Bayesian updating. In this work, we focus specifically on Markov Chain Monte Carlo (MCMC) methods (see Neal, 1993; Gilks et al., 1996; MacKay, 2003; or Robert and Casella, 2004 for more comprehensive overviews of this topic). One advantage of these methods is that non-normalized PDFs can be sampled, so that samples may be drawn from the posterior PDF without evaluating the normalizing constant in Bayes' Theorem (called the *evidence* or marginal likelihood), which can be a difficult procedure because it usually requires evaluating a high-dimensional integral over the parameter space. A remaining challenge associated with model updating by stochastic simulation is the fact that, unless the data is very sparse, the posterior PDF occupies a much smaller volume in the parameter space than the prior PDF over the parameters. This fact makes it difficult to draw samples from the posterior PDF.

One commonly-implemented MCMC method is the Gibbs sampler (Geman and Geman, 1984). When applicable, the Gibbs sampler is a powerful method for generating samples from high-dimensional posterior PDFs; for example, Ching et al. (2006) apply it to the problem of using modal data to update a stochastic linear structural model that has 312 parameters. However, the Gibbs sampler is only readily applied to model classes that produce posterior PDFs that have a special structure to them; and for this reason, it is difficult to apply to the updating of hysteretic models.

Another commonly-implemented MCMC method is the Metropolis–Hastings (M-H) algorithm (Metropolis et al., 1953; Hastings, 1970), which can be used to create samples from a Markov Chain whose stationary distribution is any specified target PDF, even a non-normalized one. Although in theory the M-H algorithm can generate samples from any posterior PDF, for higher-dimensional parameter spaces it may still be very difficult to draw samples that cover all the regions of high-probability content. For this reason, Beck and Au (2000, 2002) proposed gradual updating of the model by using the M-H algorithm to sample from a sequence of target PDFs, where each target PDF is the posterior PDF based on an increasing fraction of the available data. In this manner, the target PDF gradually converges from the broad prior PDF to the final concentrated posterior PDF. The samples from each

intermediate PDF are used to form a kernel density, which is used as a global proposal PDF in the M-H algorithm for the next "level" of sampling.

Ching and Chen (2007) modified this approach to develop what they call the Transitional Markov Chain Monte Carlo (TMCMC) method. This technique also uses a sequence of intermediate PDFs. However, rather than applying updating with part of the available data, the entire data set is used but its full effect is diluted by taking the target PDF for the mth level of the sampler to be proportional to $p\left(\mathcal{D}|\theta, \mathcal{M}\right)^{\beta_m} p\left(\theta|\mathcal{M}\right)$, where $0 \leq \beta_m \leq 1$; here, $\beta_0 = 0$ gives the initial target distribution proportional to the prior PDF and $\beta_M = 1$ for the final level of the sampler gives a target distribution proportional to the posterior PDF. Due to the conceptual similarities between this approach and the simulated annealing approach (Fishman, 1996; Neal, 1993), β_m will be referred to as the tempering parameter. In TMCMC, re-sampling is used between levels to improve the rate of convergence.

Another difference between the TMCMC algorithm and the approach of Beck and Au (2002) is in the application of the Metropolis–Hastings algorithm. Rather than using a global proposal PDF based on a kernel density constructed from the samples from the previous level, a local proposal PDF is used in what is essentially a local random walk in the parameter space.

This work focuses on the application of the TMCMC algorithm to Bayesian updating of the model parameters and to Bayesian model class selection between competing sub-classes of models, for a class of Masing hysteretic models that are believed to be well-suited to realistic modeling of the seismic behavior of structures. These models are described in the next section.

2. MASING HYSTERETIC MODELS

One fundamental approach to constructing hysteretic force–deformation relations for structural members and assemblages of members is to build them up from constitutive equations ("plasticity models") which govern material behavior at a point. However, factors such as complex stress distributions, material inhomogeneities and the large number of structural elements make this approach impractical. Also, there is no general consensus on the choice of models for cyclic plasticity under arbitrary loading.

An alternative approach is to develop simplified models that capture the essential features of the hysteretic force-deformation relationship but then, lacking a fundamental theoretical basis, these models should be validated against the observed behavior of structures. This has been done by Jayakumar (1987) for the well-known Bouc–Wen model (Wen, 1976), which is in essence a planar version of the early endochronic model (Valanis, 1971); these models are mathematically convenient, especially for random vibration studies using equivalent linearization, but when they are subjected to asymmetric cyclic loading, these models can exhibit an unphysical "drifting" behavior (Jayakumar, 1987). This behavior makes them unsuitable as a class of identification models for strong seismic response where this type of irregular loading occurs.

A simplified hysteretic model with a physical basis was presented by Masing (1926), which is based on the hypothesis that a one-dimensional hysteretic system may be viewed as a collection of ideal elasto-plastic elements (a linear spring in series with a Coulomb
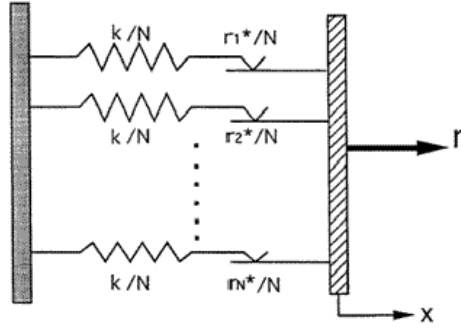
Figure 1. Conceptual sketch of the Distributed Element Model (from Chiang, 1992).

damper) with the same elastic stiffness but with a distribution of different yield strengths. This idea was used in structural dynamics by Iwan to form the Distributed Element Model (DEM), which consists of a collection of N ideal elasto-plastic elements connected in parallel (Iwan, 1966, 1967) with a common stiffness k/N for the springs but different yield strengths $r_i*/N, i = 1, \ldots N$, as shown in Figure 1. The restoring force r for a single-degree of freedom DEM subjected to a displacement x is given by:

$$r = \sum_{i=1}^{n} \frac{r_i*}{N} + kx\frac{N-n}{N} \tag{1}$$

where n is the number of elements which have yielded. Infinite collections of elasto-plastic elements can considered by introducing a yield strength distribution function $\phi(r*)$, such that restoring force $r(x)$ during initial loading is:

$$r(x) = \int_{0}^{kx} r*\phi(r*)\,dr* + kx\int_{kx}^{\infty} \phi(r*)\,dr*. \tag{2}$$

Because there is an underlying physical basis for the model, DEMs with a finite number of elements have been shown to give good representations of the hysteretic behavior of some structures, and do not exhibit the previously-discussed drifting behavior (Cifuentes, 1983; Thyagarajan, 1989). However, DEMs with an infinite number of elements are difficult to implement directly, in contrast to the finite case where the state of each element is tracked. Fortunately, there are two hysteretic rules that exactly describe the behavior of DEMs without needing to keep track of the internal behavior of the elements, which we now present.

Masing (1926) postulated that the steady-state behavior of such a hysteretic system subjected to cyclic loading could be described as follows. If the initial or "virgin" loading curve is described by the implicit relationship:
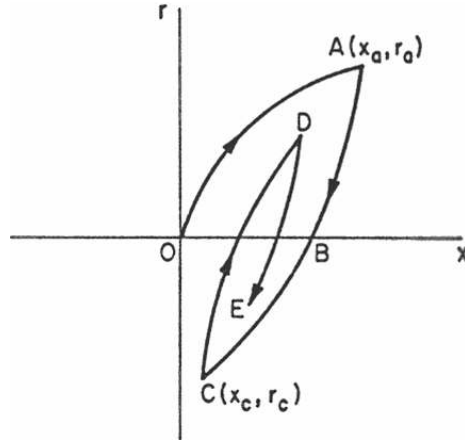
$$f(x, r) = 0 \tag{3}$$

Figure 2. Hysteresis loop for transient loading of the extended Masing model (from Jayakumar, 1987).

where x is the deflection and r is the restoring force, then each branch of the hysteresis loop between the points (x_a, r_a) and $(-x_a, -r_a)$ is given by:

$$f\left(\frac{x - x^*}{2}, \frac{r - r^*}{2}\right) = 0 \tag{4}$$

where (x^*, r^*) is the load reversal point for the branch curve: $(x^*, r^*) \equiv (x_a, r_a)$ for the unloading branch and $(x^*, r^*) \equiv (-x_a, -r_a)$ for the loading branch. This is commonly referred to as Masing's rule.

Masing's theory was extended to apply to the case of softening hysteretic systems under arbitrary loading by Jayakumar (1987) by specifying two hysteresis rules, which will henceforth be referred to as the *extended Masing rules*:

1. The equation of any hysteretic force-deformation curve can be obtained by applying the original Masing rule to the virgin loading curve using the latest point of load reversal. For example, if the virgin loading curve OA in Figure 2 is defined by equation (3) then the branch curve CD is defined by equation (4), with $(x^*, r^*) \equiv (x_C, r_C)$.
2. Once an interior curve crosses a curve from a previous load cycle under continued loading or unloading, a hysteresis loop is completed and the load-deformation curve of the previous cycle is continued. For example, if the unloading curve DE in Figure 2 is continued to point C, further unloading will follow a path that is the extension of the curve ABC.

As already mentioned, the class of models defined by these rules was shown by Jayakumar (1987) to describe the behavior of DEMs, which have an underlying physical model, and so they are expected to produce reasonable hysteretic behavior. Chiang (1992) later demonstrated the relationship between the initial loading curve $f(x, r)$ in equation (3) and the yield strength distribution function $\phi(r*)$ in equation (2), including the case of deteriorating hysteretic models.

2.1. *Masing Shear-building Model*

Jayakumar and Beck (1988) performed system identification of a full-scale six-story steel building tested pseudo-dynamically in the laboratory of the Building Research Institute in Tsukuba, Japan. Substantial yielding occurred in the lower stories during the test, which was an experimental simulation under the Taft record from the 1952 Kern County earthquake in California. For the identification, Jayakumar and Beck used a class of shear-building models in which a Masing model was used to describe the relationship between the story shear forces and the inter-story drifts. Using a least-squares output-error approach, multiple optimal models were obtained, depending on the initial choice of the model parameters given to the optimization algorithm, although all these models gave almost exactly the same good match of the measured response. This apparent non-uniqueness suggested that the chosen class of models was unidentifiable, or at least almost so. The goal of this current work is to study this same class of hysteretic shear-building models with simulated noisy seismic response data to give further insight into their identification and to demonstrate the power of Bayesian updating and model class selection, as implemented using the new TMCMC stochastic simulation algorithm.

Consider a structural model with a rigid foundation where the vector of relative displacements of the structure $x(t)$ is related to the ground acceleration $\ddot{y}(t)$ as follows:

$$M\ddot{x} + C\dot{x} + R = -Mb\ddot{y}(t) \tag{5}$$

where M and C are the mass and viscous-damping matrices, R is the restoring force vector, and b is the pseudo-static influence vector. The inter-story shear force at the ith story is given by:

$$R_i = r_i - r_{i+1} \tag{6}$$

where the relationship between inter-story shear forces and inter-story drifts is given by:

$$\dot{r}_i = K_i(\dot{x}_i - \dot{x}_{i-1})\left[1 - \left|\frac{r_i}{r_{u,i}}\right|^{\alpha_i}\right] \tag{7}$$

where K_i is the initial inter-story stiffness, $r_{u,i}$ is the story ultimate strength and the smoothness of the transition from elastic to plastic response is controlled by the positive parameter α_i. Figure 3 shows how the shape of the initial loading curve is influenced by α_i. Note that for $i = n$ in equation (6), $r_{n+1} = 0$, and for $i = 1$ in equation (7), $x_0 = 0$. This class of models represents a sub-class of Masing models; through the choice of initial loading curve, a wide variety of hysteretic models can be described by the two extended Masing rules, including many existing hysteretic models (Jayakumar, 1987; Thyagarajan, 1989).

The basic form of the force-deformation relationship given in equation (7) is similar to that of a Bouc–Wen model. The major difference is that this equation serves only as the initial loading curve that must be combined with the two extended Masing rules, rather than giving a complete description of the structural response. A potentially important advantage of the Masing shear-building model is that most of the model parameters correspond to actual physical properties (the initial stiffness and ultimate strength) and initial estimates can
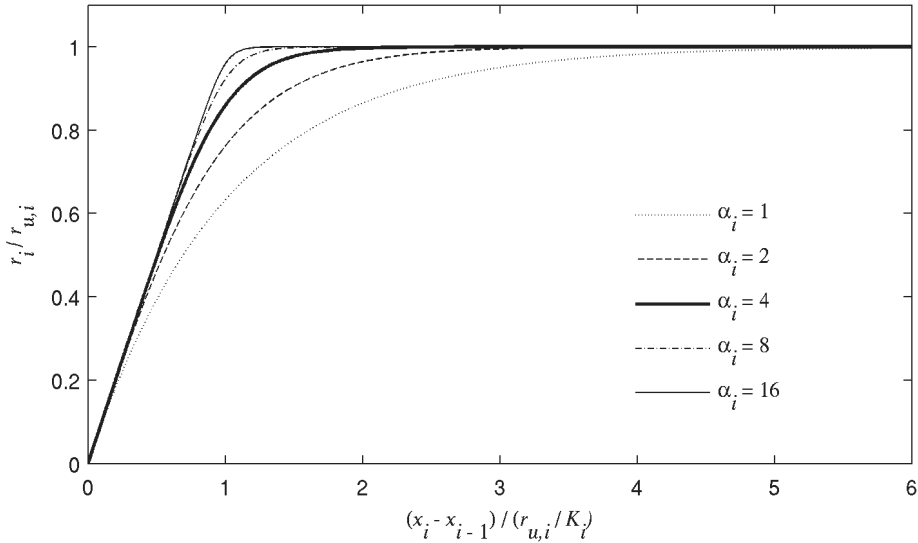
Figure 3. Plots of "virgin" loading curves of the Masing model (equation (7)) for different values of the elastic-to-plastic transition parameter α_i.

be calculated from material properties and structural drawings (Jayakumar, 1987). This information could be explicitly incorporated into the prior distribution in the Bayesian model updating procedure presented in this work.

3. BAYESIAN UPDATING AND PREDICTIONS USING STOCHASTIC SIMULATION

3.1. Bayesian Model Updating

Consider a Bayesian model class $\mathcal{M}$, which is characterized by: (i) a set of *predictive* PDFs, $p(\mathcal{D}|\theta, \mathcal{M})$, for system response $\mathcal{D}$ that is parameterized by N_p model parameters $\theta \in \Theta \in \mathbb{R}^{N_p}$; and (ii) a chosen *prior* PDF $p(\theta|\mathcal{M})$ that can incorporate existing knowledge of the system. The prior PDF is chosen to express the initial plausibility of each model in the class $\mathcal{M}$ defined by the value of the parameter vector θ.

Now suppose a set of data $\mathcal{D}$ from the system is available. The goal of Bayesian updating is to use $\mathcal{D}$ to update the probability distribution over the parameters to give the posterior PDF $p(\theta|\mathcal{D}, \mathcal{M})$ based on Bayes' theorem:

$$p(\theta|\mathcal{D}, \mathcal{M}) \propto p(\mathcal{D}|\theta, \mathcal{M}) \, p(\theta|\mathcal{M}). \tag{8}$$

Here, $p(\mathcal{D}|\theta, \mathcal{M})$ as a function of θ is called the *likelihood function*. The constant of the proportionality is the reciprocal of $p(\mathcal{D}|\mathcal{M})$, the *evidence* for model class $\mathcal{M}$, and it is

discussed later. The posterior PDF gives the updated plausibility of each model in $\mathcal{M}$ when the information in the data $\mathcal{D}$ is incorporated.

The model classes with their likelihood functions and prior PDFs that are chosen in this work are given later in the Example section. The samples are generated from the posterior PDF in Equation (8) using the TMCMC algorithm introduced by Ching and Chen (2007), which is summarized in the Appendix.

3.2. Predictions using the Posterior Robust Predictive PDF

A common use of posterior probability distributions is to make inferences about future events based on past observations (Beck, 1989; Beck and Katafygiotis, 1998). Define X_1^M as a sequence of M observations of some vector quantity x at increasing discrete time instants, that is $X_1^M = \{x_1, \ldots, x_M\}$. Now consider the case where the first N observations, X_1^N, are contained in the data set $\mathcal{D}_N$. A robust predictive PDF may be used to make inferences about the remaining observations X_{N+1}^M based on the model class $\mathcal{M}$, with each model defined by the parameter vector θ, by applying the Theorem of Total Probability:

$$p\left(X_{N+1}^M|\mathcal{D}_N, \mathcal{M}\right) = \int p\left(X_{N+1}^M|\mathcal{D}_N, \theta, \mathcal{M}\right) p\left(\theta|\mathcal{D}_N, \mathcal{M}\right) d\theta. \tag{9}$$

If the model class $\mathcal{M}$ is globally identifiable on $\mathcal{D}_N$ then the most probable value of the parameter vector, $\hat{\theta}$, may be used with Laplace's asymptotic approximation of the integral in equation (9) to obtain an estimate of the robust predictive PDF (Beck, 1989; Beck and Katafygiotis, 1998; Papadimitriou et al., 2001):

$$p\left(X_{N+1}^M|\mathcal{D}_N, \mathcal{M}\right) \approx p\left(X_{N+1}^M|\hat{\theta}, \mathcal{M}\right). \tag{10}$$

This approximation is accurate to $O(1/N)$, so the size, N, of the dataset must be reasonably large for an acceptable approximation. Alternatively, the integral in equation (9) can be estimated by using samples $\tilde{\theta}_k, k = 1, \ldots, K$, from the posterior distribution of θ, to replace the theoretical mean in equation (9) by the corresponding sample mean. Therefore, the robust predictive PDF may be approximated by:

$$p\left(X_{N+1}^M|\mathcal{D}_N, \mathcal{M}\right) \approx \frac{1}{K} \sum_{i=1}^K p\left(X_{N+1}^M|\tilde{\theta}_k, \mathcal{M}\right). \tag{11}$$

This approximation is applicable regardless of the size, N, of the dataset $\mathcal{D}_N$ and whether or not the model class $\mathcal{M}$ is identifiable based on $\mathcal{D}_N$; its accuracy depends only on the number of samples K.

4. BAYESIAN MODEL CLASS SELECTION

Bayesian model class selection (or model comparison) is essentially Bayesian updating at the model class level to make comparisons between alternative candidate model classes for predicting the response of a system (MacKay, 2003; Beck and Yuen, 2004). It has long been recognized that comparisons between model classes should factor in not only the quality of the data fit, but also the complexity of the model. Jeffreys (1961) referred to the need for a "simplicity postulate" which would give lower prior probabilities to more complex models. The basic idea is a type of Ockham's razor that simpler models that are consistent with the data should be preferred over more complex models which offer only slight improvements in the fit to the data. Box and Jenkins (1970) emphasized this idea for time series analysis, referring to it as developing parsimonious models, but a quantitative form for a Principle of Model Parsimony was not available until the pioneering work by Akaike (1974). AIC (Akaike's Information Criterion) utilized a penalty against using a larger number of uncertain (adjustable) parameters in combination with a quantification of the model data-fit based on the log likelihood of the optimal model in the model class; however, the form of the penalty term did not have a very rigorous basis. Schwartz's BIC (1978) has a modified penalty term derived from a more rigorous asymptotic analysis of Bayesian updating at the model class level. This, and subsequent work (Gull, 1989; McKay, 1992), made it clear that Bayes' Theorem at the model class level automatically enforces model parsimony without the need for ad-hoc penalty terms.

Consider a set $\mathbf{M} \equiv \left\{ \mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_{N_M} \right\}$ of N_M candidate model classes for representing a system. Given data $\mathcal{D}$, the posterior probability of each model class $P\left(\mathcal{M}_j | \mathcal{D}, \mathbf{M}\right)$, $j = 1, \ldots, N_M$, is:

$$P\left(\mathcal{M}_j | \mathcal{D}, \mathbf{M}\right) = \frac{p\left(\mathcal{D} | \mathcal{M}_j\right) P\left(\mathcal{M}_j | \mathbf{M}\right)}{\sum\limits_{i=1}^{N_M} p\left(\mathcal{D} | \mathcal{M}_i\right) P\left(\mathcal{M}_i | \mathbf{M}\right)}. \tag{12}$$

Generally, the probability of a model class is dominated by its evidence $p\left(\mathcal{D} | \mathcal{M}_j\right)$. Gull (1989) showed that the evidence can be decomposed into a "data-fit" factor and a factor that penalizes model complexity, which he referred to as the Ockham factor, in reference to Ockham's razor: "*Pluralitas non est ponenda sine neccesitate*" ("entities should not be multiplied unnecessarily"). Thus, as noted above, Bayes' Theorem at the model class level automatically incorporates a Principle of Model Parsimony without the use of any ad-hoc concepts.

4.1. *Information-Theoretic Interpretation of the Evidence for a Model Class*

Further insight into the form of this penalty against complexity can be obtained by considering the evidence from an information-theoretic point of view, as noted in Beck and Yuen (2004) and subsequently extended by Ching et al. (2005). Since the posterior probability distribution over θ_j is by definition normalized to unity,

$$\ln\left[p\left(\mathcal{D} | \mathcal{M}_j\right)\right] = \ln\left[p\left(\mathcal{D} | \mathcal{M}_j\right)\right] \int p\left(\theta_j | \mathcal{D}, \mathcal{M}_j\right) d\theta_j. \tag{13}$$

Furthermore, since the evidence is independent of θ_j, we can bring it inside the integral, and then make substitutions according to Bayes' Theorem to find that:

$$
\begin{aligned}
\ln\left[p\left(\mathcal{D}|\mathcal{M}_j\right)\right] &= \int \ln\left[\frac{p\left(\mathcal{D}|\theta_j,\mathcal{M}_j\right)p\left(\theta_j|\mathcal{M}_j\right)}{p\left(\theta_j|\mathcal{D},\mathcal{M}_j\right)}\right]p\left(\theta_j|\mathcal{D},\mathcal{M}_j\right)d\theta_j \\
&= \int \ln\left[p\left(\mathcal{D}|\theta_j,\mathcal{M}_j\right)\right]p\left(\theta_j|\mathcal{D},\mathcal{M}_j\right)d\theta_j \\
&\quad - \int \ln\left[\frac{p\left(\theta_j|\mathcal{D},\mathcal{M}_j\right)}{p\left(\theta_j|\mathcal{M}_j\right)}\right]p\left(\theta_j|\mathcal{D},\mathcal{M}_j\right)d\theta_j.
\end{aligned}
\tag{14}
$$

This formulation for the log evidence for model class $\mathcal{M}_j$, shows that it is the difference between two terms: the first term is the posterior mean of the log-likelihood function, which is a measure of the average data fit for model class $\mathcal{M}_j$, while the second term is the relative entropy between the prior and posterior distributions (Shannon, 1948; Kullback and Leibler, 1951), which is a measure of the information gained about the parameters θ_j from the data $\mathcal{D}$. Therefore, the log evidence is comprised of a data-fit term and a term which provides a penalty against more "complex" models, that is, those that extract more information from the data. This gives an intuitive understanding of why the application of Bayes' Theorem at the model class level automatically enforces Ockham's razor. Although this information-theoretic interpretation was given in Beck and Yuen (2004), they derived it using a large-sample Laplace asymptotic approximation that depended on global identifiability of the model classes. In that case, the term representing the posterior mean of the log-likelihood function is approximated by the log-likelihood evaluated at the most probable value $\hat{\theta}_j$ of the parameter vector for model class $\mathcal{M}_j$.

4.2. *Evaluation of the Evidence for a Model Class*

Using the Theorem of Total Probability, the evidence can be expressed as:

$$
p\left(\mathcal{D}|\mathcal{M}_j\right) = \int p\left(\mathcal{D}|\theta_j,\mathcal{M}_j\right)p\left(\theta_j|\mathcal{M}_j\right)d\theta_j.
\tag{15}
$$

The most straightforward approach to evaluating the evidence, in the typical case where the integral in equation (15) is too complex to be analytically integrated and of too high a dimension for numerical quadrature, is to use stochastic simulation with samples drawn from the prior PDF $p\left(\theta_j|\mathcal{M}_j\right)$. However, it is often the case that the main support regions of the prior are very different from that of the likelihood function $p\left(\mathcal{D}|\theta_j,\mathcal{M}_j\right)$, which is usually concentrated in a relatively small volume in the parameter space for $\mathcal{M}_j$. Thus, a straightforward application of common stochastic simulation methods, such as Monte Carlo simulation, performs poorly, usually giving biased estimates of the evidence.

Beck and Yuen (2004) used an approximation for the evidence based on Laplace's method of asymptotic approximation of integrals (MacKay, 1991; Beck and Katafygiotis, 1991, 1998). This approximation is valid for globally identifiable model classes when there

is a large amount of data (which ensures that the posterior PDF is very peaked). Beck and Yuen (2004) showed that this method is effective under these conditions for model class selection for dynamical systems. In cases where these criteria are not met, a method for estimating the evidence by using samples drawn from the posterior distribution was introduced by Ching et al. (2005), in conjunction with the Gibbs sampler. However, this is not applicable when using the M-H algorithm due to the presence of repeated samples as a result of the way that the algorithm works.

The TMCMC algorithm, in addition to performing updating for a model class, also provides an estimate of the evidence for the model class. As described in Step 4 in the Appendix, when generating samples for the mth target PDF, re-sampling is performed on the samples from the $(m-1)$th level, $\tilde{\theta}_k^{(m-1)}, k = 1, \ldots, N$. The re-sampling weight for each sample, $w\left(\tilde{\theta}_k^{(m-1)}\right)$, is the ratio of the target PDFs for the m th and $(m-1)$th levels, evaluated at $\tilde{\theta}_k^{(m-1)}$:

$$
w\left(\tilde{\theta}_k^{(m-1)}\right) = \frac{p\left(\mathcal{D}|\tilde{\theta}_k^{(m-1)}, \mathcal{M}\right)^{\beta_m} p\left(\tilde{\theta}_k^{(m-1)}|\mathcal{M}\right)}{p\left(\mathcal{D}|\tilde{\theta}_k^{(m-1)}, \mathcal{M}\right)^{\beta_{m-1}} p\left(\tilde{\theta}_k^{(m-1)}|\mathcal{M}\right)} = p\left(\mathcal{D}|\tilde{\theta}_k^{(m-1)}, \mathcal{M}\right)^{\beta_m - \beta_{m-1}}. \quad (16)
$$

If N is chosen to be sufficiently large, then the samples $\tilde{\theta}_k^{(m-1)}, k = 1, \ldots, N$, are distributed according to the target PDF $f_{m-1}\left(\theta^{(m-1)}\right)$ for the $(m-1)$th level that is given by normalizing $p\left(\mathcal{D}|\theta^{(m-1)}, \mathcal{M}\right)^{\beta_{m-1}} p\left(\theta^{(m-1)}|\mathcal{M}\right)$. Therefore, the expectation of $w\left(\theta^{(m-1)}\right)$ under this PDF is approximated by the sample mean S_m of $w\left(\tilde{\theta}_k^{(m-1)}\right), k = 1, \ldots, N$:

$$
\begin{aligned}
S_m &= \frac{1}{N} \sum_{k=1}^{N} w\left(\tilde{\theta}_k^{(m-1)}\right) \approx \int \frac{p\left(\mathcal{D}|\theta, \mathcal{M}\right)^{\beta_m} p\left(\theta|\mathcal{M}\right)}{p\left(\mathcal{D}|\theta, \mathcal{M}\right)^{\beta_{m-1}} p\left(\theta|\mathcal{M}\right)} f_{m-1}\left(\theta\right) d\theta \\
&\approx \frac{\int p\left(\mathcal{D}|\theta, \mathcal{M}\right)^{\beta_m} p\left(\theta|\mathcal{M}\right) d\theta}{\int p\left(\mathcal{D}|\theta, \mathcal{M}\right)^{\beta_{m-1}} p\left(\theta|\mathcal{M}\right) d\theta}.
\end{aligned} \quad (17)
$$

The products of these means of the re-sampling weights for all of the levels during one run of the TMCMC algorithm, therefore gives the approximation:

$$
\begin{aligned}
\prod_{m=1}^{M} S_m &\approx \frac{\int p^{\beta_M}\left(\mathcal{D}|\theta, \mathcal{M}\right) p\left(\theta|\mathcal{M}\right) d\theta}{\int p^{\beta_0}\left(\mathcal{D}|\theta, \mathcal{M}\right) p\left(\theta|\mathcal{M}\right) d\theta} = \frac{\int p\left(\mathcal{D}|\theta, \mathcal{M}\right) p\left(\theta|\mathcal{M}\right) d\theta}{\int p\left(\theta|\mathcal{M}\right) d\theta} \\
&= \int p\left(\mathcal{D}|\theta, \mathcal{M}\right) p\left(\theta|\mathcal{M}\right) d\theta = p\left(\mathcal{D}|\mathcal{M}\right).
\end{aligned} \quad (18)
$$

Therefore, $\prod_{m=1}^{M} S_m$ is an estimator of the evidence, which is shown by Ching and Chen (2007) to be asymptotically unbiased.

5. EXAMPLE USING SIMULATED SEISMIC RESPONSE DATA

5.1. Simulated Structure and Responses

To investigate the performance of the TMCMC algorithm with the class of Masing models for identification of hysteretic systems, we present results of a study using simulated dynamic response data. The data is generated using a three-story Masing shear-building system described by equations (5)–(7) and the two extended Masing rules. Each story has a mass of 1.25×10^5 kg and a small-amplitude inter-story stiffness of $K_i = 2.5 \times 10^8$ N/m. The ultimate strength of each story is $r_{u,i} = 1.75 \times 10^6$ N, and the value of the elastic-to-plastic transition parameter for each story is $\alpha_i = 4$. The natural frequencies of the structure (based on a linear model given by the small-amplitude stiffnesses) are 3.17 Hz, 8.88 Hz and 12.86 Hz. The structure is excited with the Sylmar ground-motion record recorded at Olive View Hospital during the 1994 Northridge earthquake in California (COSMOS virtual data center: http://db.cosmos-eq.org/). The resulting response has peak drifts of 2.77 cm, 0.85 cm and 0.50 cm for the first, second and third stories, respectively. Viscous damping is provided in the system through a Rayleigh damping matrix (sum of a term proportional to the mass matrix M and a term proportional to the small-amplitude stiffness matrix K) with coefficients $c_M = 0.293$ and $c_K = 2.64 \times 10^{-4}$ chosen to provide 1% of critical damping in the first and second modes.

Two sets of response data are simulated to provide the data for system identification. The first set corresponds to the inter-story drift time histories, referred to henceforth as data $\mathcal{D}_{drift}$. To simulate measurement noise, a small amount of Gaussian discrete white noise is added to each channel of generated data, with a standard deviation of 0.1 cm for each measurement, which is equal to about 20% of the root-mean-square value of the drift time histories. This simulated data is shown in Figure 4. Each channel corresponds to 500 data points at a sampling interval of 0.02 s. While dynamic measurements of inter-story drift are possible, and have been performed in laboratory test structures, their measurement in structures in the field is rare. Accelerometers are more commonly used for this purpose, so the second set of data, $\mathcal{D}_{acc}$, consists of simulated absolute accelerations of each floor. Again, Gaussian discrete white noise is added to the responses, this time with a standard deviation of 0.5 m/s^2, which is equal to about 20% of the root-mean-square value of the acceleration time histories. Simulated accelerations are plotted in Figure 5. The hysteresis loops generated by the Sylmar ground motion for each story are shown in Figure 6. As indicated by the figure, yielding in the simulated structure occurs primarily in the first story, with some moderately nonlinear behavior in the second story and almost linear response for the third story.

5.2. Identification Model Classes: Likelihood Functions and Prior PDFs

Four model classes are considered for system identification. All of them use the Masing shear-building model in equations (5)–(7) to generate the predicted response, $q_t^{(i)}$, $i = 1$, ..., N_d, $t = 1$, ..., N_t, for $N_d = 3$ channels of $N_t = 500$ time-points at a time-step of 0.02 s. The vector θ of uncertain model parameters include those defining the viscous-damping matrix C, that is, c_M and c_K, and the nine hysteretic structural parameters K_i, $r_{u,i}$ and α_i, $i = 1, 2, 3$, in equation (7). The prediction error, which is defined to be the difference between the uncertain system output and the identification model output, is taken as Gaussian (based
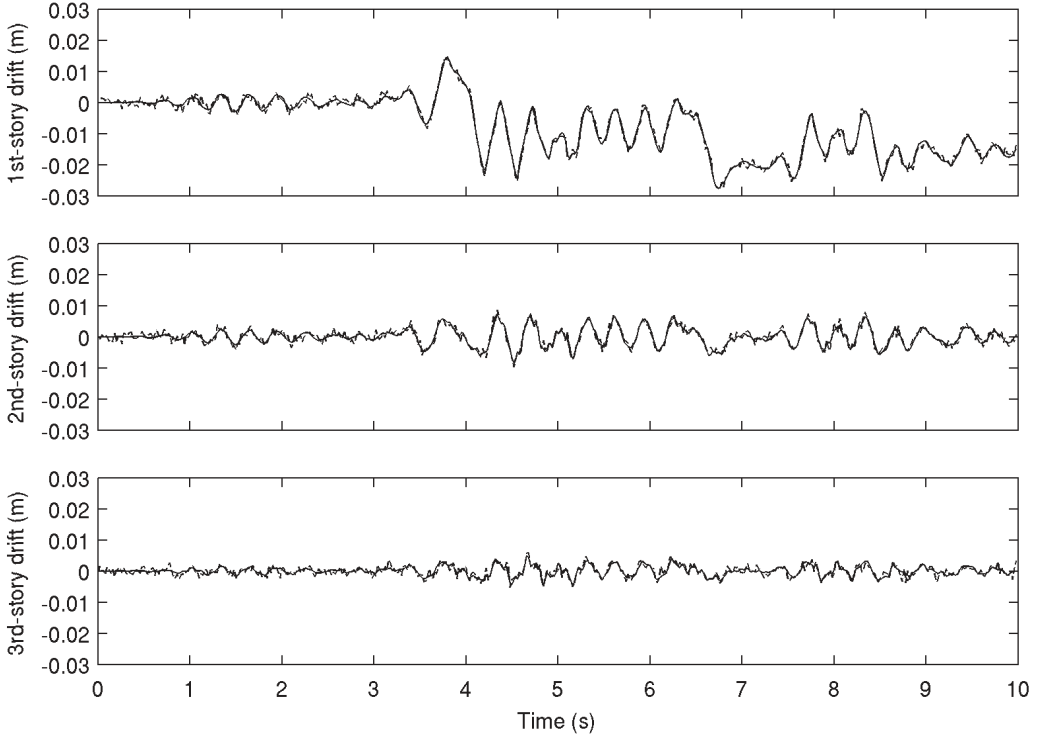
Figure 4. Simulated inter-story drift time histories. The solid plots are the actual response of the model, and the dashed plots represent the "measured" response with added white noise.

on the Principle of Maximum Information Entropy; Jaynes, 2003) with zero mean and an equal (uncertain) variance σ_q^2 for each channel. Therefore, the likelihood function is:

$$
p\left(\mathcal{D}|\theta, \mathcal{M}\right) = \frac{1}{\left(2\pi \sigma_q^2\right)^{\frac{N_t N_d}{2}}} \exp\left[-\frac{1}{2\sigma_q^2} \sum_{i=1}^{N_d} \sum_{t=1}^{N_t} \left(q_t^{(i)}(\theta) - \hat{q}_t^{(i)}\right)^2\right] \tag{19}
$$

where $\hat{q}_t^{(i)}$ is the measurement for channel i at time-point t.

The mass matrix M is equal to the actual mass matrix, which is reasonable since the mass distribution can usually be quite accurately evaluated based on structural drawings. The three small-amplitude stiffnesses and ultimate strengths of the three stories are to be estimated, along with the elastic-to-plastic transition parameters α_i, which are constrained to be equal for all three stories for model classes $\mathcal{M}_1$ and $\mathcal{M}_2$ but are allowed to vary between stories for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$. Rayleigh damping coefficients are estimated for model classes, $\mathcal{M}_2$ and $\mathcal{M}_4$, as described below.

The prior probability distributions for the small-amplitude stiffnesses K_i are taken to be independent lognormal distributions with the logarithmic mean equal to $\log\left(2.5 \times 10^8\right)$ and a logarithmic standard deviation of 0.5. The prior PDFs for the peak inter-story strengths $r_{u,i}$
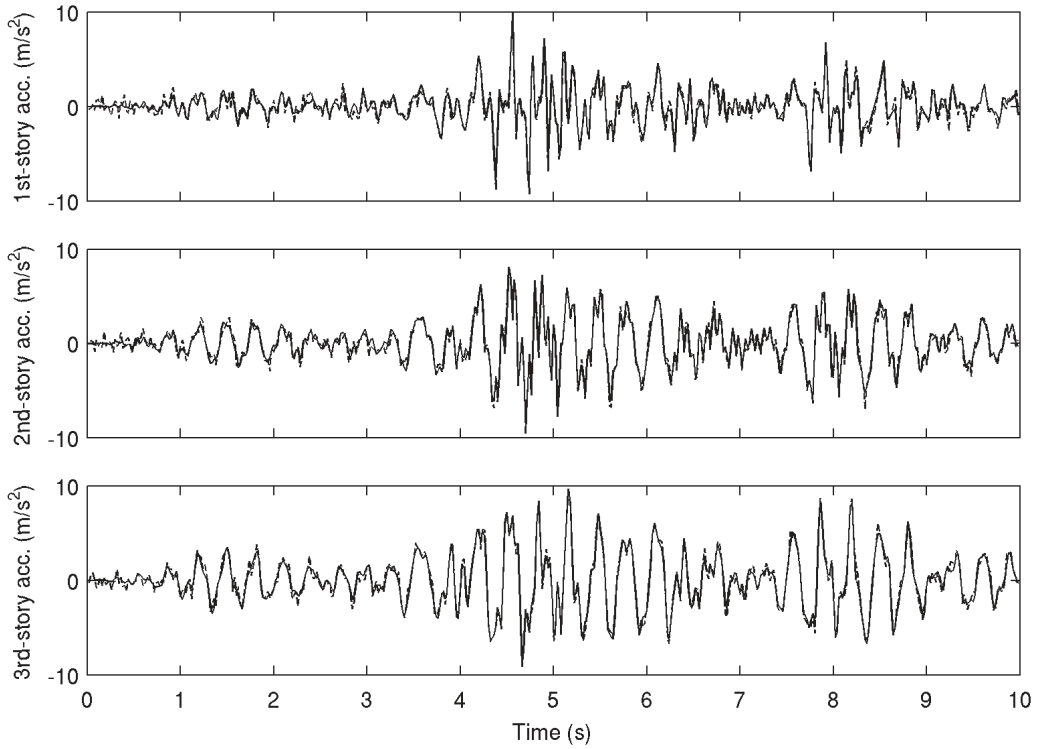
Figure 5. Simulated acceleration time histories. The solid plots are the actual response of the model, and the dashed plots represent the "measured" response with added white noise.
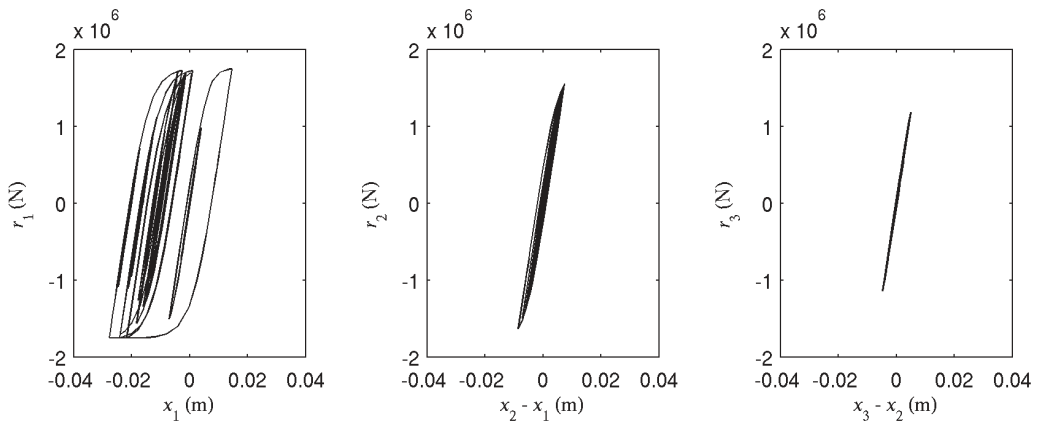

Figure 6. Simulated inter-story shear forces plotted against inter-story drifts.
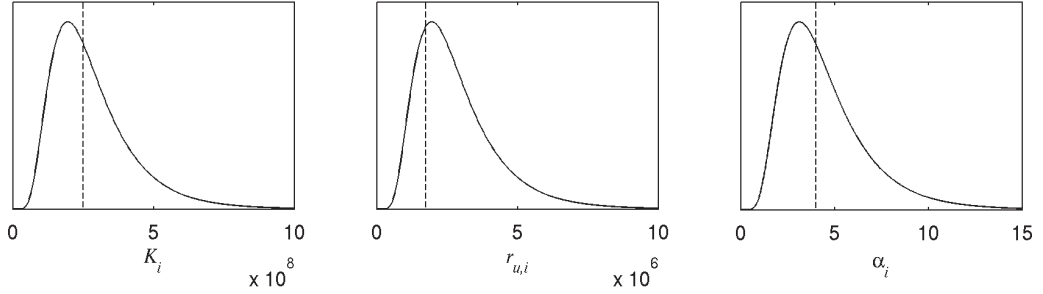
Figure 7. Prior PDFs for the small-amplitude stiffness K_i, ultimate strength $r_{u,i}$ and elastic-to-plastic transition parameter α_i for the ith story. The dashed lines indicate the values of the parameters used to generate the data.

are also independent lognormal distributions with a logarithmic means of $\log\left(2.50 \times 10^6\right)$ and a logarithmic standard deviation of 0.5. The elastic-to-plastic transition parameter α_i is also lognormally distributed with a logarithmic mean of $\log(4)$ and a logarithmic standard deviation of 0.5. Plots of the prior PDFs for K_i, $r_{u,i}$, and α_i are shown in Figure 7.

The prediction-error variance for the identification with drift records is assumed to be equal for all three stories and the prior PDF for σ_{drift}^2 is uniformly distributed between 0 and 2.5×10^{-5}, which is approximately one-half of the mean-square of the "measured" drift time histories. Similarly, the prediction-error variance for identification with acceleration records, σ_{acc}^2, is equal for all the stories and is uniformly distributed between 0 and 3, which again is approximately one-half the mean-square of the "measured" acceleration time histories.

Model classes $\mathcal{M}_1$ and $\mathcal{M}_3$ contain no Rayleigh viscous-damping matrix, whereas model classes $\mathcal{M}_2$ and $\mathcal{M}_4$ do and so damping coefficients c_M and c_K are included as uncertain parameters to be updated for those two model classes. The prior PDF for c_M is a uniform distribution between 0 and 1.5, and the prior PDF for c_K is a uniform distribution between 0 and 1.5×10^{-3}. For both coefficients, the range of the prior PDF is more than five times the actual value used to generate the data. It should be noted that the model used to generate the data is contained in model classes $\mathcal{M}_2$ and $\mathcal{M}_4$.

5.3. *Model Updating and Model Class Selection with Drift Time Histories*

For each model class, three runs were performed using the TMCMC algorithm. Each run used 1000 samples per level. Between 19 to 23 levels were needed to go from the prior PDF to the posterior PDF in a given run. Figure 8 shows how the samples in the $\left\{r_{u,1}, K_1\right\}$ space converge for a run using model class $\mathcal{M}_1$ as the tempering parameter β increases, where the model class is globally identifiable. Figure 9 shows the convergence of samples in the $\left\{r_{u,3}, K_3\right\}$ space for model class $\mathcal{M}_2$, where the model class is close to unidentifiable in the $r_{u,3}$-direction, because of the lack of third-story yielding in the data, as discussed later.

Table 1 summarizes the results of the stochastic simulations for all four model classes. The most probable values of each parameter are shown, calculated based on a lognormal fit of the aggregated samples of the three runs for each model class. This is not the same as the
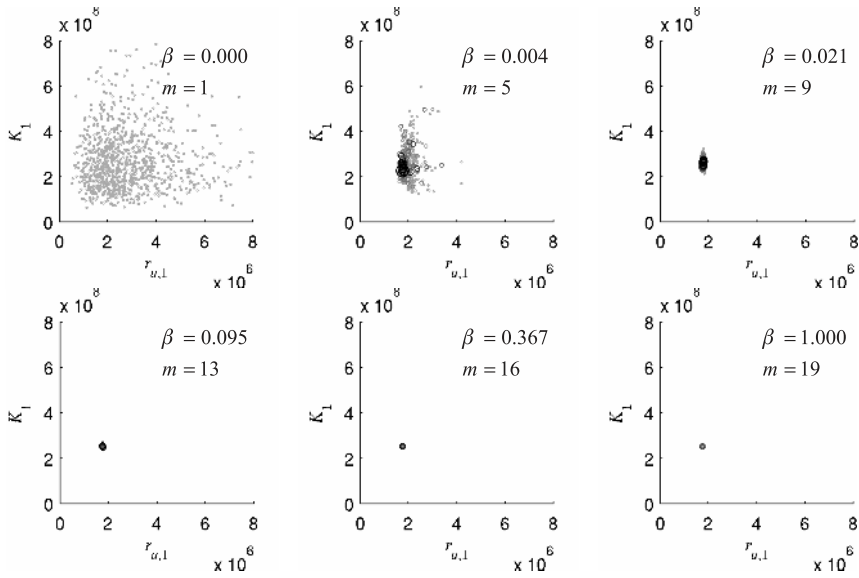
Figure 8. Plots of the samples in the $\{r_{u,1}, K_1\}$ space generated at different "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_1$ with drift data. Repeated samples are indicated by size and shade of markers.
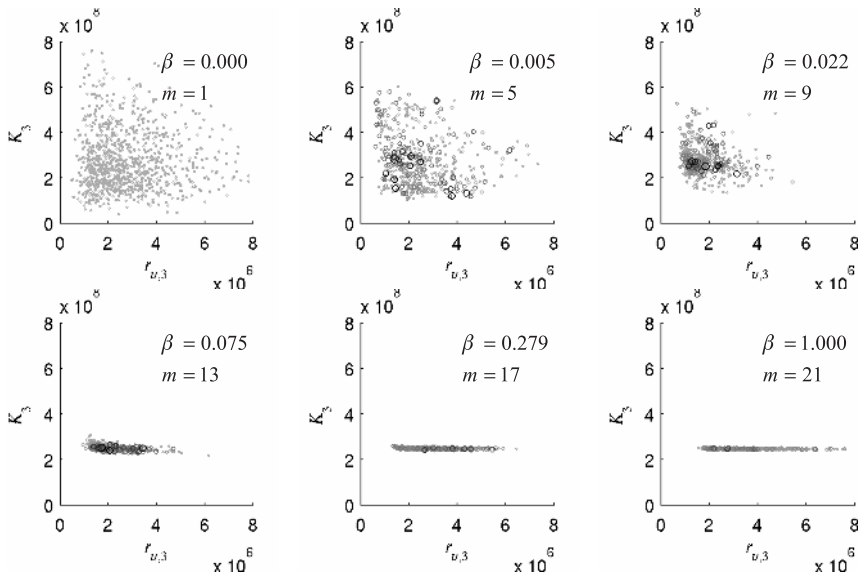

Figure 9. Plots of the samples in the $\{r_{u,3}, K_3\}$ space generated at different "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_2$ with drift data. Repeated samples are indicated by size and shade of markers.

Table 1. The most probable parameter values for model updating with drift data obtained from both a lognormal fit to the samples generated from stochastic simulation, shown with standard deviations in parentheses, and numerical optimization of the un–normali-zed posterior PDFs, shown with the difference from the simulation estimate in brackets.

	$\mathcal{M}_1$		$\mathcal{M}_2$		$\mathcal{M}_3$		$\mathcal{M}_4$	
	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.
K_1	2.523	2.525	2.501	2.504	2.531	2.525	2.510	2.507
10^8 N/m	(0.007)	[0.002]	(0.015)	[0.003]	(0.009)	[–0.006]	(0.017)	[–0.003]
K_2	2.503	2.500	2.486	2.488	2.477	2.483	2.468	2.472
10^8 N/m	(0.012)	[–0.003]	(0.021)	[0.002]	(0.019)	[0.006]	(0.026)	[0.004]
K_3	2.492	2.495	2.474	2.476	2.530	2.522	2.472	2.492
10^8 N/m	(0.013)	[0.003]	(0.017)	[–0.002]	(0.022)	[–0.008]	(0.018)	[0.020]
$r_{u,1}$	1.752	1.752	1.752	1.753	1.747	1.747	1.749	1.751
10^6 N	(0.005)	[0.000]	(0.004)	[0.001]	(0.005)	[0.000]	(0.005)	[0.002]
$r_{u,2}$	1.756	1.764	1.809	1.815	1.691	1.685	1.757	1.748
10^6 N	(0.036)	[0.008]	(0.039)	[0.006]	(0.066)	[–0.006]	(0.066)	[–0.009]
$r_{u,3}$	1.736	1.731	2.480	2.061	2.725	2.634	2.714	2.582
10^6 N	(0.129)	[–0.005]	(0.901)	[–0.419]	(0.972)	[–0.091]	(1.140)	[–0.132]
α_1	3.916	3.890	3.911	3.922	3.899	3.949	3.879	3.891
	(0.110)	[–0.026]	(0.103)	[0.011]	(0.110)	[0.060]	(0.098)	[0.012]
α_2	$=\alpha_1$	$=\alpha_1$	$=\alpha_1$	$=\alpha_1$	4.513	4.542	4.350	4.479
					(0.643)	[0.029]	(0.496)	[0.129]
α_3	$=\alpha_1$	$=\alpha_1$	$=\alpha_1$	$=\alpha_1$	1.879	2.148	3.783	2.709
					(0.445)	[0.269]	(1.571)	[–1.074]
c_M	0.000	0.000	0.293	0.294	0.000	0.000	0.261	0.299
	(fixed)	(fixed)	(0.049)	[0.001]	(fixed)	(fixed)	(0.061)	[0.038]
c_K	0.000	0.000	3.269	3.021	0.000	0.000	3.474	3.073
10^{-4}	(fixed)	(fixed)	(0.660)	[–0.248]	(fixed)	(fixed)	(1.156)	[–0.401]
σ^2	1.139	1.129	1.021	1.009	1.130	1.125	1.032	1.008
10^{-6} m^2	(0.044)	[–0.010]	(0.041)	[–0.012]	(0.043)	[–0.005]	(0.036)	[–0.024]

sample mean, although the values are close for parameters with small uncertainties. For the purposes of comparison, the most probable parameter values for each model class are also calculated by numerical optimization of the posterior PDF, and are shown in Table 1. The values obtained by optimization are within one standard deviation of the corresponding values obtained by stochastic simulation. It should be noted that convergence of the optimization algorithm was achieved using initial conditions based on the simulation results, otherwise there would have been difficulties in the optimization due to ill-conditioning (essentially a lack of identifiability), particularly for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$, as will be seen later.

For model classes $\mathcal{M}_1$ and $\mathcal{M}_2$, the estimates of the three stiffness parameters are fairly well-constrained (see Figures 8 and 9). The estimate for the first-story strength is also very tightly constrained (e.g. Figure 8 for $\mathcal{M}_1$), but the second-story strength is somewhat less pinned down and the third-story strength has a relatively large uncertainty (e.g. Figure 9 for $\mathcal{M}_2$). This behavior can be explained by looking at Figure 6; since the first story is subjected to substantial yielding, much information about its ultimate strength is available, but since no appreciable yielding occurs in the third story, the data only provides information that
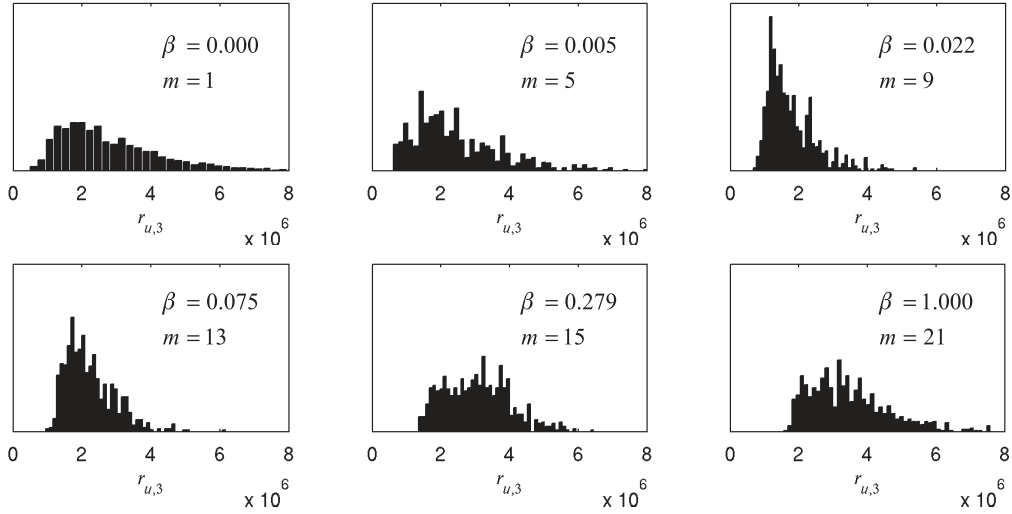
Figure 10. Normalized histograms for $r_{u,3}$ calculated using the samples shown in Figure 9 from one run of the TMCMC algorithm when updating model class $\mathcal{M}_2$ with drift data.

low values of the third-story strength are not plausible, as can by seen in Figure 9 by the way the samples move between the prior and the posterior. Figure 10 also illustrates this by showing the evolution of the histogram of $r_{u,3}$. Note that for the higher values of ultimate story strength, the histogram for the final level has the same shape as the prior PDF. This is because there is very little information in the data about the upper limit of the ultimate strength of the third story, so the posterior PDF there is limited by the prior PDF. This makes sense, as the prior information must take precedence when no useful information is provided by the data.

The estimates of the stiffness parameters for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$ are still fairly tightly constrained, as can be seen from Table 1. However, the second and third stories exhibit more complicated behavior in the $\{r_{u,i}, \alpha_i\}$ space. For model classes $\mathcal{M}_1$ and $\mathcal{M}_2$, the manifold of most probable (or plausible) models is essentially constrained to move along a curve in the parameter space where only $r_{u,3}$ varies, because the value of α_i is pinned down for all three stories by the yielding in the first story. Model class $\mathcal{M}_3$, which does not constrain the elastic-to-plastic transition parameters to be equal for all stories, has a manifold of most probable models that exhibits interaction between the ultimate story strength and elastic-to-plastic transition parameters for the second and third stories. Figure 11 shows how samples in the $\{r_{u,3}, \alpha_3\}$ space converge during one run of the TMCMC algorithm when updating model class $\mathcal{M}_3$ with the drift data. This type of geometry cannot be well-represented by an asymptotic approximation, and the numerical optimization necessary to make such an approximation is also difficult to perform without a good initial estimate of the parameters.

The samples from updating model class $\mathcal{M}_4$ are also of interest. Figure 12 shows the aggregate samples for the final level of all three simulation runs using drift data, plotted in the $\{r_{u,i}, \alpha_i\}$ space for each story. The first-story parameters are globally identifiable, as expected. The second-story parameters exhibit a manifold of most probable models similar
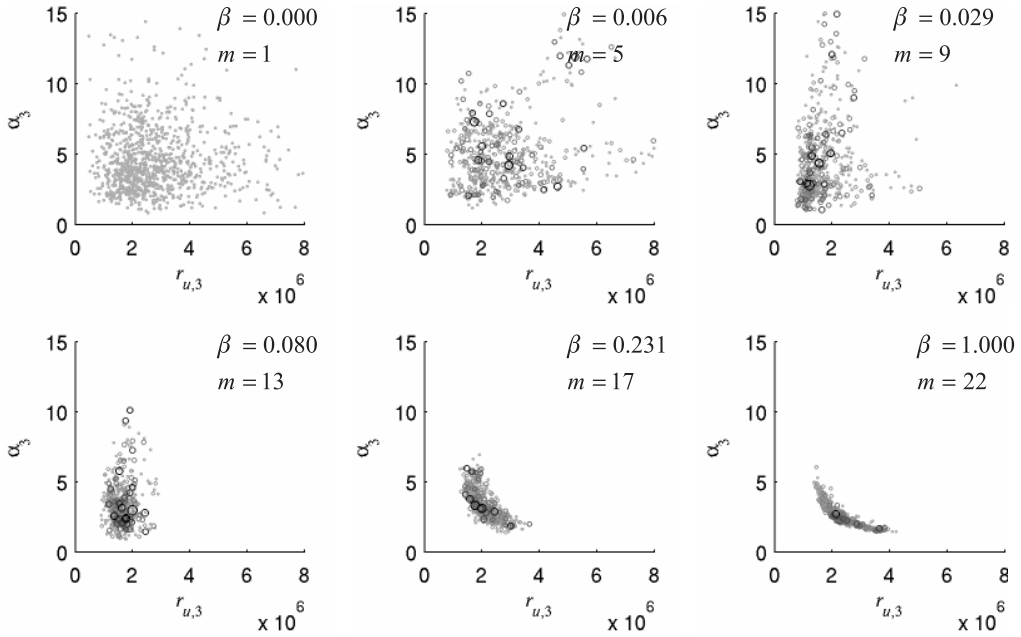
Figure 11. Plots of the samples in the $\{r_{u,3}, \alpha_3\}$ space from several "levels" of one run of the TMCMC algorithm when updating model class $\mathcal{M}_3$ with drift data. Repeated samples are indicated by size and shade of markers.

to those shown in Figure 11 for model class $\mathcal{M}_3$. However, the third story parameters are not concentrated along a manifold, but are spread through the $\{r_{u,3}, \alpha_3\}$ space. A clear lower bound on the samples is evident, but the upper bounds are essentially constrained only by the prior PDF. Since the third story experienced almost no nonlinear behavior, this lower bound on the yielding parameters is all the information that can be extracted. It is possible that for model class $\mathcal{M}_3$, the third-story samples are concentrated along this lower bound (see Figure 11) in an attempt to compensate for the lack of viscous damping, because in this region the associated hysteretic energy dissipation is largest. In model class $\mathcal{M}_4$, the identified viscous-damping parameters are fairly close to the values used to generate the data, thus no additional energy dissipation is needed, and so the parameters associated with third-story yielding are essentially free away from the "lower bound" (Figure 12). This geometry cannot be represented with any type of asymptotic approximation and it is even more challenging to perform optimization here than for model class $\mathcal{M}_3$.

Results of Bayesian model class selection for model classes updated with drift data are shown in Table 2. The log-evidence and average log-likelihood function over the posterior PDF that are estimated from stochastic simulation are shown, along with the information gain, which is not directly estimated but rather calculated from the other two quantities using equation (14). The evidence clearly favors model class $\mathcal{M}_2$, which is unsurprising as it contains the model used to generate the data and has three fewer parameters than the other model class, $\mathcal{M}_4$, containing the data-generation model. The improvement in the data-fit for
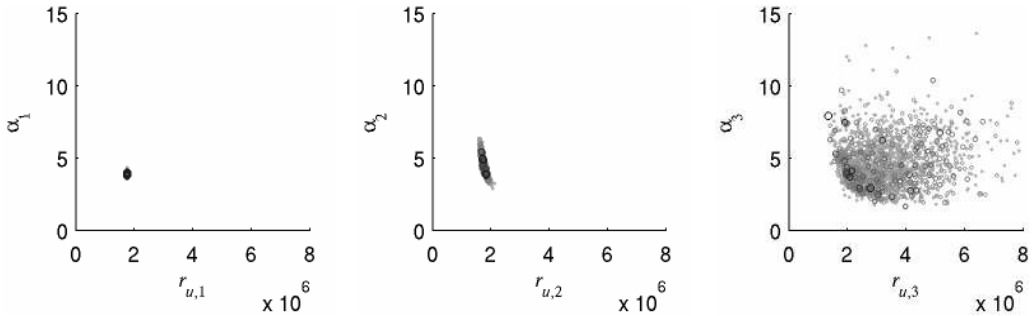
Figure 12. Aggregate samples from the final level of three runs of the TMCMC algorithm when updating model class $\mathcal{M}_4$ with drift data, plotted in the $\{r_{u,i}, \alpha_i\}$ space for all three stories.

Table 2. Bayesian model class selection results for drift data. The information gain is the difference between the average data fit and the log of the evidence, which are estimated using stochastic simulation.

Model Class	Average Data Fit	Information Gain	Log Evidence	Probability
$\mathcal{M}_1$	8138.0	43.7	8094.3	0.0000
$\mathcal{M}_2$	8219.7	45.9	8173.8	0.8022
$\mathcal{M}_3$	8140.0	50.0	8090.0	0.0000
$\mathcal{M}_4$	8220.1	47.7	8172.4	0.1978

model class $\mathcal{M}_2$ more than offsets the increased complexity of the model class (which is indicated by the larger information gain) compared to model class $\mathcal{M}_1$.

Model class $\mathcal{M}_3$ not only gives a poorer data fit than $\mathcal{M}_2$ but it also has a larger information gain, while model class $\mathcal{M}_4$ exhibits a slightly improved data fit compared to $\mathcal{M}_2$ (possibly due to over-fitting of noise), which does not balance out the extra information needed to update the additional model parameters.

The results in Table 2 also show that model class $\mathcal{M}_3$ has a larger information gain than model class $\mathcal{M}_4$, which has two additional parameters. Figures 11 and 12 may explain this somewhat counter-intuitive result. While the samples for model class $\mathcal{M}_3$ are very tightly concentrated along a manifold in the $\{r_{u,3}, \alpha_3\}$ space as seen in Figure 11, the samples for model class $\mathcal{M}_4$ are spread out through the same space, as seen in Figure 12 and as previously discussed. It seems reasonable that the extra information implied by the concentration of samples on the manifold might be greater than the information needed to identify the Rayleigh damping coefficients. This is an important point, as it shows that unlike methods such as the AIC and BIC, the penalty against model complexity in Bayesian model class selection is not based solely on the number of parameters, but rather on how much extra information is extracted from the data by the inclusion of these parameters.

Table 3. The most probable parameter values for model updating with acceleration data obtained from both a lognormal fit to the samples generated from stochastic simulation, shown with standard deviations in parentheses, and numerical optimization of the un–normalized posterior PDFs, shown with the difference from the simulation estimate in brackets.

	$\mathcal{M}_1$		$\mathcal{M}_2$		$\mathcal{M}_3$		$\mathcal{M}_4$	
	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.	Sim.	Opt.
K_1	2.503	2.501	2.500	2.500	2.523	2.526	2.500	2.499
10^8 N/m	(0.007)	[–0.002]	(0.007)	[0.000]	(0.011)	[0.003]	(0.007)	[–0.001]
K_2	2.533	2.533	2.495	2.493	2.528	2.529	2.499	2.498
10^8 N/m	(0.005)	[0.000]	(0.007)	[–0.002]	(0.006)	[–0.001]	(0.008)	[–0.001]
K_3	2.538	2.538	2.502	2.504	2.569	2.562	2.507	2.510
10^8 N/m	(0.006)	[0.000]	(0.008)	[0.002]	(0.028)	[–0.007]	(0.011)	[0.003]
$r_{u,1}$	1.808	1.806	1.747	1.747	1.814	1.814	1.746	1.746
10^6 N	(0.007)	[–0.002]	(0.007)	[0.000]	(0.006)	[0.000]	(0.024)	[0.000]
$r_{u,2}$	1.752	1.743	1.743	1.752	1.733	1.733	1.749	1.740
10^6 N	(0.019)	[–0.009]	(0.014)	[0.009]	(0.020)	[0.000]	(0.024)	[–0.009]
$r_{u,3}$	1.511	1.516	1.771	1.752	2.016	2.230	2.417	2.389
10^6 N	(0.020)	[0.005]	(0.064)	[–0.019]	(0.870)	[0.214]	(0.795)	[0.028]
α_1	3.873	3.922	4.023	4.032	3.771	3.749	4.056	4.068
	(0.087)	[0.049]	(0.075)	[0.009]	(0.080)	[–0.022]	(0.079)	[0.013]
α_2	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	4.075	4.050	3.931	4.015
					(0.193)	[–0.025]	(0.191)	[0.084]
α_3	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	$= \alpha_1$	1.826	2.014	2.456	2.690
					(0.814)	[0.188]	(0.735)	[0.234]
c_M	0.000	0.000	0.296	0.303	0.000	0.000	0.303	0.310
	(fixed)	(fixed)	(0.029)	[0.007]	(fixed)	(fixed)	(0.028)	[0.009]
c_K	0.000	0.000	2.801	2.764	0.000	0.000	2.788	2.721
10^{-4}	(fixed)	(fixed)	(0.159)	[0.037]	(fixed)	(fixed)	(0.167)	[–0.067]
σ^2	0.496	0.493	0.271	0.270	0.482	0.472	0.274	0.269
m/s^2	(0.019)	[–0.003]	(0.010)	[–0.001]	(0.018)	[–0.010]	(0.010)	[–0.005]

5.4. *Model Updating and Model Class Selection with Acceleration Time Histories*

The procedures applied to the drift response were also applied to the acceleration response. Three runs were performed using the TMCMC algorithm for each model class, each run using 1000 samples per level and needing between 19 to 23 levels to go from the prior PDF to the posterior PDF in a given run. Table 3 summarizes the results of the stochastic simulations for all four model classes, and the most probable parameter values obtained using optimization. Again, the differences in the most probable parameter values obtained from simulation and optimization are less than one standard deviation.

In general, the parameter uncertainties are less than those for the case where drift data is used for updating, possibly because the acceleration records are richer in higher frequencies, so model behavior in the elastic-to-plastic transition region plays a more significant role
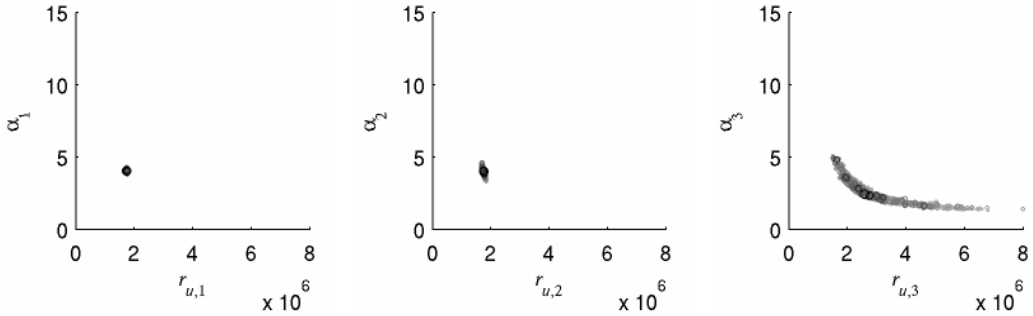
Figure 13. Aggregate samples from the final level of three runs of the TMCMC algorithm when updating model class $\mathcal{M}_4$ with acceleration data, plotted in the $\{r_{u,i}, \alpha_i\}$ space for all three stories.

Table 4. Bayesian model class selection results for acceleration data. The information gain is the difference between the average data fit and the log of the evidence, which are estimated using stochastic simulation.

Model Class	Average Data Fit	Information Gain	Log Evidence	Probability
$\mathcal{M}_1$	−1602.7	43.5	−1646.2	0.0000
$\mathcal{M}_2$	−1150.8	53.4	−1204.2	0.9989
$\mathcal{M}_3$	−1581.9	53.7	−1635.6	0.0000
$\mathcal{M}_4$	−1150.7	60.3	−1211.0	0.0011

in the response and constrains the corresponding model parameters more than for updating with drift records. Model class $\mathcal{M}_1$ exhibits a much lower identified value for the ultimate strength of the third story. This behavior is probably an attempt to compensate for the energy dissipated by the stiffness-proportional component of the viscous damping by using the hysteretic damping. The value may be lower than that obtained by updating with drift data because the offset which results from yielding, while important in the drift response, is at low frequencies and therefore is not so strongly represented in the acceleration response.

Of particular note is the identification of model class $\mathcal{M}_4$. Figure 13 shows plots of the aggregate samples plotted in the $\{r_{u,i}, \alpha_i\}$ space. Unlike the samples obtained by updating with drift records that are shown in Figure 12, the third-story yielding parameters are concentrated along a well-defined manifold in the $\{r_{u,3}, \alpha_3\}$, perhaps due to the increased importance of the high-frequency response in the acceleration records, which provides more information for the updating.

Results for model class selection using the acceleration data are shown in Table 4. Again, model class $\mathcal{M}_2$ is preferred, by an even wider margin than for updating with drift data, since the improvement in data fit for model class $\mathcal{M}_4$ is negligible. It is also of some interest that model class $\mathcal{M}_3$ is preferred to model class $\mathcal{M}_1$, while the opposite was true when drift data was used for updating. The improvement in the data fit is large enough to justify the extra information extracted from the data.

6. CONCLUDING REMARKS

Identification with the class of hysteretic Masing shear-building models and the TMCMC stochastic simulation algorithm showed good agreement with identification by optimization, where the latter is applicable, and the uncertainties associated with the parameters provided insights into the system, as did the results of the Bayesian model class selection. The example presented demonstrated the ability of the TMCMC algorithm to identify regions of high probability content that cannot be well-represented by asymptotic approximation. Additionally, TMCMC readily allows for the implementation of Bayesian model class selection, which was successfully demonstrated as a method for making quantitative comparisons between competing model classes that automatically selects models that balance complexity against consistency with the measured data. The results also illustrate the usefulness of the information-theoretic interpretation of Bayesian model class selection, as described in Section 4.1; model complexity is not merely a matter of the number of uncertain (adjustable) parameters, but rather relates to the amount of information extracted from the data by the model class.

The example also demonstrated that, although the viscous-damping forces are relatively small compared to the restoring forces, the exclusion of viscous damping from identification models can significantly alter the identified hysteretic structural parameters if viscous (i.e. rate-dependent) damping is actually present in the structure. This is problematic, because while linear viscous damping is perhaps the most commonly-used method for introducing small-amplitude energy-dissipation into structural models, it is not clear that such rate-dependent energy dissipation mechanisms are really present in civil structures. Indeed, the coefficients for a viscous-damping model are very difficult to establish from structural drawings during design. Furthermore, linear viscous-damping models, particularly Rayleigh damping, have the potential to introduce large spurious forces into the calculated structural response (Hall, 2006). In fact, the more likely source of small-amplitude energy dissipation is friction from non-structural elements, which could be well-modeled by hysteretic elements. This suggests that future research might study the development of initial loading curves for Masing models which are based on ultimate strength distributions with a larger number of elements yielding at lower force levels than in the sub-class of Masing models considered in this work, in order to create a bi-modal distribution for $\phi(r^*)$ in equation (2).

Other plans for future work in this area include a study of more complicated hysteretic systems for generating data, particularly system models that are not contained in any of the candidate model classes and, ultimately, the application of stochastic simulation methods and Masing models to real-world data from structures that have experienced significant yielding during earthquakes. In this case, since substantial yielding is often associated with damage, application of stochastic simulation techniques to the class of degrading Distributed-Element models, or the equivalent Masing models, with a finite (Cifuentes, 1983) and infinite (Chiang, 1992) number of internal elements, may be important for more realistic modeling of hysteretic structural behavior, while still employing models simple enough to be used in design, monitoring and control applications.

APPENDIX: TRANSITIONAL MARKOV CHAIN MONTE CARLO METHOD

The TMCMC algorithm, introduced by Ching and Chen (2007), is a method for sampling the posterior PDF of the model class $\mathcal{M}$, which is defined by the parameter vector θ, the prior PDF over the parameter vector, $p\left(\theta|\mathcal{M}\right)$, and the likelihood function $p\left(\mathcal{D}|\theta,\mathcal{M}\right)$, updated with data $\mathcal{D}$. This is done by sampling from a sequence of non-normalized intermediate PDFs, $f_m\left(\theta\right)$, $m = 0, \ldots, M$, given by:

$$f_m\left(\theta\right) \propto p\left(\mathcal{D}|\theta,\mathcal{M}\right)^{\beta_m} p\left(\theta|\mathcal{M}\right) \tag{A.1}$$

where the tempering parameter β_m increases monotonically with m such that $\beta_0 = 0$ and $\beta_M = 1$. The algorithm proceeds as follows:

TMCMC Algorithm:

1. $f_0\left(\theta\right) = p\left(\theta|\mathcal{M}\right)$ is chosen by the user and is assumed to be in a form that can be sampled to obtain $\tilde{\theta}_k^{(0)}$, $k = 1, \ldots, N_0$.

2. The value of the tempering parameter for the next level, β_1, is chosen such that the coefficient of variation for $\left\{ p\left(\mathcal{D}|\tilde{\theta}_k^{(0)},\mathcal{M}\right)^{\beta_1 - \beta_0}, k = 1, \ldots, N_0 \right\}$ is equal to some prescribed target value.

3. The plausibility weights for each sample, $w\left(\tilde{\theta}_k^{(0)}\right) = p\left(\mathcal{D}|\tilde{\theta}_k^{(0)},\mathcal{M}\right)^{\beta_1 - \beta_0}$ are calculated for $k = 1, \ldots, N_0$, and so is the sample mean S_1 of the N_0 plausibility weights.

4. Samples $\tilde{\theta}_k^{(1)}$, $k = 1, \ldots, N_1$ are generated by applying the Metropolis–Hastings algorithm as follows: the kth sample is drawn from a Markov Chain that starts with the lead sample, θ_{lead}, equal to one of the samples $\left\{ \tilde{\theta}_i^{(0)}, i = 1, \ldots, N_0 \right\}$ where the probability that $\theta_{lead} = \tilde{\theta}_i^{(0)}$ is given by $w\left(\tilde{\theta}_i^{(0)}\right) / \sum_{j=1}^{N_0} w\left(\tilde{\theta}_j^{(0)}\right)$. The Metropolis–Hastings algorithm is applied using a Gaussian proposal PDF that is centered at the current sample in the ith chain; that is, if $\tilde{\theta}_i^{(0)}$ is chosen as the lead sample for, say, the third time, the proposal PDF is centered at the second sample in the chain that has $\tilde{\theta}_i^{(0)}$ as the lead. The covariance matrix for the proposal PDF, Σ_m is given by

$$\Sigma_m = c_m^2 \sum_{i=1}^{N_0} w\left(\tilde{\theta}_i^{(0)}\right) \left(\tilde{\theta}_i^{(0)} - \bar{\theta}^{(0)}\right) \left(\tilde{\theta}_i^{(0)} - \bar{\theta}^{(0)}\right)^T \tag{A.2}$$

where

$$\bar{\theta}^{(0)} = \sum_{j=1}^{N_0} w\left(\tilde{\theta}_j^{(0)}\right) \tilde{\theta}_j^{(0)} / \sum_{l=1}^{N_0} w\left(\tilde{\theta}_l^{(0)}\right) \tag{A.3}$$

and c_m^2 is a control parameter that is chosen to balance the potential for large MCMC moves with maintaining a reasonable rejection rate.
5. Steps 2 through 4 are repeated until level M, where $\beta_M = 1$, is achieved.

Section 4.2 shows how the evidence $p(\mathcal{D}|\mathcal{M})$ for the model class $\mathcal{M}$ may be estimated using the TMCMC algorithm. Details and proofs concerning the statistical properties of the estimator are given in Ching and Chen (2007).

REFERENCES

Akaike, H., 1974, "A new look at the statistical identification model," *IEEE Transactions on Automatic Control* **19**, 716–723.

Ashrafi, S. A., Smyth, A.W., and Betti, R., 2005, "A parametric identification scheme for non–deteriorating and deteriorating non–linear hysteretic behavior." *Structural Control and Health Monitoring*, **13**, 108–131.

Beck, J. L., 1978, "Determining models of structures from earthquake records." Technical Report EERL 78–01, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, CA.

Beck, J. L., 1989, "Statistical system identification of structures," in *Proceedings of the 5th International Conference on Structural Safety and Reliability*, San Francisco, CA.

Beck, J. L., 1996, "System identification methods applied to measured seismic response," in *Proceedings of the 11th World Conference on Earthquake Engineering*, Acapulco, Mexico.

Beck, J. L. and Au, S.–K., 2000, "Updating robust reliability using Markov Chain simulation," in *Proceedings of the International Conference on Monte Carlo Simulation*, Monte Carlo, Monaco.

Beck, J. L., and Au, S.–K., 2002, "Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation," *Journal of Engineering Mechanics* **128**, 380–391.

Beck, J.L. and Jayakumar, P., 1996, "Class of Masing models for plastic hysteresis in structures," in *Proceedings of ASCE Structures Congress XIV*, Chicago, IL.

Beck, J. L. and Katafygiotis, L. S., 1991, "Updating of a model and its uncertainties utilizing dynamic test data," in *Proceedings of the First International Conference on Computational Stochastic Mechanics*, Corfu, Greece.

Beck, J. L. and Katafygiotis, L. S., 1998, "Updating models and their uncertainties – Bayesian statistical framework," *Journal of Engineering Mechanics* **124**, 455–461.

Beck, J.L. and Yuen, K.–V., 2004, "Model selection using response measurements: Bayesian probabilistic approach," *Journal of Engineering Mechanics* **130**, 192–203.

Benedettini, F., Capecchi, D., and Vestroni, F., 1995, "Identification of hysteretic oscillators under earthquake loading by nonparametric models," *Journal of Engineering Mechanics* **121**, 606–612.

Box, G. E. P. and Jenkins, G. M., 1970, *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, CA.

Box, G. E. P. and Tiao, G. C., 1992, *Bayesian Inference in Statistical Analysis*, Wiley, New York.

Ching, J. and Chen, Y.–J., 2007, "Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging," *Journal of Engineering Mechanics*, in press.

Ching, J., Muto, M., and Beck, J. L., 2005, "Bayesian linear structural model updating using Gibbs sampler with modal data," in *Proceedings of the 9th International Conference on Structural Safety and Reliability*, Rome, Italy,

Ching, J., Muto, M., and Beck, J. L., 2006, "Structural model updating and health monitoring with incomplete modal data using Gibbs sampler," *Computer-Aided Civil and Infrastructure Engineering* **21**, 242–257.

Chiang, D.-Y., 1992, "Parsimonious modeling of inelastic systems." Technical Report EERL 92–02, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, CA.

Cifuentes, A. O., 1984, "System identification of hysteretic structures." Technical Report EERL 84-04, California Institute of Technology, Pasadena, CA.

Cifuentes, A. O. and Iwan, W. D., "Nonlinear system identification based on modeling of restoring force behavior," *Soil Dynamics and Earthquake Engineering* **8**, 2–8.

Cox, R. T., 1961, *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore, MD.

Fishman, G. S., 1996, *Monte Carlo: Concepts, Algorithms, and Applications*, Springer, New York.

Geman, S. and Geman, D., 1984, "Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., 1996, *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Gull, S. F., 1989, "Bayesian inductive inference and maximum entropy," in *Maximum Entropy and Bayesian Methods*, J. Skilling (ed.), Kluwer, Dordrecht and Boston, MA.

Hall, J. F., 2006, "Problems encountered from the use (or misuse) of Rayleigh damping," *Earthquake Engineering and Structural Dynamics* **35**, 525–545.

Hastings, W. K., 1970, "Monte Carlo sampling methods using Markov Chains and their applications," *Biometrika* **57**, 97–109.

Iwan, W. D., 1966, "A distributed-element model for hysteresis and its steady-state dynamic response," *Journal of Applied Mechanics* **33**, 893–900.

Iwan, W. D., 1967, "On a class of models for the yielding behavior of continuous or complete systems," *Journal of Applied Mechanics* **34**, 612–617.

Jayakumar, P., 1987, "Modeling and identification in structural dynamics." Technical Report EERL 87–01, California Institute of Technology, Pasadena, CA.

Jayakumar, P. and Beck, J. L., 1988, "System identification using nonlinear structural models," in *Structural Safety Evaluation based on System Identification Approaches*, H. G. Natke and J. T. P. Yao (eds), Vieweg, Wiesbaden.

Jaynes, E. T., 2003, *Probability Theory: the Logic of Science*, Cambridge University Press, Cambridge and New York.

Jeffreys, H., 1961, *Theory of Probability*, Clarendon Press, Oxford.

Katafygiotis, L. S. and Lam, H.–F., 2002, "Tangential–projection algorithm for manifold representation in unidentifiable model updating problems," *Earthquake Engineering and Structural Dynamics* **31**, 791–812.

Kullback, S. and Leibler, R. A., 1951, "On information and sufficiency," *The Annals of Mathematical Statistics* **22**, 79–86.

Masing, G., 1926, "Eigenspannungen und verfestigung beim messing (Self–stretching and hardening for brass)," in *Proceedings of the Second International Congress on Applied Mechanics*, Zurich, Switzerland.

MacKay, D. J. C., 1991, "Bayesian methods for adaptive models," PhD thesis, California Institute of Technology, Pasadena, CA.

MacKay, D. J. C., 1992, "Bayesian interpolation," *Neural Computation* **4**, 415–447.

MacKay, D. J. C., 2003, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge and New York.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics* **21**, 1087–92.

Neal, R. M., 1993, "Probabilistic inference using Markov Chain Monte Carlo methods." Technical Report CRG–TR–93–01, Department of Computer Science, University of Toronto, Canada.

Papadimitriou, C., Beck, J. L., and Katafygiotis, L. S., 2001, "Updating robust reliability using structural test data," *Probabilistic Engineering Mechanics* **16**, 103–113.

Papadimitriou, C., and Katafygiotis, L. S., 2005, "Bayesian modeling and updating," in *Engineering Design Reliability Handbook*, D. M. Ghiocel and N. Nikolaidis (eds), CRC Press, Boca Raton, FL.

Robert, C. P. and Casella, G., 2004, *Monte Carlo Statistical Methods*, 2nd edn, Springer, New York.

Rubinstein, R. Y., 1981, *Simulation and the Monte Carlo Method*, Wiley, New York.

Schwartz, G., 1978, "Estimating the dimension of a model," *Annals of Statistics* **6**, 461–464.

Shannon, C. E., 1948, "A mathematical theory of communication." *The Bell System Technical Journal* **27**, 379–423, 623–656.

Thyagarajan, R. S., 1989, "Modeling and analysis of hysteretic structural behavior." Technical Report EERL 89–03, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, CA.

Valanis, K. C., 1971, "A theory of viscoplasticity without a yield surface," *Architectural Mechanics* **23**, 517–551.

Wen, Y.–K., 1976, "Method for random vibration of hysteretic systems," *Journal of Engineering Mechanics* **102**, 249–263.