# PRIOR AND POSTERIOR ROBUST STOCHASTIC PREDICTIONS FOR DYNAMICAL SYSTEMS USING PROBABILITY LOGIC

*James L. Beck*[1,*]   &   *Alexandros A. Taflanidis*[2]

[1] *Departments of Computing and Mathematical Sciences, and Mechanical and Civil Engineering, California Institute of Technology, Pasadena, CA 91125, USA*

[2] *Department of Civil Engineering and Geological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA*

*An overview is given of a powerful unifying probabilistic framework for treating modeling uncertainty, along with input uncertainty, when using dynamic models to predict the response of a system during its design or operation. This framework uses probability as a multivalued conditional logic for quantitative plausible reasoning in the presence of uncertainty due to incomplete information. The fundamental probability models that represent the system's uncertain behavior are specified by the choice of a stochastic system model class: a set of input–output probability models for the system and a prior probability distribution over this set that quantifies the relative plausibility of each model. A model class can be constructed from a parametrized deterministic system model by stochastic embedding which utilizes Jaynes' principle of maximum information entropy. Robust predictive analyses use the entire model class with the probabilistic predictions of each model being weighted by its prior probability, or if response data are available, by its posterior probability from Bayes' theorem for the model class. Additional robustness to modeling uncertainty comes from combining the robust predictions of each model class in a set of competing candidates weighted by the prior or posterior probability of the model class, the latter being computed from Bayes' theorem. This higher-level application of Bayes' theorem automatically applies a quantitative Ockham razor that penalizes the data-fit of more complex model classes that extract more information from the data. Robust predictive analyses involve integrals over high-dimensional spaces that usually must be evaluated numerically by Laplace's method of asymptotic approximation or by Markov chain Monte Carlo methods. These computational tools are demonstrated in an illustrative example involving the vertical dynamic response of a car being driven along a rough road.*

**KEY WORDS:** *dynamical systems, stochastic modeling, robust stochastic analysis, system identification, Bayesian updating, model class assessment, stochastic simulation*

## 1. INTRODUCTION

A common practice during dynamic design of any system is to use a single mathematical model to predict its dynamic response to prescribed external excitations (disturbances). Often this model is developed using finite-element software. The system model predictions, on their own, are not very useful, however, unless they give information about their accuracy. The response predictions will have uncertain accuracy not only because of the uncertainty in the future excitations but also because the mathematical system model will always involve approximations of the real dynamic behavior that produce uncertain affects in the predicted response; in addition, the system model will usually involve parameters whose values are uncertain. This modeling uncertainty, in addition to future excitation uncertainty, should be explicitly treated when making predictive analyses.

---

*Correspond to James L. Beck, E-mail: jimbeck@caltech.edu

In the case of an existing system where response sensor data are available, the modeling uncertainty can be reduced by updating the mathematical model of the system, thereby allowing more accurate predictions of its future response to specified excitations. This process is commonly called *system identification* and, as usually practiced, it consists primarily of *parameter estimation*. Since any system model is usually only an approximate representation of the real system behavior, there are no true values of the parameters. Therefore, determining a single "best" value for the parameter vector, such as its least-squares estimate or maximum likelihood estimate, is a questionable procedure. A further complication is that complex multiparameter system models are usually unidentifiable based on the sensor data, meaning there are multiple best estimates of the parameters.

The tasks of explicitly quantifying modeling and excitation uncertainty in response predictions during design and operation can be done in a rigorous probabilistic manner. The theory for treating excitation uncertainty, known as *random vibrations*, or more commonly in recent years, as *stochastic dynamics* (or *mechanics*), has a long history. The theory and computational tools for a probabilistic treatment of modeling uncertainty are more recent. Their development was hampered by the commonly taught restrictive interpretation of probability as the relative frequency of "inherently random" events in the "long run," which does not provide a meaning for the probability of a parameter value or a model.

Our goal in this paper is to describe a general stochastic (i.e., probabilistic) framework for handling both modeling and excitation uncertainty when predicting the dynamic response of a system based on mathematical models. It uses an interpretation of probability as a logic for quantitative plausible reasoning when there is uncertainty due to incomplete information. The foundations of probability logic are due to the physicists Cox [1, 2] and Jaynes [3, 4]. The vague and speculative concept of inherent randomness is not needed. We consider both *prior robust stochastic analysis*, which is appropriate during system design, and *posterior robust stochastic analysis*, which can be performed for an existing system if response sensor data are available. Before giving an overview of the theory for these robust predictive analyses, we first provide a brief summary of probability logic and then we define a stochastic system model class which provides the required fundamental probability models for robust stochastic analyses.

## 2. PROBABILITY LOGIC

In probability logic, probability is viewed as a multivalued conditional logic for plausible reasoning that extends binary Boolean propositional logic to the case of incomplete information. The probability $P[b|c]$ is interpreted as the degree of plausibility of the proposition (statement) $b$ based on the information in the proposition $c$ where $c$ is only conditionally asserted. This interpretation is consistent with the Bayesian perspective that probability represents a degree of belief in a proposition; indeed, probability logic provides a rigorous foundation for the Bayesian approach. Since it is not widely known, we give a brief overview to provide a basis for interpreting the probabilistic framework presented later.

For a propositional calculus of plausible reasoning involving probabilities, we need to evaluate the following probabilities in terms of more basic ones: $P[\sim b|c]$, $P[a \& b|c]$ and $P[a\ or\ b|c]$, which correspond, respectively, to the degree of plausibility based on $c$ that $b$ is not true, that both $a$ and $b$ are true, and that either $a$ or $b$ (or both) are true. Cox [1] *derived* the appropriate calculus by extending the axioms of Boolean logic which deal with the special case of complete information where the truth or falsity of $b$ is known from $c$, that is, $P[b|c] = 1$ or $P[b|c] = 0$, respectively.

Cox's results can be stated as a minimal set of axioms for probability logic. For any propositions $a$, $b$, $c$

  (i)  $P[b|c] \geq 0$

 (ii)  $P[\sim b|c] = 1 - P[b|c]$                    (Negation function)

(iii)  $P[a\&b|c] = P[a|b\&c]P[b|c]$        (Conjunction function)

Using the last two axioms and De Morgan's law from Boolean logic, we can derive

$$P[a\ or\ b|c] = P[a|c] + P[b|c] - P[a\ \&\ b|c] \qquad \text{(Disjunction function)}$$

The axioms for a probability measure $P(A)$ on subsets $A$ of a finite set $X$, as stated by Kolmogorov [5] and commonly given in textbooks on probability theory, can be derived as a special case of the probability logic axioms where

the propositions refer to uncertain membership of an object in a set [6]. For example, if $X$ denotes the set of possible values for an uncertain-valued variable $x$, then for any subset $A$ of $X$, $P(A)$ can be interpreted as $P[x \in A | \pi]$, where $\pi$ denotes the proposition that states $x \in X$ and specifies the probability model for $x$ quantifying the relative degree of plausibility of each value of $x$ in $X$. Kolmogorov also defines conditional probability in terms of unconditional probabilities, but in probability logic all probabilities are inherently conditional and so the corresponding result appears as an axiom (conjunction function).

The probability logic axioms therefore provide a calculus for handling variables whose values are uncertain because of missing information. The vague concept of *inherent* randomness, whose existence is often postulated but cannot be proved, is not needed. The axioms apply not only to variables that correspond to physical quantities but also to models and model parameters, in contrast to the relative frequency interpretation of Kolmogorov's axioms. This allows robust probabilistic predictions that account for modeling uncertainty.

## 3. STOCHASTIC SYSTEM MODEL CLASSES

### 3.1 Definition of a Model Class

In modeling the I/O (input–output) behavior of a real system, one cannot expect any chosen deterministic model to make perfect predictions, and the prediction errors of any such model will be uncertain. This motivates the introduction of a *stochastic system* (or Bayesian) *model class* $\mathcal{M}$ that consists of fundamental probability models to describe the uncertain I/O behavior of the system: a set of *I/O probability models* $\{p(\mathbf{y}|\mathbf{u},\theta,\mathcal{M}): \theta \in \Theta \subset \mathbb{R}^{N_\mathrm{p}}\}$ (sometimes called *forward models*) and a *prior probability model* $p(\theta|\mathcal{M})d\theta$ that expresses the initial probability of each model $p(\mathbf{y}|\mathbf{u},\theta,\mathcal{M})$, that is, the prior gives a measure of the initial relative plausibility of the I/O probability models corresponding to each value of the parameter vector $\theta$. Here, $\mathbf{u}$ and $\mathbf{y}$ denote the system input and output vectors that consist of discretized time histories of the excitation and corresponding system response. The prior probability model is taken to be independent of input $\mathbf{u}$ and we use $p(.)$ for a PDF (probability density function) of a variable but $P[.]$ for a probability of a statement.

The probability models defining the model class $\mathcal{M}$ are viewed as representing a state of knowledge about the system conditional on the available incomplete information and not as its inherent properties. All probabilistic predictions for the system are conditional on these chosen fundamental probability models, which we make explicit in the notation by conditioning on $\mathcal{M}$.

### 3.2 Model Class Construction by Stochastic Embedding

Any deterministic dynamic model of a system that involves uncertain parameters (e.g., a finite-element model) can be used to construct a model class $\mathcal{M}$ for the system by *stochastic embedding* [7]. Suppose that the deterministic model defines an implicit or explicit mathematical relationship $\mathbf{q}(\mathbf{u},\theta)$ between the input $\mathbf{u}$ and model output $\mathbf{q}$ where both are discretized time histories and the uncertain model parameters are denoted by $\theta$. The first step is to introduce the uncertain prediction-error time history $\mathbf{e}$ [8] as the difference between the *real system output* $\mathbf{y}$ and the *model output* $\mathbf{q}$ for the same input, i.e.,

$$\mathbf{y} = \mathbf{q} + \mathbf{e} \tag{1}$$

so $\mathbf{e}$ provides a bridge between the model world and the real world.

The next step is to establish a parametrized probability model for $\mathbf{e}$ by using the principle of maximum information entropy [3], which states that the probability model should be selected to produce the most uncertainty (largest Shannon entropy) subject to parametrized constraints that we wish to impose; the selection of any other probability model would lead to an unjustified reduction in the amount of prediction uncertainty. The maximum-entropy probability model is therefore conservative in the sense that it gives the greatest uncertainty in the prediction-error time history, and hence in the system-output time history, conditional on what one is willing to assert about the system. This is a very principled way of choosing the fundamental I/O probability model for a system in order to cover missing information about the system's behavior.

A simple choice for the probability model for **e** is produced by choosing the following constraints during entropy maximization: zero prediction-error mean at each time (any uncertain bias can be added to **q** as another uncertain parameter), and a parametrized prediction-error variance or covariance matrix at each time. The maximum entropy PDF for the prediction error **e** over an unrestricted range is then discrete-time Gaussian white noise. Therefore, the I/O probability model for the system output $y_n \in \mathbb{R}^{N_o}$ at discrete time $t_n$, conditional on the parameter vector $\theta$, is given by the following Gaussian PDF with the mean equal to the model output $q_n(\mathbf{u},\theta) \in \mathbb{R}^{N_o}$ and with a parametrized covariance matrix $\Sigma(\theta) \in \mathbb{R}^{N_o \times N_o}$:

$$p(y_n|\mathbf{u}, \theta, \mathcal{M}) = \frac{1}{(2\pi)^{N_o/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(y_n - q_n)^T \Sigma^{-1}(y_n - q_n)\right] \tag{2}$$

The I/O probability model for the system output history **y** over $N$ discrete times is then given by

$$p(\mathbf{y}|\mathbf{u}, \theta, \mathcal{M}) = \prod_{n=1}^{N} p(y_n|\mathbf{u}, \theta, \mathcal{M}) \tag{3}$$

The stochastic independence exhibited here comes from the fact that no joint moments in time are imposed as constraints during the entropy maximization. It refers to information independence which is not necessarily causal independence. It asserts that if the prediction errors at certain discrete times are given, this does not influence the plausibility of the prediction-error values at other times.

Another choice for stochastic embedding [7, 9] is to use a state-space model of the system and introduce prediction errors into the state vector evolution equation, as well as in the output equation, again modeled with a Gaussian PDF based on the principle of maximum information entropy. This alternative allows updating of the prediction-error uncertainty at unobserved points in the system, not just at the measurement points.

Either choice for the stochastic modeling of the prediction errors produces a set of parametrized I/O probability models $\{p(\mathbf{y}|\mathbf{u},\theta,\mathcal{M}): \theta \in \Theta\}$, where the uncertain parameters $\theta$ now also include those involved in specifying the probability models for the prediction errors, such as the prediction-error variances. To complete the specification of the model class $\mathcal{M}$, a prior distribution $p(\theta|\mathcal{M})$ is chosen to express the relative plausibility of each I/O probability model $p(\mathbf{y}|\mathbf{u},\theta,\mathcal{M})$ specified by the parameter vector $\theta$.

### 3.3 Bayesian Updating within a Model Class

Suppose data $\mathcal{D} = \{\hat{\mathbf{u}}, \hat{\mathbf{y}}\}$ are available from sensors on the system that consist of measured output $\hat{\mathbf{y}}$ of the system and possibly the corresponding system input $\hat{\mathbf{u}}$. These data can be used to update the relative plausibility of each I/O probability model $p(\mathbf{y}|\mathbf{u},\theta,\mathcal{M})$, $\theta \in \Theta \subset \mathbb{R}^{N_p}$, in the set defined by a model class $\mathcal{M}$ by computing the *posterior* PDF $p(\theta|\mathcal{D}, \mathcal{M})$ from Bayes' theorem:

$$p(\theta|\mathcal{D}, \mathcal{M}) = c^{-1}p(\mathcal{D}|\theta, \mathcal{M})p(\theta|\mathcal{M}) \tag{4}$$

where $c = p(\mathcal{D}|\mathcal{M}) = \int_\Theta p(\mathcal{D}|\theta,\mathcal{M})p(\theta|\mathcal{M})d\theta$ is the normalizing constant, and $p(\mathcal{D}|\theta,\mathcal{M})$, as a function of $\theta$, is the *likelihood function* which expresses the probability of getting data $\mathcal{D}$ based on the PDF $p(\mathbf{y}|\mathbf{u},\theta,\mathcal{M})$ for the system output. The constant $c = p(\mathcal{D}|\mathcal{M})$ is also called the *evidence* for the model class $\mathcal{M}$ given by data $\mathcal{D}$. Although it is a normalizing constant in (1) and so does not affect the shape of the posterior distribution, it plays an important role in computing the posterior probability of the model class, as described later.

The likelihood function should strictly be denoted by $p(\hat{\mathbf{y}}|\hat{\mathbf{u}},\theta,\mathcal{M})$ but the notation used in (4) is convenient. Furthermore, we have assumed either (a) that the measurement errors are negligible compared to the prediction errors due to modeling the system, or (b) that the sum of these errors is modeled probabilistically when constructing the likelihood function in (4) by choosing a measurement noise model in addition to the maximum entropy model for the prediction errors. These two alternatives correspond to a subtle difference in what is being probabilistically predicted: for (a), **y** denotes the predicted output time history of the actual system whereas for (b), **y** denotes the predicted output time history from the sensors.

## 4. ROBUST PREDICTIVE ANALYSIS USING A MODEL CLASS

A model class can be used to perform both prior (initial) and posterior (updated using system sensor data) robust predictive analyses based purely on the probability axioms [10]. Prior robust analyses are of importance in the robust design of systems whereas posterior robust analyses can be used to improve predictive modeling of already operating systems. These prior and posterior robust predictions correspond to a type of integrated global sensitivity analysis where the probabilistic prediction of each I/O probability model specified by the model class is considered but it is weighted by the relative plausibility of the model according to the prior or posterior PDF, respectively, in accordance with the total probability theorem.

### 4.1 Prior Robust Predictive Analysis

Based on a selected model class $\mathcal{M}$, all the probabilistic information for the prediction of the discrete response time history $\mathbf{y}$ for a specified discrete input time history $\mathbf{u}$ is contained in the prior robust predictive PDF given by the total probability theorem as

$$p(\mathbf{y}|\mathbf{u}, \mathcal{M}) = \int p(\mathbf{y}|\mathbf{u}, \boldsymbol{\theta}, \mathcal{M})p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta} \tag{5}$$

This PDF is a weighted average of the probabilistic prediction $p(\mathbf{y}|\mathbf{u},\boldsymbol{\theta},\mathcal{M})$ for each model specified by $\boldsymbol{\theta} \in \Theta$ in model class $\mathcal{M}$ where the weight is given by the prior probability $p(\boldsymbol{\theta}|\mathcal{M})d\boldsymbol{\theta}$.

Usually in assessing a system's design, the response time history $\mathbf{y}$ is not directly used but instead a *system performance measure* is selected which, because of the modeling uncertainty, is expressed as the prior expected value of some *system performance function* $\mathbf{g}(\mathbf{y})$:

$$\mathrm{E}[\mathbf{g}(\mathbf{y})|\mathbf{u}, \mathcal{M}] = \int \mathbf{g}(\mathbf{y})p(\mathbf{y}|\mathbf{u}, \mathcal{M})d\mathbf{y} \tag{6}$$

If there is also uncertainty in the excitation $\mathbf{u}$, it can be described by a stochastic input model $\mathcal{U}$ that specifies a joint PDF $p(\mathbf{u}|\mathcal{U})$ for the discrete-time input history $\mathbf{u}$. This uncertainty in the excitation can then be incorporated by evaluating the additional integral:

$$\mathrm{E}[\mathbf{g}(\mathbf{y})|\mathcal{U}, \mathcal{M}] = \int \mathrm{E}[\mathbf{g}(\mathbf{y})|\mathbf{u}, \mathcal{M}]p(\mathbf{u}|\mathcal{U})d\mathbf{u} \tag{7}$$

Combining (5), (6), and (7)

$$\mathrm{E}[\mathbf{g}(\mathbf{y})|\mathcal{U}, \mathcal{M}] = \int \mathbf{g}(\mathbf{y})p(\mathbf{y}|\mathbf{u}, \boldsymbol{\theta}, \mathcal{M})p(\mathbf{u}|\mathcal{U})p(\boldsymbol{\theta}|\mathcal{M})d\mathbf{y}d\mathbf{u}d\boldsymbol{\theta} \approx \frac{1}{K}\sum_{k=1}^{K} \mathbf{g}(\mathbf{y}^{(k)}) \tag{8}$$

where, as shown, the high-dimensional integral can be approximated using standard MCS (Monte Carlo simulation) by using samples $(\mathbf{y}^{(k)}, \mathbf{u}^{(k)}, \boldsymbol{\theta}^{(k)})$, $k = 1, 2,..., K$, where the parameter vector $\boldsymbol{\theta}^{(k)}$ is drawn from the prior $p(\boldsymbol{\theta}|\mathcal{M})$, the input time history $\mathbf{u}^{(k)}$ is drawn from the joint PDF $p(\mathbf{u}|\mathcal{U})$ implied by the stochastic input model $\mathcal{U}$, and the system output time history $\mathbf{y}^{(k)}$ is drawn from $p(\mathbf{y}|\mathbf{u}^{(k)},\boldsymbol{\theta}^{(k)},\mathcal{M})$. When there are no state prediction errors in the model, the $\mathbf{y}^{(k)}$ can be readily drawn as independent Gaussian samples at each discrete time according to (2) and (3). When both state and output prediction errors are in the model, independent Gaussian samples of them at each discrete time can be drawn and then the corresponding sample of $\mathbf{y}^{(k)}$ can be computed based on $\mathbf{u}^{(k)}$ and $\boldsymbol{\theta}^{(k)}$; an example of how to do this efficiently for a linear state-space model is given in [7, 9].

An important special case is where the system performance function $\mathbf{g}(\mathbf{y}) = \mathrm{I}_F(\mathbf{y})$, an indicator function which is equal to 1 if $\mathbf{y} \in F$ and 0 otherwise, where $F$ is a region in the response space that corresponds to *unsatisfactory system performance*, then (8) gives the *prior robust failure probability* $\mathrm{P}[F|\mathcal{U}, \mathcal{M}]$ [10]. If this failure probability is very small, a more computationally efficient algorithm than MCS should be used, such as subset simulation based on MCMC (Markov chain Monte Carlo) simulation [11, 12].

In optimal robust stochastic design, the system performance measure in (8) serves as the objective function in the optimization over the design variables specifying each design choice; for example, the performance function $\mathbf{g(y)}$ could represent the life-cycle costs for a structural design and include future uncertain economic losses from seismic damage over a specified time period as well as the up-front construction costs [13]. In this application, using stochastic simulation to evaluate the objective function in (8) for each iteration of the design variables given by an algorithm for the optimization of the system performance measure leads to a huge computational effort. A very efficient MCMC algorithm, SSO (stochastic subset Optimization), has been developed to mitigate this effort [14, 15] which uses a strategy of finding a small set of near-optimal design variables containing the optimum design, rather than trying to converge on the optimal point estimate. The study by Taflanidis and Beck [15] further examines how the model prediction error impacts optimal decisions in this robust design setting, and demonstrates that the influence can be significant.

### 4.2 Posterior Robust Predictive Analysis

If sensor data $\mathcal{D}$ are available from the system, the posterior $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ can be computed from Bayes' theorem as in Section 3.3, then the posterior robust predictive PDF is given by

$$p(\mathbf{y}|\mathbf{u},\mathcal{D},\mathcal{M}) = \int p(\mathbf{y}|\mathbf{u},\boldsymbol{\theta},\mathcal{D},\mathcal{M})p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})d\boldsymbol{\theta} \tag{9}$$

and the posterior system performance measure is given by

$$\mathrm{E}[\mathbf{g(y)}|\mathcal{D},\mathcal{U},\mathcal{M}] = \int \mathbf{g(y)}p(\mathbf{y}|\mathbf{u},\mathcal{D},\boldsymbol{\theta},\mathcal{M})p(\mathbf{u}|\mathcal{U})p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})d\mathbf{y}d\mathbf{u}d\boldsymbol{\theta} \tag{10}$$

If the system performance function $\mathbf{g(y)} = \mathrm{I}_F(\mathbf{y})$ as above, this measure is the posterior robust failure probability $\mathrm{P}[F|\mathcal{D},\mathcal{U},\mathcal{M}]$ [10].

As in the prior robust predictive case in the previous section, the required evaluation of the multidimensional integral over the parameter space usually cannot be evaluated analytically, nor evaluated numerically in a straightforward way if the number of parameters is not very small. Useful methods to approximate such integrals are Laplace's method of asymptotic approximation [8, 10, 16] and stochastic simulation methods. In contrast with the prior case, however, the evaluation of the posterior robust integral in (10) by stochastic simulation, which requires drawing samples from the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$, is much more challenging because (i) evaluation of the normalizing constant $c$ in Bayes' theorem (4) requires another challenging high-dimensional integration over the model parameter space; and (ii) the high-probability content region of $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ occupies a much smaller volume in the parameter space than that of the prior PDF and this region may be quite contorted because of the correlations between the model parameters that are induced by the data $\mathcal{D}$.

Fortunately, MCMC methods [17] can be used to draw posterior samples. MCMC algorithms that have been applied to modeling of dynamic systems include: multilevel Metropolis–Hastings algorithms with tempering or annealing [18, 19], Gibbs sampler [20], and hybrid Monte Carlo (or Hamiltonian Markov chain) simulation [21]. In these methods, the theoretical mean of the system performance function $\mathbf{g(y)}$ in (10) is still approximated by the sample mean of $\mathbf{g(y)}$:

$$\mathrm{E}[\mathbf{g(y)}|\mathcal{D},\mathcal{U},\mathcal{M}] \approx \frac{1}{K}\sum_{k=1}^{K}\mathbf{g}(\mathbf{y}^{(k)}) \tag{11}$$

where the samples $(\mathbf{y}^{(k)},\mathbf{u}^{(k)},\boldsymbol{\theta}^{(k)})$, $k = 1, 2,..., K$, are drawn as in the prior robust case in the previous section except that the $\boldsymbol{\theta}^{(k)}$ samples are now drawn from the posterior PDF $p(\boldsymbol{\theta}|\mathcal{D},\mathcal{M})$ by the MCMC algorithm. The $\boldsymbol{\theta}^{(k)}$ samples are no longer independent since they come from the stationary state of a Markov chain and so to achieve the same accuracy (e.g., same coefficient of variation on the sample mean estimator), more samples are required than in the prior case where independent samples can be drawn from the prior PDF.

## 5. ROBUST PREDICTIVE ANALYSIS USING MULTIPLE MODEL CLASSES

A set of competing candidate model classes may be chosen to deal with the uncertainty of which model class to choose to represent the dynamic behavior of a system. The probability logic axioms then lead naturally to prior and posterior hyper-robust predictive models that combine the predictions of all model classes in this set. These robust predictions are especially important when calculating failure probabilities because for reliable systems they tend to be very sensitive to the particular choice of model and this sensitivity is alleviated by considering the integrated robust or hyper-robust failure probabilities [7, 9].

If $M$ specifies a set of candidate model classes $\{\mathcal{M}_j: j = 1, 2,\ldots, N_M\}$ that is being considered for a system, together with a prior probability distribution over this set, then all the probabilistic information for the prediction of system response $\mathbf{x}$ subject to input $\mathbf{u}$ is contained in the prior hyper-robust predictive PDF based on $M$ and the total probability theorem:

$$p(\mathbf{y}|\mathbf{u},\boldsymbol{M}) = \sum_{j=1}^{N_M} p(\mathbf{y}|\mathbf{u}, \mathcal{M}_j)\mathrm{P}[\mathcal{M}_j|\boldsymbol{M}] \tag{12}$$

where the prior robust predictive PDF for each model class $\mathcal{M}_j$ from (5) is weighted by the prior probability $\mathrm{P}[\mathcal{M}_j|\boldsymbol{M}]$, which can be chosen to be $1/N_M$ if the model classes are considered equally plausible *a priori*. In the Bayesian statistical literature, (12) is called *prior model averaging*.

However, if response data $\mathcal{D}$ are available from the system, the corresponding posterior hyper-robust predictive PDF based on $\boldsymbol{M}$ (*posterior model averaging*) can be computed:

$$p(\mathbf{y}|\mathbf{u}, \mathcal{D}, \boldsymbol{M}) = \sum_{j=1}^{N_M} p(\mathbf{y}|\mathbf{u}, \mathcal{D}, \mathcal{M}_j)\mathrm{P}[\mathcal{M}_j|\mathcal{D},\boldsymbol{M}] \tag{13}$$

where the posterior robust predictive PDF for each model class $\mathcal{M}_j$ from (9) is weighted by its posterior probability $\mathrm{P}[\mathcal{M}_j|\mathcal{D}, \boldsymbol{M}]$ computed from Bayes' theorem at the model class level:

$$\mathrm{P}[\mathcal{M}_j|\mathcal{D},\boldsymbol{M}] = \frac{p(\mathcal{D}|\mathcal{M}_j)\mathrm{P}[\mathcal{M}_j|\boldsymbol{M}]}{p(\mathcal{D}|\boldsymbol{M}} \tag{14}$$

Here $p(\mathcal{D}|\mathcal{M}_j)$ is the *evidence* (sometimes called marginal likelihood) for $\mathcal{M}_j$ provided by the data $\mathcal{D}$, which is given by the total probability theorem as

$$p(\mathcal{D}|\mathcal{M}_j) = \int p(\mathcal{D}|\boldsymbol{\theta}_j, \mathcal{M}_j)p(\boldsymbol{\theta}_j|\mathcal{M}_j)d\boldsymbol{\theta}_j \tag{15}$$

The posterior probability of model class $\mathcal{M}_j$ in (14) is a measure, based on data $\mathcal{D}$, of its plausibility relative to $\boldsymbol{M}$, which states the chosen set of candidate model classes for making system response predictions and a prior probability distribution over this set. There is no implied assumption here that one of the model classes is the "correct" or "true" one.

### 5.1 Calculation of the Data-Based Evidence for a Model Class

The computation of the multidimensional integral in (15) for the evidence is nontrivial. Laplace's method of asymptotic approximation can be used when the model class is globally identifiable based on the available data $\mathcal{D}$ [7, 22], which gives

$$p(\mathcal{D}|\mathcal{M}_j) \approx p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j, \mathcal{M}_j)p(\hat{\boldsymbol{\theta}}_j|\mathcal{M}_j)(2\pi)^{N_j/2}det\left(\mathbf{H}(\hat{\boldsymbol{\theta}}_j)\right)^{-1/2} \tag{16}$$

where $N_j$ is the number of model parameters (the dimension of $\boldsymbol{\theta}_j$) for the model class $\mathcal{M}_j$ and $\mathbf{H}(\boldsymbol{\theta}_j)$ is the Hessian matrix of $-\ln p(\mathcal{D}|\boldsymbol{\theta}_j,\mathcal{M}_j)$ if the parameter estimate used in (16) is the unique MLE (maximum likelihood estimate) that maximizes $\ln p(\mathcal{D}|\boldsymbol{\theta}_j,\mathcal{M}_j)$. For a general system, Beck and Katafygiotis [8] define *global system identifiability*,

*local system identifiability*, and *system unidentifiability* based on the data in terms of whether the set of MLEs consists of a single point, discrete points, or a continuum of points in the continuous-valued parameter space, respectively. When the chosen class of models is unidentifiable based on the available data $\mathcal{D}$ so that there are multiple MLEs, only stochastic simulation methods are practical to calculate the model class evidence, such as the Markov chain Monte Carlo methods: TMCMC, an MCMC algorithm [19, 23] and the stationarity method in [24].

## 5.2 Quantitative Ockham Razor

A comparison of the posterior probability of each model class automatically implements a quantitative version of a *principle of model parsimony* or *Ockham razor* [25, 26], which states qualitatively that simpler models should be preferred over more complex models that produce only slightly better agreement with the data, although it was not completely clear how to quantify model complexity. Two common measures for this are AIC (Akaike Information Criterion) [27] and BIC (Bayesian Information Criterion) [28], which trade off a data-fit measure with a complexity measure:

$$\text{AIC}(\mathcal{M}_j|\mathcal{D}) = \ln p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j, \mathcal{M}_j) - N_j \ \ \text{and} \ \ \text{BIC}(\mathcal{M}_j|\mathcal{D}) = \ln p(\mathcal{D}|\hat{\boldsymbol{\theta}}_j, \mathcal{M}_j) - \frac{1}{2}N_j \ln N \qquad (17)$$

where $N$ is the number of data points in the system sensor data $\mathcal{D}$ (model classes with a larger AIC or BIC are to be preferred because of the scaling chosen here). Using these simplified criteria for model assessment requires caution, however, because their penalty term for model class complexity depends only on the number of uncertain parameters $N_j$, while the correct penalty term, which can be deduced by taking the logarithm of the large-$N$ asymptotic approximation of the evidence in (16), can differ greatly for two model classes with the same number of uncertain parameters [23]. Rather than using AIC and BIC to assess globally identifiable model classes, it is much better to approximate the evidence by using (16); for example, Saito and Beck [29] use this approximation to determine the data-based most probable order of ARX (auto-regressive exogenous) models for the seismic response of a high-rise building in Tokyo where AIC did not give a maximum over the model order.

A recent interesting information-theoretic result [23, 30] shows that the evidence for $\mathcal{M}_j$ explicitly builds in a trade-off between a data-fit measure of the model class and an information-theoretic measure of its complexity (the *relative entropy* or *Kullback–Liebler information* of the posterior relative to the prior), which quantifies the amount of information that the model class extracts from the data $\mathcal{D}$. This result gives a deeper understanding of why the quantitative Ockham razor based on the posterior probability for each model class, as given in (14), has a built-in mechanism against data overfitting, thereby avoiding the well-known problem that occurs when a model is judged based only on its data fit using the maximum likelihood estimates of the model parameters.

## 6. ILLUSTRATIVE EXAMPLE

As an illustrative example, Bayesian updating, model class assessment, and prior and posterior robust and hyper-robust analyses of various stochastic system model classes are considered that are based on a quarter-car model of an automobile (Fig. 1). Input to the system is the road surface elevation $u$, which will be described as function of the distance along the road $d$. The output of interest for the system is the vertical acceleration of the body of the car $\ddot{x}_s$, which is important for passenger comfort. Bayesian updating of a similar quarter-car model, as well as half-car and full-car models, was considered by Metallidis et al. [31].

## 6.1 Model Description

A quarter-car model (Fig. 1) is taken to represent the vertical dynamics of the car system. This model has two degrees of freedom, $x_s$ and $x_t$, corresponding to the vertical displacements of the body of the car with mass $m_s$ and the wheel with mass $m_t$, respectively. The two masses are connected through a spring and a dashpot that model the car suspension and shock absorbers, respectively. The suspension force is denoted by $F_s$ and the shock absorber force by
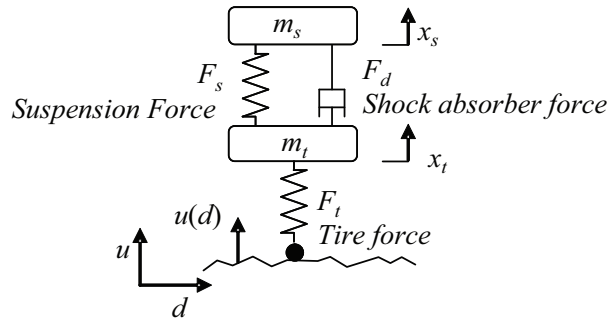
**FIG. 1:** Quarter-car model.

$F_d$. The wheel mass is connected to the road by a linear spring (modeling the tire stiffness) so the tire restoring force is

$$F_t = k_t(x_t - u) \tag{18}$$

where $k_t$ is the model parameter corresponding to the tire stiffness and $u(d)$ is the vertical elevation of the road surface above a reference level at a horizontal distance $d = vt$, where $v$ is the uniform speed of the car, taken here as 100 km/h. For notational simplicity, the dependence on time of all dynamic quantities is not explicitly denoted herein. The state evolution equation is

$$m_s\ddot{x}_s = -F_s - F_d \quad \text{and} \quad m_t\ddot{x}_t = -F_t + F_s + F_d \tag{19}$$

The model output is

$$q = \ddot{x}_s \tag{20}$$

Different models are considered for the suspension restoring force $F_s$ and for the shock absorber force $F_d$, ultimately defining different model classes. The former is described as a function of the relative displacement of the car with respect to the wheel $(x_s - x_t)$ and the latter similarly as a function of the corresponding relative velocity $(\dot{x}_s - \dot{x}_t)$.

The first model class $\mathcal{M}_1$, which is the simplest, is based on stochastic embedding of a deterministic model that is defined by linear suspension and shock absorber forces:

$$\mathcal{M}_1: \ F_s = k_{sl}(x_s - x_t) \quad \text{and} \quad F_d = c_{dl}(\dot{x}_s - \dot{x}_t) \tag{21}$$

where $k_{sl}$ and $c_{dl}$ are model parameters corresponding to the linear stiffness of the suspension and the linear damping coefficient of the shock absorber, respectively. The second and third model classes are based on models defined by, respectively, an additional nonlinear hardening component for the suspension force and an asymmetric damping component for the shock absorber force:

$$\mathcal{M}_2: \ F_s = k_{sl}(x_s - x_t) + k_{sn}(x_s - x_t)^3 \quad \text{and} \quad F_d = c_{dl}(\dot{x}_s - \dot{x}_t)$$

$$\mathcal{M}_3: \ F_s = k_{sl}(x_s - x_t) \quad \text{and} \quad F_d = c_{dl}(\dot{x}_s - \dot{x}_t) + c_{dn}\,|\dot{x}_s - \dot{x}_t| \tag{22}$$

where $k_{sn}$ and $c_{dn}$ are model parameters corresponding to the nonlinear stiffness of the suspension and the coefficient for the asymmetric damping of the shock absorber, respectively. The asymmetric damping characteristics correspond to different levels of damping forces that the absorber generates while moving upward compared to when moving downward.

The fourth model class is based on a deterministic model that is defined by incorporating all aforementioned linear and nonlinear components for the suspension and shock absorber forces:

$$\mathcal{M}_4: \ F_s = k_{sl}(x_s - x_t) + k_{sn}(x_s - x_t)^3 \quad \text{and} \quad F_d = c_{dl}(\dot{x}_s - \dot{x}_t) + c_{dn}\,|\dot{x}_s - \dot{x}_t| \tag{23}$$

The fifth and final model class is defined by adding another nonlinear part to the suspension force, with stronger hardening characteristics:

$$\mathcal{M}_5: \ F_s = k_{sl}(x_s - x_t) + k_{sn}(x_s - x_t)^3 + k_{sh}(x_s - x_t)^5 \quad \text{and} \quad F_d = c_{dl}(\dot{x}_s - \dot{x}_t) + c_{dn}\left|\dot{x}_s - \dot{x}_t\right| \tag{24}$$

where $k_{sh}$ is the model parameter introduced to model the new nonlinear stiffness component for the suspension.

Figure 2 illustrates a sample input $u$ and the corresponding response of model classes $\mathcal{M}_1$ and $\mathcal{M}_5$ (simplest and most complex, respectively) for the nominal values of the model parameters (described in the next subsection). It is observed that the differences in the predicted responses from the two models are moderate.

## 6.2 Definition of Stochastic System Model Classes

The model classes for the quarter-car model are defined through the stochastic embedding process described in Section 3.2 by extending the implicit deterministic I/O relationship defined by (19) and (20). The prediction error $e_n$ between the responses of the model $q_n$ and system $y_n$ at the $n$th time instant follows a Gaussian distribution with zero mean and variance $\sigma_e^2$ and is independent of the previous errors (from maximum entropy PDF selection). This leads then to a Gaussian white-noise sequence for the model prediction error $\{e_n\}$ and so to the I/O probability model corresponding to (2) and (3):

$$p(\mathbf{y}|\mathbf{u}, \boldsymbol{\theta}, \mathcal{M}_i) = \prod_{n=1}^{N} p(y_n|\mathbf{u}, \boldsymbol{\theta}, \mathcal{M}_i) = \frac{1}{(2\pi\sigma_e)^{N/2}} \exp\left[ -\frac{1}{2\sigma_e^2} \sum_{n=1}^{N} [y_n - q_n(\mathbf{u}, \boldsymbol{\theta}, \mathcal{M}_i)]^2 \right] \tag{25}$$

where $\mathbf{y}$ is the vector of the model response for a desired time interval consisting of $N$ discrete time samples.

The prediction-error variance $\sigma_e^2$ is an additional model parameter. The wheel mass is taken as known at 40 kg while all other model parameters are treated as uncertain quantities described by the prior probability model $p(\boldsymbol{\theta}|\mathcal{M}_i)$, $i = \{1,\ldots,5\}$, for each of the five model classes. The parameter vector for the most complex model class $\mathcal{M}_5$ is $\boldsymbol{\theta} = [m_s, k_t, k_{sl}, k_{sn}, c_{dl}, c_{dn}, k_{sh}, \sigma_e^2]^T \in \mathbb{R}_+^8$, while for the other four model classes it is a subvector of this $\boldsymbol{\theta}$. For convenience we use $\boldsymbol{\theta}$ to denote all these parameters vectors since the conditioning on $\mathcal{M}_i$ in the PDFs makes it clear which parameters are involved. The number of parameters in the $\boldsymbol{\theta}$ vector for each of the five model classes is 5, 6, 6, 7, and 8. The prior PDFs for $\boldsymbol{\theta}$ are chosen as the product of independent distributions, with $m_s$, $k_t$, $k_{sl}$, $k_{sn}$,
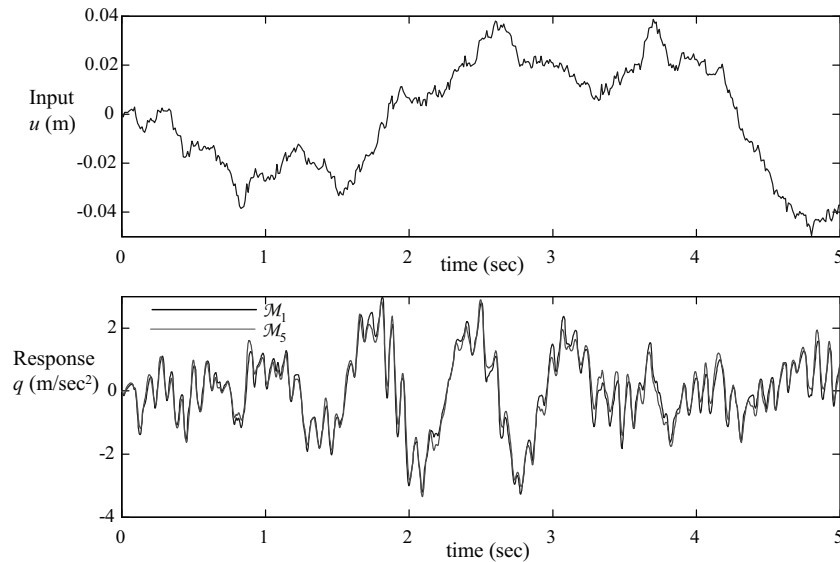


**FIG. 2:** Sample input and response for nominal models in two model classes.

$c_{dl}$, $c_{dn}$, and $k_{sh}$ each following a lognormal distribution with median equal to the nominal values 290 kg, 190 kN/m, 23.5 kN/m, 2350 kN/m$^3$, 700 N s/m, 400 N s/m, and 235 MN/m$^5$, respectively, and all with coefficient of variation 40%. These larger coefficients of variation are chosen because in reality there is a great deal of prior uncertainty about appropriate values for the model stiffness and damping parameters, and we decided to give the same value of the coefficient of variation to the car mass too. The nominal linear model used in $\mathcal{M}_1$ has modal frequencies $\omega_1 = 1.37$ Hz and $\omega_2 = 11.56$ Hz and corresponding damping ratios are $\xi_1 = 1.3\%$ and $\xi_2 = 0.4\%$. The prior PDF for $\sigma_e$ is also a lognormal distribution with median equal to 0.1 m/s$^2$ and coefficient of variation 70%. The larger coefficient of variation for the prediction error is adopted because of the limited prior knowledge about it in real problems.

This completes the definition of each model class $\mathcal{M}_i$, $i = 1,...,5$, which consists of the I/O probability model $p(\mathbf{y}|\mathbf{u},\theta,\mathcal{M}_i)$ along with the prior probability model $p(\theta|\mathcal{M}_i)$ for each I/O model within the class.

## 6.3 Prior Robust Predictive Analysis

Based on the prior information for the model classes, robust predictive analyses are performed. The investigation focuses on passenger comfort, and specifically on the probability that the vertical acceleration response is unacceptable because it exceeds some threshold $\beta$ over some time interval $[0, T]$:

$$\mathrm{P}\left[F|\mathcal{U}, \mathcal{M}\right] \triangleq \mathrm{P}\left[|y_n| > \beta \quad \text{for some} \quad t_n \in [0, T]\right] = \mathrm{P}\left[\max_{n=1,...,N} |y_n| > \beta\right] \triangleq P_F(\beta) \tag{26}$$

where $F$ denotes performance failure, $\mathcal{U}$ is the stochastic road profile model described below, and the system response is evaluated at discrete times $\{0, \Delta t, \ldots, t_n, \ldots, T\}$, with the time step $\Delta t$ properly chosen to capture the important dynamic behavior. For this study, $\Delta t$ is taken as 0.005 s and the time horizon $T$ is selected as 20 s. The system performance measure is then the prior robust failure probability $\mathrm{P}[F|\mathcal{U}, \mathcal{M}]$, which corresponds to the expected value in (8) with

$$g(\mathbf{y}) = \mathrm{I}_F(\mathbf{y}) = \begin{cases} 0 & \max_{n=1,...,N} |y_n| \le \beta \\ 1 & \text{otherwise} \end{cases} \tag{27}$$

An alternative expression for the prior robust failure probability which leads to increased efficiency in MCS can be derived by analytically integrating out the uncertainty stemming from the prediction error. This can be established by noting that when conditional on $\mathbf{u}$ and $\theta$

$$\mathrm{P}\left[\max_{n=1,...,N} |y_n| \le \beta\right] = \mathrm{P}[|y_n| \le \beta, \ \forall \ t_n] = \mathrm{P}\left[\bigcap_{n=1}^{N} |q_n + e_n| \le \beta\right] = \prod_{n=1}^{N} \mathrm{P}[|q_n + e_n| \le \beta]$$

$$= \prod_{n=1}^{N} \int_{|q_n + e_n| \le \beta} p(e_n) de_n = \prod_{n=1}^{N} \int_{-\beta - q_n}^{\beta - q_n} p(e_n) de_n = \prod_{n=1}^{n} \left[\Phi\left(\frac{\beta - q_n}{\sigma_e}\right) - \Phi\left(\frac{-\beta - q_n}{\sigma_e}\right)\right] \tag{28}$$

where $\Phi$ is the cumulative distribution function for a standard Gaussian variable and the third equality is established due to the independence of the $e_i$ when $\mathbf{u}$ and $\theta$ are given. The performance measure conditional on $\mathbf{u}$ and $\theta$ is then

$$\mathrm{P}[F|\mathbf{u}, \theta, \mathcal{M}] = \mathrm{E}[\mathbf{g}(\mathbf{y})|\mathbf{u}, \theta, \mathcal{M}] = \int \mathrm{I}_F(\mathbf{y}) p(\mathbf{y}|\mathbf{u}, \theta, \mathcal{M}) d\mathbf{y} = 1 - \prod_{n=1}^{N} \left[\Phi\left(\frac{\beta - q_n}{\sigma_e}\right) - \Phi\left(\frac{-\beta - q_n}{\sigma_e}\right)\right] \tag{29}$$

To compute $P_F(\beta) = \mathrm{P}[F|\mathcal{U}, \mathcal{M}]$ in (26), MCS is then used to integrate only over $\mathbf{u}$ and $\theta$ in the integral for the system performance measure given in (8) because the integral over $\mathbf{y}$ is replaced by the analytical expression in (29).

The road surface input $u(d)$ (see Fig. 1) is modeled as a zero-mean Gaussian stochastic process with autocorrelation function

$$\mathrm{E}[u(d_1)u(d_2)] = \sigma_r^2 \min(d_1, d_2) \tag{30}$$

where $\sigma_r$ is taken to be equal to 0.002 $(\text{m})^{1/2}$. This model implies a joint PDF on the input at the set of discrete times which is equivalently described by the following formula:

$$u_n \overset{\Delta}{=} u(vn\Delta t) = \sigma_r \sqrt{v\Delta t} \sum_{i=1}^{n} w_i; \quad u(0) = 0 \qquad (31)$$

where $\Delta t = 0.005$ s and $\{w_i\}$ is a zero-mean Gaussian white-noise sequence. Recall that distance and time are related through $d = vt$, where $v = 100$ km/h is the uniform speed of the car.

The results for the prior predictive analysis based on stochastic simulation are reported in Figs. 3–5 for the pair $P_F(\beta) - \beta$ that shows how the probability of failure changes as the threshold $\beta$ changes. In particular, the following results are reported in Fig. 3: prior robust predictive analysis for model classes $\mathcal{M}_1$ and $\mathcal{M}_5$ (the ones demonstrated in Fig. 2); predictive analysis for these model classes ignoring the model prediction error as well as the uncertainty in the model parameters (i.e., nominal model parameter values are used); and prior hyper-robust predictive analysis considering all five model classes using equal prior probability $P[\mathcal{M}_i | \boldsymbol{M}] = 1/5$ which is covered in Section 6.5. Figure 4 shows the same results as Fig. 3 but for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$. Figure 5 shows the prior robust predictive results for all model classes and the prior hyper-robust predictive results.
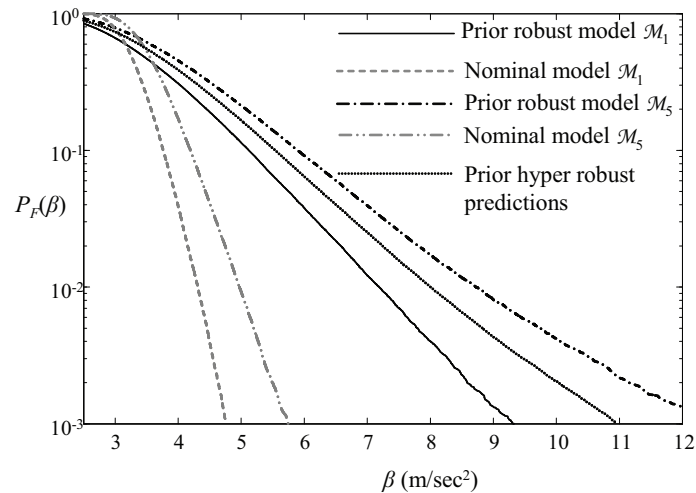


**FIG. 3:** Prior predictive analysis results for model classes $\mathcal{M}_1$ and $\mathcal{M}_5$ and prior hyper-robust predictive analysis.
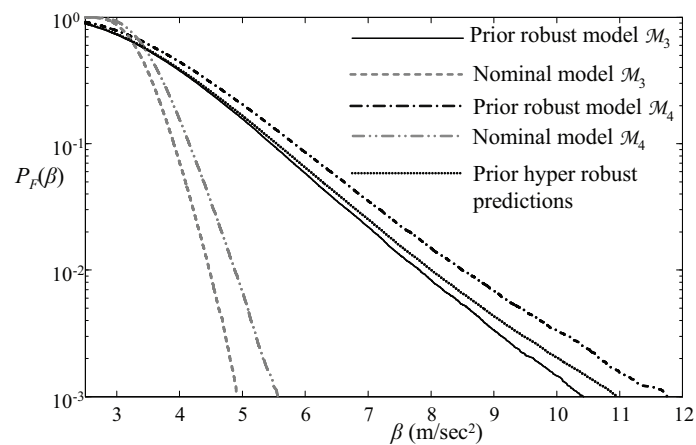


**FIG. 4:** Prior predictive analysis results for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$ and prior hyper-robust predictive analysis.
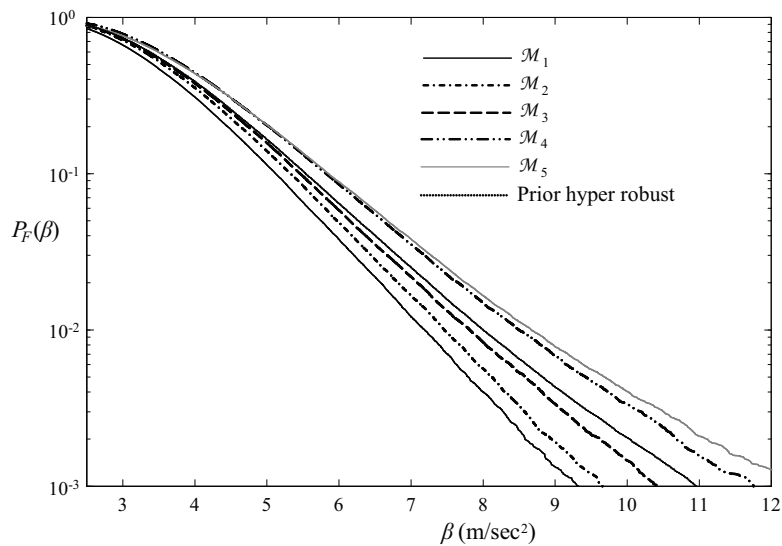
**FIG. 5:** Prior robust predictive analysis for all model classes and prior hyper-robust predictive analysis.

The results show the importance of explicitly addressing the uncertainty in modeling the system's dynamic behavior. The robust predictive analysis results in Figs. 3 and 4 are much different from the results corresponding to the nominal model predictions where the only uncertainty is in the excitation, especially for larger thresholds $\beta$ that correspond to more rare events (smaller probabilities for unacceptable behavior). The latter cases exhibit the well-known sensitivity of the tails of the probability distributions to the modeling assumptions; in particular, the robust predictive results reflect the increase in response uncertainty coming from the uncertainty in the values of the model parameters and the uncertain prediction error.

## 6.4 Bayesian Updating and Posterior Robust Predictive Analysis

A dynamic data sequence $\mathcal{D} = \{\hat{\mathbf{u}}, \hat{\mathbf{y}}\}$ (shown in Fig. 6) is now used to update each model class (Bayesian model updating). This data set consists of measurements for the car absolute acceleration for 10 s with sample interval of
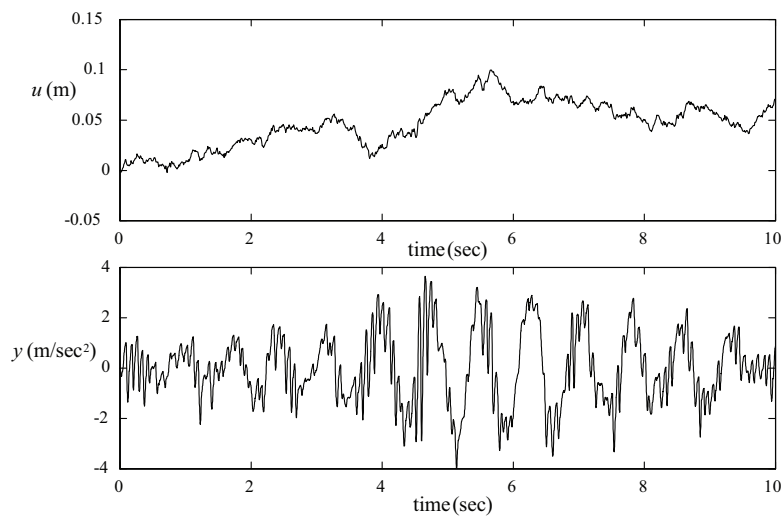


**FIG. 6:** Input-output data sequences $\mathcal{D} = \{\hat{\mathbf{u}}, \hat{\mathbf{y}}\}$ used for the Bayesian updating.

0.01 s, leading to $N_s = 1000$ data points for output $\hat{\mathbf{y}}$ along with the corresponding input $\hat{\mathbf{u}}$ for the road surface profile. These synthetic data were not generated by any of the five model classes considered; the model used to generate the data has restoring and shock absorber forces given, respectively, by

$$F_s = k_{sl}\text{sgn}(x_s - x_t)\,|x_s - x_t|^{1.14} + k_{sn}\text{sgn}(x_s - x_t)\,|x_s - x_t|^{3.25} + k_{sh}\text{sgn}(x_s - x_t)\,|x_s - x_t|^{4.85}$$

$$F_d = c_{dl}\text{sgn}(\dot{x}_s - \dot{x}_t)|\dot{x}_s - \dot{x}_t|^{0.91} + c_{dn}\,|\dot{x}_s - \dot{x}_t|^{1.05} \tag{32}$$

where sgn(.) stands for the sign function. The model parameter values chosen for generating the synthetic data were $m_s = 300$ kg, $k_t = 209$ kN/m, $k_{sl} = 18.2$ kN/m, $k_{sn} = 2212$ kN/m$^3$, $c_{dl} = 644$ N s/m, $c_{dn} = 270$ N s/m, and $k_{sh} = 210$ MN/m$^5$. Zero-mean Gaussian noise is also added to the data with standard deviation 0.03 m/s$^2$.

Based on (25), the likelihood function for the data for each model class is then given by

$$p(\hat{\mathbf{y}}|\hat{\mathbf{u}}, \boldsymbol{\theta}, M_i) = \frac{1}{(2\pi\sigma_e)^{N/2}} \exp\left[ -\frac{1}{2\sigma_e^2} \sum_{n=1}^{N} [\hat{y}_n - q_n(\hat{\mathbf{u}}, \boldsymbol{\theta}, M_i)]^2 \right] \tag{33}$$

The prior distribution $p(\boldsymbol{\theta}|\mathcal{M}_i)$ for each model class is then updated to construct the posterior distribution $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_j)$ using (4). Table 1 presents the MAP (maximum *a posteriori*) estimates $\boldsymbol{\theta}_{\text{MAP}}$ of the parameters that maximize $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_i)$ along with the coefficient of variation for each parameter. The MAP estimates are calculated by maximizing the sum of the log likelihood function and the log prior because the value of the normalizing constant in (4), which is unknown, is irrelevant to the maximization over $\boldsymbol{\theta}$. The coefficients of variation in Table 1 are calculated from the covariance matrix of the Gaussian PDF centered at $\boldsymbol{\theta}_{\text{MAP}}$ that has the same curvature as the posterior PDF at $\boldsymbol{\theta}_{\text{MAP}}$, which corresponds to taking the inverse of the Hessian matrix of $-\ln[p(\hat{\mathbf{y}}|\hat{\mathbf{u}},\boldsymbol{\theta},\mathcal{M}_i)p(\boldsymbol{\theta}|\mathcal{M}_i)]$ at $\boldsymbol{\theta}_{\text{MAP}}$. This Gaussian approximation can be viewed as the first step in Laplace's method of asymptotic approximation of the posterior robust predictive integral in (9), although we will use an MCMC algorithm instead. As expected, Table 1 shows that the system data substantially reduce the uncertainty in the description of the parameters: the coefficients of variation are much smaller than those of the prior PDFs (also shown in the table).

The robust analysis presented in Section 6.3 is repeated but now using the posterior probability information from the Bayesian updating. This provides the posterior predictive analysis results shown in Figs. 7 and 8. Samples from the posterior distribution $p(\boldsymbol{\theta}|\mathcal{D}, \mathcal{M}_j)$ for the posterior analysis are obtained through TMCMC, a Markov chain Monte Carlo algorithm proposed by Ching and Chen [19].

Figure 7 shows the prior robust predictive analysis for model classes $\mathcal{M}_1$ and $\mathcal{M}_5$ along with the posterior robust predictive analysis for the same model classes. Figure 8 repeats this analysis for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$. Comparison between the prior and posterior robust analysis results shows significant differences, which stress the importance of updating the performance predictions when monitoring data become available for the dynamic system.

**TABLE 1:** Nominal values and MAP estimates of the parameters for all model classes, along with the nominal and posterior coefficients of variation of the parameters (in parentheses)

|  | **Nominal values** | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
|---|---|---|---|---|---|---|
| $m_s$ (kg) | 290 (40%) | 302.72 (2%) | 301.32 (2.0%) | 289.61 (1.2%) | 291.44 (1.2%) | 291.15 (1.1%) |
| $k_t$ (kN/m) | 190 (40%) | 214.35 (1.2%) | 214.23 (1.3%) | 213.11 (1.3%) | 213.85 (1.3%) | 213.61 (1.2%) |
| $k_{sl}$ (kN/m) | 23.5 (40%) | 21.21 (1.2%) | 21.32 (1.2%) | 20.61 (1.0%) | 20.32 (0.9%) | 20.62 (0.9%) |
| $k_{sn}$ (kN/m$^3$) | 2350 (40%) | 0.000 (N/A) | 2658.82 (0.9%) | 0.000 (N/A) | 2619.13 (0.7%) | 2619.13 (0.7%) |
| $c_{dl}$ (N s/m) | 700 (40%) | 472.12 (3.2%) | 477.58 (3.2%) | 525.71 (2.1%) | 528.48 (2.0%) | 529.15 (2.1%) |
| $c_{dn}$ (N s/m) | 400 (40%) | 0.000 (N/A) | 0.000 (N/A) | 339.10 (0.4%) | 337.81 (0.4%) | 334.43 (0.5%) |
| $k_{sh}$ (MN/m$^5$) | 235 (40%) | 0.000 (N/A) | 0.000 (N/A) | 0.000 (N/A) | 0.000 (N/A) | 219.89 (1.6%) |
| $\sigma_e$ (m/s$^2$) | 0.1 (70%) | 0.297 (3.5%) | 0.269 (2.9%) | 0.104 (1.4%) | 0.102 (1.4%) | 0.102 (1.4%) |
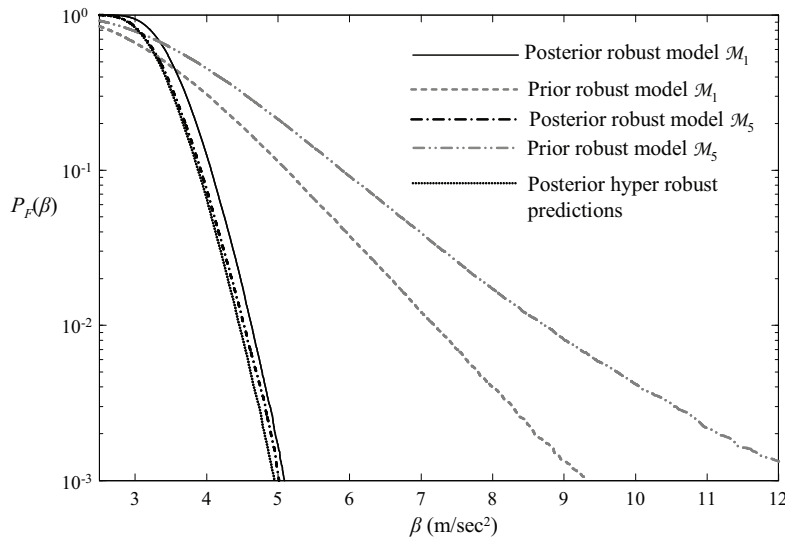
**FIG. 7:** Prior and posterior predictive analysis results for model classes $\mathcal{M}_1$ and $\mathcal{M}_5$.
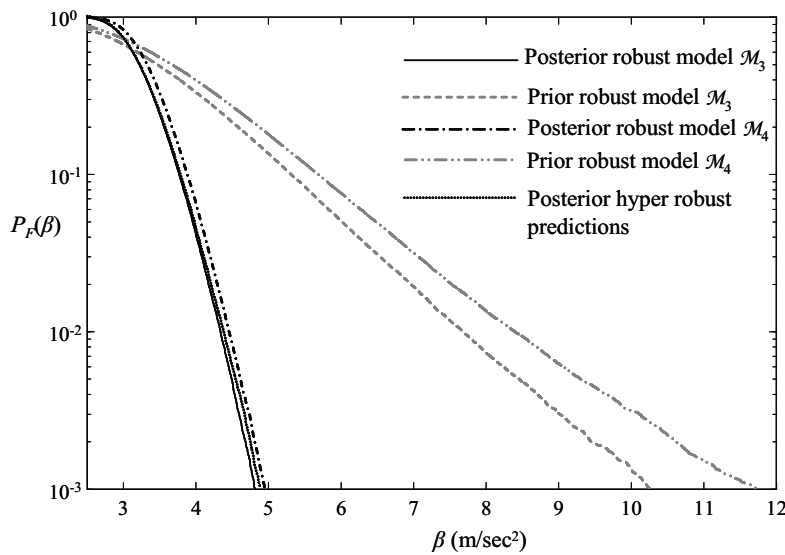


**FIG. 8:** Prior and posterior predictive analysis results for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$.

## 6.5 Model Class Assessment and Posterior Hyper-Robust Predictive Analysis

The dynamic data sequence $\mathcal{D}$ is used to calculate the posterior probability for each model class $P[\mathcal{M}_i|\mathcal{D},\boldsymbol{M}]$, $i = 1,...,5$, through (14). Table 2 presents the log evidence, $\ln p(\mathcal{D}|\mathcal{M}_i)$, and the posterior probability $P[\mathcal{M}_i|\mathcal{D},\boldsymbol{M}]$, calculated through both Laplace's asymptotic approximation (16) as well as through the stationarity method presented in Cheung and Beck [24] based on MCMC stochastic simulation. For the latter, samples from the posterior distribution $p(\theta|\mathcal{D}, \mathcal{M}_i)$ are obtained through the TMCMC algorithm [19].

The results in Table 2 show that the most probable model class is $\mathcal{M}_3$ but $\mathcal{M}_4$ also has significant probability. This illustrates the quantitative Ockham razor established by the Bayesian updating approach. Note that the most complex model class $\mathcal{M}_5$ is not the most probable one. The three model classes $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_5$ have extremely small probabilities $P[\mathcal{M}_i|\mathcal{D},\boldsymbol{M}]$ based on the data so the posterior hyper-robust prediction of the system performance

**TABLE 2:** Bayesian model class assessment results

|                         |                          | $\mathcal{M}_1$ | $\mathcal{M}_2$ | $\mathcal{M}_3$ | $\mathcal{M}_4$ | $\mathcal{M}_5$ |
|-------------------------|--------------------------|---------|---------|--------|--------|--------|
| **Laplace approximation** | $\log[p(\mathcal{D}|\mathcal{M}_j)]$ | −255.14 | −269.75 | 814.33 | 813.51 | 563.45 |
|                         | $P(\mathcal{M}_j|\mathcal{D},\boldsymbol{M})$ | 0.000 | 0.000 | 0.694 | 0.306 | 0.000 |
| **Stochastic simulation** | $\log[p(\mathcal{D}|\mathcal{M}_j)]$ | −255.32 | −277.55 | 813.71 | 813.09 | 543.22 |
|                         | $P(\mathcal{M}_j|\mathcal{D},\boldsymbol{M})$ | 0.000 | 0.000 | 0.650 | 0.350 | 0.000 |

based on (13) and all five model classes effectively consists of a weighted combination of the robust predictions given by model classes $\mathcal{M}_3$ and $\mathcal{M}_4$. Table 2 also shows that the differences are quite small between the log evidence for each $\mathcal{M}_i$ calculated by Laplace's asymptotic approximation, which is applicable here because the model classes are globally identifiable, and by stochastic simulation, which requires much more computation. The hyper-robust results presented next use the posterior probabilities given in Table 2 for the stochastic simulation approach.

Failure $F$ is again defined as exceedance of the vertical acceleration response above a threshold β over $[0, T]$, as in Section 6.3. Figure 9 shows the failure probability curves from the posterior robust predictive analysis for each model class along with the posterior hyper-robust predictive analysis results. The failure probability curve for the posterior hyper-robust case in Fig. 9 is effectively a weighted average of the curves for model classes $\mathcal{M}_3$ and $\mathcal{M}_4$ based on the posterior probabilities 0.650 and 0.350 in Table 2. Note that a smaller range of threshold β has been used in this figure compared to all previous ones, since the differences between the different model classes are smaller. This smaller difference is due to the reduction in the uncertainty of the model parameters and prediction errors coming from the information provided by the dynamic data.

Compared to the prior cases in Fig. 5, the posterior failure probabilities calculated for the different model classes in Fig. 9 are in much closer agreement. The two model classes $\mathcal{M}_3$ and $\mathcal{M}_4$ that were identified to have similar posterior probabilities $P[\mathcal{M}_i|\mathcal{D},\boldsymbol{M}]$ based on the dynamic data do not provide the same robust predictions, demonstrating the importance of a framework that can consider multiple model classes in performance reliability assessment; although $\mathcal{M}_3$ and $\mathcal{M}_4$ are both consistent with the given dynamic data, this does not necessarily mean that they will predict similar responses for every possible excitation. This further stresses the importance of posterior hyper-robust predictions that average over the robust predictions of all model classes using their relative plausibility based on the data.
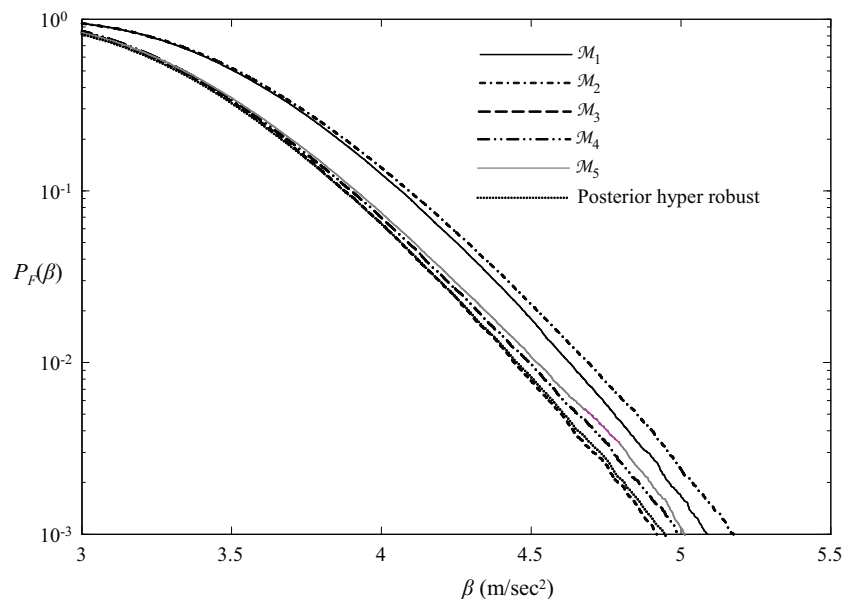


**FIG. 9:** Posterior predictive analysis for all model classes and posterior hyper-robust predictive analysis.

## 7. CONCLUDING REMARKS

A powerful unifying stochastic framework is available for treating modeling uncertainty, along with input uncertainty, when using dynamic models to predict the response during design or operation of a system. This framework is a principled one that is based solely on the probability axioms and Jaynes' principle of maximum information entropy. A key concept is a stochastic system model class which defines the fundamental probability models that allow both prior and posterior robust stochastic system analyses to be performed. Such a model class can be constructed by stochastic embedding of any deterministic model of the system's input–output behavior. There is no invocation of inherent randomness; instead, the approach is a pragmatic one that allows plausible reasoning about system behavior based on incomplete information.

The prior and posterior robust predictions of system response not only incorporate *parametric uncertainty* (uncertainty about which model in a proposed set should be used to represent the system's input–output behavior) but also *nonparametric uncertainty* due to the existence of prediction errors because of the approximate nature of any system model. Robust predictive analysis involves integrals that usually cannot be evaluated in a straightforward way. Useful computational tools are Laplace's method of asymptotic approximation and various MCMC (Markov chain Monte Carlo) algorithms.

These theoretical ideas and computational tools were demonstrated in an illustrative example involving the vertical dynamic response of a car being driven along a rough road. Multiple model classes were used to characterize the car dynamics, corresponding to different assumptions for the shock absorbers and the car suspension. The results demonstrated the importance of addressing model uncertainties when evaluating the dynamic response, as well as the benefits of using sensor data to update model descriptions to provide posterior robust and hyper-robust predictions.

## REFERENCES

1. Cox, R. T., Probability, frequency and reasonable expectation, *Am. J. Phys.*, 14:1–13, 1946.

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

3. Jaynes, E. T., *Papers on Probability, Statistics and Statistical Physics*, Rosenkrantz, R. D. (ed.), D. Reidel Publishing, Dordrecht, Holland, 1983.

4. Jaynes, E. T., *Probability Theory: The Logic of Science*, Cambridge University Press, UK, 2003.

5. Kolmogorov, A. N., *Foundations of the Theory of Probability*, Chelsea Publishing, New York, 1950.

6. Beck, J. L., Probability logic, information quantification and robust predictive system analysis, Tech. Report EERL 2008-05, Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, CA, 2008.

7. Beck, J. L., Bayesian system identification based on probability logic, *Struct. Control Health Monit.*, 17:825–847, 2010.

8. Beck, J. L. and Katafygiotis, L. S., Updating models and their uncertainties. I. Bayesian statistical framework, *J. Eng. Mech.*, 124(4):455–461, 1998.

9. Cheung, S. H., *Stochastic analysis, model and reliability updating of complex systems with applications to structural dynamics*, PhD Thesis in Civil Engineering, California Institute of Technology, Pasadena, CA, 2009.

10. Papadimitriou, C., Beck, J. L., and Katafygiotis, L. S., Updating robust reliability using structural test data, *Probab. Eng. Mech.*, 16:103–113, 2001.

11. Au, S. K. and Beck, J. L., Estimation of small failure probabilities in high dimensions by subset simulation, *Probab. Eng. Mech.*, 16:263–277, 2001.

12. Au, S. K. and Beck, J. L., Subset simulation and its application to seismic risk based on dynamic analysis, *J. Eng. Mech.*, 29:901–917, 2003.

13. Taflanidis, A. A. and Beck, J. L., Life-cycle cost optimal design of passive dissipative devices, *Struct. Safety.*, 31:508–522, 2009.

14. Taflanidis, A. A. and Beck, J. L., An efficient framework for optimal robust stochastic system design using stochastic simulation, *Comput. Methods Appl. Mech. Eng.*, 198:88–101, 2008.

15. Taflanidis, A. A. and Beck, J. L., Reliability-based design using two-stage stochastic optimization with a treatment of model prediction errors, *J. Eng. Mech.*, 136(12):1460–1473, 2010.

16. Yuen, K. V., *Bayesian Methods for Structural Dynamics and Civil Engineering*, John Wiley and Sons, New York, 2010.

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

18. Beck, J. L. and Au, S. K., Bayesian updating of structural models and reliability using Markov Chain Monte Carlo simulation, *J. Eng. Mech.*, 128:380–391, 2002.

19. Ching, J. and Chen, Y. J., Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging, *J. Eng. Mech.*, 133:816–832, 2007.

20. Ching, J., Muto, M., and Beck, J. L., Structural model updating and health monitoring with incomplete modal data using Gibbs Sampler, *Comput. Aided Civ. Infrastruct. Eng.*, 21:242–257, 2006.

21. Cheung, S. H. and Beck, J. L., Bayesian model updating using Hybrid Monte Carlo simulation with application to structural dynamic models with many uncertain parameters, *J. Eng. Mech.*, 135:243–255, 2009.

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

23. Muto, M. and Beck, J. L., Bayesian updating of hysteretic structural models using stochastic simulation, *J. Vib. Control*, 14:7–34, 2008.

24. Cheung, S. H. and Beck, J. L., Calculation of the posterior probability for Bayesian model class assessment and averaging from posterior samples based on dynamic system data, *Computer-Aided Civil Infrastructure Eng.*, 25:304–321, 2010.

25. Gull, S. F., Bayesian inductive inference and maximum entropy, *Maximum Entropy and Bayesian Methods*, Skilling, J. (ed.), Kluwer, Dordrecht, Holland, 1989.

26. Mackay, D. J. C., *Bayesian methods for adaptive models*, PhD Thesis in Computation and Neural Systems, California Institute of Technology, Pasadena, CA, 1992.

27. Akaike, H, A new look at the statistical model identification, *IEEE Trans. Autom. Control*, 19:716–723 1974.

28. Schwarz, G., Estimating the dimension of a model, *Ann. Stat.*, 6:461–464, 1978.

29. Saito, T. and Beck, J. L., Bayesian model selection for ARX models and its application to structural health monitoring, *Earthquake Eng. Struct. Dyn.*, 39:1737–1759, 2010.

30. Ching, J., Muto, M., and Beck, J. L., Bayesian linear structural model updating using Gibbs sampler with modal data, *Proc. Int. Conf. on Structural Safety and Reliability*, Rome, Italy, 2005.

31. Metallidis, P, Verros, G., Natsiavas, S., and Papadimitriou, C., Fault detection and optimal sensor location in vehicle suspensions, *J. Vib. Control*, 9:337–359, 2003.