Adaptive Bayesian Inference Framework for Joint Model and Noise Identification

Mansureh-Sadat Nabiyan, Ph.D.[1]; Hamed Ebrahimian, M.ASCE[2]; Babak Moaveni, M.ASCE[3]; and Costas Papadimitriou, M.ASCE[4]

Abstract: Model updating, the process of inferring a model from data, is prone to the adverse effects of modeling error, which is caused by simplification and idealization assumptions in the mathematical models. In this study, an adaptive recursive Bayesian inference framework is developed to jointly estimate model parameters and the statistical characteristics of the prediction error that includes the effects of modeling error and measurement noise. The prediction error is usually modeled as a Gaussian white noise process in a Bayesian model updating framework. In this study, the prediction error is assumed to be a nonstationary Gaussian process with an unknown and time-variant mean vector and covariance matrix to be estimated. This allows one to better account for the effects of time-variant model uncertainties in the model updating process. The proposed approach is verified numerically using a 3-story 1-bay nonlinear steel moment frame excited by an earthquake. Comparison of the results with those obtained from a classical nonadaptive recursive Bayesian model updating method shows the efficacy of the proposed approach in the estimation of the prediction error statistics and model parameters. **DOI: 10.1061/(ASCE)EM.1943-7889.0002084.**

Author keywords: Bayesian inference; Model updating; Noise identification; Modeling error; System identification; Adaptive Kalman filter.

Introduction

Model updating has emerged as a powerful tool for system identification, parameter estimation, damage identification, response reconstruction, and virtual sensing. In the application of model updating, the unknown model parameters and/or input loads are estimated by minimizing the mismatch between the measured and model-predicted responses (Friswell and Mottershead 2013; Ebrahimian 2015). This mismatch, referred to as the prediction error, encapsulates the measurement noise and the effects of model uncertainties, which include model parameter uncertainties and modeling error (Beck and Yuen 2004; Goller and Schueller 2011; Soize 2017). Model parameter uncertainties can be reduced through the model updating process given favorable identifiability conditions (Ebrahimian et al. 2019). Modeling error is caused by inherent mathematical idealizations, approximations, and simplifications in the numerical model. If not accounted for properly, modeling error can cause estimation bias resulting in incorrect and/or inaccurate model updating outcomes. There are different methodologies in the literature to account for model uncertainties, including parametric (Ghanem and Pellissetti 2002; Soize and Ghanem 2004) and nonparametric probabilistic approaches (Desceliers et al. 2004; Soize 2005).

[1]Research Affiliate, Dept. of Civil and Environmental Engineering, Tufts Univ., 419 Boston Ave., Medford, MA 02155.

[2]Assistant Professor, Dept. of Civil and Environmental Engineering, Univ. of Nevada, Virginia St., Reno, NV 89557 (corresponding author). ORCID: https://orcid.org/0000-0003-1992-6033. Email: hebrahimian@unr.edu

[3]Professor, Dept. of Civil and Environmental Engineering, Tufts Univ., 419 Boston Ave., Medford, MA 02155. ORCID: https://orcid.org/0000-0002-8462-4608

[4]Professor, Dept. of Mechanical Engineering, Univ. of Thessaly, Argonafton & Filellinon, Volos 38221, Greece.

Note. This manuscript was submitted on June 25, 2021; approved on November 11, 2021; published online on December 28, 2021. Discussion period open until May 28, 2022; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Engineering Mechanics***, © ASCE, ISSN 0733-9399.

The recursive Bayesian inference framework provides a mathematical basis to formulate model updating in the presence of uncertainties and noise. In this framework, the prediction error is modeled as a random process characterized by a joint probability distribution function (PDF). Generally, the prediction error can be a correlated, nonstationary, nonwhite, and non-Gaussian random process. However, in the classic nonadaptive recursive Bayesian model updating formulation, also referred to as nonadaptive Bayesian filtering, the prediction error is often assumed to be an independent Gaussian white noise process (i.e., a stationary, zero-mean, uncorrelated Gaussian random process) for mathematical simplicity (Chatzi et al. 2010; Ebrahimian 2015; Ebrahimian et al. 2015; Erazo and Nagarajaiah 2017; Nabiyan et al. 2020). This assumption results in a zero-mean Gaussian PDF for the prediction error with a time-invariant covariance matrix. Nevertheless, this limiting assumption can be violated in real-world conditions, perhaps most commonly due to the effects of modeling error. Modeling error can be a bias error and not referenced to a zero mean. It might produce a shifted response that is not necessarily centered on zero error (Sanayei et al. 2001). Incorrect characterization of the prediction error statistics will affect model updating performance and can result in biased estimation or divergence of updating parameters (Xu et al. 2019). It can also result in incorrect uncertainty quantification of model parameters, i.e., although the parameters may be estimated with reasonable accuracy, their uncertainty bounds may not be realistic (Law and Stuart 2012; Ernst et al. 2015).

To address this issue, Bayesian inference methods for joint model parameters and noise identification have been proposed in the literature for structural and mechanical engineering applications. Yuen and Kuok (2016) proposed a Bayesian method for the estimation of the diagonal entries of the prediction error covariance matrix. The estimated covariance matrix is then fed into an extended Kalman filter (EKF) for joint state-parameter estimation. Astroza et al. (2019a) proposed a dual adaptive filtering method to handle the effects of modeling error. The dual filtering method consists of a Kalman filter to estimate the diagonal entries of the prediction error covariance matrix based on a covariance-matching technique (Mehra 1972) and an unscented Kalman filter (UKF) (Wu and Smyth 2007) to estimate the unknown model parameters. The aforementioned studies assume that the prediction error is uncorrelated in space, and therefore, noise identification is limited to the estimation of the diagonal entries of the prediction error covariance matrix. To remove this limiting assumption, Song et al. (2020) proposed an adaptive Kalman filter using two types of covariance-matching methods, i.e., the forgetting factor method (Akhlaghi et al. 2017) and the moving window method (Mehra 1972; Almagbile et al. 2010), to estimate the full covariance matrix of the prediction error jointly with the unknown model parameters. They demonstrated the effectiveness of the approach in the presence of modeling error. Amini Tehrani et al. (2020) combined the Kalman filter method with a covariance-matching method to estimate the full covariance matrix of the prediction error jointly with the state vector and unknown model parameters.

Although these methods alleviated some of the limiting assumptions for the prediction error (e.g., stationary and uncorrelated in space), the zero-mean Gaussian assumption still remains. To resolve this limitation, Kontoroupi and Smyth (2016) developed a Bayesian method for joint estimation of the mean vector and covariance matrix of the prediction error. In this approach, the mean vector of the prediction error is assumed to have a Gaussian distribution, while an inverse-Wishart distribution is considered for the covariance matrix of the prediction error. First, the distributions are updated based on Bayesian inference, and then, mean estimates of the updated distributions are used in the UKF algorithm for joint parameter and state estimation. Their work is capable of estimating a biased (nonzero-mean) prediction error. However, it considers the mean vector and covariance matrix of prediction error as time-invariant.

In this paper, a recursive Bayesian inference formulation is developed to jointly estimate the unknown model parameters and the statistical characteristics (mean vector and covariance matrix) of the prediction error. The prediction error is modeled as a nonstationary Gaussian random process with a time-variant mean vector and covariance matrix to be estimated iteratively and jointly with the unknown model parameters. The evolution of unknown model parameters in time is modeled as a random walk process using a zero-mean Gaussian process noise. The prior PDF of unknown model parameters is assumed to be Gaussian. It is also assumed that the vectors of initial unknown model parameters, the process noise, and the prediction error at each time step are all mutually independent. The estimation problem is solved using a two-step marginal maximum *a posteriori* (MAP) estimation approach. The proposed method is still based on a Gaussian distribution assumption for the prediction error, which may be violated in the presence of a modeling error. Nevertheless, developing the capability to estimate the time-variant mean vector and covariance matrix of the prediction error is an advancement with respect to the state of the art.

The paper is organized as follows. The formulation of the proposed method is presented first and is followed by a verification study using a nonlinear model of a steel moment frame structure under earthquake excitation. Finally, conclusions are summarized based on the observed results.

Bayesian Inference Formulation for Joint Model and Noise Identification

Problem Statement

Model updating using input-output measurements can be formulated as a Bayesian filtering problem for parameter-only estimation as follows (Haykin 2004; Ebrahimian et al. 2015). The *filter* terminology refers to a stochastic estimator (Besançon 2007), which means an auxiliary static or dynamic system that often runs in parallel to the real system under investigation to estimate desired system quantities (Ritter 2020). Following this analogy, a model updating problem can be regarded as a filtering problem

$$\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k-1} + \boldsymbol{\gamma}_{k-1}; \quad \boldsymbol{\gamma}_{k-1} \sim N(\mathbf{0}, \mathbf{Q}) \tag{1}$$

$$\mathbf{y}_k = \mathbf{h}(\boldsymbol{\theta}_k, \mathbf{f}_{1:k}) + \boldsymbol{\omega}_k; \quad \boldsymbol{\omega}_k \sim N(\boldsymbol{\mu}_k, \mathbf{R}_k) \tag{2}$$

where $\boldsymbol{\theta}_k \in \mathbb{R}^{n_{\boldsymbol{\theta}}\times 1}$ = unknown model parameter vector modeled as a random process; $\mathbf{y}_k \in \mathbb{R}^{n_{\mathbf{y}}\times 1}$ = vector of measured responses; and $n_{\boldsymbol{\theta}}$ and $n_{\mathbf{y}}$ = number of unknown model parameters and measurement channels, respectively. The $\mathbf{h}(\boldsymbol{\theta}_k, \mathbf{f}_{1:k}) \in \mathbb{R}^{n_{\mathbf{y}}\times 1}$ is the response function of the numerical model, which in this paper is assumed to be a finite-element (FE) model, to an input force-time history from time step 1 to k, $\mathbf{f}_{1:k}$. In this study, we assume that the input forces are measurable and known, and for the sake of notation brevity, $\mathbf{h}(\boldsymbol{\theta}_k, \mathbf{f}_{1:k})$ is replaced by $\mathbf{h}_k(\boldsymbol{\theta}_k)$ henceforth. The $\mathbf{x} \sim N(\bar{\mathbf{x}}, \boldsymbol{\Sigma})$ denotes random vector $\mathbf{x}$ following a Gaussian (or Normal) distribution with mean vector $\bar{\mathbf{x}}$ and covariance matrix $\boldsymbol{\Sigma}$, and the PDF of $\mathbf{x}$ is shown as $p(\mathbf{x}) = N(\mathbf{x}|\bar{\mathbf{x}}, \boldsymbol{\Sigma})$. The terms $\boldsymbol{\gamma}_k \in \mathbb{R}^{n_{\boldsymbol{\theta}}\times 1}$ and $\boldsymbol{\omega}_k \in \mathbb{R}^{n_{\mathbf{y}}\times 1}$ denote process noise and prediction error vectors, respectively. The process noise vector is characterized by a zero-mean Gaussian white noise process with covariance matrix $\mathbf{Q}$, which is usually one of the filter tuning parameters (Astroza et al. 2019b). The prediction error accounts for the mismatch between the measured and FE model-predicted responses of the structure and is modeled as a nonstationary Gaussian random process with unknown and time-variant mean vector $\boldsymbol{\mu}_k$ and covariance matrix $\mathbf{R}_k$ to be estimated recursively and jointly with the unknown model parameter $\boldsymbol{\theta}_k$. It should be mentioned that the initial unknown model parameter vector $\boldsymbol{\theta}_0$ and the process noise and prediction error at each time step, $\{\boldsymbol{\theta}_0, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots\}$, are all assumed to be mutually independent.

To proceed with the Bayesian inference formulation, $\boldsymbol{\theta}_k$ and $\boldsymbol{\mu}_k$ are modeled as random vectors, and $\mathbf{R}_k$ is modeled as a random matrix. Similar to other recursive filtering approaches, the proposed framework includes *prediction* and *correction* steps at each time step k (Simon 2006; Krishnan 2015). In the prediction step, the joint distribution of updating parameters, $\boldsymbol{\theta}_k$, $\boldsymbol{\mu}_k$, and $\mathbf{R}_k$, is propagated from the previous to the current time step through a dynamic model. In the correction step, prior estimates of the updating parameters, denoted as $\hat{\boldsymbol{\theta}}_k^-$, $\hat{\boldsymbol{\mu}}_k^-$, and $\hat{\mathbf{R}}_k^-$, are corrected to posterior estimates, denoted as $\hat{\boldsymbol{\theta}}_k^+$, $\hat{\boldsymbol{\mu}}_k^+$, and $\hat{\mathbf{R}}_k^+$, through the Bayes' theorem by absorbing the information in the new measurement $\mathbf{y}_k$. The process is further described in the following sections.

Bayesian Inference Formulation

Using the Bayes' theorem, the joint posterior distribution of updating parameters, $\boldsymbol{\theta}_k$, $\boldsymbol{\mu}_k$, and $\mathbf{R}_k$, given the measured responses from time steps 1 to k, $\mathbf{y}_{1:k}$, can be derived

$$p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1})p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1}) \tag{3}$$

where $p(\mathbf{y}_k|\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1})$ = likelihood function; and $p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1})$ = joint prior distribution of $\boldsymbol{\theta}_k$, $\boldsymbol{\mu}_k$, and $\mathbf{R}_k$ at time step k. The normalizing (evidence) term is ignored in Eq. (3); therefore, the sign $\propto$ denoting *proportional to* is used. Because the proposed estimation approach is recursive (Simon 2006), only the new data point $\mathbf{y}_k$ is used for updating parameters at each time step. Now, we expand the terms on the right-hand side of this equation.

Based on Eq. (2), the likelihood function follows a Gaussian distribution

$$p(\mathbf{y}_k|\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1}) = p(\boldsymbol{\omega}_k) = N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) \tag{4}$$

The joint prior distribution, $p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1})$, can be written in a hierarchical form (Behmanesh et al. 2015; Xu et al. 2019)

$$p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1}) = p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1})p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1}) \tag{5}$$

where $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1})$ is referred to as a hyperprior with hyperparameters of $\boldsymbol{\mu}_k$ and $\mathbf{R}_k$ (Huang and Beck 2015). By substituting Eq. (5) into Eq. (3), it can be followed that

$$p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1})p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1}) \times p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1}) \tag{6}$$

The prior distribution of $\boldsymbol{\theta}_k$ is approximated as Gaussian (Moore and Anderson 1979); in other words

$$p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1}) = N(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_k^-, \mathbf{P}_{\boldsymbol{\theta},k}^-) \tag{7}$$

where mean vector $\hat{\boldsymbol{\theta}}_k^-$ = prior estimate for $\boldsymbol{\theta}_k$; and $\mathbf{P}_{\boldsymbol{\theta},k}^-$ = prior covariance matrix of $\boldsymbol{\theta}_k$. Furthermore, the prior distribution of $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1})$ is assumed to follow a Normal-Inverse-Wishart (NIW) distribution. In Bayesian statistics, NIW distribution is often used as the joint conjugate prior for the mean vector and covariance matrix of a Gaussian distribution (Murphy 2007). The conjugacy guarantees the same functional form for the posterior and prior distributions (O'Hagan and Forster 2004). NIW distribution is the product of a Normal (or Gaussian) distribution and an Inverse-Wishart (IW) distribution. Therefore, the prior distribution of $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1})$ in Eq. (6) can be expressed

$$\begin{aligned} p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1}) &= NIW(\boldsymbol{\mu}_k, \mathbf{R}_k|\hat{\boldsymbol{\mu}}_k^-, \lambda_k^-, v_k^-, \mathbf{V}_k^-) \\ &= N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-}\right) \times IW(\mathbf{R}_k|v_k^-, \mathbf{V}_k^-) \end{aligned} \tag{8}$$

where $\hat{\boldsymbol{\mu}}_k^- \in \mathbb{R}^{n_\mathbf{y}\times 1}$ = prior mean vector of $\boldsymbol{\mu}_k$ (also considered as prior estimate for $\boldsymbol{\mu}_k$); $\lambda_k^- > 0$ is the confidence parameter; $v_k^- > n_\mathbf{y} - 1$ is the degree of freedom parameter; and $\mathbf{V}_k^- \in \mathbb{R}^{n_\mathbf{y}\times n_\mathbf{y}}$ = symmetric positive definite scale matrix (Xu et al. 2019). Considering that $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1}) = p(\boldsymbol{\mu}_k|\mathbf{R}_k, \mathbf{y}_{1:k-1}) \times p(\mathbf{R}_k|\mathbf{y}_{1:k-1})$, the right-hand side of Eq. (8) can be separated as follows

$$p(\boldsymbol{\mu}_k|\mathbf{R}_k, \mathbf{y}_{1:k-1}) = N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-}\right) \tag{9a}$$

$$p(\mathbf{R}_k|\mathbf{y}_{1:k-1}) = IW(\mathbf{R}_k|v_k^-, \mathbf{V}_k^-) \tag{9b}$$

The Normal and IW distributions in Eq. (9) are defined subsequently, ignoring normalizing terms (Gelman et al. 2013)

$$N(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{10a}$$

$$IW(\boldsymbol{\Sigma}|v, \mathbf{V}) \propto |\boldsymbol{\Sigma}|^{-(v+n_\mathbf{y}+1)/2} \exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{V}\boldsymbol{\Sigma}^{-1})\right) \tag{10b}$$

The terms $|.|$ and $\mathrm{tr}(.)$ in these equations denote matrix determinant and matrix trace, respectively.

Dynamic Models

The evolution of updating parameters $\boldsymbol{\theta}_k$, $\boldsymbol{\mu}_k$, and $\mathbf{R}_k$ from time step k-1 to k are characterized by a series of dynamic models discussed in this study. Following the random walk model considered for the unknown model parameter vector $\boldsymbol{\theta}_k$ in Eq. (1), its mean vector and covariance matrix are transferred from each estimation time step to the next as follows

$$\hat{\boldsymbol{\theta}}_k^- = \hat{\boldsymbol{\theta}}_{k-1}^+ \tag{11a}$$

$$\mathbf{P}_{\boldsymbol{\theta},k}^- = \mathbf{P}_{\boldsymbol{\theta},k-1}^+ + \mathbf{Q} \tag{11b}$$

Although the unknown model parameter vector $\boldsymbol{\theta}$ is considered to be time-invariant in this study, and thus, $\mathbf{Q}$ can be set to zero, a small process noise has been shown to improve the parameter estimation process (Song et al. 2020).

An explicit presentation of a dynamic model for the covariance matrix of the prediction error, $\mathbf{R}_k$, is difficult (Sarkka and Nummenmaa 2009). Therefore, the following two dynamic models for the statistical parameters of the prior IW distribution are used in this study based on the heuristic model proposed by Sarkka (Sarkka and Hartikainen 2013)

$$v_k^- = \rho(v_{k-1}^+ + n_\mathbf{y} + 1) - n_\mathbf{y} - 1 \tag{12a}$$

$$\mathbf{V}_k^- = \rho\mathbf{V}_{k-1}^+ \tag{12b}$$

where $\rho \in (0, 1]$ = forgetting factor. If $\rho = 1$, Eq. (12) results in a stationary prediction error covariance matrix, while smaller values for ρ allow for larger time variations in statistical properties of the covariance matrix (Sarkka and Hartikainen 2013). The mode (most probable) value of the prior IW distribution, considered as a prior estimate for $\mathbf{R}_k$, is defined as $\hat{\mathbf{R}}_k^- = [\mathbf{V}_k^-/(v_k^- + n_\mathbf{y} + 1)]$, while the posterior estimate for $\mathbf{R}_{k-1}$ is defined as $\hat{\mathbf{R}}_{k-1}^+ = [\mathbf{V}_{k-1}^+/(v_{k-1}^+ + n_\mathbf{y} + 1)]$ (O'Hagan and Forster 2004). Based on the dynamic models defined in Eq. (12), it can be concluded that $\hat{\mathbf{R}}_k^- = \hat{\mathbf{R}}_{k-1}^+$.

Similar to the dynamic model considered for $\mathbf{R}_k$, a heuristic model is considered for propagating the uncertainties of the $\boldsymbol{\mu}_k$ through time

$$\hat{\boldsymbol{\mu}}_k^- = \hat{\boldsymbol{\mu}}_{k-1}^+ \tag{13a}$$

$$\lambda_k^- = \rho'\lambda_{k-1}^+ \tag{13b}$$

where $\rho' \in (0, 1]$ = another forgetting factor. Similar to the case presented for ρ, $\rho' = 1$ represents a stationary $\boldsymbol{\mu}_k$, while smaller values for ρ' results in larger time variations in statistical properties of the prediction error mean.

Two-Step Marginal MAP Estimation

This section outlines an approach to find posterior estimates of updating parameters $\boldsymbol{\theta}_k$, $\boldsymbol{\mu}_k$, and $\mathbf{R}_k$ using a MAP estimation method. The objective is to maximize the joint posterior distribution $p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k})$ to find MAP estimates as follows

$$\{\hat{\boldsymbol{\theta}}_k^+, \hat{\boldsymbol{\mu}}_k^+, \hat{\mathbf{R}}_k^+\} = \underset{\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k}{\mathrm{argmax}}\, p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) \tag{14}$$

The joint posterior distribution $p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k})$ can be factored into two marginal distributions as follows using the product rule (Schum 2001)

$$p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) = p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k})p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) \tag{15}$$

J. Eng. Mech., 2022, 148(3): 04021165

Based on the marginal MAP estimation method (Haykin 2004), the two terms on the right-hand side of Eq. (15) are maximized separately to find the MAP estimates, in other words

$$\{\hat{\boldsymbol{\theta}}_k^+\} = \underset{\boldsymbol{\theta}_k}{\operatorname{argmax}}\, p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k}) \tag{16}$$

$$\{\hat{\boldsymbol{\mu}}_k^+, \hat{\mathbf{R}}_k^+\} = \underset{\boldsymbol{\mu}_k, \mathbf{R}_k}{\operatorname{argmax}}\, p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) \tag{17}$$

Therefore, the correction step in the Bayesian inference formulation is divided into two separate MAP estimation problems that should be solved iteratively to converge to the MAP estimates of the joint posterior distribution $p(\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k})$. For this purpose, first, the MAP estimate of $p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k})$ is derived given the $\hat{\boldsymbol{\mu}}_k^+$ and $\hat{\mathbf{R}}_k^+$ obtained from the previous iteration; then, the MAP estimate of $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k})$ is obtained based on the estimated $\hat{\boldsymbol{\theta}}_k^+$, and this process is repeated iteratively until convergence. The solutions to these separate MAP estimation problems are provided in the following sections.

Marginal MAP Estimate of θ_k

The MAP estimation problem in Eq. (16) is familiar and seen in the nonadaptive Bayesian model updating formulations (Ebrahimian et al. 2015). Based on the Bayes' theorem, it can be observed that

$$p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\boldsymbol{\theta}_k, \boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1})p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1}) \tag{18}$$

Because both the likelihood and the prior distributions for $\boldsymbol{\theta}_k$ are Gaussian according to Eqs. (4) and (7), respectively, the posterior distribution will also be Gaussian (Ebrahimian et al. 2018), i.e., $p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k}) = N(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_k^+, \mathbf{P}_{\boldsymbol{\theta},k}^+)$, in which the mean vector $\hat{\boldsymbol{\theta}}_k^+$ (considered the same as the MAP estimate for $\boldsymbol{\theta}_k$) and the covariance matrix $\mathbf{P}_{\boldsymbol{\theta},k}^+$ can be obtained as follows—derivation details are provided in Appendix I

$$\hat{\boldsymbol{\theta}}_k^+ = \hat{\boldsymbol{\theta}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^-) - \hat{\boldsymbol{\mu}}_k^+)) \tag{19}$$

$$\mathbf{P}_{\boldsymbol{\theta},k}^+ = \mathbf{P}_{\boldsymbol{\theta},k}^- - \mathbf{K}_k\mathbf{P}_{\mathbf{yy},k}\mathbf{K}_k^T \tag{20}$$

where $\mathbf{K}_k = \mathbf{P}_{\boldsymbol{\theta}\mathbf{y},k}(\mathbf{P}_{\mathbf{yy},k})^{-1}$; $\mathbf{P}_{\boldsymbol{\theta}\mathbf{y},k} = \mathbf{P}_{\boldsymbol{\theta},k}^-(\mathbf{C}_k^-)^T$; and $\mathbf{P}_{\mathbf{yy},k} = \mathbf{C}_k^-\mathbf{P}_{\boldsymbol{\theta},k}^-(\mathbf{C}_k^-)^T + \hat{\mathbf{R}}_k^+$. The term $\mathbf{C}_k^- = \{[\partial\mathbf{h}_k(\boldsymbol{\theta}_k)]/\partial\boldsymbol{\theta}_k\}|_{\boldsymbol{\theta}_k=\hat{\boldsymbol{\theta}}_k^-}$ is the model response sensitivity matrix with respect to $\boldsymbol{\theta}_k$ at $\boldsymbol{\theta}_k = \hat{\boldsymbol{\theta}}_k^-$. The terms $\hat{\boldsymbol{\mu}}_k^+$ and $\hat{\mathbf{R}}_k^+$ are the MAP estimates of the mean vector and covariance matrix of the prediction error at time step k; they will be estimated by solving Eq. (17), as will be outlined in the next section. It should be noted that Eqs. (19) and (20) are similar to the equations used in the nonadaptive Bayesian model updating methods with the exception of $\hat{\boldsymbol{\mu}}_k^+$, which is now present in Eq. (19) to account for the nonzero-mean prediction error.

Marginal MAP Estimates of μ_k and R_k

Now, we proceed with the second MAP problem in Eq. (17). The term $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k})$ can be written using the Bayes' rule

$$p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{y}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k-1})p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1}) \tag{21}$$

According to Eq. (8), the prior distribution $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k-1})$ has a NIW distribution, which is a conjugate prior for the Gaussian likelihood. Therefore, the posterior is also NIW, i.e., $p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) = NIW(\boldsymbol{\mu}_k, \mathbf{R}_k|\hat{\boldsymbol{\mu}}_k^+, \lambda_k^+, v_k^+, \mathbf{V}_k^+)$, with updated parameters $\hat{\boldsymbol{\mu}}_k^+, \lambda_k^+, v_k^+, \mathbf{V}_k^+$. By substituting Eqs. (4) and (8) into Eq. (21), the following updating equations can be derived—refer to Appendix II for derivation details

$$\hat{\boldsymbol{\mu}}_k^+ = \frac{\lambda_k^-}{1+\lambda_k^-}\hat{\boldsymbol{\mu}}_k^- + \frac{1}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+)) \tag{22a}$$

$$\lambda_k^+ = 1 + \lambda_k^- \tag{22b}$$

$$v_k^+ = 1 + v_k^- \tag{22c}$$

$$\mathbf{V}_k^+ = \mathbf{V}_k^- + \frac{\lambda_k^-}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-)(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-)^T \tag{22d}$$

$$\hat{\mathbf{R}}_k^+ = \frac{\mathbf{V}_k^+}{v_k^+ + n_{\mathbf{y}} + 1} \tag{22e}$$

In this solution, the mode of the posterior NIW distribution is selected as the MAP estimates for $\boldsymbol{\mu}_k$ and $\mathbf{R}_k$, i.e., $\hat{\boldsymbol{\mu}}_k^+$ and $\hat{\mathbf{R}}_k^+$, as shown in Eqs. (22a) and (22e).

The solution of the MAP estimation problem in Eq. (14) will result in solving the coupled Eqs. (19), (22a), and (22e) simultaneously. The coupled nonlinear equations cannot be solved analytically; therefore, a fixed-point iteration algorithm (Hoffman and Frankel 2018) is used in this study. First, $\hat{\boldsymbol{\theta}}_k^+$ is estimated using Eq. (19) given the $\hat{\boldsymbol{\mu}}_k^+$ and $\hat{\mathbf{R}}_k^+$ obtained from the previous iteration; then, $\hat{\boldsymbol{\mu}}_k^+$ and $\hat{\mathbf{R}}_k^+$ are updated using Eqs. (22a) and (22e) based on the estimated $\hat{\boldsymbol{\theta}}_k^+$, and this process is repeated iteratively until convergence. It should be mentioned that to avoid estimating nonphysical values for unknown model parameters, a constraint correction method based on PDF truncation, similar to one implemented by Ebrahimian et al. (2018), can be used in this study. Three convergence criteria are considered based on the relative L2 norm of the difference between two consecutive estimations of $\hat{\boldsymbol{\theta}}_k^+$, $\hat{\boldsymbol{\mu}}_k^+$, and $\hat{\mathbf{R}}_k^+$. A convergence tolerance of 0.01 is considered in this study. This value can be adjusted by balancing accuracy versus the computational cost. Note that the L2 norm of a matrix is equal to its largest singular value (Golub and Van Loan 1996). Fig. 1 provides the flowchart of the proposed two-step marginal MAP estimation approach.

Verification Study: 3-story 1-bay Steel Moment Frame

Model Description and Data Simulation

In this section, the performance of the proposed Bayesian inference framework for joint estimation of model parameters and noise (noise or prediction error = modeling error + measurement noise) is evaluated with a nonlinear model of a 3-story 1-bay steel moment frame under earthquake excitation, as shown in Fig. 2(a). A two-dimensional (2D) nonlinear FE model of the structure is developed in the open-source FE analysis platform OpenSees (McKenna 2000). Columns are made of A992 steel with a W14 × 311 cross-section, and beams are made of A36 steel with a W24 × 68 cross-section. Nodal mass of 80 metric tons is considered at beam-column nodes shown by black circles in Fig. 2(a). Columns and beams are modeled using force-based beam-column elements with fiber sections. Seven integration points are considered for numerical integration along the length of each element using the Gauss-Lobatto quadrature rule. Column and beam webs are discretized into 10 fibers along the

Initialize: $\hat{\boldsymbol{\theta}}_0^+$, $\mathbf{P}_{\boldsymbol{\theta},0}^+$, $\hat{\boldsymbol{\mu}}_0^+$, λ_0^+, v_0^+, and $\mathbf{V}_0^+$

Set $\mathbf{Q}$, ρ, and ρ'

For each time step k ($k=1, 2, \ldots, N$)

- Prediction step:

$$\hat{\boldsymbol{\theta}}_k^- = \hat{\boldsymbol{\theta}}_{k-1}^+$$
$$\mathbf{P}_{\boldsymbol{\theta},k}^- = \mathbf{P}_{\boldsymbol{\theta},k-1}^+ + \mathbf{Q}$$
$$\hat{\boldsymbol{\mu}}_k^- = \hat{\boldsymbol{\mu}}_{k-1}^+$$
$$\lambda_k^- = \rho'\lambda_{k-1}^+$$
$$v_k^- = \rho(v_{k-1}^+ + n_{\mathbf{y}} + 1) - n_{\mathbf{y}} - 1$$
$$\mathbf{V}_k^- = \rho\mathbf{V}_{k-1}^+$$

  - Run model with $\hat{\boldsymbol{\theta}}_k^-$ to find response vector $\mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^-)$, and compute response sensitivity matrix $\mathbf{C}_k^-$.

- Correction step:

$$\hat{\boldsymbol{\theta}}_k^{+(0)} = \hat{\boldsymbol{\theta}}_k^- \quad , \hat{\boldsymbol{\mu}}_k^{+(0)} = \hat{\boldsymbol{\mu}}_k^- \quad , \mathbf{V}_k^{+(0)} = \mathbf{V}_k^-$$
$$\lambda_k^+ = 1 + \lambda_k^-$$
$$v_k^+ = 1 + v_k^-$$
$$\mathbf{P}_{\boldsymbol{\theta}\mathbf{y},k} = \mathbf{P}_{\boldsymbol{\theta},k}^-(\mathbf{C}_k^-)^T$$
$$\mathbf{L}_k^- = \mathbf{C}_k^-\mathbf{P}_{\boldsymbol{\theta},k}^-(\mathbf{C}_k^-)^T$$

  - Iterate ($j = 0, 1, \ldots$):

$$\hat{\mathbf{R}}_k^{+(j)} = \frac{\mathbf{V}_k^{+(j)}}{v_k^+ + n_{\mathbf{y}} + 1}$$
$$\mathbf{P}_{\mathbf{yy},k}^{(j)} = \mathbf{L}_k^- + \hat{\mathbf{R}}_k^{+(j)}$$
$$\mathbf{K}_k^{(j)} = \mathbf{P}_{\boldsymbol{\theta}\mathbf{y},k}(\mathbf{P}_{\mathbf{yy},k}^{(j)})^{-1}$$
$$\hat{\boldsymbol{\theta}}_k^{+(j+1)} = \hat{\boldsymbol{\theta}}_k^- + \mathbf{K}_k^{(j)}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^-) - \hat{\boldsymbol{\mu}}_k^{+(j)})$$

    - Run model with $\boldsymbol{\theta}_k^{+(j+1)}$ to find response vector $\mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^{+(j+1)})$.

$$\hat{\boldsymbol{\mu}}_k^{+(j+1)} = \frac{\lambda_k^-}{1+\lambda_k^-}\hat{\boldsymbol{\mu}}_k^- + \frac{1}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^{+(j+1)}))$$
$$\mathbf{V}_k^{+(j+1)} = \mathbf{V}_k^- + \frac{\lambda_k^-}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^{+(j+1)}) - \hat{\boldsymbol{\mu}}_k^-)(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^{+(j+1)}) - \hat{\boldsymbol{\mu}}_k^-)^T$$

    - Check for convergence: If $\dfrac{\left\|\hat{\boldsymbol{\theta}}_k^{+(j+1)} - \hat{\boldsymbol{\theta}}_k^{+(j)}\right\|}{\left\|\hat{\boldsymbol{\theta}}_k^{+(j)}\right\|} < 0.01$, $\dfrac{\left\|\hat{\boldsymbol{\mu}}_k^{+(j+1)} - \hat{\boldsymbol{\mu}}_k^{+(j)}\right\|}{\left\|\hat{\boldsymbol{\mu}}_k^{+(j)}\right\|} < 0.01$, and $\dfrac{\left\|\hat{\mathbf{R}}_k^{+(j+1)} - \hat{\mathbf{R}}_k^{+(j)}\right\|}{\left\|\hat{\mathbf{R}}_k^{+(j)}\right\|} < 0.01$, then set $\hat{\boldsymbol{\theta}}_k^+ = \hat{\boldsymbol{\theta}}_k^{+(j)}$, $\hat{\boldsymbol{\mu}}_k^+ = \hat{\boldsymbol{\mu}}_k^{+(j)}$, $\mathbf{V}_k^+ = \mathbf{V}_k^{+(j)}$, find $\mathbf{P}_{\boldsymbol{\theta},k}^+ = \mathbf{P}_{\boldsymbol{\theta},k}^- - \mathbf{K}_k^{(j)}\mathbf{P}_{\mathbf{yy},k}^{(j)}(\mathbf{K}_k^{(j)})^T$, and move to next time step; otherwise, $j = j+1$ and iterate again.

Fig. 1. Two-step marginal MAP estimation approach for joint model and noise identification.

height and one fiber across the width. Their flanges are also discretized into one fiber along the height and 3 fibers along the width. The uniaxial material model for steel fibers is based on the Giuffre-Menegotto-Pinto (GMP) constitutive model (Filippou et al. 1983). Rayleigh damping is considered to model the structural damping assuming a 2% damping ratio for the first two modes. The Newmark average acceleration method (Chopra 2017) is used to integrate the equations of motion using a time step size of $\Delta t = 0.02$ obtained using a convergence study.

The 1989 Loma Prieta earthquake (0° component at Los Gatos station) is selected as the base excitation, as illustrated in Fig. 2(b). The horizontal absolute acceleration response time histories at each floor (referred to as true/nominal responses and denoted by $\mathbf{y}^{\text{true}}$) are simulated and contaminated with measurement noise to represent the sensor measurements (denoted by $\mathbf{y}$). The measurement locations are shown by black boxes in Fig. 2(a). The measurement noise is considered as a nonstationary Gaussian random process with a time-variant mean vector and covariance matrix denoted by $\boldsymbol{\mu}_k^{\text{true}}$ and $\mathbf{R}_k^{\text{true}}$, respectively. The measurement noises are assumed statistically uncorrelated; therefore, the off-diagonal entries of $\mathbf{R}_k^{\text{true}}$ at each time step k are equal to zero. Sine functions are considered to model a variation of $\boldsymbol{\mu}_k^{\text{true}}$ and the diagonal entries of $\mathbf{R}_k^{\text{true}}$ in time

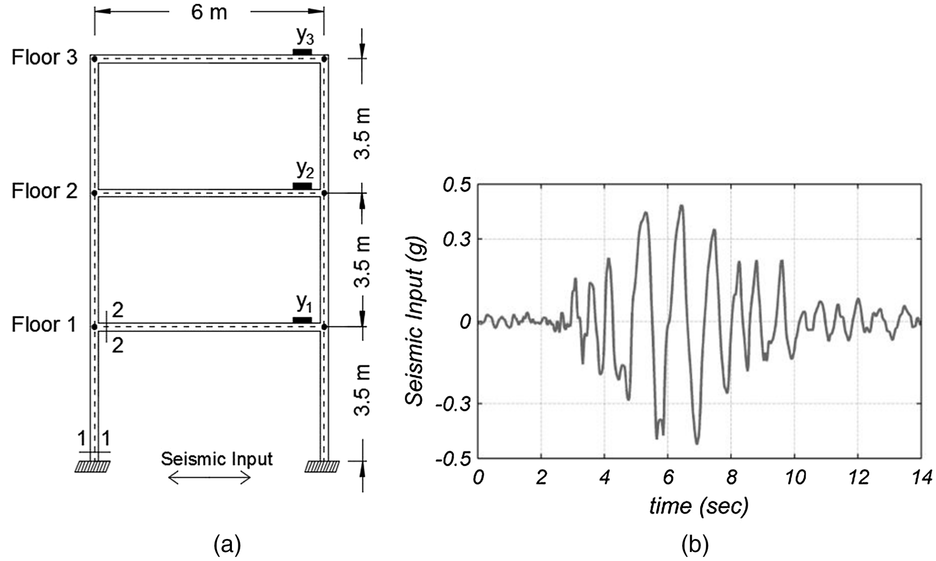

Fig. 2. (a) 3-story 1-bay steel moment frame; and (b) ground acceleration time history of 1989 Loma Prieta earthquake recorded at Los Gatos station in the 0° component.

$$\boldsymbol{\mu}_k^{\mathrm{true}} = \tilde{\boldsymbol{\mu}}^{\mathrm{noise}} \sin\left(\frac{4\pi}{N}k\right) \tag{23}$$

$$\mathrm{diag}(\mathbf{R}_k^{\mathrm{true}}) = \tilde{\mathbf{r}}^{\mathrm{noise}}\left(\sin\left(\frac{\pi}{N}k\right)+1\right)^2 \tag{24}$$

where N = number of time steps; and $\tilde{\boldsymbol{\mu}}^{\mathrm{noise}}$ and $\tilde{\mathbf{r}}^{\mathrm{noise}}$ = constant-coefficient vectors and defined as follows

$$\tilde{\boldsymbol{\mu}}^{\mathrm{noise}} = 50 \times \mathrm{mean}(\mathbf{y}^{\mathrm{true}}) = [1.97, 5.41, 7.50]^T \times 10^{-2} g \tag{25}$$

$$\tilde{\mathbf{r}}^{\mathrm{noise}} = (0.1 \times \mathrm{RMS}(\mathbf{y}^{\mathrm{true}}))^2 = [1.66, 2.88, 7.65]^T \times 10^{-4} g^2 \tag{26}$$

Three parameters of the GMP model for beams and columns are considered as the unknown (updating) model parameters, including Young's modulus E, yield stress σ_y, and strain-hardening ratio b. The parameters are specified by subscripts b and c for beams and columns, respectively. The nominal (or true) values of these parameters used for simulation are reported in Table 1. The mass and stiffness proportional components of Rayleigh damping, i.e., α and β, respectively, are also considered as the unknown model parameters to be estimated. Their true or nominal values are considered as $\alpha^{\mathrm{true}} = 0.1718\ \mathrm{s}^{-1}$ and $\beta^{\mathrm{true}} = 0.0014$ s. Therefore, the unknown model parameter vector $\boldsymbol{\theta}$ includes eight parameters, and each one is normalized by its true value, in other words

$$\boldsymbol{\theta} = [E_c/E_c^{\mathrm{true}}, \sigma_{yc}/\sigma_{yc}^{\mathrm{true}}, b_c/b_c^{\mathrm{true}}, E_b/E_b^{\mathrm{true}}, \sigma_{yb}/\sigma_{yb}^{\mathrm{true}}, b_b/b_b^{\mathrm{true}}, \alpha/\alpha^{\mathrm{true}}, \beta/\beta^{\mathrm{true}}]^T$$

Identification Results

The proposed method is applied to estimate the unknown model parameter vector $\boldsymbol{\theta}$ together with the mean vector and covariance matrix of the prediction error. The initial value of $\boldsymbol{\theta}$ is selected $\hat{\boldsymbol{\theta}}_0^+ = [0.7, 0.7, 1.2, 0.8, 1.3, 0.8, 1.2, 0.7]^T$ as with the initial covariance matrix $\mathbf{P}_{\boldsymbol{\theta},0}^+ = \mathrm{diag}(0.2\hat{\boldsymbol{\theta}}_0^+)^2$, i.e., a 20% initial coefficient of variation (COV) is considered to characterize the prior covariance matrix. The process noise covariance matrix is assumed as $\mathbf{Q} = \mathrm{diag}(10^{-4}\hat{\boldsymbol{\theta}}_0^+)^2$, and the forgetting factor parameters are selected as $\rho = 0.95$ and $\rho' = 0.95$. Based on our study, the parameter estimation results are acceptable for $0.8 \leq \rho < 1$ and $0.7 \leq \rho' < 1$.

Table 1. True/nominal values of the three parameters of the GMP model for columns and beams used for measurement simulations

Frame member	E^{true} (GPa)	σ_y^{true} (MPa)	b^{true}
Columns	200	350	0.08
Beams	200	250	0.05

However, choosing lower values for ρ and ρ' may deteriorate the performance of the model updating process. In this example, a finite-difference method is used to calculate the sensitivity matrix $\mathbf{C}$ at each time step. The initial mean vector and covariance matrix of the prediction error is assumed as $\hat{\boldsymbol{\mu}}_0^+ = \mathbf{0}$ and $\hat{\mathbf{R}}_0^+ = \mathrm{diag}(\tilde{\mathbf{r}}^{\mathrm{noise}})/100$, respectively, where $\tilde{\mathbf{r}}^{\mathrm{noise}}$ vector is defined in Eq. (26). Other initial parameters considered for the NIW distribution are $\lambda_0^+ = 1$, $v_0^+ = 4.1$, and $\mathbf{V}_0^+ = (v_0^+ + n_{\mathbf{y}} + 1)\hat{\mathbf{R}}_0^+$, with $n_{\mathbf{y}} = 3$. In this verification study, the results of the proposed joint model and noise identification method are compared with a classic nonadaptive recursive Bayesian method (Ebrahimian et al. 2015), in which the prediction error $\boldsymbol{\omega}_k$ is assumed to be a zero-mean Gaussian white noise with a time-invariant diagonal covariance matrix, i.e., $\boldsymbol{\omega}_k \sim N(\mathbf{0}, \mathbf{R}_k = \hat{\mathbf{R}}_0^+)$.

The estimated model parameters are shown in Fig. 3. As can be seen, the parameters E_b and E_c start to update and converge to their true values much earlier than the other two material model parameters because the structural responses are sensitive to the stiffness-related material parameter from the beginning of the excitation.

J. Eng. Mech., 2022, 148(3): 04021165

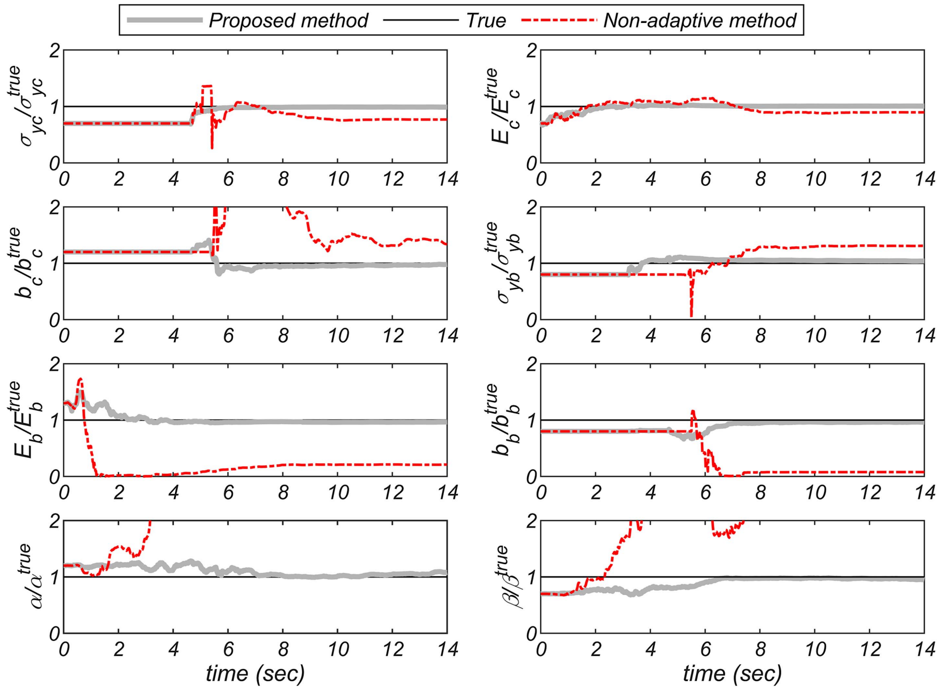

Fig. 3. Estimated model parameters using the proposed and nonadaptive Bayesian model updating methods.

Upon entering the strong-motion part of the excitation, the response of the beams and then the columns enter the nonlinear region, and thus, structural responses become sensitive to the yield strength, σ_y, and strain hardening ratio, b (Ebrahimian et al. 2015). There is also small response sensitivity with respect to the Rayleigh damping coefficients from the beginning of the excitation, which results to slow convergence of these two parameters. Fig. 3 also compares the estimation results with those obtained by a nonadaptive Bayesian model updating method. As can be seen, model parameter estimates are biased or incorrect for the nonadaptive Bayesian model updating method. The nonadaptive Bayesian model updating method can even result in divergence of the model updating process or producing nonphysical model parameter estimates, e.g., zero estimated values for E_b. Fig. 4 and Fig. 5 show the estimated mean and covariance of the prediction error, respectively, at each time step. It can be seen that the proposed method accurately tracks the trend of the true mean vector and covariance matrix in time.

The estimated absolute acceleration responses at each floor—using the final estimated model parameters—are compared with the true/nominal counterparts in Fig. 6. The noticeable discrepancies

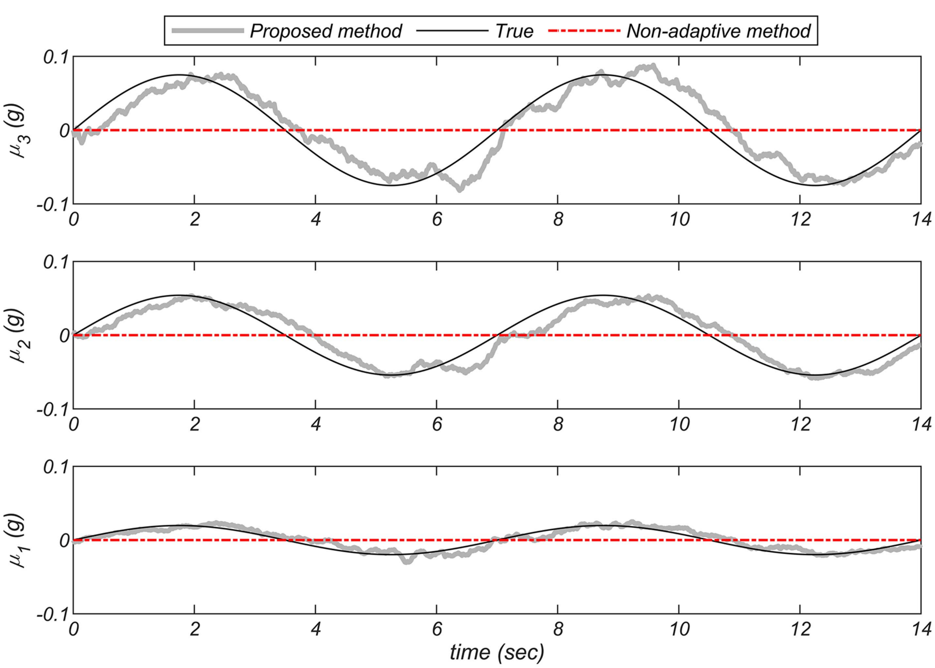

Fig. 4. Estimated components of the prediction error mean vector using the proposed Bayesian model updating method. The mean vector of the prediction error is not estimated and thus remains constant in the nonadaptive Bayesian model updating method. Note that $\boldsymbol{\mu} = [\mu_1, \mu_2, \mu_3]^T$.

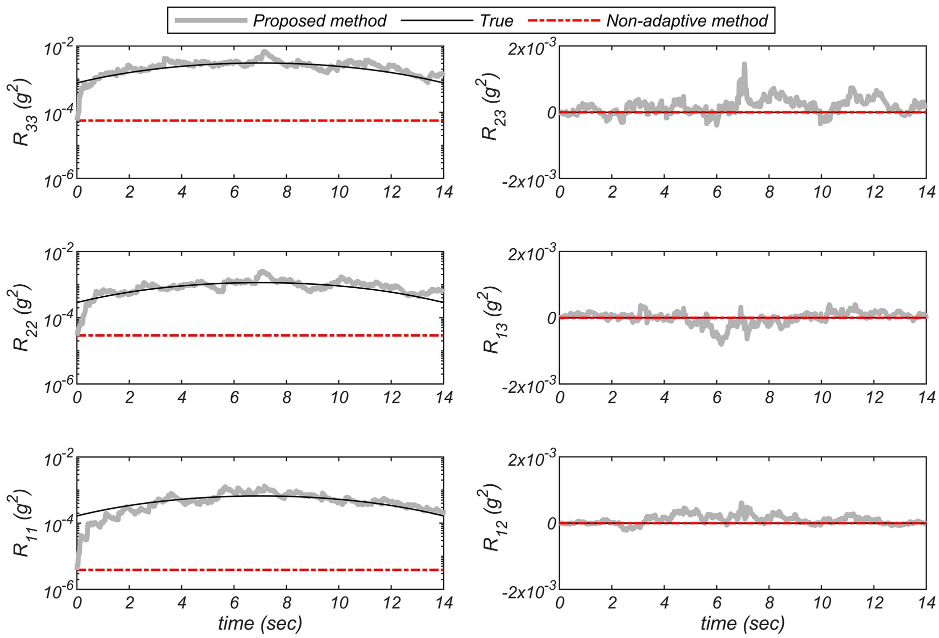

Fig. 5. Estimated components of the prediction error covariance matrix using the proposed Bayesian model updating method. The covariance matrix of the prediction error is not estimated and thus remains constant in the nonadaptive Bayesian model updating method.

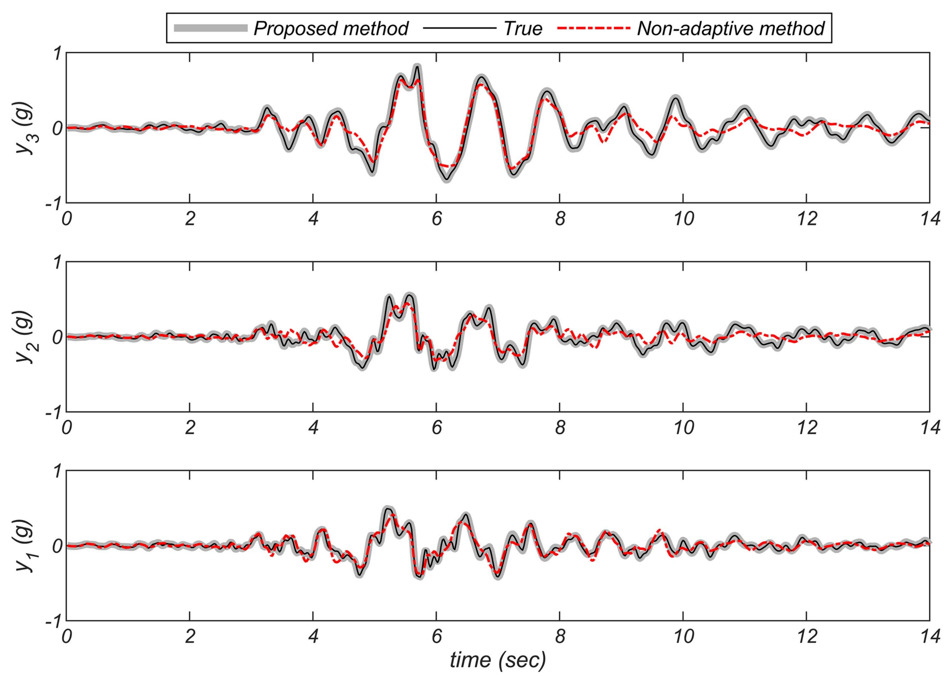

Fig. 6. Comparison of the true and estimated absolute acceleration responses using the proposed and nonadaptive Bayesian model updating methods.

between the estimated and nominal responses for the nonadaptive Bayesian model updating method, which is due to the biased model parameter estimates, clearly show the incapability of the nonadaptive Bayesian model updating method to perform in the presence of a time-variant prediction error. The acceleration responses predicted using the proposed joint model and noise identification method match the nominal responses well.

Finally, the moment-curvature response at the base section of the left column—Section 1-1 in Fig. 2(a)—and the stress-strain response of the top flange of the first-floor beam at the plastic hinge location—Section 2-2 in Fig. 2(a)—are estimated from the updated models using the nonadaptive and proposed model updating methods and presented in Fig. 7. The figure clearly shows the detrimental effects of the time-variant noise in the response prediction and virtual sensing capability of the nonadaptive Bayesian model updating. This is why the proposed method can handle the time-variant noise effects and provide accurate virtual sensing capability.

Summary and Conclusion

This paper presented an adaptive recursive Bayesian inference framework for joint model and noise identification using a two-step

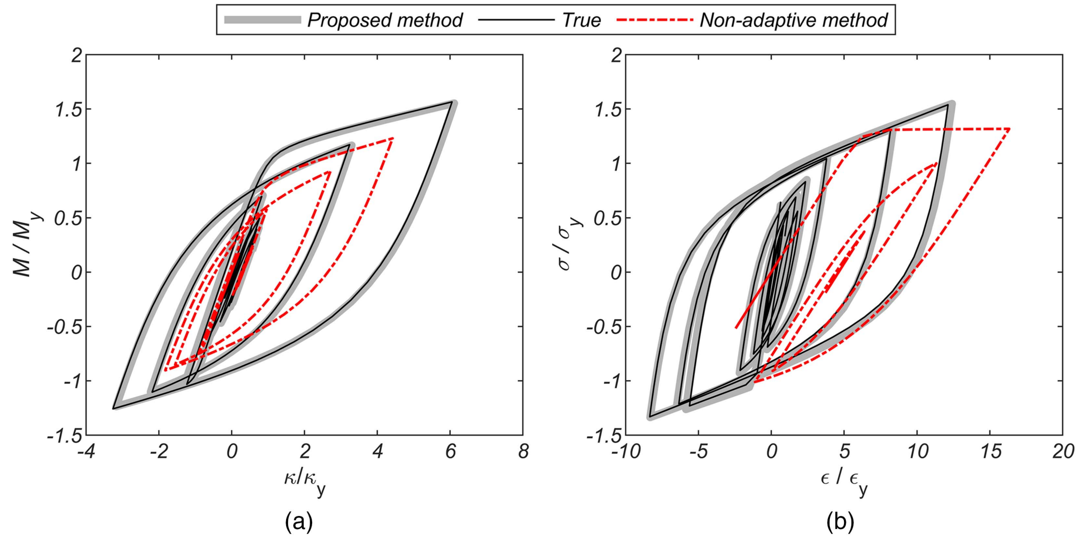

Fig. 7. Comparison of the true and estimated (a) moment-curvature response at the base section of the left column shown in Fig. 2(a); and (b) stress-strain response of the top flange of the first-floor beam at the plastic hinge location shown in Fig. 2(a). The responses are obtained from the updated models from the proposed and nonadaptive Bayesian model updating methods.

marginal maximum a posteriori (MAP) estimation approach. Noise or prediction error can include the measurement noise and the effects of modeling error. This approach results in two separate MAP estimation problems that should be solved iteratively: one to estimate the unknown model parameters and the other to estimate the mean vector and covariance matrix of the prediction error. The proposed approach was verified using numerically simulated data obtained from a nonlinear steel moment frame model subjected to earthquake excitation. The absolute acceleration responses at each floor were simulated and polluted by a nonstationary Gaussian noise with a time-variant mean vector and covariance matrix. Eight model parameters, including six parameters characterizing the constitutive models of the beams and columns steel material, and two Rayleigh damping coefficients were considered as unknown and estimated using the proposed approach and a nonadaptive recursive Bayesian model updating approach for comparison. The verification study demonstrated the detrimental effects of time-variant noise on the nonadaptive Bayesian model updating results, in which the estimated model parameters were significantly biased. However, the proposed Bayesian inference framework for joint model and noise identification was able to estimate the model parameters correctly due to its capability to estimate the time-variant mean and covariance of the prediction error. Considering the statistical characteristics of prediction error as unknowns to be estimated provides additional degrees of freedom in the recursive Bayesian model updating approach and can alleviate the biased or incorrect model parameter estimation results. Therefore, the proposed framework can provide a step forward to account for the effects of modeling error in finite-element model updating. Further efforts are underway to validate the proposed framework through real-world case studies.

Appendix I. Derivation of MAP Estimates of Unknown Model Parameters

This appendix presents the derivation of the MAP estimate of the unknown model parameter, $\boldsymbol{\theta}_k$, given $\hat{\boldsymbol{\mu}}_k^+$, and $\hat{\mathbf{R}}_k^+$. By substituting Eqs. (4) and (7) into Eq. (18), it can be followed that

$$p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k}) \propto N(\boldsymbol{\omega}_k|\hat{\boldsymbol{\mu}}_k^+, \hat{\mathbf{R}}_k^+) \times N(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_k^-, \mathbf{P}_{\boldsymbol{\theta},k}^-) \tag{27}$$

Following Lemma 1 of Appendix III, distribution $N(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_k^-, \mathbf{P}_{\boldsymbol{\theta},k}^-)$ can be written in the canonical form

$$N(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_k^-, \mathbf{P}_{\boldsymbol{\theta},k}^-) = \exp\left(-\frac{1}{2}\boldsymbol{\theta}_k^T\boldsymbol{\Lambda}_1\boldsymbol{\theta}_k + \boldsymbol{\eta}_1^T\boldsymbol{\theta}_k + \xi_1\right) \tag{28}$$

where $\boldsymbol{\Lambda}_1 = (\mathbf{P}_{\boldsymbol{\theta},k}^-)^{-1}$; $\boldsymbol{\eta}_1 = (\mathbf{P}_{\boldsymbol{\theta},k}^-)^{-1}\hat{\boldsymbol{\theta}}_k^-$; and $\xi_1 = -\frac{1}{2}(n_{\boldsymbol{\theta}}\ln(2\pi) + \ln|\mathbf{P}_{\boldsymbol{\theta},k}^-| + (\hat{\boldsymbol{\theta}}_k^-)^T(\mathbf{P}_{\boldsymbol{\theta},k}^-)^{-1}\hat{\boldsymbol{\theta}}_k^-)$. Also, the term $N(\boldsymbol{\omega}_k|\hat{\boldsymbol{\mu}}_k^+, \hat{\mathbf{R}}_k^+)$ in Eq. (28) can be expanded

$$N(\boldsymbol{\omega}_k|\hat{\boldsymbol{\mu}}_k^+, \hat{\mathbf{R}}_k^+) = \frac{1}{(2\pi)^{n_y/2}|\hat{\mathbf{R}}_k^+|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{\omega}_k - \hat{\boldsymbol{\mu}}_k^+)^T(\hat{\mathbf{R}}_k^+)^{-1}(\boldsymbol{\omega}_k - \hat{\boldsymbol{\mu}}_k^+)\right) \tag{29}$$

Following Eq. (2), it is clear that $\boldsymbol{\omega}_k = \mathbf{y}_k - \mathbf{h}_k(\boldsymbol{\theta}_k)$. The term $\mathbf{h}_k(\boldsymbol{\theta}_k)$, which is the model response function can be linearized using a first-order Taylor expansion about $\hat{\boldsymbol{\theta}}_k^-$, i.e., $\mathbf{h}_k(\boldsymbol{\theta}_k) \simeq \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^-) + \mathbf{C}_k^-(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-)$ in which $\mathbf{C}_k^- = \{[\partial\mathbf{h}_k(\boldsymbol{\theta}_k)]/\partial\boldsymbol{\theta}_k\}|_{\boldsymbol{\theta}_k=\hat{\boldsymbol{\theta}}_k^-}$. Therefore, it can be followed that

$$\begin{aligned}\boldsymbol{\omega}_k - \hat{\boldsymbol{\mu}}_k^+ &= \mathbf{y}_k - \mathbf{h}_k(\boldsymbol{\theta}_k) - \hat{\boldsymbol{\mu}}_k^+ \\ &= \mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^-) - \mathbf{C}_k^-(\boldsymbol{\theta}_k - \hat{\boldsymbol{\theta}}_k^-) - \hat{\boldsymbol{\mu}}_k^+ \\ &= \boldsymbol{\delta}_k - \mathbf{C}_k^-\boldsymbol{\theta}_k\end{aligned} \tag{30}$$

where

$$\boldsymbol{\delta}_k = \mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^-) + \mathbf{C}_k^-\hat{\boldsymbol{\theta}}_k^- - \hat{\boldsymbol{\mu}}_k^+ \tag{31}$$

By substituting Eq. (30) into Eq. (29), the following canonical form can be derived

$$N(\boldsymbol{\omega}_k|\hat{\boldsymbol{\mu}}_k^+, \hat{\mathbf{R}}_k^+) = \exp\left(-\frac{1}{2}\boldsymbol{\theta}_k^T\boldsymbol{\Lambda}_2\boldsymbol{\theta}_k + \boldsymbol{\eta}_2^T\boldsymbol{\theta}_k + \xi_2\right) \tag{32}$$

J. Eng. Mech., 2022, 148(3): 04021165

where $\boldsymbol{\Lambda}_2 = (\mathbf{C}_k^-)^T(\hat{\mathbf{R}}_k^+)^{-1}\mathbf{C}_k^-$; $\boldsymbol{\eta}_2 = (\mathbf{C}_k^-)^T(\hat{\mathbf{R}}_k^+)^{-1}\boldsymbol{\delta}_k$; and $\xi_2 = -\frac{1}{2}(n_\mathbf{y}\ln(2\pi) + \ln|\hat{\mathbf{R}}_k^+| + \boldsymbol{\delta}_k^T(\hat{\mathbf{R}}_k^+)^{-1}\boldsymbol{\delta}_k)$. Based on Lemma 2 of Appendix III, the product of two Gaussian distributions on the right-hand side of Eq. (27) is a Gaussian, i.e., $p(\boldsymbol{\theta}_k|\boldsymbol{\mu}_k, \mathbf{R}_k, \mathbf{y}_{1:k}) = N(\boldsymbol{\theta}_k|\hat{\boldsymbol{\theta}}_k^+, \mathbf{P}_{\boldsymbol{\theta},k}^+)$, and its mean vector and covariance matrix can be derived as follows

$$\hat{\boldsymbol{\theta}}_k^+ = (\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2)^{-1}(\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2) = ((\mathbf{P}_{\boldsymbol{\theta},k}^-)^{-1} + (\mathbf{C}_k^-)^T(\hat{\mathbf{R}}_k^+)^{-1}\mathbf{C}_k^-)^{-1} \times ((\mathbf{P}_{\boldsymbol{\theta},k}^-)^{-1}\hat{\boldsymbol{\theta}}_k^- + (\mathbf{C}_k^-)^T(\hat{\mathbf{R}}_k^+)^{-1}\boldsymbol{\delta}_k) \tag{33a}$$

$$\mathbf{P}_{\boldsymbol{\theta},k}^+ = (\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2)^{-1} = ((\mathbf{P}_{\boldsymbol{\theta},k}^-)^{-1} + (\mathbf{C}_k^-)^T(\hat{\mathbf{R}}_k^+)^{-1}\mathbf{C}_k^-)^{-1} \tag{33b}$$

By substituting Eq. (31) into Eq. (33a) and defining the Kalman gain matrix as $\mathbf{K}_k = ((\mathbf{P}_{\boldsymbol{\theta},k}^-)^{-1} + (\mathbf{C}_k^-)^T(\hat{\mathbf{R}}_k^+)^{-1}\mathbf{C}_k^-)^{-1}(\mathbf{C}_k^-)^T(\hat{\mathbf{R}}_k^+)^{-1}$, Eq. (33a) results in

$$\hat{\boldsymbol{\theta}}_k^+ = \hat{\boldsymbol{\theta}}_k^- + \mathbf{K}_k(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^-) - \hat{\boldsymbol{\mu}}_k^+)) \tag{34}$$

Alternatively, it can be shown that the Kalman gain can be derived as follows (Simon 2006)

$$\mathbf{K}_k = \mathbf{P}_{\boldsymbol{\theta}\mathbf{y},k}(\mathbf{P}_{\mathbf{yy},k})^{-1} \tag{35}$$

The posterior covariance matrix can be derived (Simon 2006)

$$\mathbf{P}_{\boldsymbol{\theta},k}^+ = \mathbf{P}_{\boldsymbol{\theta},k}^- - \mathbf{K}_k\mathbf{P}_{\mathbf{yy},k}\mathbf{K}_k^T \tag{36}$$

where $\mathbf{P}_{\boldsymbol{\theta}\mathbf{y},k} = \mathbf{P}_{\boldsymbol{\theta},k}^-(\mathbf{C}_k^-)^T$; and $\mathbf{P}_{\mathbf{yy},k} = \mathbf{C}_k^-\mathbf{P}_{\boldsymbol{\theta},k}^-(\mathbf{C}_k^-)^T + \hat{\mathbf{R}}_k^+$.

Appendix II. Derivation of MAP Estimates of Mean Vector and Covariance Matrix of Prediction Error

This appendix provides the derivation of MAP estimates of the mean vector and covariance matrix of prediction error, i.e., $\boldsymbol{\mu}_k$ and $\mathbf{R}_k$, given $\hat{\boldsymbol{\theta}}_k^+$. By substituting Eqs. (4) and (8) into Eq. (21), it can be followed that

$$p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) \propto N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) \times NIW(\boldsymbol{\mu}_k, \mathbf{R}_k|\hat{\boldsymbol{\mu}}_k^-, \lambda_k^-, v_k^-, \mathbf{V}_k^-) \propto N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) \times N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-}\right) \times IW(\mathbf{R}_k|v_k^-, \mathbf{V}_k^-) \tag{37}$$

Based on the conjugacy property of NIW distribution with respect to the Normal likelihood function, the posterior has the same distribution as the prior. Therefore, the posterior is also NIW, in other words

$$p(\boldsymbol{\mu}_k, \mathbf{R}_k|\mathbf{y}_{1:k}) = NIW(\boldsymbol{\mu}_k, \mathbf{R}_k|\hat{\boldsymbol{\mu}}_k^+, \lambda_k^+, v_k^+, \mathbf{V}_k^+) = N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^+, \frac{\mathbf{R}_k}{\lambda_k^+}\right) \times IW(\mathbf{R}_k|v_k^+, \mathbf{V}_k^+) \tag{38}$$

Combining Eqs. (37) and (38) results in

$$\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^+, \frac{\mathbf{R}_k}{\lambda_k^+}\right) \times IW(\mathbf{R}_k|v_k^+, \mathbf{V}_k^+) \propto N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) \times N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-}\right) \times IW(\mathbf{R}_k|v_k^-, \mathbf{V}_k^-) \tag{39}$$

The Gaussian distribution $N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k)$ in Eq. (39) can be expanded

$$N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) = \frac{1}{(2\pi)^{n_\mathbf{y}/2}|\mathbf{R}_k|^{1/2}}\exp\left(-\frac{1}{2}(\boldsymbol{\omega}_k - \boldsymbol{\mu}_k)^T\mathbf{R}_k^{-1}(\boldsymbol{\omega}_k - \boldsymbol{\mu}_k)\right) \tag{40}$$

Considering that $\boldsymbol{\omega}_k = \mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+)$, Eq. (40) can be written in a canonical form for $\boldsymbol{\mu}_k$ using Lemma 1 of Appendix III as follows

$$N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) = \exp\left(-\frac{1}{2}\boldsymbol{\mu}_k^T\boldsymbol{\Lambda}_1\boldsymbol{\mu}_k + \boldsymbol{\eta}_1^T\boldsymbol{\mu}_k + \xi_1\right) \tag{41}$$

where

$$\boldsymbol{\Lambda}_1 = \mathbf{R}_k^{-1} \tag{42a}$$

$$\boldsymbol{\eta}_1 = \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+)) \tag{42b}$$

$$\xi_1 = -\frac{1}{2}(n_\mathbf{y}\ln(2) + \ln|\mathbf{R}_k| + (\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+))^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+))) \tag{42c}$$

The Gaussian distribution $N(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-})$ on the right-hand side of Eq. (39) can also be written in a canonical form

$$N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-}\right) = \exp\left(-\frac{1}{2}\boldsymbol{\mu}_k^T\boldsymbol{\Lambda}_2\boldsymbol{\mu}_k + \boldsymbol{\eta}_2^T\boldsymbol{\mu}_k + \xi_2\right) \tag{43}$$

where

$$\boldsymbol{\Lambda}_2 = \left(\frac{\mathbf{R}_k}{\lambda_k^-}\right)^{-1} \tag{44a}$$

$$\boldsymbol{\eta}_2 = \left(\frac{\mathbf{R}_k}{\lambda_k^-}\right)^{-1}\hat{\boldsymbol{\mu}}_k^- \tag{44b}$$

$$\xi_2 = -\frac{1}{2}\left(n_\mathbf{y}\ln(2\pi) + \ln\left|\frac{\mathbf{R}_k}{\lambda_k^-}\right| + (\hat{\boldsymbol{\mu}}_k^-)^T\left(\frac{\mathbf{R}_k}{\lambda_k^-}\right)^{-1}\hat{\boldsymbol{\mu}}_k^-\right) \tag{44c}$$

The Inverse-Wishart (IW) distribution $IW(\mathbf{R}_k|v_k^-, \mathbf{V}_k^-)$ on the right-hand side of Eq. (39) can be expanded

$$IW(\mathbf{R}_k|v_k^-, \mathbf{V}_k^-) = c_1|\mathbf{R}_k|^{-(v_k^- + n_\mathbf{y} + 1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{V}_k^-\mathbf{R}_k^{-1})\right) \tag{45}$$

where c_1 = constant term.

Using Lemma 2 of Appendix III, the product of two Gaussian distributions $N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k)$ and $N(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-})$ on the right-hand side of Eq. (39) is also Gaussian and can be written in the canonical form for $\boldsymbol{\mu}_k$. Following Eqs. (41) and (43), it can be observed that

$$N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) \times N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-}\right) = \exp\left(-\frac{1}{2}\boldsymbol{\mu}_k^T\boldsymbol{\Lambda}\boldsymbol{\mu}_k + \boldsymbol{\eta}^T\boldsymbol{\mu}_k + \xi\right) \times \exp(\xi_1 + \xi_2 - \xi) \tag{46}$$

where $\boldsymbol{\Lambda}$, $\boldsymbol{\eta}$, and ξ can be calculated using Lemma 2 of Appendix III and Eqs. (42) and (44) as follows

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2 = \mathbf{R}_k^{-1} + \left(\frac{\mathbf{R}_k}{\lambda_k^-}\right)^{-1} = (1 + \lambda_k^-)\mathbf{R}_k^{-1} \tag{47a}$$

J. Eng. Mech., 2022, 148(3): 04021165

$$\boldsymbol{\eta} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2 = \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+)) + \left(\frac{\mathbf{R}_k}{\lambda_k^-}\right)^{-1}\hat{\boldsymbol{\mu}}_k^-$$
$$= \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) + \lambda_k^-\hat{\boldsymbol{\mu}}_k^-) \quad (47b)$$

$$\xi = -\frac{1}{2}(n_{\mathbf{y}}\ln(2\pi) + \ln|\boldsymbol{\Lambda}^{-1}| + \boldsymbol{\eta}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta})$$
$$= -\frac{1}{2}\left(n_{\mathbf{y}}\ln(2\pi) + \ln\left|\frac{\mathbf{R}_k}{1+\lambda_k^-}\right| + (\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) + \lambda_k^-\hat{\boldsymbol{\mu}}_k^-)^T \times \frac{\mathbf{R}_k^{-1}}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) + \lambda_k^-\hat{\boldsymbol{\mu}}_k^-)\right) \quad (47c)$$

The expression $\xi_1 + \xi_2 - \xi$ in Eq. (46) can be expanded as follows

$$\xi_1 + \xi_2 - \xi = -\frac{1}{2}\left(\ln|\mathbf{R}_k| + (\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+))^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+)) + (\hat{\boldsymbol{\mu}}_k^-)^T\left(\frac{\mathbf{R}_k}{\lambda_k^-}\right)^{-1}\hat{\boldsymbol{\mu}}_k^- - (\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) + \lambda_k^-\hat{\boldsymbol{\mu}}_k^-)^{\boldsymbol{T}} \times \frac{\mathbf{R}_k^{-1}}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) + \lambda_k^-\hat{\boldsymbol{\mu}}_k^-)\right) + c_2$$
$$= -\frac{1}{2}\left(\ln|\mathbf{R}_k| + \frac{\lambda_k^-}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-)^T \times \mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-)\right) + c_2 \quad (48)$$

where c_2 = constant term; and tr(.) = matrix trace. Therefore, the term $\exp(\xi_1 + \xi_2 - \xi)$ on the right-hand side of Eq. (46) can be expressed as follows considering Lemma 3 of Appendix III

$$\exp(\xi_1 + \xi_2 - \xi) = c_3|\mathbf{R}_k|^{-\frac{1}{2}} \times \exp\left(-\frac{1}{2}\mathrm{tr}\left(\frac{\lambda_k^-}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-) \times (\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-)^T\mathbf{R}_k^{-1}\right)\right) \quad (49)$$

where c_3 = constant term. Therefore, the right-hand side of Eq. (39) can be obtained using Eqs. (45) and (46) as follows

$$N(\boldsymbol{\omega}_k|\boldsymbol{\mu}_k, \mathbf{R}_k) \times N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^-, \frac{\mathbf{R}_k}{\lambda_k^-}\right) \times IW(\mathbf{R}_k|v_k^-, \mathbf{V}_k^-)$$
$$= \exp\left(-\frac{1}{2}\boldsymbol{\mu}_k^T\boldsymbol{\Lambda}\boldsymbol{\mu}_k + \boldsymbol{\eta}^T\boldsymbol{\mu}_k + \xi\right) \times \exp(\xi_1 + \xi_2 - \xi) \times c_1|\mathbf{R}_k|^{-(v_k^- + n_{\mathbf{y}} + 1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{V}_k^-\mathbf{R}_k^{-1})\right) \quad (50)$$

Comparing Eqs. (39) and (50), and considering that the term $\exp(\xi_1 + \xi_2 - \xi)$ is a function of $\mathbf{R}_k$ based on Eq. (49), it can be observed that

$$N\left(\boldsymbol{\mu}_k|\hat{\boldsymbol{\mu}}_k^+, \frac{\mathbf{R}_k}{\lambda_k^+}\right) = \exp\left(-\frac{1}{2}\boldsymbol{\mu}_k^T\boldsymbol{\Lambda}\boldsymbol{\mu}_k + \boldsymbol{\eta}^T\boldsymbol{\mu}_k + \xi\right) \quad (51a)$$

$$IW(\mathbf{R}_k|v_k^+, \mathbf{V}_k^+) = \exp(\xi_1 + \xi_2 - \xi) \times c_1|\mathbf{R}_k|^{-(v_k^- + n_{\mathbf{y}} + 1)/2}\exp\left(-\frac{1}{2}\mathrm{tr}(\mathbf{V}_k^-\mathbf{R}_k^{-1})\right) \quad (51b)$$

where the updated mean vector $\hat{\boldsymbol{\mu}}_k^+$ and covariance matrix $\mathbf{R}_k/\lambda_k^+$ can be obtained using Lemma 1 of Appendix III as follows

$$\hat{\boldsymbol{\mu}}_k^+ = \boldsymbol{\Lambda}^{-1}\boldsymbol{\eta} = \frac{\mathbf{R}_k}{1+\lambda_k^-}(\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) + \lambda_k^-\hat{\boldsymbol{\mu}}_k^-)) \quad (52a)$$

$$\frac{\mathbf{R}_k}{\lambda_k^+} = \boldsymbol{\Lambda}^{-1} = \frac{\mathbf{R}_k}{1+\lambda_k^-} \quad (52b)$$

These equations can be simplified as follows

$$\hat{\boldsymbol{\mu}}_k^+ = \frac{\lambda_k^-}{1+\lambda_k^-}\hat{\boldsymbol{\mu}}_k^- + \frac{1}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+)) \quad (53a)$$

$$\lambda_k^+ = 1 + \lambda_k^- \quad (53b)$$

The updated parameters of IW distribution are obtained by matching the terms on the left- and right-hand side of Eq. (51b) as follows

$$v_k^+ = 1 + v_k^- \quad (54a)$$

$$\mathbf{V}_k^+ = \mathbf{V}_k^- + \frac{\lambda_k^-}{1+\lambda_k^-}(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-)(\mathbf{y}_k - \mathbf{h}_k(\hat{\boldsymbol{\theta}}_k^+) - \hat{\boldsymbol{\mu}}_k^-)^T \quad (54b)$$

Appendix III. Three Useful Lemmas

This appendix presents three Lemmas that are used in Appendices I and II.

Lemma 1: Canonical representation of a multivariate Gaussian distribution function

The multivariate Gaussian distribution of a random vector $\mathbf{x} \in \mathbb{R}^{n_{\mathbf{x}}\times 1}$ with a distribution function of $\mathbf{x} \sim N(\bar{\mathbf{x}}, \boldsymbol{\Sigma})$ can be written in the canonical form (Wu 2005)

$$p(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Lambda}\mathbf{x} + \boldsymbol{\eta}^T\mathbf{x} + \xi\right) \quad (55)$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}$; $\boldsymbol{\eta} = \boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}}$, $\xi = -\frac{1}{2}(n_{\mathbf{x}}\ln(2\pi) + \ln|\boldsymbol{\Sigma}| + \bar{\mathbf{x}}^T\boldsymbol{\Sigma}^{-1}\bar{\mathbf{x}})$ or $\xi = -\frac{1}{2}(n_{\mathbf{x}}\ln(2\pi) + \ln|\boldsymbol{\Lambda}^{-1}| + \boldsymbol{\eta}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta})$; and $|.|$ presents the determinant operator.

Lemma 2: Product of two multivariate Gaussian distribution functions

Consider two independent random vectors of size $n_{\mathbf{x}}$ with Gaussian distributions as $\mathbf{x}_1 \sim N(\bar{\mathbf{x}}_1, \boldsymbol{\Sigma}_1)$ and $\mathbf{x}_2 \sim N(\bar{\mathbf{x}}_2, \boldsymbol{\Sigma}_2)$. Using the canonical form presented in Lemma 1, it can be followed that

$$p_1(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Lambda}_1\mathbf{x} + \boldsymbol{\eta}_1^T\mathbf{x} + \xi_1\right)$$
$$p_2(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathbf{x}^T\boldsymbol{\Lambda}_2\mathbf{x} + \boldsymbol{\eta}_2^T\mathbf{x} + \xi_2\right) \quad (56)$$

Therefore, the product of two distribution functions $p_1(\mathbf{x})$ and $p_2(\mathbf{x})$ can be derived

$$p_1(\mathbf{x}) \times p_2(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathbf{x}^T(\boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2)\mathbf{x} + (\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2)^T\mathbf{x} + (\xi_1 + \xi_2)\right) \quad (57)$$

By defining the following new terms

$$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_1 + \boldsymbol{\Lambda}_2$$
$$\boldsymbol{\eta} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2$$
$$\xi = -\frac{1}{2}(n_{\mathbf{x}}\ln(2\pi) + \ln|\boldsymbol{\Lambda}^{-1}| + \boldsymbol{\eta}^T\boldsymbol{\Lambda}^{-1}\boldsymbol{\eta}) \quad (58)$$

it can be followed that

$$p_1(\mathbf{x}) \times p_2(\mathbf{x}) = \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{\Lambda}\mathbf{x} + \mathbf{\eta}^T\mathbf{x} + \xi\right) \times \exp(\xi_1 + \xi_2 - \xi) \tag{59}$$

By comparing Eqs. (55) and (59), the product of the two Gaussian distributions will be a scaled Gaussian distribution $\mathbf{x} \sim N(\bar{\mathbf{x}}, \mathbf{\Sigma})$ with the scale factor of $\exp(\xi_1 + \xi_2 - \xi)$. The mean vector and the covariance matrix of the resulting Gaussian distribution can be defined as follows (Deisenroth et al. 2020)

$$\begin{aligned}\bar{\mathbf{x}} &= \mathbf{\Lambda}^{-1}\mathbf{\eta} = (\mathbf{\Lambda}_1 + \mathbf{\Lambda}_2)^{-1}(\mathbf{\eta}_1 + \mathbf{\eta}_2) \\ &= (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}(\mathbf{\Sigma}_1^{-1}\bar{\mathbf{x}}_1 + \mathbf{\Sigma}_2^{-1}\bar{\mathbf{x}}_2) \\ \mathbf{\Sigma} &= \mathbf{\Lambda}^{-1} = (\mathbf{\Lambda}_1 + \mathbf{\Lambda}_2)^{-1} = (\mathbf{\Sigma}_1^{-1} + \mathbf{\Sigma}_2^{-1})^{-1}\end{aligned} \tag{60}$$

Lemma 3: Trace of a quadratic form

If $\mathbf{x} \in \mathbb{R}^{n_\mathbf{x} \times 1}$ is a vector, and $\mathbf{A} \in \mathbb{R}^{n_\mathbf{x} \times n_\mathbf{x}}$ is a matrix, then (Kollo and von Rosen 2006)

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathrm{tr}(\mathbf{A}\mathbf{x}\mathbf{x}^T) = \mathrm{tr}(\mathbf{x}\mathbf{x}^T\mathbf{A}) \tag{61}$$

where $\mathrm{tr}(.)$ denotes a matrix trace.

Data Availability Statement

All data, models, and code that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgments

The third author acknowledges partial support of this study through the National Science Foundation Grant 1903972. The opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily represent the views of the sponsors.

References

Akhlaghi, S., N. Zhou, and Z. Huang. 2017. "Adaptive adjustment of noise covariance in Kalman filter for dynamic state estimation." In *Proc., 2017 IEEE Power & Energy Society General Meeting*. New York: IEEE.

Almagbile, A., J. Wang, and W. Ding. 2010. "Evaluating the performances of adaptive Kalman filter methods in GPS/INS integration." *J. Global Positioning Syst.* 9 (1): 33–40. https://doi.org/10.5081/jgps.9.1.33.

Amini Tehrani, H., A. Bakhshi, and T. T. Y. Yang. 2021. "Online probabilistic model class selection and joint estimation of structures for post-disaster monitoring." *J. Vib. Control* 27 (Aug): 1860–1878. https://doi.org/10.1177/1077546320949115.

Astroza, R., A. Alessandri, and J. P. Conte. 2019a. "A dual adaptive filtering approach for nonlinear finite element model updating accounting for modeling uncertainty." *Mech. Syst. Signal Process.* 115 (Jan): 782–800. https://doi.org/10.1016/j.ymssp.2018.06.014.

Astroza, R., H. Ebrahimian, and J. P. Conte. 2019b. "Performance comparison of Kalman− based filters for nonlinear structural finite element model updating." *J. Sound Vib.* 438 (Jan): 520–542. https://doi.org/10.1016/j.jsv.2018.09.023.

Beck, J. L., and K.-V. Yuen. 2004. "Model selection using response measurements: Bayesian probabilistic approach." *J. Eng. Mech.* 130 (2): 192–203. https://doi.org/10.1061/(ASCE)0733-9399(2004)130:2(192).

Behmanesh, I., B. Moaveni, G. Lombaert, and C. Papadimitriou. 2015. "Hierarchical Bayesian model updating for structural identification." *Mech. Syst. Signal Process.* 64 (Dec): 360–376. https://doi.org/10.1016/j.ymssp.2015.03.026.

Besançon, G. 2007. *Nonlinear observers and applications*. Berlin: Springer.

Chatzi, E. N., A. W. Smyth, and S. F. Masri. 2010. "Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty." *Struct. Saf.* 32 (5): 326–337. https://doi.org/10.1016/j.strusafe.2010.03.008.

Chopra, A. K. 2017. *Dynamics of structures: Theory and applications to earthquake engineering*. New York: Prentice-Hall.

Deisenroth, M. P., A. A. Faisal, and C. S. Ong. 2020. *Mathematics for machine learning*. Cambridge, MA: Cambridge University Press.

Desceliers, C., C. Soize, and S. Cambier. 2004. "Non-parametric–parametric model for random uncertainties in non-linear structural dynamics: Application to earthquake engineering." *Earthquake Eng. Struct. Dyn.* 33 (3): 315–327. https://doi.org/10.1002/eqe.352.

Ebrahimian, H. 2015. *Nonlinear finite element model updating for nonlinear system and damage identification of civil structures*. San Diego: Univ. of California.

Ebrahimian, H., R. Astroza, and J. P. Conte. 2015. "Extended Kalman filter for material parameter estimation in nonlinear structural finite element models using direct differentiation method." *Earthquake Eng. Struct. Dyn.* 44 (10): 1495–1522. https://doi.org/10.1002/eqe.2532.

Ebrahimian, H., R. Astroza, J. P. Conte, and R. R. Bitmead. 2019. "Information-theoretic approach for identifiability assessment of nonlinear structural finite-element models." *J. Eng. Mech.* 145 (7): 04019039. https://doi.org/10.1061/(ASCE)EM.1943-7889.0001590.

Ebrahimian, H., M. Kohler, A. Massari, and D. Asimaki. 2018. "Parametric estimation of dispersive viscoelastic layered media with application to structural health monitoring." *Soil Dyn. Earthquake Eng.* 105 (Feb): 204–223. https://doi.org/10.1016/j.soildyn.2017.10.017.

Erazo, K., and S. Nagarajaiah. 2017. "An offline approach for output-only Bayesian identification of stochastic nonlinear systems using unscented Kalman filtering." *J. Sound Vib.* 397 (Jun): 222–240. https://doi.org/10.1016/j.jsv.2017.03.001.

Ernst, O. G., B. Sprungk, and H.-J. Starkloff. 2015. "Analysis of the ensemble and polynomial chaos Kalman filters in Bayesian inverse problems." *SIAM/ASA JUQ.* 3 (1): 823–851. https://doi.org/10.1137/140981319.

Filippou, F. C., E. P. Popov, and V. V. Bertero. 1983. *Effects of bond deterioration on hysteretic behavior of reinforced concrete joints*. Berkeley, CA: Earthquake Engineering Research Center.

Friswell, M., and J. E. Mottershead. 2013. *Finite element model updating in structural dynamics*. Berlin: Springer.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2013. *Bayesian data analysis*. London: CRC press.

Ghanem, R., and M. Pellissetti. 2002. "Adaptive data refinement in the spectral stochastic finite element method." *Commun. Numer. Methods Eng.* 18 (2): 141–151. https://doi.org/10.1002/cnm.476.

Goller, B., and G. Schueller. 2011. "Investigation of model uncertainties in Bayesian structural model updating." *J. Sound Vib.* 330 (25): 6122–6136. https://doi.org/10.1016/j.jsv.2011.07.036.

Golub, G. H., and C. F. Van Loan. 1996. "Matrix computations." In *Johns Hopkins studies in the mathematical sciences*. Baltimore: Johns Hopkins University Press.

Haykin, S. 2004. *Kalman filtering and neural networks*. New York: Wiley.

Hoffman, J. D., and S. Frankel. 2018. *Numerical methods for engineers and scientists*. London: CRC Press.

Huang, Y., and J. L. Beck. 2015. "Hierarchical sparse Bayesian learning for structural health monitoring with incomplete modal data." *Int. J. Uncertainty Quantif.* 5 (2): 15. https://doi.org/10.1615/Int.J.UncertaintyQuantification.2015011808.

Kollo, T., and D. von Rosen. 2006. *Advanced multivariate statistics with matrices*. Berlin: Springer.

Kontoroupi, T., and A. W. Smyth. 2016. "Online noise identification for joint state and parameter estimation of nonlinear systems." *J. Risk Uncertainty Eng. Syst. A: Civ. Eng.* 2 (3): B4015006. https://doi.org/10.1061/AJRUA6.0000839.

Krishnan, V. 2015. *Probability and random processes*. New York: Wiley.

Law, K. J., and A. M. Stuart. 2012. "Evaluating data assimilation algorithms." *Mon. Weather Rev.* 140 (11): 3757–3782. https://doi.org/10.1175/MWR-D-11-00257.1.

McKenna, F., G. L. Fenves, and M. H. Scott. 2000. *Open system for earthquake engineering simulation*. Berkeley: Univ. of California.

Mehra, R. 1972. “Approaches to adaptive filtering.” *IEEE Trans. Autom. Control* 17 (5): 693–698. https://doi.org/10.1109/TAC.1972.1100100.
Moore, J. B., and B. Anderson. 1979. *Optimal filtering*. New York: Prentice-Hall.
Murphy, K. P. 2007. *Conjugate Bayesian analysis of the Gaussian distribution*. Technical Rep. Vancouver, BC, Canada: Univ. of British Columbia.
Nabiyan, M. S., F. Khoshnoudian, B. Moaveni, and H. Ebrahimian. 2020. “Mechanics-based model updating for identification and virtual sensing of an offshore wind turbine using sparse measurements.” *Struct. Control Health Monit.* 28 (2): e2647. https://doi.org/10.1002/stc.2647.
O’Hagan, A., and J. J. Forster. 2004. *Kendall’s advanced theory of statistics, volume 2B: Bayesian inference*. London: Edward Arnold.
Ritter, B. 2020. *Nonlinear state estimation and noise adaptive Kalman filter design for wind turbines*. Berlin: Epubli.
Sanayei, M., B. Arya, E. M. Santini, and S. Wadia-Fascetti. 2001. “Significance of modeling error in structural parameter estimation.” *Comput.-Aided Civ. Inf.* 16 (1): 12–27. https://doi.org/10.1111/0885-9507.00210.
Sarkka, S., and J. Hartikainen. 2013. “Non-linear noise adaptive Kalman filtering via variational Bayes.” In *Proc., 2013 IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP)*. New York: IEEE.
Sarkka, S., and A. Nummenmaa. 2009. “Recursive noise adaptive Kalman filtering by variational Bayesian approximations.” *IEEE Trans. Autom. Control* 54 (3): 596–600. https://doi.org/10.1109/TAC.2008.2008348.
Schum, D. A. 2001. *The evidential foundations of probabilistic reasoning*. Evanston, IL: Northwestern University Press.
Simon, D. 2006. *Optimal state estimation: Kalman, H infinity, and nonlinear approaches*. New York: Wiley.
Soize, C. 2005. “A comprehensive overview of a non-parametric probabilistic approach of model uncertainties for predictive models in structural dynamics.” *J. Sound Vib.* 288 (3): 623–652. https://doi.org/10.1016/j.jsv.2005.07.009.
Soize, C. 2017. *Uncertainty quantification*. Berlin: Springer.
Soize, C., and R. Ghanem. 2004. “Physical systems with random uncertainties: Chaos representations with arbitrary probability measure.” *SIAM J. Sci. Comput.* 26 (2): 395–410. https://doi.org/10.1137/S1064827503424505.
Song, M., R. Astroza, H. Ebrahimian, B. Moaveni, and C. Papadimitriou. 2020. “Adaptive Kalman filters for nonlinear finite element model updating.” *Mech. Syst. Signal Process.* 143 (Sep): 106837. https://doi.org/10.1016/j.ymssp.2020.106837.
Wu, J. 2005. “Some properties of the normal distribution.” Accessed February 12, 2020. http://www.cc.gatech.edu/~wujx/paper/Gaussian.pdf.
Wu, M., and A. W. Smyth. 2007. “Application of the unscented Kalman filter for real-time nonlinear structural system identification.” *Struct. Control Health Monit.* 14 (7): 971–990. https://doi.org/10.1002/stc.186.
Xu, D., Z. Wu, and Y. Huang. 2019. “A new adaptive Kalman filter with inaccurate noise statistics.” *Circuits Syst. Signal Process.* 38 (9): 4380–4404. https://doi.org/10.1007/s00034-019-01053-w.
Yuen, K.-V., and S.-C. Kuok. 2016. “Online updating and uncertainty quantification using nonstationary output-only measurement.” *Mech. Syst. Signal Process.* 66 (Jan): 62–77. https://doi.org/10.1016/j.ymssp.2015.05.019.

J. Eng. Mech., 2022, 148(3): 04021165