

# Nonlinear Model Updating through a Hierarchical Bayesian Modeling Framework

Xinyu Jia<sup>1</sup>, Omid Sedehi<sup>2</sup>, Costas Papadimitriou<sup>1,\*</sup>, Lambros S. Katafygiotis<sup>2</sup>, Babak Moaveni<sup>3</sup>

<sup>1</sup>*Department of Mechanical Engineering, University of Thessaly, Volos, Greece*

<sup>2</sup>*Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, China*

<sup>3</sup>*Department of Civil and Environmental Engineering, Tufts University, MA, USA*

## ABSTRACT

A new time-domain probabilistic technique based on hierarchical Bayesian modeling (HBM) framework is proposed for calibration and uncertainty quantification of hysteretic type nonlinearities of dynamical systems. Specifically, probabilistic hyper models are introduced respectively for material hysteretic model parameters as well as prediction error variance parameters, aiming to consider both the uncertainty of the model parameters as well as the prediction error uncertainty due to unmodelled dynamics. A new asymptotic approximation is developed to simplify the process of nonlinear model updating and substantially reduce the computational burden of the HBM framework. Two numerical examples are conducted to verify the accuracy and performance of the proposed method considering Bouc-Wen (BW) hysteretic type nonlinearities. Model error is manifested as uncertainty due to variability in the measured data from multiple datasets. Results from a five-story numerical structure indicate that the model error is the main source of error that can affect the uncertainty in the model parameters due to the variability in the experimental data. It is also demonstrated that the parameter uncertainty due to the variability arising from model error depends on the sensor locations. It is shown that the proposed approach is robust for not only quantifying uncertainties of structural parameters and prediction error parameters, but also predicting the system quantities of interests (QoI) with reasonable accuracy and providing reliable uncertainty bounds, as opposed to the conventional Bayesian approach which often severely underestimates the uncertainty bounds.

**Keywords:** hierarchical Bayesian modeling; nonlinear system; Bouc-Wen model; structural parameter uncertainty; prediction error uncertainty; time-domain response

## 1. Introduction

Finite element (FE) models are extensively employed for representing structural systems and predicting their responses to dynamic loads [1–4]. Discrepancies between the predicted responses from FE models and the measured responses from the physical structures are often inevitable. To achieve a more authentic model, model updating has received considerable attentions in recent decades using deterministic (e.g. [5–8]) and probabilistic approaches (e.g. [9–12]). Updating linear models has been widely applied and is shown to achieve a great progress in the field of structural dynamics [6,9,13]. However, most physical structures are inherently characterized by nonlinear behaviours with higher uncertainties when subjected to large loads due to material and/or geometric nonlinearities. Linear systems often neglect such nonlinearities and therefore cause a considerably large modeling error between the real structures and the updated models. Characterizations of such nonlinearities may provide more information for accurate and efficient representations of real structures. To this end, updating nonlinear models is essential for accurate response and reliability predictions [14–16] and assessment of structures subjected to large loads such as earthquakes [17–19].

The core idea of model updating techniques is to find the most probable values of the structural model parameters by minimizing the difference between the FE-predicted and the measured responses [20]. The non-probabilistic approaches, or refereed as deterministic methods, can be applied to address such problems. Several studies have already demonstrated a good performance of the deterministic strategies [21,22]. However, a common shortcoming exists in those methods as well. Although the most plausible values of the model parameters can be estimated, the effect of the parameter uncertainties is often neglected. Such uncertainties can be arisen from the model error, the measurement noise or the changing environmental and ambient conditions. Quantifying the uncertainties is necessary for understanding the statistical characteristics of model parameters as well as propagating those uncertainties to predict the system quantities of interests (QoI). A remedy for the

---

\* Corresponding Author  
Email Address: [costasp@uth.gr](mailto:costasp@uth.gr)

deterministic approaches is to apply the probabilistic means in model updating process. Due to their rigorous probabilistic framework, Bayesian inference methods have been widely used for quantifying and propagating the uncertainties in model updating [23–27]. Several contributions based on Bayesian strategies have already been proposed for nonlinear model updating. Muto and Beck developed a Bayesian updating method for estimating the hysteretic material model parameters using stochastic simulation [28]. Ebrahimian et al. presented a framework for damage identification of dynamical structures with material nonlinearities using batch Bayesian estimations [29]. Song et al. proposed a Bayesian model updating methodology for dynamical systems with geometric nonlinearities based on the nonlinear normal modes extracted from broadband vibration data [30]. Ceravolo et al. employed a Bayesian uncertainty quantification framework for the identification of hysteretic parameters with consideration of the model discrepancy in seismic structural health monitoring [31]. More investigations for nonlinear model updating based on the Bayesian techniques can be found in the literature [32–39].

Bayesian inference provides a powerful probabilistic tool for updating nonlinear models and handling the uncertainties of nonlinear model parameters. However, the conventional Bayesian inference framework cannot properly account for an underlying variability in model parameters and uncertainties arising from multiple data sets under different excitations, operational, environmental and experimental conditions. The variability in the model parameters can originate from the presence of model and experimental error [40]. The uncertainty of the model parameters due to these variabilities is irreducible, in contrast to the identification uncertainty which is usually inversely proportional to the amount of data considered in a data set. To properly account for this irreducible variability, uncertainties can be embedded in the model parameters by assigning a parameterized hyper prior distribution. The hyper parameters in this distribution are assumed to be unknown quantities and to be estimated from the available multiple data sets.

A hierarchical Bayesian modeling (HBM) framework has recently been introduced in various engineering fields [41–47]. In the field of structural dynamics, it was initially proposed by Behmanesh et al. for structural identification based upon a mainly full simulation HBM approach [42]. For improving the efficiency of the HBM framework, Sedehi et al. and Jia et al. respectively developed asymptotic approximation-included HBM approaches for model updating based on time-domain [48,49] and frequency-domain linear models [50]. Patsialis et al. applied the HBM framework for reduced order structural models in earthquake engineering [51]. Such improvements have been successfully applied in updating linear models.

In this paper, a new time-domain hierarchical Bayesian modeling framework is developed for the identification of nonlinear models, aiming to quantifying the uncertainties of the nonlinear model parameters and the prediction error parameters, and further propagating the overall uncertainties to the system output QoI. The contributions of this work are in the following aspects. A methodology is proposed for nonlinear model updating using response time history data which can characterize the nonlinear behaviors in the real structures. More importantly, the proposed methodology can capture the uncertainties due to model error by embedding uncertainties in the model and prediction error parameters. This is achieved by assigning a parameterized distribution in these parameters with hyper parameters to be estimated from the multiple datasets. Moreover, the presented methodology adopts a novel asymptotic approximation approach which can significantly improve the computational efficiency of the HBM framework.

The paper is organized as follows. Section 2 presents the detailed mathematical formulation of the proposed HBM framework and the approximations used. Section 3 applies the HBM framework to identify parameters of a nonlinear system with nonlinearities modelled by the Bouc-Wen hysteresis law. Section 4 provides two numerical examples to demonstrate the effectiveness of the proposed approach to account for uncertainties due to model error. Section 5 reports the conclusions of this study.

## 2. Hierarchical Bayesian Nonlinear Model Updating

### 2.1. Proposed hierarchical models

Suppose that  $N_D$  data sets of measured vibration time histories  $\mathbf{D} = \{D_i, i = 1, \dots, N_D\}$  subjected to  $N_D$  known input loadings are available from a nonlinear structure. Let  $D_i = \{\hat{\mathbf{Y}}_i(j) \in R^{N_0}, j = 1, 2, \dots, N_i\}$  be the  $i$ -th experimental data set consisting of a sequence of response data measured at  $N_0$  degrees of freedom (DOF), where the notation  $j$  corresponds to a time instant  $t_j = j\Delta t_i$  and  $N_i$  is the number of the sampled data using the sampling rate  $\Delta t_i$ . Let also  $\mathbf{X}_i = \{\mathbf{X}_i(j), j = 1, 2, \dots, N_i\}$  be the  $i$ -th input loading which corresponds to the  $i$ -th data set  $D_i$ . Consider a parameterized class of nonlinear model  $M(\boldsymbol{\theta})$  that is used to

characterize the nonlinear behavior (e.g. material nonlinearities) of dynamical systems, where  $\boldsymbol{\theta} \in R^{N_\theta}$  is the set of material/structural model parameters to be estimated using the measured response time histories, and  $N_\theta$  is the total number of the unknown parameters in the set  $\boldsymbol{\theta}$ . Let also  $\mathbf{g}_i(\boldsymbol{\theta}) = \{\mathbf{g}_i(j; \boldsymbol{\theta}) \in R^{N_s}, j = 1, 2, \dots, N_s\}$  be the response time histories under the same input loading  $\mathbf{X}_i$  and the same sampling rate  $\Delta t_i$  predicted from the model  $M(\boldsymbol{\theta})$ , where  $N_s$  denotes the number of DOFs. Subsequently, the discrepancy  $\boldsymbol{\varepsilon}_i$  between the  $i$ -th experimental data set and the  $i$ -th predicted response time histories can be defined based on the prediction error equation:

$$\boldsymbol{\varepsilon}_i = \hat{\mathbf{Y}}_i(j) - L\mathbf{g}_i(j; \boldsymbol{\theta}) \quad (1)$$

where  $L$  is a selection matrix, usually containing elements of zeroes and ones, that associates the DOF of the model with the measured DOF.

In the analysis that follows, the prediction error is modelled by a random vector with Gaussian probability distribution with zero mean and covariance matrix  $\boldsymbol{\Sigma}$ . A probabilistic model  $p(\boldsymbol{\Sigma})$  is used herein to describe the uncertainty of the prediction error term. Realization of the prediction error parameters is free to vary across the different data sets, with the realization  $\boldsymbol{\Sigma}_i$  that corresponds to the data set  $D_i$  considered to be an independent sample of the distribution  $p(\boldsymbol{\Sigma})$ . For the  $i$ -th data set  $D_i$ ,  $\boldsymbol{\Sigma}_i$  is assumed as a diagonal matrix

with  $l$ -th diagonal element  $(\sigma_i a_{i,l})^2$ , where  $a_{i,l} = \sqrt{\frac{1}{N} \sum_{j=1}^N \mathbf{g}_{i,l}^2(j; \boldsymbol{\theta})}$  denotes the intensity of the model predictions in the  $l$ -th DOF corresponding to the  $i$ -th data set. Thus, the probabilistic model  $p(\boldsymbol{\Sigma})$  can be equally represented as the probabilistic model  $p(\sigma^2)$ , and it is modelled by the inverse gamma (IG) distribution given by:

$$p(\sigma^2) = \text{IG}(\sigma^2 | \lambda_1, \lambda_2) = \frac{(\lambda_2)^{\lambda_1}}{\Gamma(\lambda_1)} (\sigma^2)^{-\lambda_1-1} \exp\left(-\frac{\lambda_2}{\sigma^2}\right) \quad (2)$$

where the parameters  $\lambda_1$  and  $\lambda_2$  are the shape hyper-parameter and scale hyper-parameter, respectively.

Similarly, the uncertainty of model parameters  $\boldsymbol{\theta}$  is probabilistically modeled using a Gaussian distribution [42,48]:

$$p(\boldsymbol{\theta} | \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) = N(\boldsymbol{\theta} | \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta) \quad (3)$$

with unknown mean  $\boldsymbol{\mu}_\theta \in R^{N_\theta}$  and covariance matrix  $\boldsymbol{\Sigma}_\theta \in R^{N_\theta \times N_\theta}$ . The parameter set  $\{\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta\}$  is considered to be an uncertain hyper parameter set to be estimated using the available multiple data sets. Realization of  $\boldsymbol{\theta}$  from the Gaussian distribution  $N(\boldsymbol{\theta} | \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$  can vary across the different data set, where the realization  $\boldsymbol{\theta}_i$  is considered to be an independent sample of the distribution  $N(\boldsymbol{\theta} | \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta)$  that corresponds to the data set  $D_i$ .

This constitutes a hierarchy model that has two classes of parameters. The first class of model parameters comprises the  $N_D(N_\theta + 1)$  experiment-specific parameters  $\boldsymbol{\theta}_i$  and  $\sigma_i^2$ ,  $i = 1, \dots, N_D$ , while the second class model parameters comprises at most  $N_\theta + 2 + N_\theta(N_\theta + 1)$  hyper-parameters involved in  $\boldsymbol{\mu}_\theta$ ,  $\boldsymbol{\Sigma}_\theta$ ,  $\lambda_1$  and  $\lambda_2$ . The number of parameters in the first set increases linearly with the number of datasets making the parameter estimation problem challenging when the number of data sets increases. The full set  $\Theta$  of all parameters to be identified is  $\Theta = \{\{\boldsymbol{\theta}_i\}_{i=1}^{N_D}, \{\sigma_i^2\}_{i=1}^{N_D}, \boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta, \lambda_1, \lambda_2\}$ .

The graphical representation of the proposed hierarchical Bayesian nonlinear modeling framework, showing the hierarchical structure, is depicted in Fig. 1. The arrows show the conditional dependence of parameters. For example, the parameters  $\boldsymbol{\theta}_i$  are conditional on  $\boldsymbol{\mu}_\theta$  and  $\boldsymbol{\Sigma}_\theta$ , the prediction error parameters  $\sigma_i^2$  are conditional on  $\lambda_1$  and  $\lambda_2$ , while the model predictions  $\mathbf{g}_{i,l}$  are conditioned on  $\boldsymbol{\theta}_i$ . First the unknown hyper-parameters consisting of the set of variables  $\{\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta, \lambda_1, \lambda_2\}$  are identified, where the first two variables describe the uncertainty of the model parameters  $\boldsymbol{\theta}$  while the other two parameters capture the uncertainty corresponding to the prediction error parameter  $\sigma^2$ . The structural parameter uncertainty along with the prediction error uncertainty can then be propagated to the predictions of quantities of interest (QoI). The

theoretical details are formulated in the following section.

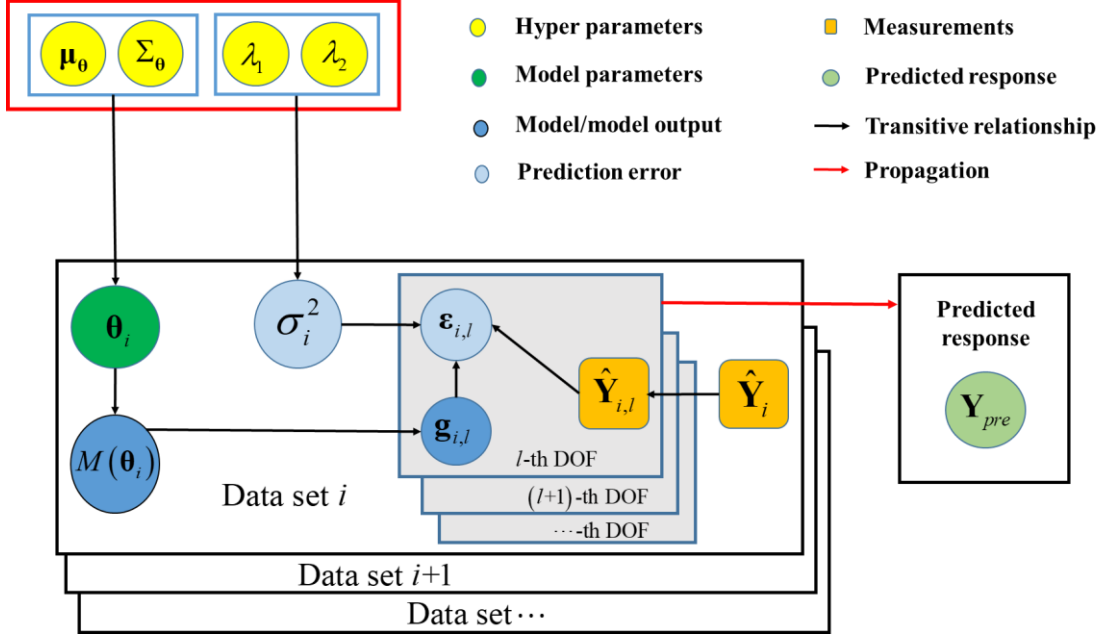


Fig. 1 Graphical representation of the proposed hierarchical Bayesian modeling framework for nonlinear model updating

## 2.2. Formulation for the proposed HBM framework

### 2.2.1. Posterior distribution of full set of parameters

The full parameters include the model parameters  $\{\boldsymbol{\theta}_i\}_{i=1}^{N_D}$ , the prediction error parameters  $\{\sigma_i^2\}_{i=1}^{N_D}$  and the hyper-parameters  $\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2\}$ . In the case of considering  $N_D$  independent data sets, the joint prior distribution of the full parameters is expressed as:

$$p(\Theta) = p(\{\boldsymbol{\theta}_i\}_{i=1}^{N_D}, \{\sigma_i^2\}_{i=1}^{N_D}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) = p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) \prod_{i=1}^{N_D} p(\boldsymbol{\theta}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) p(\sigma_i^2 | \lambda_1, \lambda_2) \quad (4)$$

where  $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2)$  denotes the prior distribution of hyper-parameters, and  $p(\boldsymbol{\theta}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  and  $p(\sigma_i^2 | \lambda_1, \lambda_2)$  are introduced in Eqs. (3) and (2), respectively. In developing Eq. (4), the hierarchy structure in Fig. 1 is assumed. Specifically, the experiment-specific parameter set  $\boldsymbol{\theta}_i$  and  $\sigma_i^2$  are independent, the distribution of  $\boldsymbol{\theta}_i$  is independent of the values of the prediction error hyper parameters  $\lambda_1$  and  $\lambda_2$ , while the distribution for  $\sigma_i^2$  is independent of the structural model hyper parameters  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\Sigma}_0$ . According to Bayes' theorem, the posterior distribution of full parameters is proportional to the prior distribution and the likelihood function:

$$p(\Theta | \mathbf{D}) \propto p(\mathbf{D} | \Theta) p(\Theta) \quad (5)$$

Due to the independence of individual data set, and the fact that the  $i$ -th data set  $D_i$  depends only on the  $i$ -th model parameter  $\boldsymbol{\theta}_i$  and its prediction error parameter  $\sigma_i^2$ , the likelihood function can be simplified as:

$$p(\mathbf{D} | \Theta) = p(\mathbf{D} | \{\boldsymbol{\theta}_i\}_{i=1}^{N_D}, \{\sigma_i^2\}_{i=1}^{N_D}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) = \prod_{i=1}^{N_D} p(D_i | \boldsymbol{\theta}_i, \sigma_i^2) \quad (6)$$

Herein  $p(D_i | \boldsymbol{\theta}_i, \sigma_i^2)$  is the likelihood function for a specific data set  $i$ , which can be readily obtained based on the prediction error equation in Eq. (1):

$$p(D_i | \boldsymbol{\theta}_i, \sigma_i^2) \propto (\sigma_i^2)^{-\frac{N_0 N_i}{2}} \exp\left\{-\frac{N_0 N_i}{2\sigma_i^2} J(\boldsymbol{\theta}_i)\right\} \quad (7)$$

where  $J(\boldsymbol{\theta}_i)$  is stated as:

$$J(\boldsymbol{\theta}_i) = \frac{1}{N_0 N_i} \sum_{l=1}^{N_0} \frac{1}{a_{i,l}^2} \sum_{j=1}^{N_i} (\hat{\mathbf{Y}}_i(j) - \mathbf{g}_i(j; \boldsymbol{\theta}_i))^2 \quad (8)$$

Note that  $J(\boldsymbol{\theta}_i)$  tends to a finite value representing the average discrepancy between the measurements and the model predictions. Subsequently, using the expressions in Eqs. (2) and (3), and substituting the simplified likelihood function from Eq. (6) into Eq. (5) yields:

$$\begin{aligned} p(\boldsymbol{\Theta} | \mathbf{D}) &= p(\{\boldsymbol{\theta}_i\}_{i=1}^{N_D}, \{\sigma_i^2\}_{i=1}^{N_D}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) \\ &\propto p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) \prod_{i=1}^{N_D} p(D_i | \boldsymbol{\theta}_i, \sigma_i^2) N(\boldsymbol{\theta}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) IG(\sigma_i^2 | \lambda_1, \lambda_2) \end{aligned} \quad (9)$$

### 2.2.2. Marginal posterior distribution of hyper-parameters

For obtaining the posterior distribution of hyper parameters, the computational procedure proposed by the previous study [48] will be followed in this section, namely the joint distribution in Eq. (9) firstly will be marginalized over  $\sigma_i^2$ 's and then over  $\boldsymbol{\theta}_i$ 's.

Marginalizing the joint distribution of full parameters over  $\sigma_i^2$ 's, one can obtain:

$$p(\{\boldsymbol{\theta}_i\}_{i=1}^{N_D}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) \propto p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) \prod_{i=1}^{N_D} \left[ N(\boldsymbol{\theta}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \int_{\sigma_i^2} p(D_i | \boldsymbol{\theta}_i, \sigma_i^2) IG(\sigma_i^2 | \lambda_1, \lambda_2) d\sigma_i^2 \right] \quad (10)$$

The integral in Eq. (10) is evaluated analytically in Appendix A, resulting in:

$$\int_{\sigma_i^2} p(D_i | \boldsymbol{\theta}_i, \sigma_i^2) IG(\sigma_i^2 | \lambda_1, \lambda_2) d\sigma_i^2 \propto \frac{(\lambda_2)^{\lambda_1}}{\Gamma(\lambda_1)} \Gamma(f(\lambda_1)) \left[ \frac{N_0 N_i}{2} J(\boldsymbol{\theta}_i) + \lambda_2 \right]^{-f(\lambda_1)} \quad (11)$$

where  $\Gamma(\cdot)$  is the Gamma function and  $f(\lambda_1) = (N_0 N_i + 2\lambda_1) / 2$ . Hence, Eq. (10) can be rewritten as:

$$p(\{\boldsymbol{\theta}_i\}_{i=1}^{N_D}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) \propto p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) \left[ \frac{(\lambda_2)^{\lambda_1}}{\Gamma(\lambda_1)} \Gamma(f(\lambda_1)) \right]^{N_D} \prod_{i=1}^{N_D} \left[ N(\boldsymbol{\theta}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \exp(-L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)) \right] \quad (12)$$

where  $L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)$  is defined in the form:

$$L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2) = f(\lambda_1) \ln \left[ \frac{N_0 N_i}{2} J(\boldsymbol{\theta}_i) + \lambda_2 \right] \quad (13)$$

To simplify the analysis and derive analytical expressions for the posterior PDF of the hyper parameters, a key asymptotic approximation, valid for large number of data, is next introduced. The function  $\exp(-L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2))$  can be approximated by using Taylor expansion when a large number of data points are available [41]:

$$\exp(-L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)) \approx \exp \left( -L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2) - \frac{1}{2} (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i)^T \mathbf{H}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2) (\boldsymbol{\theta}_i - \hat{\boldsymbol{\theta}}_i) \right) \quad (14)$$

Herein  $\hat{\boldsymbol{\theta}}_i$  is the most probable values (MPV) computed by minimizing the objective function  $L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)$ , while  $L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)$  and  $\mathbf{H}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)$  are respectively the function value and the hessian matrix of the function  $L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)$  with respect to the parameters  $\boldsymbol{\theta}_i$  evaluated at the MPV  $\hat{\boldsymbol{\theta}}_i$ . The calculation of the MPV  $\hat{\boldsymbol{\theta}}_i$  and the hessian matrix  $\mathbf{H}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)$  are derived in Appendix B and are shown to be given by

$$\hat{\boldsymbol{\theta}}_i = \arg \min_{\boldsymbol{\theta}_i} (J(\boldsymbol{\theta}_i)) \quad (15)$$

$$\mathbf{H}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2) = \frac{N_0 N_i f(\lambda_1)}{N_0 N_i J(\hat{\boldsymbol{\theta}}_i) + 2\lambda_2} \mathbf{H}(\hat{\boldsymbol{\theta}}_i) \quad (16)$$

where  $\mathbf{H}(\hat{\boldsymbol{\theta}}_i) = \nabla \nabla^T J(\boldsymbol{\theta}_i) \Big|_{\boldsymbol{\theta}_i = \hat{\boldsymbol{\theta}}_i}$  is the hessian matrix of the measure of fit function  $J(\boldsymbol{\theta}_i)$  evaluated at  $\hat{\boldsymbol{\theta}}_i$ . It is noted that the MPV value can be readily solved by minimizing the function  $J(\boldsymbol{\theta}_i)$  which is independent of the unknown hyper parameters  $\lambda_1$  and  $\lambda_2$ . Moreover, an explicit expression is found for the hessian matrix

$\mathbf{H}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)$  in terms of the hessian of the function  $J(\boldsymbol{\theta}_i)$  and the hyper-parameters  $\lambda_1$  and  $\lambda_2$ . Based on the calculations of Eq. (15) and Eq. (16), Eq. (14) can be rewritten as:

$$\exp(-L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)) \propto \left[ \frac{N_0 N_i}{2} J(\hat{\boldsymbol{\theta}}_i) + \lambda_2 \right]^{-f(\lambda_1)} \sqrt{|\boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)|} N(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_i, \boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)) \quad (17)$$

where the covariance matrix

$$\boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2) = \frac{1}{N_0 N_i} \frac{N_0 N_i J(\hat{\boldsymbol{\theta}}_i) + 2\lambda_2}{f(\lambda_1)} \mathbf{H}^{-1}(\hat{\boldsymbol{\theta}}_i) = \frac{2}{N_0 N_i} \frac{J(\hat{\boldsymbol{\theta}}_i) + \frac{2}{N_0 N_i} \lambda_2}{1 + \frac{2}{N_0 N_i} \lambda_1} \mathbf{H}^{-1}(\hat{\boldsymbol{\theta}}_i) \quad (18)$$

is the inverse of the hessian matrix  $\mathbf{H}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)$ . One can view expressions (15) and (18) as quantifying the most probable value and the identification uncertainty corresponding to Bayesian parameter inference of  $\boldsymbol{\theta}_i$  using the dataset  $D_i$  under uniform prior distribution for  $\boldsymbol{\theta}_i$ .

It is noted that for large number of data points  $N_0 N_i$ , Eq. (18) can be approximately written as:

$$\boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2) = \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \approx \frac{2}{N_0 N_i} J(\hat{\boldsymbol{\theta}}_i) \mathbf{H}^{-1}(\hat{\boldsymbol{\theta}}_i) \quad (19)$$

demonstrating that the covariance matrix  $\boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2) = \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i)$  is independent of the hyper parameters  $\lambda_1$  and  $\lambda_2$ . More importantly, Eq. (19) shows the uncertainty in the estimates of  $\boldsymbol{\theta}_i$  is inversely proportional to the square root of the number of data points, reflecting the fact that the identification uncertainty decreases as the number of data points in a dataset increases. Substituting the expression from Eq. (17) into Eq. (12) yields:

$$p(\{\boldsymbol{\theta}_i\}_{i=1}^{N_D}, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) \propto p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) T(\hat{\boldsymbol{\theta}}, \lambda_1, \lambda_2) \prod_{i=1}^{N_D} \left[ N(\boldsymbol{\theta}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) N(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_i, \boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)) \right] \quad (20)$$

Herein the function  $T(\hat{\boldsymbol{\theta}}, \lambda_1, \lambda_2)$  is defined as:

$$T(\hat{\boldsymbol{\theta}}, \lambda_1, \lambda_2) = \left[ \frac{(\lambda_2)^{\lambda_1}}{\Gamma(\lambda_1)} \Gamma(f(\lambda_1)) \right]^{N_D} \prod_{i=1}^{N_D} \left[ \frac{N_0 N_i}{2} J(\hat{\boldsymbol{\theta}}_i) + \lambda_2 \right]^{-f(\lambda_1)} \sqrt{|\boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)|} \quad (21)$$

Subsequently, by marginalizing the distribution in Eq. (20) over  $\boldsymbol{\theta}_i$ 's space for  $i=1, \dots, N_D$ , and noting that

$$\int_{\boldsymbol{\theta}_i} N(\boldsymbol{\theta}_i | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) N(\boldsymbol{\theta}_i | \hat{\boldsymbol{\theta}}_i, \boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)) d\boldsymbol{\theta}_i = N(\boldsymbol{\mu}_0 | \hat{\boldsymbol{\theta}}_i, \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2))$$

one readily obtains the posterior distribution of the hyper parameters as follows:

$$p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) \propto p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) T(\hat{\boldsymbol{\theta}}, \lambda_1, \lambda_2) \prod_{i=1}^{N_D} N(\boldsymbol{\mu}_0 | \hat{\boldsymbol{\theta}}_i, \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_L(\hat{\boldsymbol{\theta}}_i, \lambda_1, \lambda_2)) \quad (22)$$

As seen from Eq. (22), the hyper parameters corresponding to both model parameters and prediction error hyper parameters can be computed together through the proposed framework. Any Markov Chain Monte Carlo (MCMC) algorithm such as the transitional MCMC (TMCMC) [52,53] or the nested sampling algorithm [54] can be used to draw the samples from the posterior distribution of the hyper parameters. Note that sampling from the posterior distribution  $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D})$  is not a time consuming operation since it does no longer require computationally expensive model runs. The model runs are required only to estimate and store the values of  $\hat{\boldsymbol{\theta}}_i$  and  $\mathbf{H}(\hat{\boldsymbol{\theta}}_i)$  before the sampling approach is initiated.

### 2.2.3. Posterior distribution of hyper parameters for large number of data

For large number of data ( $N_0 N_i$  is large) within each data set, the values of  $f(\lambda_1) = N_0 N_i / 2$  and  $\Gamma(f(\lambda_1)) = \Gamma(N_0 N_i / 2)$  are independent of the parameter  $\lambda_1$ . Using Eq. (19) and assuming that prior to data the hyper parameter sets  $\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\}$  and  $\{\lambda_1, \lambda_2\}$  are independent, i.e.  $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) = p(\lambda_1, \lambda_2) p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ , the posterior distribution in Eq. (22) takes the simplified form

$$p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) \approx p(\lambda_1, \lambda_2 | \mathbf{D}) p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 | \mathbf{D}) \quad (23)$$

where

$$p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 | \mathbf{D}) \propto \prod_{i=1}^{N_D} N(\boldsymbol{\mu}_0 | \hat{\boldsymbol{\theta}}_i, \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i)) p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \quad (24)$$

and

$$p(\lambda_1, \lambda_2 | \mathbf{D}) \propto T(\hat{\boldsymbol{\theta}}, \lambda_1, \lambda_2) \propto \left[ \frac{(\lambda_2)^{\lambda_1}}{\Gamma(\lambda_1)} \right]^{N_D} \prod_{i=1}^{N_D} \left[ \frac{N_0 N_i}{2} J(\hat{\boldsymbol{\theta}}_i) + \lambda_2 \right]^{-N_0 N_i / 2} p(\lambda_1, \lambda_2) \quad (25)$$

According to Eq. (23) the hyper parameter sets  $\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\}$  and  $\{\lambda_1, \lambda_2\}$  remain independent given the data sets. This independence is useful to analytically derive the most probable values and the uncertainties of the hyper parameters  $\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\}$  and  $\{\lambda_1, \lambda_2\}$  by separately considering the distribution  $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 | \mathbf{D})$  and  $p(\lambda_1, \lambda_2 | \mathbf{D})$ .

It has been shown in [55] that the form in Eq. (24) of the posterior distribution of the hyper parameters  $\{\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0\}$  yields the most probable values of the hyper parameters to be

$$\hat{\boldsymbol{\mu}}_0 = \frac{1}{N_D} \sum_{i=1}^{N_D} \hat{\boldsymbol{\theta}}_i \quad (26)$$

and using the assumption that all  $\tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i)$  are approximately equal, the hyper-parameter covariance matrix to be

$$\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{N_D} \sum_{i=1}^{N_D} (\hat{\boldsymbol{\mu}}_0 - \hat{\boldsymbol{\theta}}_i)(\hat{\boldsymbol{\mu}}_0 - \hat{\boldsymbol{\theta}}_i)^T - \boldsymbol{\Sigma}_0 \quad (27)$$

where  $\boldsymbol{\Sigma}_0 = \frac{1}{N_D} \sum_{i=1}^{N_D} \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i)$  is taken as the average of the identification uncertainty of each dataset. The uncertainty in the hyper parameters, approximated by the inverse of the hessian of  $-\ln p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 | \mathbf{D})$  evaluated at the most probable values  $\{\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0\}$ , can be derived to be

$$\boldsymbol{\Sigma}(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0) = \frac{1}{N_D} \mathbf{H}_h^{-1}(\hat{\boldsymbol{\mu}}_0, \hat{\boldsymbol{\Sigma}}_0) \quad (28)$$

where  $\mathbf{H}_h(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$  is a 2x2 block matrix with the individual blocks given by [55]

$$\begin{aligned} \mathbf{H}_{\boldsymbol{\mu}\boldsymbol{\mu}}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) &= \frac{1}{N_D} \sum_{i=1}^{N_D} \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} \\ \mathbf{H}_{\boldsymbol{\mu}\boldsymbol{\Sigma}}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) &= -\frac{1}{N_D} \sum_{i=1}^{N_D} \left[ \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} \otimes \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} (\boldsymbol{\mu}_0 - \hat{\boldsymbol{\theta}}_i) \right] \\ \mathbf{H}_{\boldsymbol{\Sigma}\boldsymbol{\Sigma}}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) &= -\frac{1}{2N_D} \sum_{i=1}^{N_D} \left[ \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} \otimes \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} \right] \\ &\quad + \frac{1}{N_D} \sum_{i=1}^{N_D} \left[ \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} \otimes \left[ \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} (\boldsymbol{\mu}_0 - \hat{\boldsymbol{\theta}}_i)(\boldsymbol{\mu}_0 - \hat{\boldsymbol{\theta}}_i)^T \left[ \boldsymbol{\Sigma}_0 + \tilde{\boldsymbol{\Sigma}}_L(\hat{\boldsymbol{\theta}}_i) \right]^{-1} \right] \right] \end{aligned} \quad (29)$$

where  $\otimes$  denotes the Kronecker tensor product of two matrices. Note that the matrices in the left-hand-side of Eq. (29) tend to finite values as the number of datasets  $N_D$  increases.

Using Eq. (25), the most probable values  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$  can be obtained by minimizing the  $-\ln p(\lambda_1, \lambda_2 | \mathbf{D})$  with respect to the parameters  $\lambda_1$  and  $\lambda_2$ , while an asymptotic estimate of the uncertainty in the estimates of  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$  can readily be derived by calculating the inverse of the hessian of  $-\ln p(\lambda_1, \lambda_2 | \mathbf{D})$ , evaluated at the most probable values  $\hat{\lambda}_1$  and  $\hat{\lambda}_2$ , in the form:

$$\boldsymbol{\Sigma}_\lambda(\hat{\lambda}_1, \hat{\lambda}_2) = \frac{1}{N_D} \mathbf{H}_\lambda^{-1}(\hat{\lambda}_1, \hat{\lambda}_2) \quad (30)$$

where the elements of the 2x2 hessian matrix are

$$\begin{aligned}
\mathbf{H}_{\lambda,11}(\lambda_1, \lambda_2) &= \frac{d^2 \ln \Gamma(\lambda_1)}{d\lambda_1^2} = \psi'(\lambda_1) = \psi(1, \lambda_1) \\
\mathbf{H}_{\lambda,12}(\lambda_1, \lambda_2) &= -\frac{1}{\lambda_2} \\
\mathbf{H}_{\lambda,22}(\lambda_1, \lambda_2) &= \frac{\lambda_1}{\lambda_2^2} - \frac{2}{N_0 N_i} \frac{1}{N_D} \sum_{i=1}^{N_D} \frac{1}{\left[ J_i(\hat{\boldsymbol{\theta}}_i) + \frac{2}{N_0 N_i} \lambda_2 \right]^2} \approx \frac{\lambda_1}{\lambda_2^2}
\end{aligned} \tag{31}$$

where  $\psi(\cdot)$  is the Digamma function and  $\psi(k, \lambda_1)$  is the  $k$ -th derivative of the digamma function at  $\lambda_1$ , which is readily evaluated by a Matlab-based function ‘psi’. It is noted that Eq. (31) does not explicitly depend on the estimates  $\hat{\boldsymbol{\theta}}_i$  for large number of data per dataset (large  $N_0 N_i$ ).

The aforementioned formulations (28) and (30) for the covariance matrices of the estimates of all hyper parameters to be inversely proportional to the number  $N_D$  of the datasets suggest that the uncertainties in the hyper parameter estimates reduces as the number of datasets increases. Also, the forms (28) and (30) for the covariance matrices suggest that the posterior uncertainty in the hyper parameters tends to be approximated by a Gaussian distribution as the number  $N_D$  increases.

#### 2.2.4. Probability distribution function (PDF) of model and prediction error parameters

The PDF of model parameters  $\boldsymbol{\theta}$  given the data  $D$  can be simplified to

$$p(\boldsymbol{\theta} | \mathbf{D}) = \int \int \int \int p(\boldsymbol{\theta} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) d\boldsymbol{\mu}_0 d\boldsymbol{\Sigma}_0 d\lambda_1 d\lambda_2 \tag{32}$$

where we used the fact that  $p(\boldsymbol{\theta} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2) = p(\boldsymbol{\theta} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ , i.e. the conditional distribution of the model parameter  $\boldsymbol{\theta}$  given the values of the hyper parameters  $\boldsymbol{\mu}_0$  and  $\boldsymbol{\Sigma}_0$  is independent of hyper-parameters  $\lambda_1, \lambda_2$ . Sampling estimates can be employed to compute the integral in the form:

$$p(\boldsymbol{\theta} | \mathbf{D}) \approx \frac{1}{M} \sum_{k=1}^M N(\boldsymbol{\theta} | \boldsymbol{\mu}_0^{(k)}, \boldsymbol{\Sigma}_0^{(k)}) \tag{33}$$

by drawing the samples  $\boldsymbol{\mu}_0^{(k)}, \boldsymbol{\Sigma}_0^{(k)}, \lambda_1^{(k)}, \lambda_2^{(k)}$  from the distribution  $p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D})$ , where  $M$  is the number of samples.

Similarly, the PDF of prediction error parameter  $\sigma^2$  can be computed by following the same procedure:

$$p(\sigma^2 | \mathbf{D}) = \int \int \int \int p(\sigma^2 | \lambda_1, \lambda_2) p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, \lambda_1, \lambda_2 | \mathbf{D}) d\boldsymbol{\mu}_0 d\boldsymbol{\Sigma}_0 d\lambda_1 d\lambda_2 \approx \frac{1}{M} \sum_{k=1}^M IG(\sigma^2 | \lambda_1^{(k)}, \lambda_2^{(k)}) \tag{34}$$

Thus, the samples of structural parameters and prediction error parameter can be obtained according to Eqs. (33) and (34), respectively using any MCMC algorithm.

#### 2.2.5. Predictions of output QoI

After the structural model is calibrated, the uncertainty of the structural parameters and the uncertainty from the prediction error parameter can be used to predict the uncertainty in output QoI using Monte Carlo simulations. Specifically, let  $\boldsymbol{\theta}^{(q)}$  denote the  $q$ -th sample of model parameter  $\boldsymbol{\theta}$  generated from its distribution  $p(\boldsymbol{\theta} | D)$  in Eq. (33),  $\sigma^{2(q)}$  is the  $q$ -th sample of prediction error variance parameter which can be generated from its distribution  $p(\sigma^2 | D)$  in Eq. (34), and  $\boldsymbol{\varepsilon}_l^{(q)}$  is the prediction error term sampled from a Gaussian distribution with zero mean and covariance matrix  $\boldsymbol{\Sigma} = (\sigma^{2(q)} a_{pre,l}^2) \mathbf{I}$ , where  $a_{pre,l}$  is the intensity of the predicted response time histories in the  $l$ -th DOF. The uncertainty in a response time history  $\mathbf{Y}_{pre}(j)$  can be obtained by analyzing the samples generated according to the following expression:

$$\mathbf{Y}_{pre,l}^{(q)}(j) = \mathbf{g}_l(j; \boldsymbol{\theta}^{(q)}) + \boldsymbol{\varepsilon}_l^{(q)} \tag{35}$$

where  $l$  defines the  $l$ -th DOF of the system and  $j$  is the data point. Eq. (34) is applied to both the observed and unobserved QoI. Observations of a specific type of response time histories (e.g. accelerations) at a limited



number of DOF provides estimates of the prediction error parameter  $\sigma^2$  which can then be used in (35) to make predictions of response time histories (e.g. accelerations) at unobserved DOF of the model. For the unobserved QoI of different type (e.g. velocities, strain, displacements) an estimate of the variance  $\sigma^2$  involved in the definition of the zero mean Gaussian error term  $\boldsymbol{\varepsilon}_i$  is not available from the observations and has to be subjectively postulated. Here, the responses to unobserved QoI of different type than that used for parameter estimation are computed based on only the structural parameters uncertainty. This is reasonable as the estimates of the prediction error are not available for such unobserved QoI.

### 2.3. Computational procedure of the proposed algorithm

The procedure for the parameter identification along with the predictions of QoI is summarized in Algorithm 1. The computational process of the proposed HBM framework includes two steps for the identification of the hyper parameters. In the first step the MPV and the hessian matrix are obtained according to each data set. In this step it requires the model runs and thus it is the most computationally expensive step. While in the second step, it does not need the model runs, and only the MPV and the Hessian matrix evaluated from the step are used for sampling the posterior distribution of the hyper parameters.

It is reminded that an alternative full sampling (FS)-based HBM approach [43,51] can be also utilized to obtain the samples of the hyper parameters. This is achieved by drawing the samples of the model and prediction error parameters from the likelihood function for each data set in the first step, and then use all the available samples to compute the posterior distribution of the hyper parameters in the second step. Therefore, the computational cost of the FS approach is more expensive than the proposed asymptotic approximation approach. The comparison of the computational effort between the FS and the proposed methods is conducted in the second application of Section 4, where both methods are used for estimating the posterior distributions of the hyper parameters.

Algorithm 1:

Proposed HBM framework for parameter identification and response prediction

#### 1. Identify hyper parameters

**First Step:** Find MPV and hessian of function  $L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)$  for each data set

- 1.1) Minimize  $J(\boldsymbol{\theta}_i)$  to compute MPV  $\hat{\boldsymbol{\theta}}_i$  using Eq. (15)
- 1.2) Evaluate Hessian matrix  $\mathbf{H}(\hat{\boldsymbol{\theta}}_i)$  at MPV  $\hat{\boldsymbol{\theta}}_i$

**Second Step:** Compute posterior distribution  $p(\boldsymbol{\mu}_0, \Sigma_0, \lambda_1, \lambda_2 | \mathbf{D})$  using Eq. (22)

#### 2. Identify model parameters and prediction error parameter

- 2.1) Compute PDF  $p(\boldsymbol{\theta} | \mathbf{D})$  of model parameters using Eq. (33)
- 2.2) Compute PDF  $p(\sigma^2 | \mathbf{D})$  of prediction error parameter using Eq. (34)

#### 3. Predictions of QoI

- 3.1) Draw samples  $\boldsymbol{\theta}^{(q)} (q=1, \dots, N_q)$  from  $p(\boldsymbol{\theta} | \mathbf{D})$ , and  $\sigma^{2(q)}$  from  $p(\sigma^2 | \mathbf{D})$
- 3.2) Draw samples from  $N(\mathbf{0}, \sigma^{2(q)} \mathbf{a}_{pre,i}^2 \mathbf{I})$  for  $\boldsymbol{\varepsilon}_i$
- 3.3) Calculate the predictions using Eq. (35)
- 3.4) Estimate the statistical properties using Monte Carlo simulations

### 3. Application to Nonlinear Systems using Bouc-Wen hysteresis

The Bouc-Wen (BW) model is widely used in dynamical structures to represent the hysteretic behavior of nonlinear systems [22,31]. It was initially proposed by Bouc [56], subsequently modified by Wen [57] and thereafter extended by other researchers in the literature [58–60]. Details of the formulation in civil infrastructure can be found in references [59,61]. Herein, for demonstration purposes, it will be applied to a multistory building represented by a shear model with BW hysteretic-type inter-story nonlinearities.

Specifically, the differential equation of motion is written in the form:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{F}_R(\mathbf{u}(t), \mathbf{z}(t)) = \mathbf{P}(t) \quad (36)$$

where  $\mathbf{M}$  and  $\mathbf{C}$  are the mass and the viscous damping matrices,  $\mathbf{u}(t)$  is the relative displacement

response,  $\mathbf{F}_R(\mathbf{u}(t), \mathbf{z}(t))$  is the nonlinear restoring force vector at time  $t$ , and  $\mathbf{P}(t)$  is the input force vector. According to the shear building model, the  $i$ -th component of the vector  $\mathbf{F}_R(\mathbf{u}(t), \mathbf{z}(t))$  is given by [28,59]:

$$F_R^i = F_r^i - F_r^{i+1} \quad (37)$$

$i = 1, \dots, N_0$ , with  $F_r^{i+1} = 0$  for  $i = N_0$ . According to the BW model the quantity  $F_r^i$  in Eq. (37) is the nonlinear inter-story restoring force given as [59]:

$$F_r^i = \alpha_i k_i u_i(t) + (1 - \alpha_i) k_i z_i(t) \quad (38)$$

where  $\alpha_i k_i u_i(t)$  corresponds to the elastic component whereas  $(1 - \alpha_i) k_i z_i(t)$  represents to the hysteretic component,  $k_i$  denotes the stiffness,  $\alpha_i$  defines the share of linear part while  $1 - \alpha_i$  define the share of nonlinear hysteretic part, and  $z_i(t)$  is the virtual hysteretic displacement which comprises the hysteretic component of the system. Without including the pinching effect and degradation functions (stiffness or strength degradation), the formulation of the implemented hysteretic displacement  $z_i(t)$  can be simplified as:

$$\dot{z}_i(t) = A_i \dot{u}_i(t) - \beta_i |\dot{u}_i(t)| |z_i(t)| |z_i(t)|^{n_i-1} - \gamma_i \dot{u}_i(t) |z_i(t)|^{n_i} \quad (39)$$

where the parameter  $A_i$  determines the tangent stiffness, and the parameters  $\beta_i, \gamma_i$  and  $n_i$  affect the shape and smoothness of the hysteretic model, respectively.

For solving the differential equation of the BW model, the state vector in state space form is given as:

$$\mathbf{y}(t) = \begin{Bmatrix} \mathbf{y}_1(t) \\ \mathbf{y}_2(t) \\ \mathbf{y}_3(t) \end{Bmatrix} = \begin{Bmatrix} \mathbf{u}(t) \\ \dot{\mathbf{u}}(t) \\ \mathbf{z}(t) \end{Bmatrix} \quad (40)$$

The derivative of the state vector  $\mathbf{y}(t)$  is readily obtained as:

$$\dot{\mathbf{y}}(t) = \begin{Bmatrix} \dot{\mathbf{y}}_1(t) \\ \dot{\mathbf{y}}_2(t) \\ \dot{\mathbf{y}}_3(t) \end{Bmatrix} = \begin{Bmatrix} \mathbf{y}_2(t) \\ \mathbf{M}^{-1} \mathbf{P}(t) - \mathbf{M}^{-1} (\mathbf{C} \mathbf{y}_2(t) + \mathbf{F}_R(\mathbf{y}_1(t), \mathbf{y}_3(t))) \\ \mathbf{A} \mathbf{y}_2(t) - \boldsymbol{\beta} |\mathbf{y}_2(t)| \otimes \mathbf{y}_3(t) \otimes |\mathbf{y}_3(t)|^{n_i-1} - \boldsymbol{\gamma} \mathbf{y}_2(t) \otimes |\mathbf{y}_3(t)|^{n_i} \end{Bmatrix} \quad (41)$$

The ordinary differential equations in (41) can then be solved numerically using the fourth-fifth order Runge-Kutta method [62]. The implemented BW model at each structural element can be fully parameterized by 6 parameters, namely the linear stiffness parameter  $\mathbf{k} = \{k_i\}_{i=1}^n$  and the nonlinear parameters  $\boldsymbol{\alpha} = \{\alpha_i\}_{i=1}^n$ ,  $\mathbf{A} = \{A_i\}_{i=1}^n$ ,  $\boldsymbol{\beta} = \{\beta_i\}_{i=1}^n$ ,  $\boldsymbol{\gamma} = \{\gamma_i\}_{i=1}^n$ ,  $\mathbf{n} = \{n_i\}_{i=1}^n$ . Selected parameters can be incorporated into the model parameters set  $\boldsymbol{\theta}$  and subsequently identified based on the proposed HBM framework.

In order to find the most probable values  $\hat{\boldsymbol{\theta}}_i$  needed in the HBM formulation in Eq. (22), one needs to solve an optimization problem of minimizing  $J(\boldsymbol{\theta}_i)$  in Eq. (8) with respect to the model parameters  $\boldsymbol{\theta}_i$ . Gradient based optimization techniques are used which require the gradient of  $J(\boldsymbol{\theta}_i)$ . These gradients require the knowledge of the derivatives  $\mathbf{y}_\theta(t)$  of the response vector  $\mathbf{y}(t)$  in (41) with respect to a model parameter  $\theta$  in the set  $\boldsymbol{\theta}_i$ . Using the form in Eq.(41), analytical expressions can readily be developed for these derivatives. Specifically, the expressions for the derivatives  $\mathbf{y}_\theta(t)$  constitute a set of differential equations that is also solved using the fourth-fifth order Runge-Kutta method. The formulation for the analytical derivatives is not included in this work, however, similar study can be found in [63].

## 4. Illustrative examples

### 4.1. Case study 1: calibration of a SDOF nonlinear system using a linear model

This case study aims to demonstrate that uncertainties from processing multiple data sets arise due to model error and to investigate the effectiveness of the HBM framework for quantifying uncertainties in the presence of model error. To introduce model error the physical system is assumed to be nonlinear and the model of the system to be linear. Specifically, a single degree of freedom (SDOF) nonlinear structure modeled by Bouc-Wen

hysteretic type nonlinearity is employed as the physical system, as shown in Fig. 2(a). The mass  $m$  and the viscous damping ratio  $\zeta$  are assigned the nominal values of 1 kg and 2%, respectively. The initial stiffness  $k$  which corresponds to the linear component is set to the nominal value of 40 N/m. The values of the parameters associated with the nonlinear component are set to  $\alpha=0.5, \beta=2, \gamma=2, n=1$  and  $A=1$ . The excitations are Gaussian sequences with mean zero and standard deviation of 5. Recorded measured data consisting of acceleration time histories that are simulated from the nonlinear model. Specifically, 100 data sets consisting of acceleration time histories are simulated from the nominal nonlinear model using different realizations of the white noise excitation. For each data set, the sampling rate is taken as 0.01s corresponding to a sampling frequency 100Hz for a total of 10 seconds. The linear model, shown in Fig. 2(b), used to represent the nonlinear system is parameterized with 2 model parameters  $\theta=(\theta_1, \theta_2)^T$  associated with the stiffness and damping ratio. The parameters  $\theta_1$  and  $\theta_2$  multiply the nominal stiffness and damping ratio, respectively.

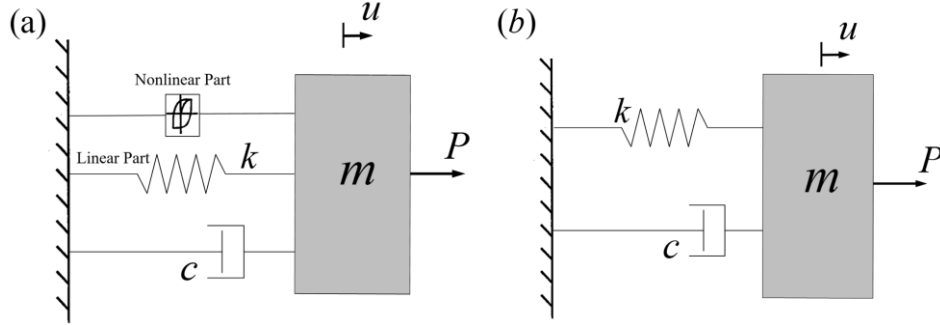


Fig. 2 (a) SDOF system with BW hysteresis (b) SDOF linear model of the system in (a)

The proposed HBM framework is next applied to identify the hyper parameters of the model and to predict the responses of the linear model. The MPVs of model parameters  $\theta_i$  for each data set is computed according to Eq. (15), and the hessian matrix evaluated at  $\theta_i$  is then calculated. The posterior distribution of the hyper-parameters  $\{\mu_0, \Sigma_0, \lambda_1, \lambda_2\}$  can then be estimated using Eq. (22). Nested sampling algorithm [54] is employed here to draw the samples from the posterior distribution. The number of the initial samples and the tolerance value in the sampler are set to 500 and 0.01, respectively. The results of the posterior distribution of the hyper-parameters corresponding to  $N_D=100$  is shown in Fig. 3, where the diagonal figures show the marginal PDF of each hyper parameter, and the lower diagonal sub-figures show the contour plots for each pair of the hyper parameters. The mean of the hyper parameters are also reported in Table 1. It can be seen from Fig. 3 that a clear peak appears for all the hyper parameters. The first two hyper parameters (hyper mean) have mean values of 0.9455 and 2.2607, respectively. These values deviate from their nominal values of  $\mu_0=(1,1)^T$  especially for the second hyper parameter due to the existence of substantial model error. Regarding the uncertainty of the model parameters, the variability of damping ratio with a coefficient of variation (cov) of 13.7% is much larger than the variability of the stiffness parameter with a cov of 1.47%. This is reasonable since the model response is more sensitive to the stiffness parameter than the damping ratio parameter. Furthermore, the marginal distribution of the hyper parameters in Fig. 3 provide the uncertainty in the estimates of the hyper parameters  $\{\mu_0, \Sigma_0\}$  with cov, computed from the samples, equal to (0.0015, 0.0138) for  $\mu_0$  and (0.0722, 0.0708) for the diagonal elements of  $\Sigma_0$ . Such uncertainties are large for the elements of the hyper covariance  $\Sigma_0$  and are expected to affect propagation of uncertainties into output QoI. As mentioned in the theoretical formulation, the uncertainty in such estimates is inversely proportional to the square root of the number of datasets and thus is expected to be higher for smaller number of datasets used. The last two parameters  $\lambda_1$  and  $\lambda_2$  aim to capture the uncertainty of the prediction error parameter. It is obvious that the two parameters exhibit a strong correlation as shown in Fig. 3 by the subplot corresponding to the projection of samples and contour to the two parameter space  $(\lambda_1, \lambda_2)$ . With those parameters one can calculate the probability of the prediction error based on Eq. (34) and the mean of the prediction error is also reported in Table 1.

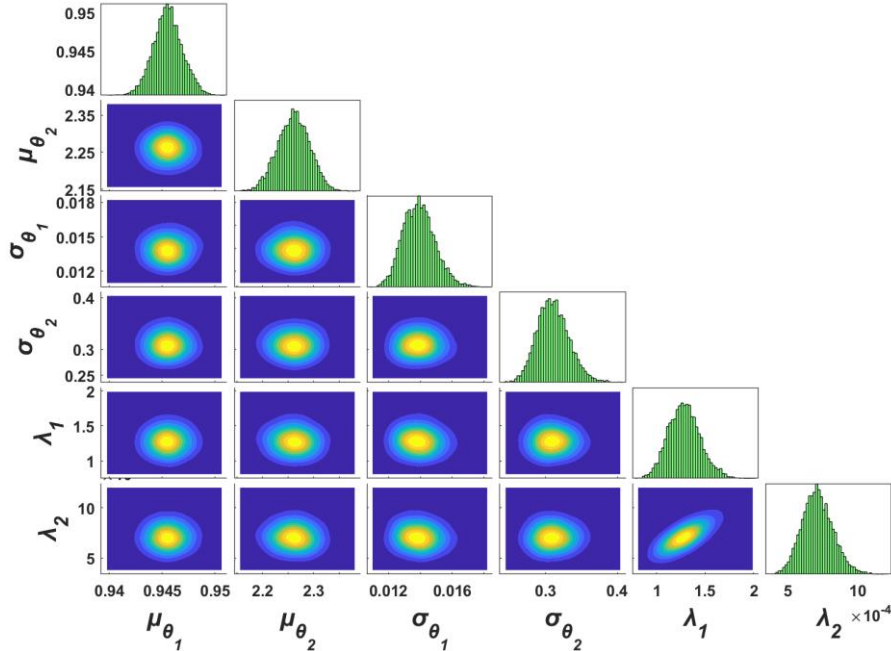


Fig. 3 Posterior distribution of the hyper-parameters

Table 1

Estimates of mean and standard deviation of the model parameters and prediction error parameters

	Model parameters						Prediction error parameters		
	$\hat{\mu}_{\theta_1}$	$\hat{\sigma}_{\theta_1}$	Cov, $\frac{\hat{\sigma}_{\theta_1}}{\hat{\mu}_{\theta_1}}$	$\hat{\mu}_{\theta_2}$	$\hat{\sigma}_{\theta_2}$	Cov, $\frac{\hat{\sigma}_{\theta_2}}{\hat{\mu}_{\theta_2}}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\sigma}$
HBM	0.9455	0.0139	1.47%	2.2607	0.3105	13.7%	1.2863	0.0007	0.0260
CBM	0.9390	0.0001	0.01%	2.4410	0.0019	0.078%	-	-	0.0690

For the purpose of the comparisons between the proposed HBM approach and the conventional Bayesian method (CBM), 100 data sets used in HBM are incorporated in a single data set for investigating the performance of the CBM. Table 1 includes results of the mean estimates and the identification uncertainty of the model parameters obtained from the CBM. As seen, the CBM provides estimates of the mean of the model parameters that are quite close to the proposed HBM approach. However, a substantial difference is revealed regarding the extent of variability of the model parameters as quantified by the standard deviations in the two approaches. Specifically, the proposed HBM framework offers larger uncertainty estimates as opposed to the CBM which yields extremely small evaluations of the model parameter standard deviations. This feature of the proposed HBM approach is beneficial for predictions of realistic uncertainty bounds of the unobserved quantities of interest. Moreover, the proposed approach can predict the hyper parameters of the prediction error, elucidating explicitly a probability distribution of the prediction error parameter. It should be noted that the mean of the prediction error computed by CBM is larger than that by HBM. This is because the prediction error in CBM include both the expected variability to the model parameters due to multiple datasets and uncertainty for each data set due to the model error. In contrast, the proposed HBM split the two uncertainties, embedding part of the uncertainty in the model parameters, and therefore present a more reasonable framework for model predictions and structure assessments.

The identification uncertainty of model parameter in each data set can be estimated according to Eq.(18), where the parameters  $\lambda_1$  and  $\lambda_2$  are assumed to be their mean evaluations from Table 1. Fig. 4 shows the plots of the identification uncertainty obtained based on individual data set and the ensemble uncertainty observed over multiple data sets for the model and prediction error parameters as a function of the number of data sets  $N_D$  ranging from 1 to 100. Apparently, the identification uncertainty for each dataset is fairly small which is attributed to the fact that a large amount of data points is used here in measurements. Such uncertainties can be negligible for further predictions. However, the ensemble uncertainty arising from the variability due to model error, is large and irreducible, tending to a constant value after approximately 10 datasets. More

importantly, this uncertainty covers almost all the mean values of the individual model parameters, representing that the model parameter values have significant variabilities due to the presence of model error.

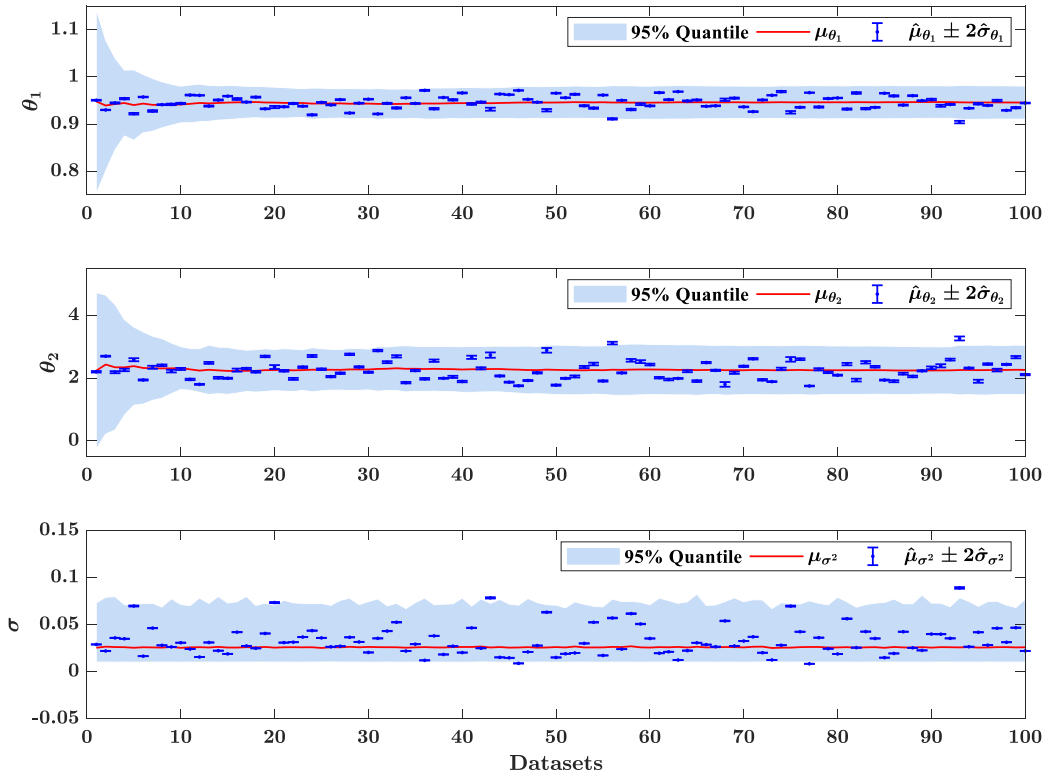


Fig. 4 Estimates of identification uncertainty and ensemble uncertainty

The above mentioned uncertainties are propagated to predict output quantities of interest. The predictions are performed for observed and unobserved QoI in order to highlight the importance of the different uncertainties. For comparison purpose, the CBM is also employed to predict the QoI. Fig. 5 shows the predictions of acceleration (observed QoI) where the uncertainties from both the model parameters and the prediction error parameter are considered. Despite the conceptual differences in HBM and CBM in quantifying and handling uncertainties, it is seen that by accounting for the overall uncertainties both methods can provide reliable uncertainty bounds (UB) and the actual response (as shown in the red line of Fig. 5(a)) which generates from a new input excitation falls inside of the UB. However, a significant difference between the HBM and CBM is shown in Fig. 6 for the predictions of displacement (unobserved QoI). Due to the fact that the prediction error is available only for the observed QoI, only the uncertainty of the model parameters is propagated to the predictions of the unobserved QoI. It is evident that owing to an underestimation of the identification uncertainty of model parameters, the CBM provides a thin UB if only the uncertainties of model parameters are considered, and the actual response falls outside of the thin UB. In contrast, the proposed HBM approach provides reasonable UB by accounting for only the uncertainties of the model parameters as most parts of the actual displacement falls within the produced UB, as depicted in Fig. 6.

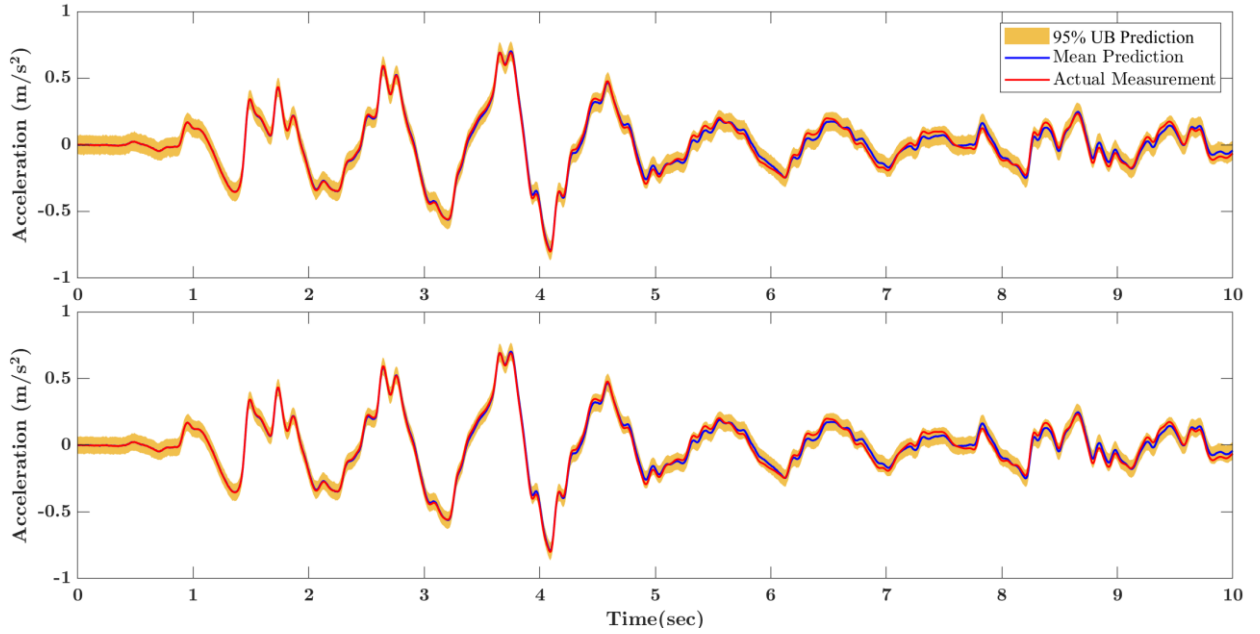


Fig. 5 Predictions of acceleration using CBM (top) and HBM (bottom)

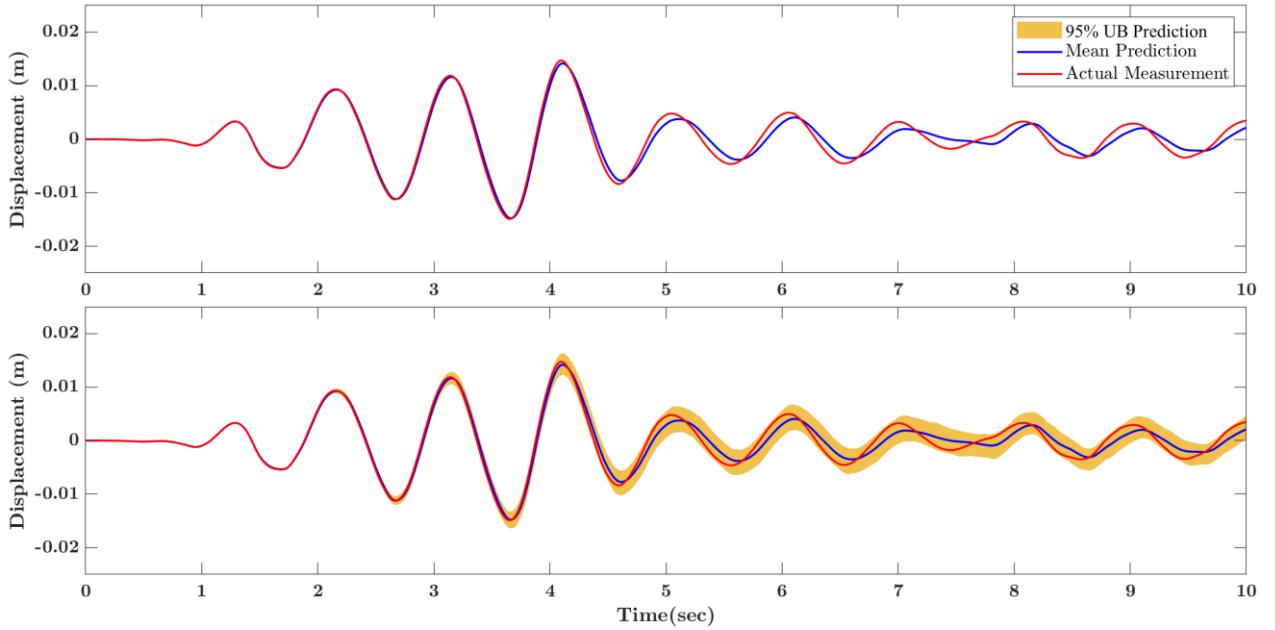


Fig. 6 Predictions of displacement using CBM (top) and HBM (bottom)

#### 4.2. Case study 2: calibration of a 5 DOF nonlinear system with BW hysteretic model

Fig. 7(a) shows a 5-DOF shear building model of a structure with mass  $m=1$  kg, stiffness  $k_0=500$  N/m, and hysteretic nonlinearities for the inter-story stiffness of each floor. The nominal values of the parameters of the nonlinear BW hysteretic model for each story are assumed to be  $\alpha_0=0.1$ ,  $\beta_0=1$ ,  $\gamma_0=1$ ,  $A=2$  and  $n=1$ , respectively. The physical system is assumed to be the model shown in Fig. 7(a) with the values of the model parameters corresponding to mass  $m$ , stiffness  $k$  and the nonlinear parameters  $\alpha, \beta$  and  $\gamma$  in all five floors of the structure independently perturbed by 5% from their nominal values. The perturbed value of each parameter is different for each floor of the building. The other two parameters  $A$  and  $n$  are set to be their nominal values. The system is excited at the base with  $N_D=25$  different earthquake excitations taken from the engineering strong motion (ESM) database [64]. One of the earthquake excitations is shown in Fig. 2(b). 25 data sets of acceleration time histories are generated and used as measurements for identifying the model of the system. Most of the analysis and results that follow are based on time history measurement obtained from a single acceleration sensor located at the first story.

The model of the system is parameterized using the four unknown parameters  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$  representing normalized  $k, \alpha, \beta$  and  $\gamma$ , denoted with red in Fig. 7(a). The parameters  $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$  are assumed to be the same for all stories and are normalized by the nominal values of the corresponding model properties such as  $k = \theta_1 k_0$ ,  $\alpha = \theta_2 \alpha_0$ ,  $\beta = \theta_3 \beta_0$  and  $\gamma = \theta_4 \gamma_0$ . Therefore, the identification will involve estimating the stiffness parameter  $\theta_1$  and evaluating the material nonlinear parameters  $(\theta_2, \theta_3, \theta_4)$ . The other quantities are assumed to be deterministic with their values set to their nominal values.

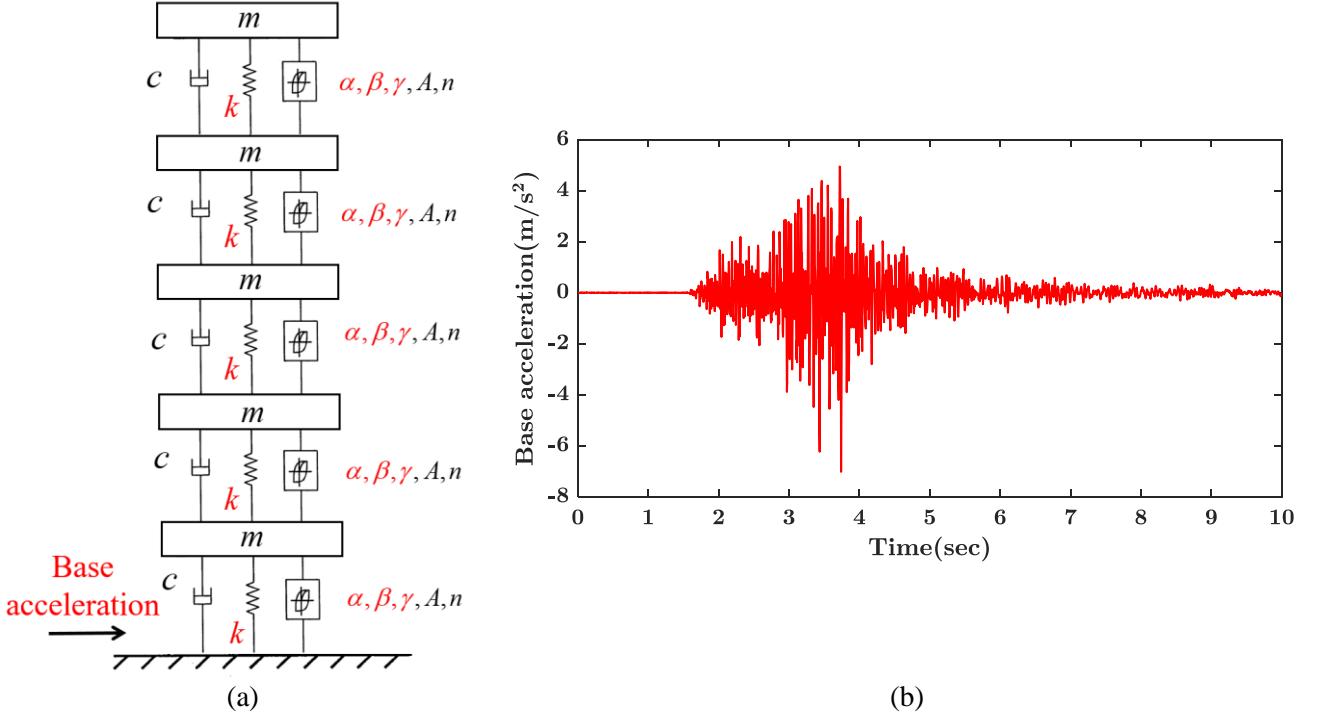


Fig. 7 (a) 5-DOF shear model of a building system (b) Base earthquake excitation

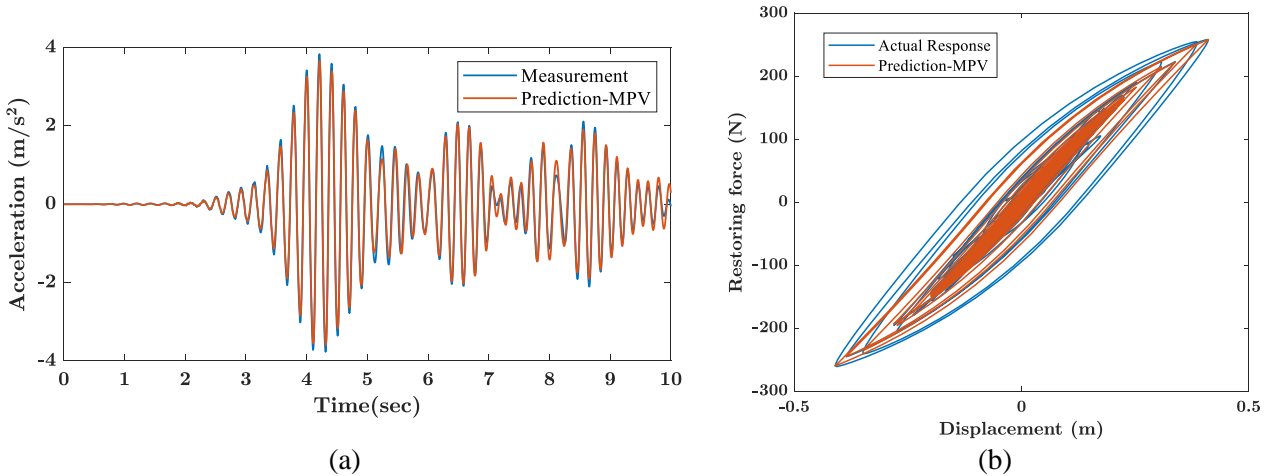


Fig. 8 (a) Comparisons between the measured acceleration response and prediction and (b) the actual hysteresis loop and prediction from optima model of the first floor.

The proposed hierarchical Bayesian nonlinear model updating approach is applied to identify the stiffness and the parameters of the hysteretic nonlinearity. The MPV of the model parameters are estimated from each data set according to the first step of the proposed algorithm. To identify whether the updated model can match the real structure, the measured accelerations subjected to the base excitation in Fig. 7(b) and the corresponding hysteresis loops of the first floor are plotted along with the responses from its optimal calibrated model, as shown in Fig. 8. As seen, there is a mismatch between the measurements and the predictions due to the presence of model error. Such discrepancy will be captured by the prediction error term.

Subsequently, the posterior distributions of the hyper-parameters corresponding to 25 data sets are computed based on Eq. (22), and the results by using a single acceleration sensor located at the first story are shown in Fig. 9. The deviation of the estimated hyper mean values  $\hat{\mu}_0$  from unity indicates their departure from the nominal values of the real structure. It is indicated that the hyper means of the model parameters (linear and nonlinear) are not fairly close to the nominal values of the real structure due to model error, since the updated model cannot match the simulated measurements from the real structure perfectly. It is also noted that the values of hyper standard deviations are considerable for all the model parameters, corresponding to coefficient of variations  $\hat{\sigma}_{\theta_i} / \hat{\mu}_{\theta_i}$  ranging from 1.21% to 13.65%, as shown in Table 2. These values indicate a relatively large variability in the model parameters from data set to data set. For the hyper parameters of prediction error, again a strong relationship between  $\lambda_1$  and  $\lambda_2$  is observed since the posterior distribution of prediction error depends on both of them.

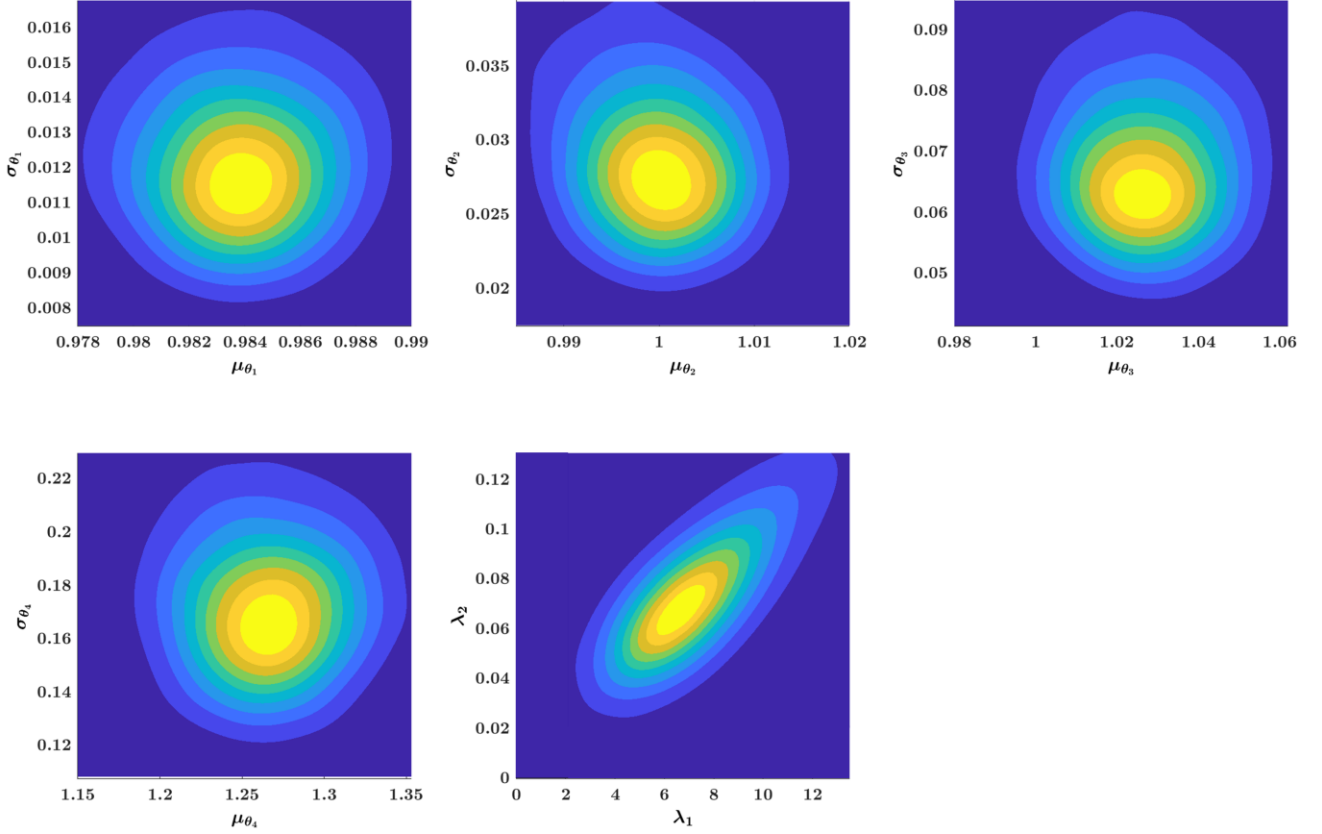


Fig. 9 Posterior distributions of hyper parameters computed based on Eq. (22)

Table 2

Estimates of means of hyper parameters

$\hat{\mu}_{\theta_1}$	0.9838	$\hat{\mu}_{\theta_2}$	1.0004	$\hat{\mu}_{\theta_3}$	1.0268	$\hat{\mu}_{\theta_4}$	1.2661	$\hat{\lambda}_1$	7.2965
$\hat{\sigma}_{\theta_1}$	0.0121	$\hat{\sigma}_{\theta_2}$	0.0286	$\hat{\sigma}_{\theta_3}$	0.0675	$\hat{\sigma}_{\theta_4}$	0.1728	$\hat{\lambda}_2$	0.0750
$\frac{\hat{\sigma}_{\theta_1}}{\hat{\mu}_{\theta_1}}$	1.23%	$\frac{\hat{\sigma}_{\theta_2}}{\hat{\mu}_{\theta_2}}$	2.86%	$\frac{\hat{\sigma}_{\theta_3}}{\hat{\mu}_{\theta_3}}$	6.57%	$\frac{\hat{\sigma}_{\theta_4}}{\hat{\mu}_{\theta_4}}$	13.65%	-	-

The posterior distributions of the hyper parameters are also computed according to the analytical solutions derived in Eqs. (26)-(31), as shown in Fig. 10. Results from the analytical solutions are in good agreement with the ones from Eq. (22). It is also noted that a slightly difference for the uncertainty bounds of the hyper parameters can be found between Fig. 9 and Fig. 10. This is due to the fact that the posterior distribution of the hyper parameters computed by the analytical solutions are under the conditions where the posterior distributions are approximated as a Gaussian distribution. The uncertainties of the hyper parameters are neglected in the analytical solution while the uncertainties of sampling based on Eq. (22) are considered for the hyper



parameters. However, based on the theoretical results developed in this work, such differences are expected to be reduced for large number of data sets.

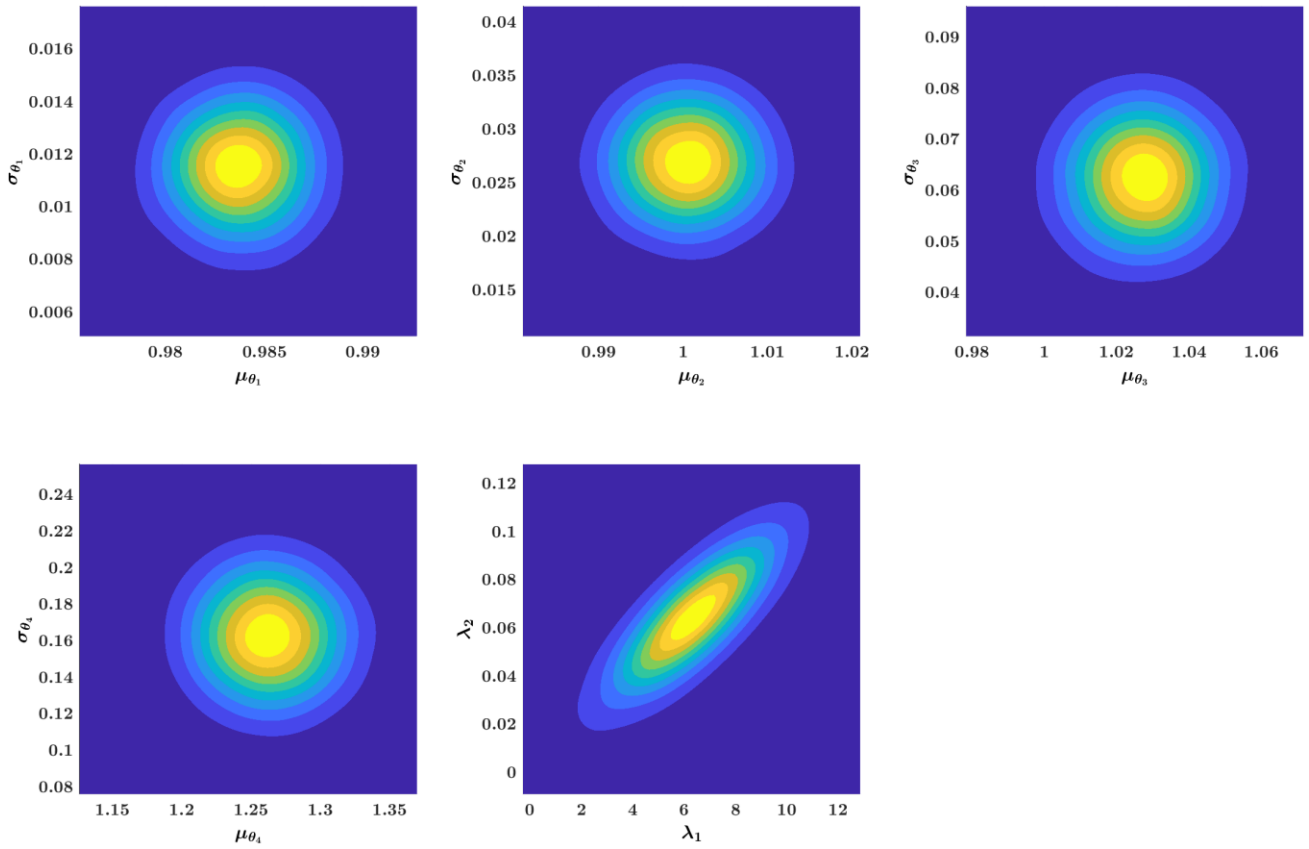


Fig. 10 Posterior distributions of hyper parameters computed based on the analytical solutions

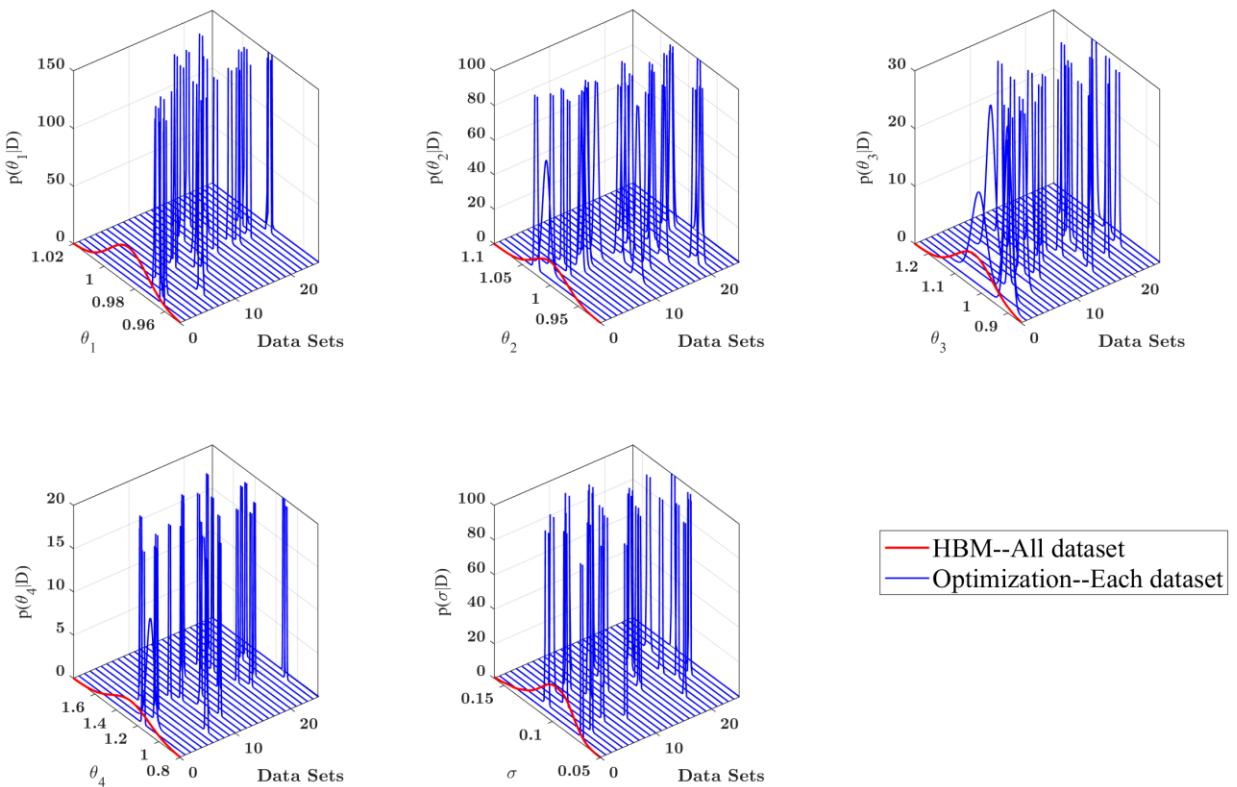


Fig. 11 Estimates of identification uncertainty in each dataset and uncertainty due to variability overall datasets

Moreover, the identification uncertainty for each data set calculated by the optimization procedure is compared to the uncertainty due to variability computed by the whole HBM framework. Results for the marginal PDF of the model parameters for each dataset and the marginal PDF of the model parameters estimated from HBM are presented in Fig. 11. It is shown that the identification uncertainty of the model and prediction error parameters in each data set (with blue) is much smaller than the ensemble uncertainty over all data sets (with red), and therefore the identification uncertainty can be negligible with sufficient number of data points. It is also seen that the mean of the parameters in each data set varies over different data sets due to the presence of model error. However, most of the mean values fall within the ensemble uncertainty bounds derived from the proposed HBM framework.

For the purpose of comparing the proposed method with the full sampling (FS) method [43,51] as well as the conventional Bayesian method (CBM), results of the mean values of the hyper parameters alongside the prediction error parameters are depicted in Fig. 12. It is clear that the proposed approach can capture the ensemble uncertainty of the model parameters and provide clear estimates of the model error while the conventional Bayesian approach fails to reach the same values. Although the CBM offers a relatively good accuracy in terms of the means of the model parameters, it severely underestimates the uncertainty of the model parameters, incorporating such uncertainties into the prediction error term together with the identification uncertainty and the model error uncertainty. As a result, CBM overestimates the value of the prediction error parameter. The proposed method can accurately estimate almost the same values as the FS approach but at a significantly reduced computational effort. The computational effort for the proposed and the FS method, carried out in a computer with a 32-core processor, are reported in Table 3. It is clear that the computational effort of the proposed method is around 10 times faster than that with the FS approach in the first step, and around 5 times faster in the second step. Therefore, the proposed method can not only guarantee the computational accuracy but also improve the computational efficiency to a great extent.

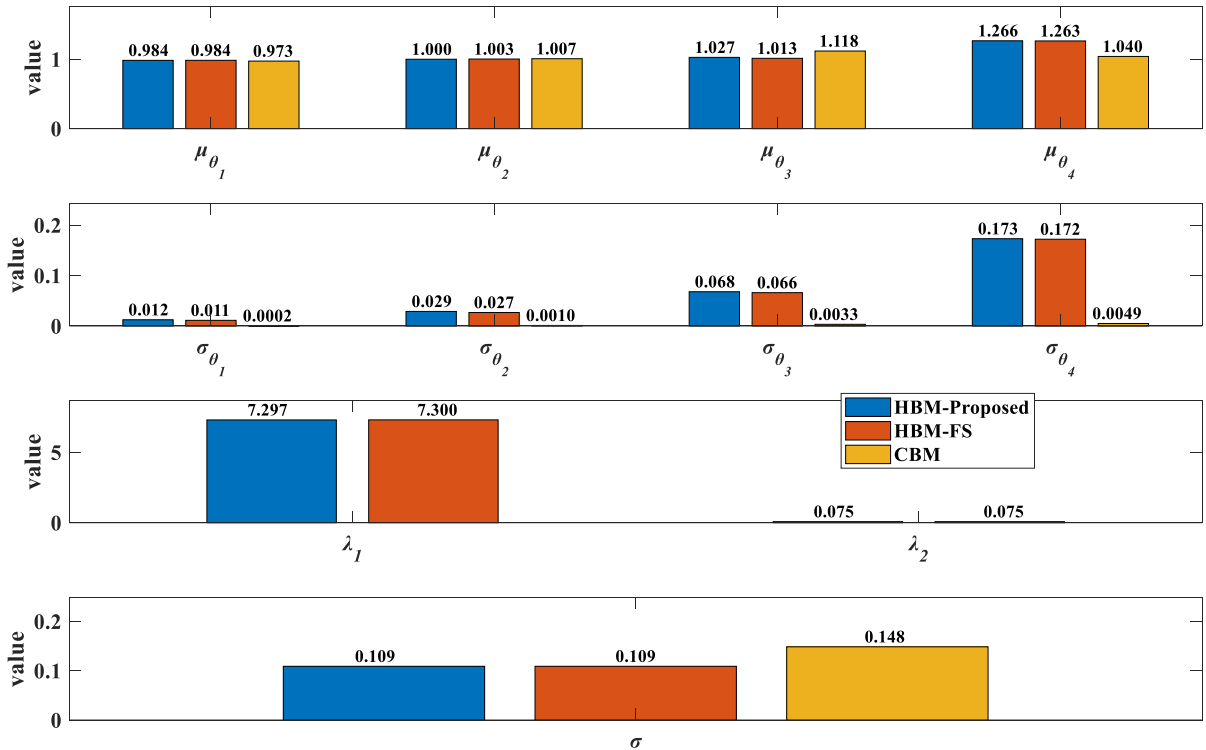


Fig. 12 Mean estimates of the hyper parameters and prediction error parameters by using the proposed method, FS and CBM

Table 3  
Computational effort of the proposed method and FS method

	Step 1	Step 2	Total
Proposed	52s	23s	75s
FS	652s	115s	766s

The estimated uncertainty of the model parameters and the uncertainty from prediction error parameters are next used to predict output QoI. Results are presented for two cases. In the first case, relevant to observed QoI such as accelerations in this example, the uncertainty in both the structural model and prediction error parameters is propagated. In the second case, relevant to unobserved quantities of interest such as displacement, only the uncertainty in the structural model parameters is propagated to the responses since the prediction error term and its uncertainty is not known. Fig. 13 depicts the predicted accelerations along with the measured accelerations of the first floor using CBM and HBM. As expected, both methods can offer a good accuracy and most of the actual response is contained within the uncertainty bounds. However, it should be noted that the sources of uncertainties from both CBM and HBM are conceptually different, with significant part of the uncertainty in the HBM method to be embedded in the structural model parameter, while most of the uncertainty in the CBM to be quantified in the model prediction term. Fig. 14 shows uncertainty propagation results of considering only the structural model parameter uncertainty for the predictions. The displacements of the first floor are predicted as the unobserved quantities using CBM and HBM. Fig. 15 shows the results for the unobserved displacement of the nonlinear part  $z_i$  of the response using CBM and the proposed HBM. It is observed that the proposed method delivers a reasonably accurate uncertainty bound which contains most of the measurements while the conventional Bayesian method provides an extremely thin uncertainty bound where the measurement falls outside.

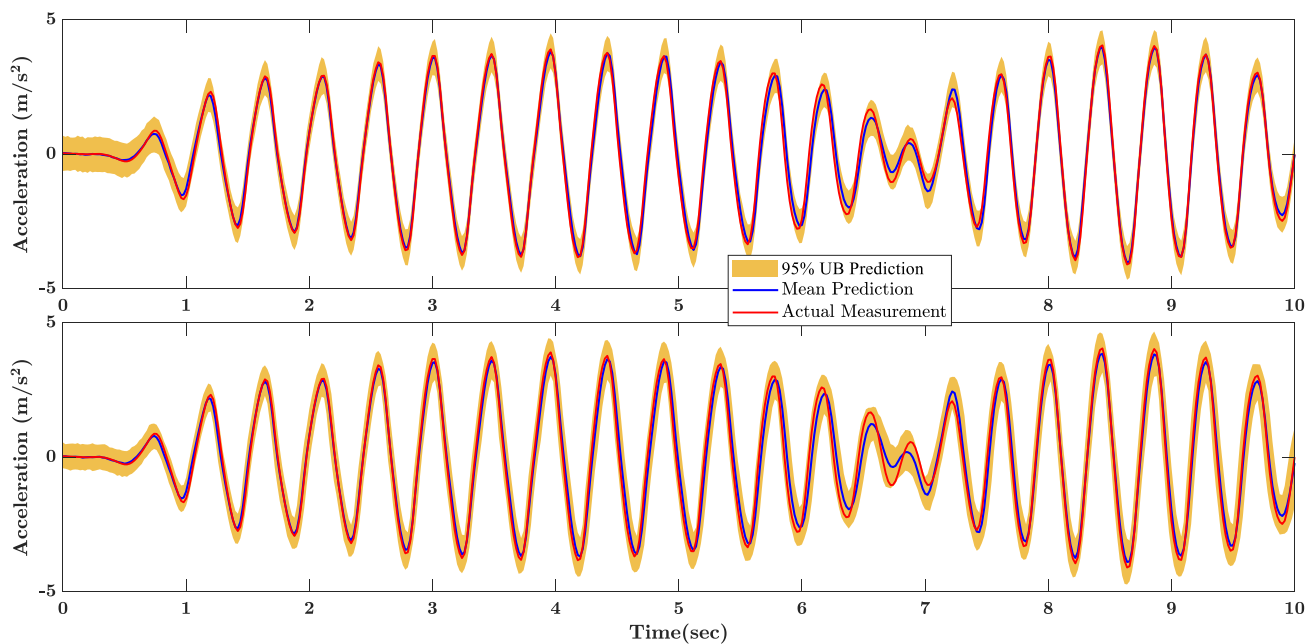


Fig. 13 Predictions of accelerations of the first floor using CBM (top) and HBM (bottom)

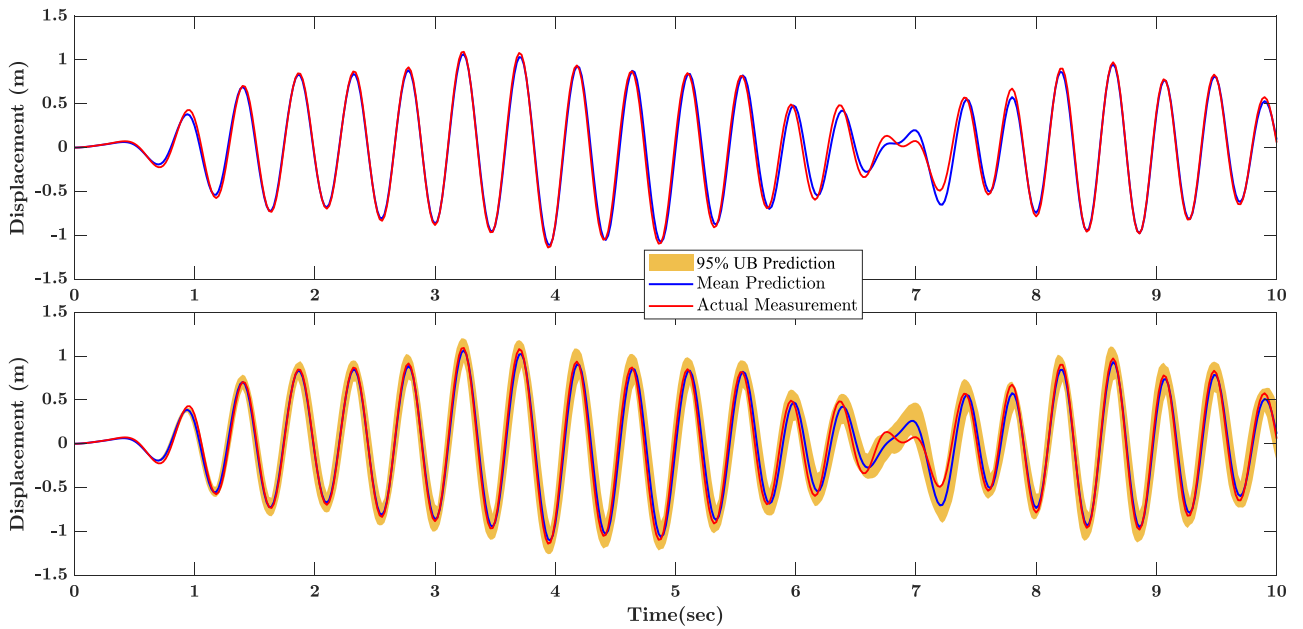


Fig. 14 Predictions of displacement of the first floor CBM (top) and HBM (bottom)

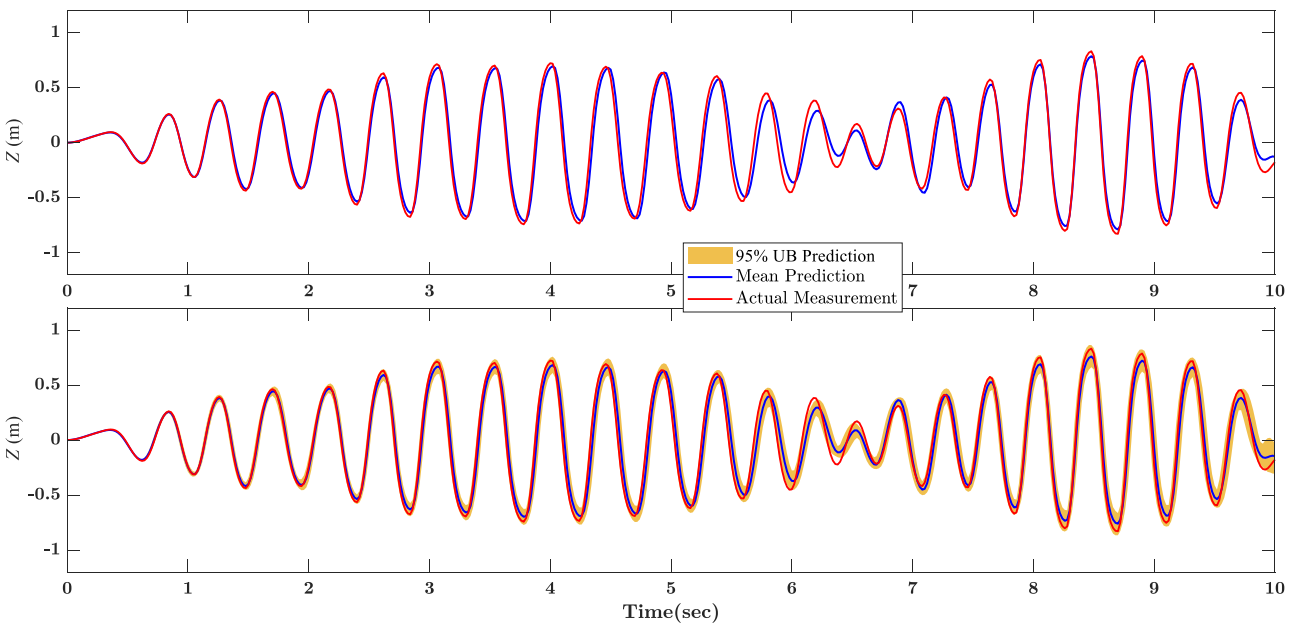


Fig. 15 Predictions of nonlinear displacements of the first floor CBM (top) and HBM (bottom)

Finally, the effect of sensor locations and number of sensors on the estimation of the parameters and uncertainties is investigated. Results for the most probable values of the hyper parameters are reported in Table 4 for one acceleration sensor placed either at first or fifth floor, 2 acceleration sensors placed at first and fifth floors, as well as 5 acceleration sensors placed at all the floors. It can be seen that different sensor configurations affect the estimates of the hyper parameters. These differences in the estimates from different sensor configurations are expected to affect uncertainties in structural and prediction error model parameters. Results for the posterior PDF of the model and prediction error parameters, computed according to Eqs. (33) and (34), are shown in Fig. 16 for the different sensor configurations considered in Table 4. These results suggest that the mean estimate of the parameters as well as the spread of uncertainty in the parameters depend on the number and location of sensors. Measurements from different sensor locations may cause different prediction errors and thus affect the parameter uncertainty due to the presence of model error. Placing a single sensor at the first floor reduces the prediction error. Increasing the number of sensors and thus the number of measurements, although

is expected to reduce the identification accuracy by reducing the uncertainties in the estimates of the model parameters for each data set, it does not significantly affect the spread of uncertainty due to the variability from multiple datasets.

Table 4

Statistical information of model parameters and prediction error parameter corresponding to Fig. 16

Sensor(s)	$\mu_{\theta_1}$	$\sigma_{\theta_1}$	$\mu_{\theta_2}$	$\sigma_{\theta_2}$	$\mu_{\theta_3}$	$\sigma_{\theta_3}$	$\mu_{\theta_4}$	$\sigma_{\theta_4}$	$\sigma$
1 <sup>st</sup> floor	0.9838	0.0121	1.0004	0.0284	1.0261	0.0653	1.2605	0.1728	0.1040
5 <sup>th</sup> floor	1.0159	0.0173	0.9201	0.0376	0.8594	0.0991	1.8660	0.2486	0.1615
1 <sup>st</sup> and 5 <sup>th</sup> floors	1.0078	0.0139	0.9728	0.0320	0.9253	0.0720	1.8204	0.2399	0.1582
All the floors	1.0085	0.0102	0.9880	0.0197	0.9822	0.0926	1.8730	0.2278	0.1686

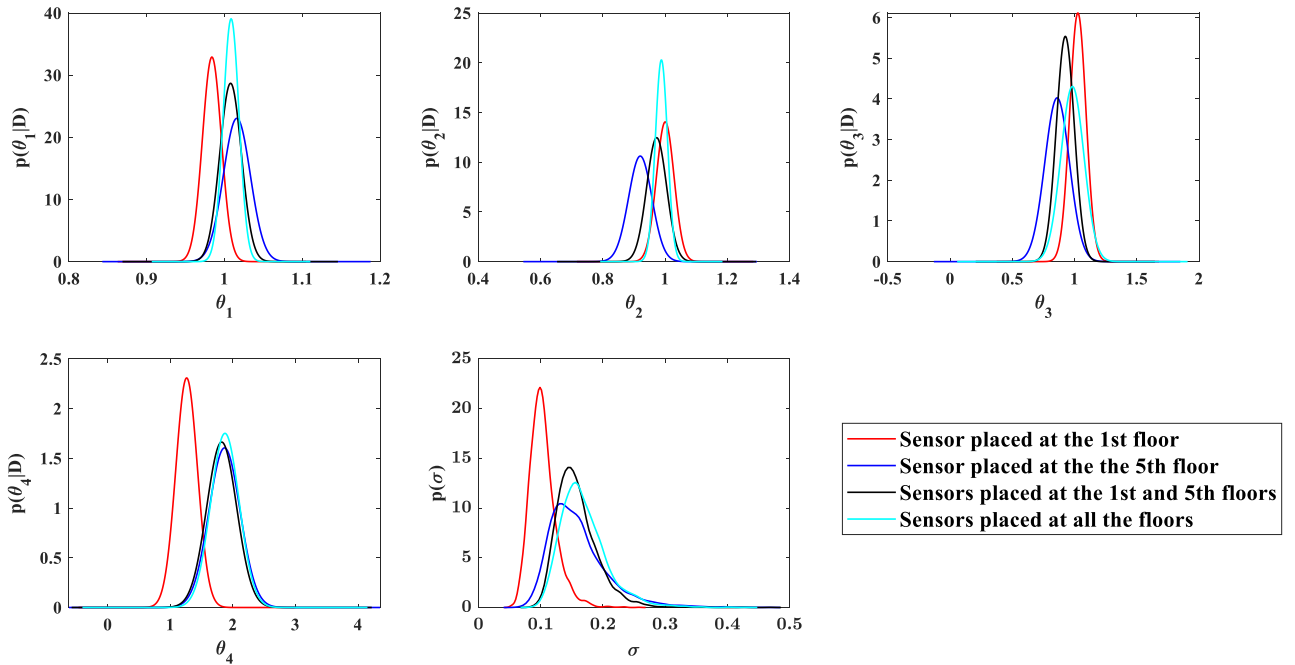


Fig. 16 Probabilistic distributions of the model and prediction error parameters

## 5. Conclusions

A nonlinear model updating strategy based on HBM framework is proposed and evaluated when applied to two numerical models of nonlinear systems characterized by Bouc-Wen hysteresis type nonlinearities. According to the HBM framework, uncertainties due to the variability arising from multiple datasets are embedded in the model parameters by assigning a Gaussian distribution to the model parameters with hyper parameters defined as the mean and the covariance matrix of the Gaussian distribution. The unmodelled dynamics are included in Gaussian prediction error by assigning a hyper distribution to the prediction error parameter. Asymptotic approximations are introduced to obtain insightful analytical expression for the posterior distribution of the hyper parameters. The accuracy of the asymptotic approximations is guaranteed due to the large number of data involved in the time history measurements for each dataset. Sampling from this posterior distribution of the hyper parameters is computationally very efficient since there are no expensive model runs involved at this stage. The resulting analytical expressions provide valuable insight into the mean and covariance of the hyper parameters and their dependence of the most probable values and identification uncertainty of the model parameters estimated for each dataset, as well as the variability in the model parameter estimates arising from multiple datasets. In particular, the analytical expressions were used to show that the structural model

hyper parameters are independent from the model prediction error hyper parameters for large number of data within each data set.

Two numerical examples are used to demonstrate the effectiveness and applicability of the proposed HBM framework. It is clearly demonstrated that the irreducible uncertainty arising from the variability of the multiple datasets is due to model error, while the identification uncertainty for each dataset can be reduced as the number of data in each dataset increases. The HBM framework is capable of accounting for the irreducible uncertainty, while conventional Bayesian inference fails to account for such uncertainty, resulting to underestimation of uncertainty bounds for the structural model and prediction error parameters. In particular, embedding the uncertainties due to variability from multiple datasets to the structural model parameters has the effect of obtaining reasonable uncertainty bounds for unobserved response QoI as opposed to significantly underestimated uncertainty bounds obtained from the classical Bayesian inference framework. Moreover, although the number and location of sensors seems to slightly affect the values of the hyper parameters and thus affect the model parameter estimates and uncertainties, increasing the number of sensors does not substantially affect the irreducible uncertainty arising from the variability due to multiple datasets. Finally, the asymptotic approximation is shown to provide one order of magnitude reduction of the computational effort compared to existing full sampling approach for the considered case study.

### Acknowledgement

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764547.

### Appendix A. Calculation for the integral in Eq. (10)

Substituting the likelihood function  $p(\mathbf{D}_i | \boldsymbol{\theta}_i, \sigma_i^2)$  from Eq. (7) and the inverse gamma distribution form Eq. (6) to the integral in Eq. (10) yields:

$$\int_{\sigma_i^2} p(\mathbf{D}_i | \boldsymbol{\theta}_i, \sigma_i^2) IG(\sigma_i^2 | \lambda_1, \lambda_2) d\sigma_i^2 \propto \frac{(\lambda_2)^{\lambda_1}}{\Gamma(\lambda_1)} \int_{\sigma_i^2} \sigma_i^{-N_0 N_i - 2\lambda_1 - 2} \exp \left\{ -\frac{\frac{N_0 N_i}{2} J(\boldsymbol{\theta}_i) + \lambda_2}{\sigma_i^2} \right\} d\sigma_i^2 \quad (\text{A1})$$

The integral in Eq. (A1) can be obtained analytically as:

$$\int_{\sigma_i^2} \sigma_i^{-N_0 N_i - 2\lambda_1 - 2} \exp \left\{ -\frac{\frac{N_0 N_i}{2} J(\boldsymbol{\theta}_i) + \lambda_2}{\sigma_i^2} \right\} d\sigma_i^2 = \Gamma(f(\lambda_1)) \left[ \frac{N_0 N_i}{2} J(\boldsymbol{\theta}_i) + \lambda_2 \right]^{-f(\lambda_1)} \quad (\text{A2})$$

where  $\Gamma(\bullet)$  is the Gamma function and  $f(\lambda_1)$  is defined as:

$$f(\lambda_1) = \frac{N_0 N_i + 2\lambda_1}{2} \quad (\text{A3})$$

Substituting (A2) into (A1) one derives Eq. (11).

### Appendix B. Calculation for the MPV and hessian matrix in Eq. (15) and Eq. (16)

The first order derivative of function  $L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)$  in Eq. (13) with respect to parameters  $\boldsymbol{\theta}_i$  can be computed as:

$$\nabla_{\boldsymbol{\theta}_i} L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2) = \frac{N_0 N_i f(\lambda_1)}{N_0 N_i J(\boldsymbol{\theta}_i) + 2\lambda_2} \nabla_{\boldsymbol{\theta}_i} J(\boldsymbol{\theta}_i) \quad (\text{B1})$$

Setting the derivatives equal to zero one can obtain the MPV of parameters  $\boldsymbol{\theta}_i$  as:

$$\hat{\boldsymbol{\theta}}_i = \arg \min_{\boldsymbol{\theta}_i} (J(\boldsymbol{\theta}_i)) \quad (\text{B2})$$

The hessian of the function  $L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2)$  can be also calculated as follows:

$$\nabla_{\boldsymbol{\theta}_i} \nabla_{\boldsymbol{\theta}_i}^T L(\boldsymbol{\theta}_i, \lambda_1, \lambda_2) = N_0 N_i f(\lambda_1) \frac{\mathbf{H}(\boldsymbol{\theta}_i) (N_0 N_i J(\boldsymbol{\theta}_i) + 2\lambda_2) - N_0 N_i \nabla_{\boldsymbol{\theta}_i} J(\boldsymbol{\theta}_i) \nabla_{\boldsymbol{\theta}_i}^T J(\boldsymbol{\theta}_i)}{(N_0 N_i J(\boldsymbol{\theta}_i) + 2\lambda_2)^2} \quad (\text{B3})$$

where  $\mathbf{H}(\boldsymbol{\theta}_i) = \nabla_{\boldsymbol{\theta}_i} \nabla_{\boldsymbol{\theta}_i}^T J(\boldsymbol{\theta}_i)$  is the hessian of the function  $J(\boldsymbol{\theta}_i)$ . Substituting  $\hat{\boldsymbol{\theta}}_i$  calculated from Eq. (B2) into Eq. (B3) and noting that  $\nabla_{\boldsymbol{\theta}_i} J(\hat{\boldsymbol{\theta}}_i) = \mathbf{0}$  one can get the hessian matrix evaluated at the MPV  $\hat{\boldsymbol{\theta}}_i$  of the parameters in the form given in Eq. (16).

## References

- [1] T.J.R. Hughes, The finite element method: linear static and dynamic finite element analysis, Courier Corporation, 2012.
- [2] O.C. Zienkiewicz, R.L. Taylor, Finite Element Method for Solid and Structural Mechanics, Elsevier, 2005.
- [3] V. Papadopoulos, M. Papadrakakis, G. Deodatis, Analysis of mean and mean square response of general linear stochastic finite element systems, *Comput. Methods Appl. Mech. Eng.* 195 (2006) 5454–5471.
- [4] J. Guggenberger, H. Grundmann, Stochastic response of large FEM models with hysteretic behaviour in beam elements, *Comput. Methods Appl. Mech. Eng.* 194 (2005) 1739–1756.
- [5] J.E. Mottershead, M.I. Friswell, Model updating in structural dynamics: A survey, *J. Sound Vib.* 167 (1993) 347–375.
- [6] M. Friswell, J.E. Mottershead, Finite element model updating in structural dynamics, Springer Science & Business Media, 2013.
- [7] N. Distefano, A. Rath, System identification in nonlinear structural seismic dynamics, *Comput. Methods Appl. Mech. Eng.* 5 (1975) 353–372.
- [8] N. Distefano, A. Rath, Sequential identification of hysteretic and viscous models in structural seismic dynamics, *Comput. Methods Appl. Mech. Eng.* 6 (1975) 219–232.
- [9] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, *J. Eng. Mech.* 124 (1998) 455–461.
- [10] L.S. Katafygiotis, J.L. Beck, Updating models and their uncertainties. II: Model identifiability, *J. Eng. Mech.* 124 (1998) 463–467.
- [11] E. Simoen, C. Papadimitriou, G. Lombaert, On prediction error correlation in Bayesian model updating, *J. Sound Vib.* 332 (2013) 4136–4152.
- [12] H. Jensen, C. Papadimitriou, Bayesian Finite Element Model Updating, in: *Sub-Structure Coupling Dyn. Anal.*, Springer, 2019: pp. 179–227.
- [13] M. Song, S. Yousefianmoghadam, M.E. Mohammadi, B. Moaveni, A. Stavridis, R.L. Wood, An application of finite element model updating for damage assessment of a two-story reinforced concrete building and comparison with lidar, *Struct. Heal. Monit.* 17 (2018) 1129–1150.
- [14] M. Papadrakakis, V. Papadopoulos, N.D. Lagaros, Structural reliability analysis of elastic-plastic structures using neural networks and Monte Carlo simulation, *Comput. Methods Appl. Mech. Eng.* 136 (1996) 145–163.
- [15] H.A. Jensen, C. Vergara, C. Papadimitriou, E. Millas, The use of updated robust reliability measures in stochastic dynamical systems, *Comput. Methods Appl. Mech. Eng.* 267 (2013).
- [16] S. Bansal, S.H. Cheung, A new stochastic simulation algorithm for updating robust reliability of linear structural dynamic systems subjected to future Gaussian excitations, *Comput. Methods Appl. Mech. Eng.* 326 (2017) 481–504.
- [17] E. Asgari, B. Moaveni, A. Stavridis, Nonlinear finite element model updating of an infilled frame

- based on identified time-varying modal parameters during an earthquake, *J. Sound Vib.* 333 (2014) 6057–6073.
- [18] Z.C. Wang, Y. Xin, W.X. Ren, Nonlinear structural joint model updating based on instantaneous characteristics of dynamic responses, *Mech. Syst. Signal Process.* 76–77 (2016) 476–496.
- [19] J. Ma, G. Chen, L. Ji, L. Qian, S. Dong, A general methodology to establish the contact force model for complex contacting surfaces, *Mech. Syst. Signal Process.* 140 (2020) 106678.
- [20] W.-X. Ren, H.-B. Chen, Finite element model updating in structural dynamics by using the response surface method, *Eng. Struct.* 32 (2010) 2455–2465.
- [21] M. Kurt, M. Eriten, D.M. McFarland, L.A. Bergman, A.F. Vakakis, Methodology for model updating of mechanical components with local nonlinearities, *J. Sound Vib.* 357 (2015) 331–348.
- [22] E. Asgariéh, B. Moaveni, A.R. Barbosa, E. Chatzi, Nonlinear model calibration of a shear wall building using time and frequency data features, *Mech. Syst. Signal Process.* 85 (2017) 236–251.
- [23] J.L. Beck, S.-K. Au, Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation, *J. Eng. Mech.* 128 (2002) 380–391.
- [24] K.V. Yuen, *Bayesian Methods for Structural Dynamics and Civil Engineering*, 2010.
- [25] K.V. Yuen, Recent developments of Bayesian model class selection and applications in civil engineering, *Struct. Saf.* 32 (2010) 338–346.
- [26] W. Betz, I. Papaioannou, J.L. Beck, D. Straub, Bayesian inference with Subset Simulation: Strategies and improvements, *Comput. Methods Appl. Mech. Eng.* 331 (2018) 72–93.
- [27] I. Babuška, F. Nobile, R. Tempone, A systematic approach to model validation based on Bayesian updates and prediction related rejection criteria, *Comput. Methods Appl. Mech. Eng.* 197 (2008) 2517–2539.
- [28] M. Muto, J.L. Beck, Bayesian updating and model class selection for hysteretic structural models using stochastic simulation, *J. Vib. Control.* 14 (2008) 7–34.
- [29] H. Ebrahimian, R. Astroza, J.P. Conte, R.A. de Callafon, Nonlinear finite element model updating for damage identification of civil structures using batch Bayesian estimation, *Mech. Syst. Signal Process.* 84 (2017) 194–222.
- [30] M. Song, L. Renson, J.P. Nođ, B. Moaveni, G. Kerschen, Bayesian model updating of nonlinear systems using nonlinear normal modes, *Struct. Control Heal. Monit.* 25 (2018) 1–20.
- [31] R. Ceravolo, A. Faraci, G. Miraglia, Bayesian calibration of hysteretic parameters with consideration of the model discrepancy for use in seismic structural health monitoring, *Appl. Sci.* 10 (2020).
- [32] K.-V. Yuen, J.L. Beck, Updating Properties of Nonlinear Dynamical Systems with Uncertain Input, *J. Eng. Mech.* 129 (2003) 9–20.
- [33] J. Ching, J.L. Beck, K.A. Porter, Bayesian state and parameter estimation of uncertain dynamical systems, *Probabilistic Eng. Mech.* 21 (2006) 81–96.
- [34] E.N. Chatzi, A.W. Smyth, S.F. Masri, Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty, *Struct. Saf.* 32 (2010) 326–337.
- [35] E.N. Chatzi, Andrew W. Smyth, The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing, *Struct. Control Heal. Monit.* (2011) 99–123.
- [36] M. Wu, A.W. Smyth, Application of the unscented Kalman filter for real-time nonlinear structural system identificatio, *Struct. Control Heal. Monit.* 14(7) (2007) 971–990.
- [37] P. Liu, S.K. Au, Bayesian parameter identification of hysteretic behavior of composite walls, *Probabilistic Eng. Mech.* 34 (2013) 101–109.
- [38] R. Sandhu, C. Pettit, M. Khalil, D. Poirel, A. Sarkar, Bayesian model selection using automatic relevance



- determination for nonlinear dynamical systems, *Comput. Methods Appl. Mech. Eng.* 320 (2017) 237–260.
- [39] M. Ebrahimzadeh Hassanabadi, A. Heidarpour, S. Eftekhari Azam, M. Arashpour, Recursive principal component analysis for model order reduction with application in nonlinear Bayesian filtering, *Comput. Methods Appl. Mech. Eng.* 371 (2020) 113334.
- [40] M. Song, I. Behmanesh, B. Moaveni, C. Papadimitriou, Accounting for Modeling Errors and Inherent Structural Variability through a Hierarchical Bayesian Model Updating Approach: An Overview, *Sensors*. 20 (2020) 3874.
- [41] G.C. Ballesteros, P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, Bayesian hierarchical models for uncertainty quantification in structural dynamics, in: *Vulnerability, Uncertainty, Risk Quantif. Mitigation, Manag.*, 2014: pp. 1615–1624.
- [42] I. Behmanesh, B. Moaveni, G. Lombaert, C. Papadimitriou, Hierarchical Bayesian model updating for structural identification, *Mech. Syst. Signal Process.* 64–65 (2015) 360–376.
- [43] S. Wu, P. Angelikopoulos, G. Tauriello, C. Papadimitriou, P. Koumoutsakos, Fusing heterogeneous data for the calibration of molecular dynamics force fields using hierarchical Bayesian models, *J. Chem. Phys.* 145 (2016) 244112.
- [44] I. Behmanesh, B. Moaveni, Accounting for environmental variability, modeling errors, and parameter estimation uncertainties in structural identification, *J. Sound Vib.* 374 (2016) 92–110.
- [45] X.-W. Liu, D.-G. Lu, Survival analysis of fatigue data: Application of generalized linear models and hierarchical Bayesian model, *Int. J. Fatigue*. 117 (2018) 39–46.
- [46] S. Wu, P. Angelikopoulos, J.L. Beck, P. Koumoutsakos, Hierarchical Stochastic Model in Bayesian Inference for Engineering Applications: Theoretical Implications and Efficient Approximation, *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.* 5 (2019).
- [47] M. Song, B. Moaveni, C. Papadimitriou, A. Stavridis, Accounting for amplitude of excitation in model updating through a hierarchical Bayesian approach: Application to a two-story reinforced concrete building, *Mech. Syst. Signal Process.* 123 (2019) 68–83.
- [48] O. Sedehi, C. Papadimitriou, L.S. Katafygiotis, Probabilistic hierarchical Bayesian framework for time-domain model updating and robust predictions, *Mech. Syst. Signal Process.* 123 (2019) 648–673.
- [49] O. Sedehi, C. Papadimitriou, L.S. Katafygiotis, Data-driven uncertainty quantification and propagation in structural dynamics through a hierarchical Bayesian framework, *Probabilistic Eng. Mech.* 60 (2020) 103047.
- [50] X. Jia, O. Sedehi, C. Papadimitriou, L.S. Katafygiotis, B. Moaveni, Hierarchical Bayesian Modeling Framework for Model Updating and Robust Predictions in Structural Dynamics using Modal Properties, (2021). <https://zenodo.org/record/5078051#.YOV0LrUzbD4>.
- [51] D. Patsialis, A.P. Kyprioti, A.A. Taflanidis, Bayesian calibration of hysteretic reduced order structural models for earthquake engineering applications, *Eng. Struct.* 224 (2020) 111204.
- [52] J. Ching, Y.-C. Chen, Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging, *J. Eng. Mech.* 133 (2007) 816–832.
- [53] S. Wu, P. Angelikopoulos, C. Papadimitriou, P. Koumoutsakos, Bayesian annealed sequential importance sampling: an unbiased version of transitional Markov chain Monte Carlo, *ASCE-ASME J. Risk Uncertain. Eng. Syst. Part B Mech. Eng.* 4 (2018).
- [54] M. Coughlin, N. Christensen, J. Gair, S. Kandhasamy, E. Thrane, Method for estimation of gravitational-wave transient model parameters in frequency-time maps, *Class. Quantum Gravity*. 31 (2014).
- [55] O. Sedehi, L.S. Katafygiotis, C. Papadimitriou, Hierarchical Bayesian operational modal analysis: Theory and computations, *Mech. Syst. Signal Process.* 140 (2020) 106663.

- [56] R. Bouc, A mathematical model for hysteresis, *Acta Acust. United with Acust.* 24 (1971) 16–25.
- [57] Y.-K. Wen, Method for random vibration of hysteretic systems, *J. Eng. Mech. Div.* 102 (1976) 249–263.
- [58] F. Ikhouane, V. Mañosa, J. Rodellar, Dynamic properties of the hysteretic Bouc-Wen model, *Syst. Control Lett.* 56 (2007) 197–205.
- [59] F. Ikhouane, J. Rodellar, *Systems with Hysteresis: Analysis, Identification and Control using the Bouc-Wen Model*, 2007.
- [60] P. Sengupta, B. Li, Modified Bouc-Wen model for hysteresis behavior of RC beam-column joints with limited transverse reinforcement, *Eng. Struct.* 46 (2013) 392–406.
- [61] M. Ismail, F. Ikhouane, J. Rodellar, The hysteresis Bouc-Wen model, a survey, *Arch. Comput. Methods Eng.* 16 (2009) 161–188.
- [62] J.R. Cash, A.H. Karp, A variable order Runge-Kutta method for initial value problems with rapidly varying right-hand sides, *ACM Trans. Math. Softw.* 16 (1990) 201–222.
- [63] Z. Cao, Q. Fei, D. Jiang, D. Zhang, H. Jin, R. Zhu, Dynamic sensitivity-based finite element model updating for nonlinear structures using time-domain responses, *Int. J. Mech. Sci.* 184 (2020).
- [64] Luzi L., Lanzano G., Felicetta C., D’Amico M. C., Russo E., Sgobba S., Pacor, F., & ORFEUS Working Group 5 (2020). *Engineering Strong Motion Database (ESM) (Version 2.0)*. Istituto Nazionale di Geofisica e Vulcanologia (INGV). <https://doi.org/10.13127/ESM.2>