Operational Modal Analysis using Variational Bayes

Binbin Li
PhD Student, Dept. of Civil and Environmental Engineering, University of California, Berkeley, USA

Armen Der Kiureghian
Taisei Professor, Dept. of Civil and Environmental Engineering, University of California, Berkeley, USA

ABSTRACT: Operational modal analysis is the primary tool for modal parameters identification in civil engineering. Bayesian statistics offers an ideal framework for analyzing uncertainties associated with the identified modal parameters. However, the exact Bayesian analysis is usually intractable due to the high computation demanding in obtaining the posterior distributions of modal parameters. In this paper, the variational Bayes is employed to provide an approximated solution. Working with the state space representation of a dynamic system, the joint distribution of the state transition matrix and observation matrix as well as the joint distribution of the process noise and measurement error are firstly obtained analytically using conjugate priors, then the distributions of modal parameters are extracted from these obtained joint distributions based on sampling because no closed form solution exists. A numerical simulation example demonstrates the performance of the proposed approach. The variational Bayes yields a consistent estimation of modal parameters although the variability is slightly under-estimated. Moreover, the variational Bayes is more flexible than the Laplace approximation and much more efficient than Monte Carlo sampling.

1. INTRODUCTION

Modal parameters (i.e. modal frequencies, damping ratios and mode shapes) are the characteristic properties of linear structural model. They are directly related to the mass and stiffness distribution, which are influenced by changes in support and continuity conditions as well as material properties. The identification of modal parameters has become a standard tool for model updating, damage detection and condition assessment.

Controlled dynamic test usually yields accurate estimation of modal parameters, but it is not often applicable to civil structures due to their huge size, high load capacity and noisy operational conditions. Therefore, the operational modal analysis (OMA), which utilizes only the stochastic dynamic response, becomes the primary modal testing method in civil engineering. The basic assumption in OMA is that sources of excitations are broad-band stochastic processes adequately modeled by band-limited white noise. Since the main source of random vibration in civil structures comes from wind, traffic, ground tremor and low magnitude of earthquake, this assumption could be generally satisfied.

Modal identification has been a subject of extensive research, and a recent review can be found in Reynders (2012). Since the signal-to-noise ratio cannot be directly controlled and various uncertainties impact on different stages of OMA, the uncertainty associated with the identified modal parameters should be a primary concern. The impact of uncertainties on operational modal identification has been seriously studied, e.g. Ciloglu, et al. (2006), and they showed that both the aleatory and epistemic uncertainties have significant influence on OMA. Realizing this, many researchers began to improve some original deterministic algorithms to provide the uncertainty information using perturbation method, e.g. Reynders, et al. (2008), Lam and Mevel (2011). These perturbation-based methods are limited to estimating variance but the variance itself is insufficient to determine confidence intervals if the modal parameters exhibit non-normal distributions. In addition, Bayesian approach is also applied to determine the uncertainty in identified modal parameters. Both Yuen and Katafygiotis (2001) and Au (2013) applied the Laplace approximation to identify modal parameters from the time domain and frequency domain, respectively.

In this paper, a more flexible approximation called variational Bayes (VB) is employed for the purpose of modal identification. The VB is a deterministic approach to analytically approximate the posterior distribution of parameters, and the fundamental principle is to find a surrogate distribution close to the true posterior while making inference tractable. In statistics, Beal (2003) applied the VB framework to identify of state space model (SSM), in which, however, the prior setting does not respect the invariance under coordinate transformation and is not allowed online identification. In control-related area, Fujimoto, et al. (2011) introduced a conjugate prior for the full covariance matrix to fix the aforementioned drawbacks for the input-output system identification. However, in all previous formulations, the state transition matrix and observation matrix are assumed to be statistically independent as well as the process and measurement noise. Since the SSM is invariant under coordinate transformation, these assumptions seems inappropriate. In addition, the VB works in an iterative way and the prior is also updated in each iteration in previous setting, which can incur the under-estimation of variability. Therefore, focusing on output-only modal identification, we formulate a new VB algorithm to include the aforementioned statistical correlations by introducing the matrix variate distribution, and modify the updating procedure to alleviate the under-estimation of variability.

2. VARIATIONAL BAYES

This section is intended to provide a broad overview of the variational Bayes algorithm in a general setting. The fundamental idea of VB is to approximate the posterior distribution by a tractable surrogate distribution, and it is suitable to infer the latent variable model. To be more precise, let's see a simple linear model,

$$\boldsymbol{Y} = \boldsymbol{\Theta X} + \boldsymbol{\epsilon} \tag{1}$$

in which $\boldsymbol{Y} \in \mathbb{R}^m$ is the observed variable, $\boldsymbol{X} \in \mathbb{R}^n$ is the latent variable, $\boldsymbol{\epsilon} \in \mathbb{R}^m$ represents the measurement error, $\boldsymbol{\Theta} \in \mathbb{R}^{m \times n}$ is the model parameter, which is a random matrix. The above model can also be represented as a probabilistic graphic model (PGM) shown Figure 1, in which each node is associated with a random variable and directed link refers to a conditional dependence.

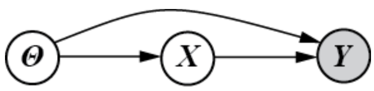

Figure 1: Probabilistic Graphical Model.

The problem is to infer the posterior distribution $p_{\boldsymbol{\Theta}|\boldsymbol{Y}}(\boldsymbol{\theta}|\boldsymbol{y})$ given the data $\boldsymbol{y}$ and the prior distribution $\pi_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x})$. Since $\boldsymbol{\Theta}$ and $\boldsymbol{X}$ are coupled in this model, there are rare cases that can be solved analytically. Instead, we try to find a surrogate distribution $q^*_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x})$ from a predetermined family $\mathbb{Q}$ to minimize the distance from the true posterior distribution $p_{\boldsymbol{\Theta X}|\boldsymbol{Y}}(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})$ in the measure of Kullback-Leibler (KL) divergence. Formally,

$$q^*_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x}) = \min_{q_{\boldsymbol{\Theta X}}} \mathrm{KL}\,[q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x}) || p_{\boldsymbol{\Theta X}|\boldsymbol{Y}}(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})]$$
$$= \min_{q_{\boldsymbol{\Theta X}}} \iint q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x}) \ln \frac{q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x})}{p_{\boldsymbol{\Theta X}|\boldsymbol{Y}}(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})} d\boldsymbol{\theta} d\boldsymbol{x} \tag{2}$$

Usually, it's hard to evaluate the KL divergence in Eq. (2), but we can find an alternative way to minimize it without direct evaluation. The log marginal likelihood of the observed variable $\boldsymbol{y}$ gives

$$\ln p_{\boldsymbol{Y}}(\boldsymbol{y}) = \iint q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x}) \ln p_{\boldsymbol{Y}}(\boldsymbol{y})\, d\boldsymbol{\theta} d\boldsymbol{x}$$
$$= \iint q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta}, \boldsymbol{x}) \ln \frac{p_{\boldsymbol{\Theta X Y}}(\boldsymbol{\theta}, \boldsymbol{x}, \boldsymbol{y})}{p_{\boldsymbol{\Theta X}|\boldsymbol{Y}}(\boldsymbol{\theta}, \boldsymbol{x}|\boldsymbol{y})} d\boldsymbol{\theta} d\boldsymbol{x}$$

$$= \iint q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x}) \ln \frac{q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x})}{p_{\boldsymbol{\Theta X}|\boldsymbol{Y}}(\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y})} d\boldsymbol{\theta} d\boldsymbol{x} + \iint q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x}) \ln \frac{p_{\boldsymbol{Y}|\boldsymbol{\Theta X}}(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\pi_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x})}{q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x})} d\boldsymbol{\theta} d\boldsymbol{x}$$
$$\triangleq \mathrm{KL}[q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x})||p_{\boldsymbol{\Theta X}|\boldsymbol{Y}}(\boldsymbol{\theta},\boldsymbol{x}|\boldsymbol{y})] + \mathrm{F}[q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x})] \quad (3)$$

Since the log marginal likelihood keeps constant given the model, minimizing the KL divergence is equivalent to maximizing the free energy $\mathrm{F}[q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x})]$. On the other hand, the free energy works as a lower bound of the log marginal likelihood since the KL divergence is always non-negative; therefore, we can regard the variational Bayes as maximizing the marginal likelihood by increasing its lower bound. Because $\boldsymbol{\Theta}$ and $\boldsymbol{X}$ are still coupled in Eq. (2), mean field approximation are further introduced to make an independence assumption between $\boldsymbol{\Theta}$ and $\boldsymbol{X}$, i.e. $q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x}) = q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})q_{\boldsymbol{X}}(\boldsymbol{x})$. Although this assumption may seem abrupt, one may think of it as replacing stochastic dependence between $\boldsymbol{\Theta}$ and $\boldsymbol{X}$ by deterministic dependencies between relevant moments of two sets of variables. If we further specify independent priors for $\boldsymbol{\Theta}$ and $\boldsymbol{X}$, Eq. (3) can be rewritten as

$$\mathrm{F}[q_{\boldsymbol{\Theta X}}(\boldsymbol{\theta},\boldsymbol{x})] \approx \mathrm{F}[q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), q_{\boldsymbol{X}}(\boldsymbol{x})]$$
$$= \iint q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})q_{\boldsymbol{X}}(\boldsymbol{x}) \ln p_{\boldsymbol{Y}|\boldsymbol{\Theta X}}(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\, d\boldsymbol{\theta} d\boldsymbol{x} - \int q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) \ln \frac{q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}{\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta})} d\boldsymbol{\theta} - \int q_{\boldsymbol{X}}(\boldsymbol{x}) \ln \frac{q_{\boldsymbol{X}}(\boldsymbol{x})}{\pi_{\boldsymbol{X}}(\boldsymbol{x})} d\boldsymbol{x}$$
$$= \mathrm{E}_{\boldsymbol{\Theta X}}[\ln p_{\boldsymbol{Y}|\boldsymbol{\Theta X}}(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})] - \mathrm{KL}[q_{\boldsymbol{\Theta}}(\boldsymbol{\theta})||\pi_{\boldsymbol{\Theta}}(\boldsymbol{\theta})] - \mathrm{KL}[q_{\boldsymbol{X}}(\boldsymbol{x})||\pi_{\boldsymbol{X}}(\boldsymbol{x})] \quad (4)$$

where $\mathrm{E}_{\boldsymbol{\Theta X}}[\cdot]$ means the expectation operator with respect to the joint distribution of $\boldsymbol{\Theta}$ and $\boldsymbol{X}$. Then in order to maximizing the free energy $\mathrm{F}[q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), q_{\boldsymbol{X}}(\boldsymbol{x})]$, we have to reduce the KL divergences between the approximating distributions and prior distributions and simultaneously increase the expected log conditional likelihood, which can be done using the following theorem, and the updating procedures are illustrated in Figure 2.

Theorem 1. Variational Bayesian EMP (VB-EMP)
Let $\mathcal{M}$ be a statistical model with parameter $\boldsymbol{\Theta}$ giving rise to an observed data set $\boldsymbol{y}$ with corresponding latent variable $\boldsymbol{X}$. The free energy $\mathrm{F}[q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), q_{\boldsymbol{X}}(\boldsymbol{x})]$ given in Eqn. (4) can be iteratively optimized by performing the following updates, using superscript t to denote iteration step:

VB Expectation step (VBE): $q_{\boldsymbol{X}}^{(t+1)}(\boldsymbol{x}) = Z_{\boldsymbol{X}}^{-1}\pi_{\boldsymbol{X}}^{(t)}(\boldsymbol{x}) \exp[\int q_{\boldsymbol{\Theta}}^{(t)}(\boldsymbol{\theta}) \ln p_{\boldsymbol{Y}|\boldsymbol{\Theta X}}(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\, d\boldsymbol{\theta}]$

VB Maximization step (VBM):
$$q_{\boldsymbol{\Theta}}^{(t+1)}(\boldsymbol{\theta}) = Z_{\boldsymbol{\Theta}}^{-1}\pi_{\boldsymbol{\Theta}}^{(t)}(\boldsymbol{\theta}) \exp[\int q_{\boldsymbol{X}}^{(t)}(\boldsymbol{x}) \ln p_{\boldsymbol{Y}|\boldsymbol{\Theta X}}(\boldsymbol{y}|\boldsymbol{\theta},\boldsymbol{x})\, d\boldsymbol{x}]$$

VB Prior-updating step (VBP):
$\pi_{\boldsymbol{\Theta}}^{(t+1)}(\boldsymbol{\theta}) = q_{\boldsymbol{\Theta}}^{(t+1)}(\boldsymbol{\theta})$; $\pi_{\boldsymbol{X}}^{(t)}(\boldsymbol{x}) = q_{\boldsymbol{X}}^{(t+1)}(\boldsymbol{x})$.
where $Z_{\boldsymbol{X}}$ and $Z_{\boldsymbol{\Theta}}$ are normalizing constants. Moreover, each step will increase $\mathrm{F}[q_{\boldsymbol{\Theta}}(\boldsymbol{\theta}), q_{\boldsymbol{X}}(\boldsymbol{x})]$ monotonically, and converge to a local maximum.

Due the limited length of the paper, all proofs are neglected. The preceding theorem shows that the VB-EMP algorithm is guaranteed to converge to local maximum monotonically, and in particular, for the exponential family, Wang & Titterington (2004) have proved that it is consistent, i.e. converging to the true value in asymptotic sense, whenever the starting value is sufficiently near the true. It also can be seen that the priors are updated in each iteration, so that it becomes narrower and narrower, resulting in the under-estimation of variability.

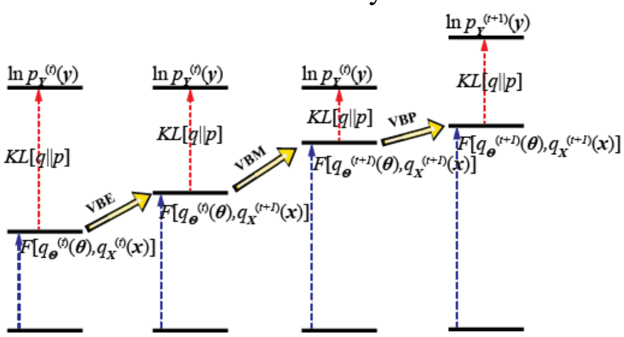

Figure 2: Variational Bayes EMP.

3. STOCHASTIC SSM

The equation of motion of a discrete, linear, and time-invariant dynamical system with N_d degrees of freedom (DOFs) under stochastic external force and ground motion is described by

$$\boldsymbol{M}\ddot{\boldsymbol{u}}(t)+\boldsymbol{C}_d\dot{\boldsymbol{u}}(t)+\boldsymbol{K}\boldsymbol{u}(t)=\boldsymbol{P}\boldsymbol{f}(t)-\boldsymbol{M}\boldsymbol{l}\ddot{\boldsymbol{u}}_g(t);$$
$$\boldsymbol{u}(0)=\boldsymbol{u}_0,\ \dot{\boldsymbol{u}}(0)=\dot{\boldsymbol{u}}_0 \quad (5)$$

where $\boldsymbol{M}\in\mathbb{R}^{N_d\times N_d}$, $\boldsymbol{C}_d\in\mathbb{R}^{N_d\times N_d}$ and $\boldsymbol{K}\in\mathbb{R}^{N_d\times N_d}$ are the mass, damping and stiffness matrices, respectively; $\boldsymbol{u}(t)\in\mathbb{R}^{N_d}$, $\dot{\boldsymbol{u}}(t)\in\mathbb{R}^{N_d}$ and $\ddot{\boldsymbol{u}}(t)\in\mathbb{R}^{N_d}$ are the nodal displacement, velocity and acceleration responses relative to the ground motion; $\boldsymbol{f}(t)\in\mathbb{R}^{N_f}$ is the external force vector; $\boldsymbol{P}\in\mathbb{R}^{N_d\times N_f}$ is the load coefficient matrix representing the spatial influence of the external force; $\ddot{\boldsymbol{u}}_g(t)\in\mathbb{R}^{N_g}$ is the ground acceleration vector, and $\boldsymbol{l}\in\mathbb{R}^{N_d\times N_g}$ is the corresponding influence matrix. $\boldsymbol{u}_0\in\mathbb{R}^{N_d}$ and $\dot{\boldsymbol{u}}_0\in\mathbb{R}^{N_d}$ are the initial relative displacement and velocity vectors. Note that we consider both of external force and ground motion to represent the general source of random excitation. For system identification, it's convenient to convert Eq. (5) into a state-space model. After discretization in time, the state space equation of motion takes the form (Reynders, 2012),

$$\begin{bmatrix}\boldsymbol{X}_{k+1}\\ \boldsymbol{Y}_k\end{bmatrix}=\begin{bmatrix}\boldsymbol{A}\\ \boldsymbol{C}\end{bmatrix}\boldsymbol{X}_k+\begin{bmatrix}\boldsymbol{w}_k\\ \boldsymbol{v}_k\end{bmatrix} \quad (6)$$

where $\boldsymbol{X}_k=\left[\boldsymbol{u}_k^{\mathrm{T}}\ \dot{\boldsymbol{u}}_k^{\mathrm{T}}\right]^{\mathrm{T}}\in\mathbb{R}^{N_s}$ $(N_s=2N_d)$ is the state vector, $\boldsymbol{Y}_k\in\mathbb{R}^{N_o}$ is the recorded response at time kT_s for $k=1,2,\dots,N$, T_s is the sampling period. $\boldsymbol{w}_k$ represents the joint effect of model error and unknown input and $\boldsymbol{v}_k$ stands for the joint effect of measurement error and input uncertainty, and they can be modeled as zero-mean Gaussian band-limited white noise with unknown covariance $\boldsymbol{\Sigma}=[\boldsymbol{Q},\boldsymbol{S};\boldsymbol{S}^{\mathrm{T}},\boldsymbol{R}]$ .More specifically, the state transition matrix $\boldsymbol{A}\in\mathbb{R}^{N_s\times N_s}$ and the observation matrix $\boldsymbol{C}\in\mathbb{R}^{N_o\times N_s}$ have the following form if only accelerations are recorded:

$$\boldsymbol{A}=\exp\left(T_s\begin{bmatrix}\boldsymbol{0}_{N_d} & \boldsymbol{I}_{N_d}\\ -\boldsymbol{M}^{-1}\boldsymbol{K} & -\boldsymbol{M}^{-1}\boldsymbol{C}_d\end{bmatrix}\right)$$
$$\boldsymbol{C}=\boldsymbol{S}_0[-\boldsymbol{M}^{-1}\boldsymbol{K}\quad -\boldsymbol{M}^{-1}\boldsymbol{C}_d] \quad (7)$$

in which $\boldsymbol{C}\in\mathbb{R}^{N_o\times N_d}$ is a selection matrix to represent we can only measure $N_o(\le N_s)$ DOFs of the structure.

It is well-known that the state transition matrix $\boldsymbol{A}$ and observation matrix $\boldsymbol{C}$ are only unique up to similarity transformation, i.e. for one input-output pair or output only, there are possibly infinite pairs of $\boldsymbol{A}$'s and $\boldsymbol{C}$'s satisfying Eq. (6). However, the modal parameters are invariant under similarity transformation, therefore they could be uniquely determined from the measured data.The modal parameters can be extracted from $\boldsymbol{A}$ and $\boldsymbol{C}$ using the following steps:

(1) Take the eigenvalue decomposition: $\boldsymbol{A}=\boldsymbol{\Psi}\boldsymbol{\Lambda}\boldsymbol{\Psi}^{-1}$;

(2) Partition the eigenvalue and eigenvector matrices: $\boldsymbol{\Lambda}=\begin{bmatrix}\boldsymbol{\lambda} & \boldsymbol{0}\\ \boldsymbol{0} & \bar{\boldsymbol{\lambda}}\end{bmatrix}$, $\boldsymbol{\Psi}=[\boldsymbol{\psi}\quad \bar{\boldsymbol{\psi}}]$

(3) Calculate modal parameters: $\omega_i=|\ln\lambda_i|/T_s$, $\xi_i=Re(\ln\lambda_i)/|\ln\lambda_i|$, $\boldsymbol{\phi}_i=\boldsymbol{C}\boldsymbol{\psi}_i$

where ω_i , ξ_i and $\boldsymbol{\phi}_i$ are the i th modal frequencies, damping ratios and mode shapes; λ_i and $\boldsymbol{\psi}_i$ are the i th diagonal element and i th column of $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$; $\bar{\boldsymbol{\lambda}}$ and $\bar{\boldsymbol{\psi}}$ are the complex conjugates of $\boldsymbol{\lambda}$ and $\boldsymbol{\psi}$, respectively.

4. APPICATION OF VB IN SSM

Since modal parameters can be extracted from SSM, the problem becomes how to identify the model parameters $\boldsymbol{\Gamma}=[\boldsymbol{A}^{\mathrm{T}}\ \boldsymbol{C}^{\mathrm{T}}]^{\mathrm{T}}$ and covariance matrix $\boldsymbol{\Sigma}$ given the observed sequence $\boldsymbol{y}_{1:N}\triangleq\{\boldsymbol{y}_1,\dots,\boldsymbol{y}_N\}$. Note that matrices $\boldsymbol{A}$ and $\boldsymbol{C}$ in Eq. (7) share some common terms, leading them to be statistically correlated; therefore, the parameter $\boldsymbol{\Gamma}$ is considered as a matrix random variable in our formulation. If the covariance matrix $\boldsymbol{\Sigma}$ is also a matrix random variable, we can use the PGM shown in Figure 3 to represent the stochastic SSM in Eq. (6). As is the usual case in latent variable model, the parameter $\boldsymbol{\Theta}\triangleq\{\boldsymbol{\Gamma},\boldsymbol{\Sigma}\}$ and latent variable $\boldsymbol{X}_{1:N+1}\triangleq\{\boldsymbol{X}_1,\dots,\boldsymbol{X}_{N+1}\}$ are coupled, leading the exact Bayesian inference analytically intractable. Now, we apply the variational Bayes framework to crack this difficulty, i.e. to find the factorized distributions $q_{\boldsymbol{\Gamma}\boldsymbol{\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma})q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})$ to approximate the true posterior $p_{\boldsymbol{\Gamma}\boldsymbol{\Sigma}\boldsymbol{X}_{1:N+1}|\boldsymbol{Y}_{1:N}}(\boldsymbol{\gamma},\boldsymbol{\sigma},\boldsymbol{x}_{1:N+1}|\boldsymbol{y}_{1:N})$.

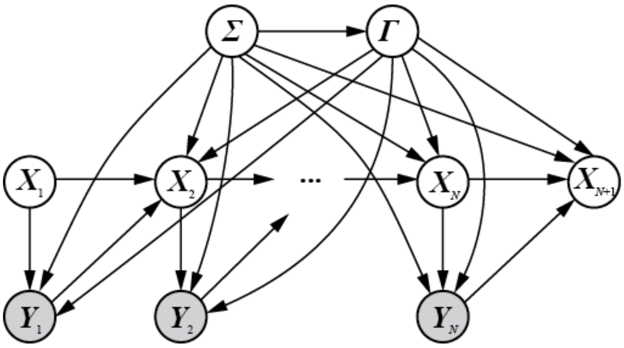

Figure 3: PGM representation for stochastic SSM

Before applying the theorem 1, let's define conjugate priors for parameters $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$. Conjugate priors are commonly used in Bayesian statistics in order to get a closed-form solution for posteriors. For the PGM in Figure 3, we assume a matrix normal (MN), inverse Wishart (IW) distribution, i.e.

$$\begin{aligned}\pi_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma}) &= \pi_{\boldsymbol{\Gamma}|\boldsymbol{\Sigma}}(\boldsymbol{\gamma}|\boldsymbol{\sigma})\pi_{\boldsymbol{\Sigma}}(\boldsymbol{\sigma}) \\ &= MN(\boldsymbol{\mu_\Gamma},\boldsymbol{\Sigma},\boldsymbol{\Pi})IW(d,\boldsymbol{D}) \qquad (8)\end{aligned}$$

where $MN(\boldsymbol{\mu_\Gamma},\boldsymbol{\Sigma},\boldsymbol{\Pi})$ is the probability density function (PDF) of matrix normal distribution with mean matrix $\boldsymbol{\mu_\Gamma} \in \mathbb{R}^{(N_s+N_s)\times N_s}$, left and right covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{(N_s+N_o)\times(N_s+N_o)}$ and $\boldsymbol{\Pi} \in \mathbb{R}^{N_s\times N_s}$, respectively. $IW(d,\boldsymbol{D})$ means the PDF of inverse Wishart distribution with d DOFs and scale matrix $\boldsymbol{D}$. For the distribution of latent variables $\boldsymbol{X}_{1:N+1}$, we only need to specify a prior for $\boldsymbol{X}_1$, then this information will propagate to $\boldsymbol{X}_{2:N+1}$ automatically through the chain structure. Then, we specify a multivariable normal distribution with unknown mean $\boldsymbol{\mu}_1$ and covariance $\boldsymbol{P}_1$, i.e. $p_{\boldsymbol{X}_1} = N(\boldsymbol{\mu}_1,\boldsymbol{P}_1)$. As stated in Section 2, the uncertainty would be underestimated if the priors are updated in each iteration, so that we choose not to update the prior of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ but only the prior of $\boldsymbol{X}_1$. Similar to the Theorem 1, we can apply the basic procedure of variational Bayes to the more complicated stochastic SSM, and this is achieved by the following theorem.

Theorem 2. Considering the PGM shown in Figure 3 and prior distributions of parameters $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ in Eq. (8), and denote $\boldsymbol{Z}_k = \left[\boldsymbol{X}_{k+1}^{\mathrm{T}}\ \boldsymbol{Y}_k^{\mathrm{T}}\right]^{\mathrm{T}}$, the following statements hold:

(1) The free energy function $\mathrm{F}[q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma}), q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})]$ is given by

$$\begin{aligned}&\mathrm{F}\left[q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma}), q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})\right] \\ &= -\iint q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma}) \ln\frac{q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma})}{\pi_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma})} d\boldsymbol{\gamma} d\boldsymbol{\sigma} \\ &- \int q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1}) \ln\frac{q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})}{\pi_{\boldsymbol{X}_1}(\boldsymbol{x}_1)} d\boldsymbol{x}_{1:N+1} \\ &+ \iiint q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma}) q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1}) \\ &\sum_{k=1}^{N} \ln p_{\boldsymbol{Z}_k|\boldsymbol{\Gamma\Sigma X}_k}(z_k|\boldsymbol{\gamma},\boldsymbol{\sigma},\boldsymbol{x}_{1:N+1})\, d\boldsymbol{\gamma} d\boldsymbol{\sigma} d\boldsymbol{x}_{1:N+1}\end{aligned}$$

(2) $\mathrm{F}[q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma}), q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})]$ can be iteratively optimized by performing the following updates, using superscript t to denote iteration step:

VBE step:

$$\begin{aligned}&q_{\boldsymbol{X}_{1:N+1}}^{(t+1)}(\boldsymbol{x}_{1:N+1}) \\ &\propto \exp\Big\{-\tfrac{1}{2}\Big[\Big(\boldsymbol{x}_1-\boldsymbol{\mu}_1^{(t)}\Big)^{\mathrm{T}} \boldsymbol{P}_1^{(t)}\Big(\boldsymbol{x}_1-\boldsymbol{\mu}_1^{(t)}\Big) + \\ &d^{(t)}\sum_{k=1}^{N}\Big(\boldsymbol{z}_k-\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(t)}\boldsymbol{x}_k\Big)^{\mathrm{T}} \boldsymbol{D}^{(t)^{-1}}\Big(\boldsymbol{z}_k-\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(t)}\boldsymbol{x}_k\Big) + \\ &(N_s+N_o)\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{\Pi}^{(t)}\boldsymbol{x}_k\Big]\Big\}\end{aligned}$$

VBM step:

$$\begin{aligned}&q_{\boldsymbol{\Gamma\Sigma}|\boldsymbol{Y}_{1:N}}^{(t+1)}(\boldsymbol{\gamma},\boldsymbol{\sigma}|\boldsymbol{y}_{1:N}) \\ &= \mathrm{MN}\big(\boldsymbol{\mu}^{(t+1)},\boldsymbol{\Sigma},\boldsymbol{\Pi}^{(t+1)}\big)\mathrm{IW}\big(d^{(t+1)},\boldsymbol{D}^{(t+1)}\big)\end{aligned}$$

with

$$\begin{aligned}&\boldsymbol{\Pi}^{(t+1)} = \left[\boldsymbol{\Pi}^{-1} + \sum_{\boldsymbol{k=1}}^{\boldsymbol{N}} \mathrm{E}_{\boldsymbol{X}_k|\boldsymbol{Y}_{1:N}}\big(\boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{T}}\big)\right]^{-1}, \\ &\boldsymbol{\mu_\Gamma}^{(t+1)} = \Big[\sum_{\boldsymbol{k=1}}^{\boldsymbol{N}} \mathrm{E}_{\boldsymbol{X}_{k:k+1}|\boldsymbol{Y}_{1:N}}\big(\boldsymbol{z}_k\boldsymbol{x}_k^{\mathrm{T}}\big) + \\ &\boldsymbol{\mu_\Gamma}\boldsymbol{\Pi}^{-1}\Big]\boldsymbol{\Pi}^{(t+1)},\ d^{(t+1)} = d+N, \\ &\boldsymbol{D}^{(t+1)} = \boldsymbol{D} + \sum_{\boldsymbol{k=1}}^{\boldsymbol{N}} \mathrm{E}_{\boldsymbol{X}_{k+1}|\boldsymbol{Y}_{1:N}}\big(\boldsymbol{z}_k\boldsymbol{x}_k^{\mathrm{T}}\big) + \\ &\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(t)}\boldsymbol{\Pi}^{(t)^{-1}}\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(t)^{\mathrm{T}}} - \boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(t+1)}\boldsymbol{\Pi}^{(t+1)^{-1}}\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(t+1)^{\mathrm{T}}}.\end{aligned}$$

VBH step: $\boldsymbol{\mu}_1^{(t+1)} = \mathrm{E}_{\boldsymbol{X}_1|\boldsymbol{Y}_{1:N}}(\boldsymbol{x}_1)$, $\boldsymbol{P}_1^{(t+1)} = \mathrm{E}_{\boldsymbol{X}_1|\boldsymbol{Y}_{1:N}}[(\boldsymbol{x}_1-\widehat{\boldsymbol{\mu}}_1^{(t+1)})(\boldsymbol{x}_1-\widehat{\boldsymbol{\mu}}_1^{(t+1)})^{\mathrm{T}}]$.

Initialized by setting $\boldsymbol{\mu}_1^{(1)} = \boldsymbol{\mu}_1$, $\boldsymbol{P}_1^{(1)} = \boldsymbol{P}_1$, $\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(1)} = \boldsymbol{\mu_\Gamma}$, $\boldsymbol{\Pi}^{(1)} = \boldsymbol{\Pi}$, $d^{(1)} = d$ and $\boldsymbol{D}^{(1)} = \boldsymbol{D}$.

Moreover, each step will increase $\mathrm{F}[q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma},\boldsymbol{\sigma}), q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})]$ monotonically, and converge to a local maximum.

Theorem 2 provides the foundation to iteratively get the posterior distribution of parameters and latent variable. However, the expectation quantities in the VBM step have not been evaluated explicitly so far. Due to the special structure of $q_{\boldsymbol{X}_{1:N+1}}(\boldsymbol{x}_{1:N+1})$, the standard Kalman smoother is not directly applicable, and some transformations are needed in advance. The basic steps are as follows:

(1) Make the following definitions and the Cholesky decomposition

$\boldsymbol{\mu}_{\boldsymbol{\Gamma}}^{(t)} = \begin{bmatrix} \overline{\boldsymbol{A}} \\ \overline{\boldsymbol{C}} \end{bmatrix}, \frac{\boldsymbol{D}^{(t)}}{d^{(t)} - N_s - N_o - 1} = \begin{bmatrix} \overline{\boldsymbol{Q}} & \overline{\boldsymbol{S}} \\ \overline{\boldsymbol{S}}^{\mathrm{T}} & \overline{\boldsymbol{R}} \end{bmatrix},$

$(N_s + N_o)\boldsymbol{\Pi}^{(t)} = \boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}$

(2) Construct new parameters for SSM:

$\widetilde{\boldsymbol{y}}_k = \begin{bmatrix} \boldsymbol{y}_k \\ \boldsymbol{0}_{N_s} \end{bmatrix}, \ \widetilde{\boldsymbol{C}} = \begin{bmatrix} \overline{\boldsymbol{C}} \\ \boldsymbol{U} \end{bmatrix}, \ \widetilde{\boldsymbol{R}} = \begin{bmatrix} \overline{\boldsymbol{R}} & \boldsymbol{0}_{N_d \times N_s} \\ \boldsymbol{0}_{N_s \times N_d} & \boldsymbol{I}_{N_s} \end{bmatrix},$

$\widetilde{\boldsymbol{A}} = \overline{\boldsymbol{A}} - \overline{\boldsymbol{S}}\overline{\boldsymbol{R}}^{-1}\overline{\boldsymbol{C}}, \widetilde{\boldsymbol{Q}} = \overline{\boldsymbol{Q}} - \overline{\boldsymbol{S}}\overline{\boldsymbol{R}}^{-1}\overline{\boldsymbol{S}}^{\boldsymbol{T}}$

(3) Based on these new parameters, the square-root filtering algorithm (Gibson & Ninness, 2005) is applied to calculate expectation terms.

Although the posterior distribution $q_{\boldsymbol{\Gamma\Sigma}}(\boldsymbol{\gamma}, \boldsymbol{\sigma})$ can be obtained exactly using Theorem 2, there is no closed form solution for the distribution of modal parameters because the eigenvalue decomposition is involved from $\boldsymbol{\Gamma}$ to modal parameters. Therefore, the Monte Carlo sampling is applied to approximate the distributions of modal parameters. This step is straightforward because it is easy to sample the posterior distribution of $\boldsymbol{\Gamma}$ which can be proved to be a matrix variate-t distribution.

5. EMPIRICAL STUDY

In order to demonstrate the performance of the proposed algorithm, the modal identification of a simulated 8-DOF mass-spring system (Figure 4) is provided in this section. The model parameters are as follows: $m = 1$, $k_i = 800i$ for $i = 1, \dots, 9$ and Rayleigh damping matrix $\boldsymbol{C}_d = 0.69\boldsymbol{M} + 1.743 \times 10^{-4}\boldsymbol{K}$. The same input, modeled by band-limited white noise, is applied to all DOFs. The acceleration responses at DOFs 2, 4, 6 and 8 are recorded with sampling frequency 25 Hz for 200 seconds. Then, these responses are contaminated by independent Gaussian white noise with variance equal to 25 % of the largest acceleration variance. Besides the VB algorithm, the Expectation-Maximization algorithm (Gibson & Ninness, 2005) and the Gibbs sampling (Shumway & Stoffer, 2011) are also implemented for comparison, and all are initialized by the same parameters identified from stochastic subspace identification.

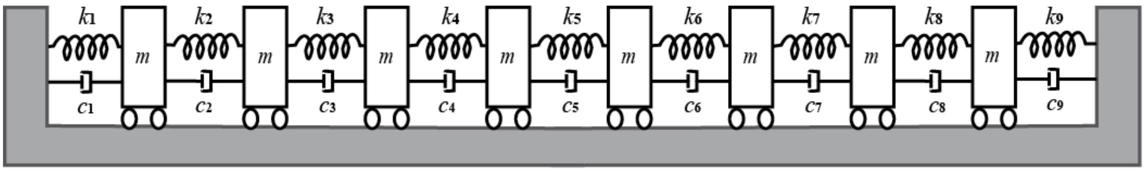

Figure 4: 8-DOF mass-spring system.

Table 1 Identified Modal Parameters using EM, VB and Gibbs Sampler

Mode	Frequencies						Damping Ratios (%)					
	True	EM	VB		Gibbs		True	EM	VB		Gibbs	
			Mean	COV(%)	Mean	COV(%)			Mean	COV(%)	Mean	COV(%)
1	2.94	2.92	2.94	0.21	2.93	0.34	2.00	2.00	1.93	9.95	1.87	18.25
2	5.87	5.88	5.87	0.19	5.88	0.21	1.24	1.16	1.30	14.59	1.29	15.79
3	8.60	8.57	8.56	0.14	8.56	0.16	1.10	1.07	1.09	12.74	1.03	15.27
4	11.19	11.21	11.22	0.11	11.20	0.13	1.09	0.79	0.78	12.55	0.84	15.05
5	13.78	13.77	13.74	0.12	13.75	0.15	1.15	1.22	1.24	9.98	1.33	12.04
6	16.52	16.54	16.54	0.11	16.55	0.13	1.23	1.07	1.20	9.33	1.12	11.82
7	19.54	19.54	19.53	0.13	19.56	0.17	1.35	1.34	1.78	7.38	1.33	12.94
8	23.12	23.09	23.08	0.12	23.06	0.14	1.50	1.56	1.77	7.03	1.44	10.54

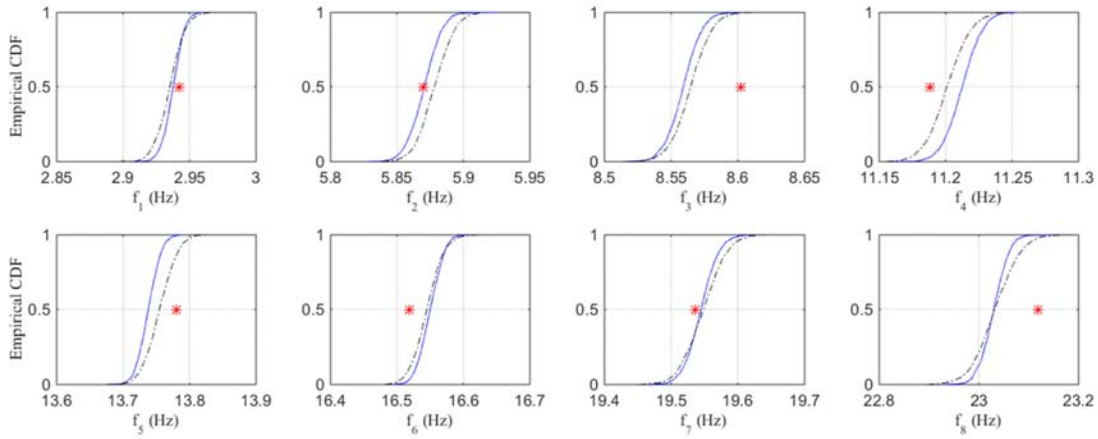

Figure 5: Empirical CDF of identified modal frequencies
Solid line: variational Bayes; Dash-dot line: Gibbs sampler; Star: true value.

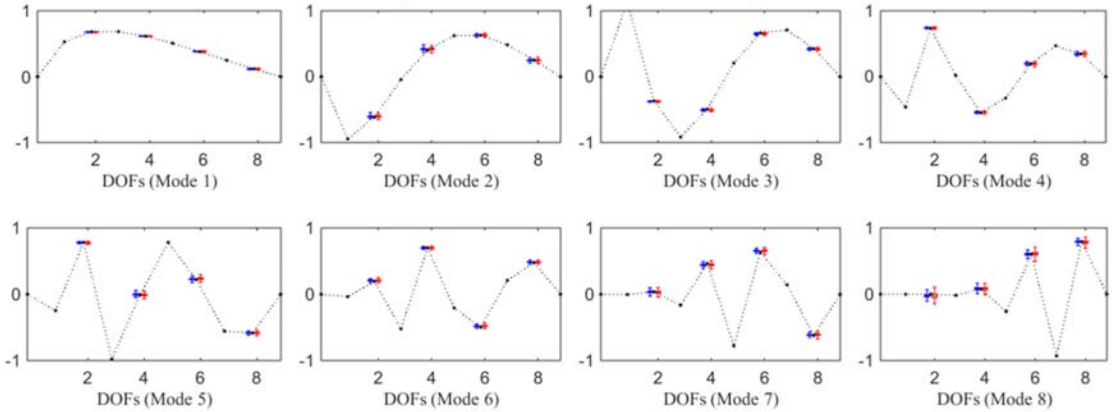

Figure 6: Boxplot of identified mode shapes
Blue: variational Bayes; Red: Gibbs sampler; Dot line: true value.

Table 1 lists the identified modal frequencies and damping ratios as well as the coefficient of variation (COV) for variational Bayes and Gibbs samplers (1800 samples). By comparing the point estimators (MLE or posterior means), all three methods are consistent, although slight bias exists. Of all the three methods, the EM seems the most accurate, while the VB and Gibbs give similar performance. The COVs of variational Bayes are always smaller than those of Gibbs simpler, about 15% less. The comparison of empirical cumulative distribution functions (CDF) of modal frequencies are shown in Figure 5. Again, we can see the distributions estimated from VB shows less variant than those estimated from Gibbs sampler, but both are relatively close in most cases. The same phenomena can be oberserved from the empirical CDF of damping ratios, thus these plots are ignored. The boxplot of identified mode shapes are illustrated in Figure 6, in which both the VB and Gibbs sample provide accurate estimation of mode shapes, while the variations in identified mode shapes by VB are still smaller than those by Gibbs sampler.

In SSM-based modal identification, another issue encountered is how to identify the physical modes from the estimated model. In this paper, we will not solve this problem thoroughly, but provide an empirical evidence that the physical modes can be identified based on the uncertainty information. For this 8-DOF mass-spring system, an order of 24 is specified in the SSM, so that there are 4 spurious modes. Figure 7 shows the COV of identified frequencies and standard deviation of damping ratios for the VB. The interesting observation is that the spurious modes corresponds to the peaks in the graph, i.e. the spurious modes have significantly larger uncertainties than the physical modes. The intuitive explanation behind this phenomenon, similar to that of stabilization diagram (Reynders, et al., 2012), is that the physical modes should be consistent in different realization while the spurious computational and noise modes have no reason to be focused.

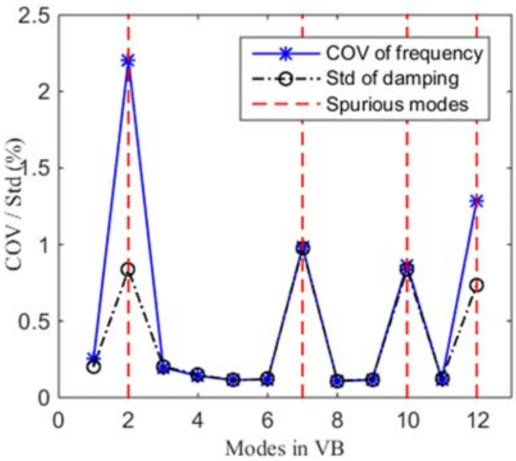

Figure 7: Spurious modes identification.

6. CONCLUSIONS

An operational modal analysis approach based on variational Bayes was introduced in this paper. Through the empirical study the OMA of an 8-DOF mass-spring system, the following conclusions can be derived:

(1) The posterior mean of VB estimator is consistent for modal parameters, and VB estimator tends to under-estimate the variability comparing with the Gibbs sampler.

(2) The variational Bayes approach works in the sense of coordinate ascent and neither gradient nor Hessian matrix are required in optimization.

(3) The uncertainty information provided by VB can be used to remove the spurious modes, and this is a good point worth further study.

(4) Of the three approaches, the VB approach is recommended, because the EM algorithm only gives a point estimation, while the computation demanding of Gibbs sampler is too high to be applied in practice.

7. REFERENCES

Au, S, Zhang, F. and Ni, Y. (2013). "Bayesian operational modal analysis: Theory, computation, practice." *Computers and Structures* 126: 3–14.

Beal, M. (2003). "Variational Algorithms for Approximate Bayesian Inference." University College London, UK, PhD Dissertation.

Ciloglu, S. (2006). “The Impact of Uncertainty in Operational Modal Analysis For Structural Identification of Constructed Systems”. Drexel University, USA, PhD disseration.

Fujimoto, K., Satoh, A. and Fukunaga S. (2011). "System Identification Based on Variational Bayes Method and The Invariance under Coordinate Transformations." *50th IEEE CDC-ECC.* Orlando, FL, USA. 3882-3888.

Gibson, S., and Ninness, B. 2005. "Robust maximum-likelihood estimation of multivariable dynamic systems." *Automatica* 4: 1667–1682.

Lam, X., and Mevel, L. (2011). "Uncertainty Quantification for Eigensystem-Realization-Algorithm, A Class of Subspace System Identification." *Proceedings of the 18th IFAC World Congress.* Milan, Italy.

Reynders, E. (2012). "System Identification Methods for (Operational) Modal Analysis: Review and Comparison." *Archives of Computational Methods in Engineering* 19: 51-124.

Reynders, E., Pintelon, R., and De Roeck, G. (2008). "Uncertainty bounds on modal parameters obtained from stochastic subspace identification." *Mechanical Systems and Signal Processing* 22: 948 – 969.

Shumway, R., and Stoffer, D. (2011). *Time Series Analysis and Its Applications: With R Examples.* 3rd. New York, USA: Springer.

Wang, B., and Titterington, D. (2004). "Convergence and asymptotic normality of variational Bayesian approximations for exponential family models with missing values." Banff, Canada: AUAI Press. 577-584.

Yuen, K, and Katafygiotis L. (2001). "Bayesian time-domain approach for modal updating using ambient data." *Probabilitic Engineering Mechanics* 16: 219-231.