Mechanical Systems and Signal Processing 29 (2012) 328–342

Contents lists available at SciVerse ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Connecting Bayesian and frequentist quantification of parameter uncertainty in system identification

Siu-Kui Au

Department of Civil and Architectural Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong

ARTICLE INFO

Article history:
Received 26 August 2011
Received in revised form
28 November 2011
Accepted 10 January 2012
Available online 7 February 2012

Keywords:
Bayesian method
Modal identification
Modeling error
Operational modal analysis
System identification

ABSTRACT

In Bayesian system identification, results are often described in terms of the 'most probable value' (MPV) of the model parameters and their posterior uncertainty in terms of the 'posterior covariance matrix'. The MPV carries the same connotation as 'the best parameter' or the 'optimal parameter' (or variants) in non-Bayesian methods. In a 'frequentist' perspective, parameter uncertainty is often described in terms of the ensemble covariance matrix of the MPVs among statistically identical experiment trials. The (Bayesian) posterior covariance of parameter and the (frequentist) covariance matrix of the MPV need not coincide, but intuition suggests that they should be consistent. This paper investigates the relationship between these two quantifications of parameter uncertainty in system identification. A second order theory is presented, which shows that in a weighted sense the ensemble average of the posterior covariance matrix is equal to the ensemble covariance matrix of MPV. Focusing on ambient modal identification, we investigate using synthetic data the accuracy of the relationships suggested based on intuition and those based on the second order theory. It is found that neither one is exact but both give a good approximation when there is little or no modeling error. The investigation is also extended to laboratory and field data, where greater discrepancy is observed. The results highlight the role of the (Bayesian) posterior covariance matrix as a quality measure of identification results in the context of the assumed model and a given set of data. The (frequentist) covariance of MPVs, in reality when estimated on a sample basis, is an aggregate quantity that also reflects possible system changes and/or modeling errors among experiment trials.

1. Introduction

Bayesian system identification involves inferring probabilistic information about a set of model parameters $\boldsymbol{\theta}=[\theta_1,\ldots,\theta_{n_p}]$ using data $\mathbf{z}$ measured from a subject system [1–3]. The inference outcome is encapsulated in the 'posterior distribution' of $\boldsymbol{\theta}$, which is conditional on the model M used and available data $\mathbf{z}$. Using Bayes' Theorem, it is given by

$$p(\boldsymbol{\theta}|\mathbf{z},M)=c(\mathbf{z})p(\mathbf{z}|\boldsymbol{\theta},M)p(\boldsymbol{\theta}|M) \tag{1}$$

where $c(\mathbf{z})=p(\mathbf{z}|M)^{-1}$ is a normalizing factor that does not depend on $\boldsymbol{\theta}$; $p(\boldsymbol{\theta}|M)$ is called the 'prior distribution' that reflects one's knowledge about $\boldsymbol{\theta}$ in the absence of data; $p(\mathbf{z}|\boldsymbol{\theta},M)$ is called the 'likelihood function' that reflects the plausibility of obtaining the data $\mathbf{z}$ when the system is modeled by M with parameters assuming the value $\boldsymbol{\theta}$. The likelihood

E-mail address: siukuiau@cityu.edu.hk

0888-3270/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ymssp.2012.01.010

function contains vital information of the problem and must be derived specifically for a particular type of system defined by M. The model M includes all deterministic and probabilistic assumptions required to completely define the probabilistic relationship between the data $\mathbf{z}$ and the set of model parameters $\boldsymbol{\theta}$.

In problems where the parameters are identifiable [1] and there is a sufficiently large amount of data, the variation of the posterior distribution is dominated by that of the likelihood function, rendering them directly proportional to each other. The posterior distribution can be characterized by its 'most probable value' (MPV) $\hat{\boldsymbol{\theta}}$ and covariance matrix $\hat{\mathbf{C}}$. The MPV maximizes the posterior distribution and is often the reported value that best represents the system taking into account the information from the measured data. The posterior covariance matrix quantitatively reflects the uncertainty associated with $\boldsymbol{\theta}$ in the presence of the data.

In a non-Bayesian context, the MPV $\hat{\boldsymbol{\theta}}$ carries the same connotation as the 'best parameter value' or 'sample estimate'. The accuracy of the identified parameters is often quantified by their ensemble covariance matrix

$$\mathbf{E}_0 = E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T|\boldsymbol{\theta}_0,M] = \int [\hat{\boldsymbol{\theta}}(\mathbf{z})-\boldsymbol{\theta}_0][\hat{\boldsymbol{\theta}}(\mathbf{z})-\boldsymbol{\theta}_0]^T p(\mathbf{z}|\boldsymbol{\theta}_0,M)d\mathbf{z} \tag{2}$$

In the expectation in (2), the MPV $\hat{\boldsymbol{\theta}}$ implicitly depends on the data $\mathbf{z}$. It has been assumed that there is no modeling error so that the distribution of $\mathbf{z}$ indeed results from the model M with parameter $\boldsymbol{\theta}_0$. This quantification of parameter uncertainty involves the idea of the 'true' or 'actual' parameter $\boldsymbol{\theta}_0$, based on which the data is assumed to result. It also involves the notion of the 'ensemble average' $E[\,\cdot\,|\boldsymbol{\theta}_0,M]$ that is conceptually the average over statistically identical experiments. We refer this quantification of parameter uncertainty as a 'frequentist' description. In reality, even if one accepts the existence of the 'true parameter' $\boldsymbol{\theta}_0$, the value is unknown and is often replaced by the sample estimate in the derived formula of $\mathbf{E}_0$. When such formula is not available, which is often the case in inference problems of complicated systems, $\mathbf{E}_0$ is estimated by sample statistics $\hat{\mathbf{E}}$ over different trials of experiments, e.g.,

$$\hat{\mathbf{E}} = \frac{1}{N_D}\sum_{r=1}^{N_D}(\hat{\boldsymbol{\theta}}_r-\overline{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}}_r-\overline{\boldsymbol{\theta}})^T \tag{3}$$

where

$$\overline{\boldsymbol{\theta}} = \frac{1}{N_D}\sum_{r=1}^{N_D}\hat{\boldsymbol{\theta}}_r \tag{4}$$

is the sample mean of the parameter estimates $\{\hat{\boldsymbol{\theta}}_r : r=1,\ldots,N_D\}$ from N_D trials of experiment. In this context, $\overline{\boldsymbol{\theta}}$ is a proxy for $\boldsymbol{\theta}_0$.

It is clear that $\mathbf{E}_0$ is a frequentist counterpart to the posterior covariance matrix $\hat{\mathbf{C}}$. There is no strict mathematical relationship equating these two quantities but intuition expects that when there is no or little modeling error they should be consistent in describing the uncertainty of $\boldsymbol{\theta}$ in the presence of data. Since $\hat{\mathbf{C}}$ depends on the data $\mathbf{z}$ which is modeled probabilistically, an intuitive and practical guess is whether its ensemble average, $E[\hat{\mathbf{C}}|\boldsymbol{\theta}_0,M]$, is equal to $\mathbf{E}_0$, where

$$E[\hat{\mathbf{C}}|\boldsymbol{\theta}_0,M] = \int \hat{\mathbf{C}}(\mathbf{z})p(\mathbf{z}|\boldsymbol{\theta}_0,M)d\mathbf{z} \tag{5}$$

It should be noted that, in contrast to the (frequentist) sample covariance matrix, issues regarding how close the (Bayesian) posterior covariance matrix calculated using a given data set is to the (often unknown) ensemble average are irrelevant in application because the posterior covariance matrix reflects precisely the unresolved uncertainty that is a function of the modeling assumptions and the data given. In other words there is no 'true' or 'universal' quantity of uncertainty to be estimated [4].

In this work, we investigate the relationship between the frequentist and Bayesian description of parameter uncertainty in inference problems. This is motivated by empirical observations on the proximity between the calculated posterior uncertainty of model parameters and the setup-to-setup variability of the MPV in recent Bayesian modal identification studies [5,6]. It is also of practical interest to ascertain the meaning and the role of the two statistics so as to allow their proper use in applications.

The paper is outlined as follows. We first derive a mathematical relationship that connects the frequentist and Bayesian statistics in the absence of modeling error. The relationship is based on a second order approximation of the log-likelihood function and is therefore approximate. It reduces to intuitive guess, upon consideration of some heuristics. A simple example is then given to illustrate that the approximation of the second order theory depends on the particular problem setup. Focusing on ambient modal identification, we then investigate the accuracy of the derived results. To examine the theoretical accuracy, we use simulated data with little or no modeling error. It will be demonstrated that neither the second order theory nor intuition gives exact results, but they provide a good approximation of similar quality. Our investigation is also extended to real experimental data, spanning from well-controlled laboratory conditions to field structures where the environment can hardly be controlled. Greater discrepancy is observed, which is believed to be attributed to modeling error. An approximate relationship is also suggested to interpret the discrepancy between the frequentist and Bayesian statistics observed in repeated experiments.

2. Second order theory

With little loss of generality, we shall assume a uniform prior distribution, so that the posterior distribution is directly proportional to the likelihood function. For analysis purpose it is convenient to work with the 'negative log-likelihood function' (NLLF):

$$L(\boldsymbol{\theta},\mathbf{z}) = -\ln p(\mathbf{z}|\boldsymbol{\theta},M) \tag{6}$$

so that

$$p(\mathbf{z}|\boldsymbol{\theta},M) = \exp[-L(\boldsymbol{\theta},\mathbf{z})] \tag{7}$$

By construction, the NLLF attains its minimum at the MPV $\hat{\boldsymbol{\theta}}(\mathbf{z})$, where we have emphasized that the MPV depends on the data $\mathbf{z}$. On the other hand, the posterior covariance matrix can be approximated by the inverse of the Hessian of the NLLF at the MPV. Specifically, for a given $\mathbf{z}$ consider a second order Taylor expansion of L with respect to $\boldsymbol{\theta}$ about the MPV $\hat{\boldsymbol{\theta}}(\mathbf{z})$

$$L(\boldsymbol{\theta},\mathbf{z}) \approx L(\hat{\boldsymbol{\theta}}(\mathbf{z}),\mathbf{z}) + \frac{1}{2}[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})]^T \hat{H}(\mathbf{z})[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})] \tag{8}$$

where the first order term vanishes due to the minimizing nature of the MPV; $\hat{H}(\mathbf{z})$ is the Hessian of L with respect to $\boldsymbol{\theta}$ and evaluated at $\hat{\boldsymbol{\theta}}(\mathbf{z})$.Substituting (8) into (7) gives a Gaussian approximation of the posterior distribution

$$p(\boldsymbol{\theta}|\mathbf{z},M) \propto p(\mathbf{z}|\boldsymbol{\theta},M) \propto \exp\left\{-\frac{1}{2}[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})]^T \hat{H}(\mathbf{z})[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})]\right\} \tag{9}$$

Comparing (9) with a Gaussian distribution, one finds that the posterior covariance matrix of $\boldsymbol{\theta}$ is given by the inverse of the Hessian, i.e.,

$$\hat{\mathbf{C}}(\mathbf{z}) = \hat{H}(\mathbf{z})^{-1} \tag{10}$$

Having clarified the mathematical relationship of $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{C}}$ with the NLLF, we next derive a relationship that connects the frequentist and Bayesian statistics. The derivation is based on the fundamental property of the likelihood function $p(\mathbf{z}|\boldsymbol{\theta},M)$ as a legitimate probability distribution for $\mathbf{z}$, namely, for any $\boldsymbol{\theta}$

$$\int p(\mathbf{z}|\boldsymbol{\theta},M)d\mathbf{z} = 1 \tag{11}$$

In terms of the NLLF, this means

$$\int \exp[-L(\boldsymbol{\theta},\mathbf{z})]d\mathbf{z} = 1 \tag{12}$$

2.1. Relationship for MPV

Since (11) holds for any $\boldsymbol{\theta}$, all its derivatives with respect to $\boldsymbol{\theta}$ are identically equal to zero. In particular, for the first derivatives,

$$\int \nabla_{\boldsymbol{\theta}} p(\mathbf{z}|\boldsymbol{\theta},M)d\mathbf{z} = \mathbf{0} \tag{13}$$

where $\nabla_{\boldsymbol{\theta}}$ denotes the gradient (a 1-by-n_p row vector) with respect to $\boldsymbol{\theta}$. From (7),

$$\nabla_{\boldsymbol{\theta}}^T p(\mathbf{z}|\boldsymbol{\theta},M) = -[\nabla_{\boldsymbol{\theta}}^T L(\boldsymbol{\theta},\mathbf{z})]\exp[-L(\boldsymbol{\theta},\mathbf{z})] = -[\nabla_{\boldsymbol{\theta}}^T L(\boldsymbol{\theta},\mathbf{z})]p(\mathbf{z}|\boldsymbol{\theta},M) \tag{14}$$

From (8), the gradient of L with respect to $\boldsymbol{\theta}$ is given by, after considering symmetry of $\hat{H}(\mathbf{z})$,

$$\nabla_{\boldsymbol{\theta}}^T L(\boldsymbol{\theta},\mathbf{z}) \approx \hat{H}(\mathbf{z})[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})] = \hat{\mathbf{C}}(\mathbf{z})^{-1}[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})] \tag{15}$$

Substituting (15) into (14) gives

$$\nabla_{\boldsymbol{\theta}}^T p(\mathbf{z}|\boldsymbol{\theta},M) \approx -\hat{\mathbf{C}}(\mathbf{z})^{-1}[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})]p(\mathbf{z}|\boldsymbol{\theta},M) \tag{16}$$

Substituting (16) into (13) gives

$$\int \hat{\mathbf{C}}(\mathbf{z})^{-1}[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})]p(\mathbf{z}|\boldsymbol{\theta},M)d\mathbf{z} \approx \mathbf{0} \tag{17}$$

That is, evaluating at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ and interpreting $\int \cdot\, p(\mathbf{z}|\boldsymbol{\theta}_0,M)d\mathbf{z}$ as the ensemble average $E[\cdot\,|\boldsymbol{\theta}_0,M]$,

$$E[\hat{\mathbf{C}}^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)|\boldsymbol{\theta}_0,M] \approx \mathbf{0} \tag{18}$$

We postpone discussion of this equation after deriving the counterpart result for the posterior covariance matrix.

2.2. Relationship for posterior covariance matrix

To explore the results for the posterior covariance matrix, we make use of the fact that the Hessian of (11) with respect to $\boldsymbol{\theta}$ is equal to a zero matrix

$$\int \nabla^2_{\boldsymbol{\theta}} p(\mathbf{z}|\boldsymbol{\theta},M)d\mathbf{z} = \mathbf{0} \tag{19}$$

The Hessian of the likelihood function in (7) is given by

$$\nabla^2_{\boldsymbol{\theta}} p(\mathbf{z}|\boldsymbol{\theta},M) = -\hat{H}(\mathbf{z})p(\mathbf{z}|\boldsymbol{\theta},M)+[\nabla^T_{\boldsymbol{\theta}} L(\boldsymbol{\theta},\mathbf{z})][\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta},\mathbf{z})]p(\mathbf{z}|\boldsymbol{\theta},M) \tag{20}$$

Substituting (10) and (15) into (20) and using (19) gives, after rearranging,

$$\int \hat{\mathbf{C}}(\mathbf{z})^{-1}[\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})][\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}(\mathbf{z})]^T\hat{\mathbf{C}}(\mathbf{z})^{-1}p(\mathbf{z}|\boldsymbol{\theta},M)d\mathbf{z} \approx \int \hat{\mathbf{C}}(\mathbf{z})^{-1}p(\mathbf{z}|\boldsymbol{\theta},M)d\mathbf{z} \tag{21}$$

That is, evaluating at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$,

$$E[\hat{\mathbf{C}}^{-1}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M] \approx E[\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M] \tag{22}$$

In the above derivation, the major assumption lies in the quadratic approximation of the NLLF. In this regard, one notes that we approximate the logarithm of the likelihood function rather than itself directly. The former is preferred because the functions encountered in practice are more akin to an exponential form. In particular, a quadratic form for the likelihood function is always unbounded for large deviation from the mean, while an exponential form goes to zero. Naturally, the results in (18) and (22) are not universal as they depend on the form of approximation adopted in the derivation. The conditioning statement in the expectations also reminds us that the data is assumed to result from the model $(\boldsymbol{\theta}_0,M)$ without modeling error.

2.3. Weighted ensemble average

Eqs. (18) and (22) are as far as we can say regarding a general Bayesian inference problem with quadratic approximation in the NLLF and in the absence of modeling error. Eq. (18) does not indicate that the ensemble average of the MPV is equal to the actual parameter. Similarly, (22) does not indicate that the ensemble covariance matrix is equal to the ensemble average of the posterior covariance matrix. We next express them in a form that has more intuitive linkage to ensemble average.

Define a random matrix:

$$\mathbf{W}(\mathbf{z}) = E[\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M]^{-1}\hat{\mathbf{C}}(\mathbf{z})^{-1} \tag{23}$$

It is clear that the diagonal entries of $\mathbf{W}$ are dimensionless. The ensemble average of $\mathbf{W}$ is equal to the identity matrix because

$$E[\mathbf{W}|\boldsymbol{\theta}_0,M] = E[\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M]^{-1}E[\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M] = \mathbf{I} \tag{24}$$

For this reason we call $\mathbf{W}$ a 'weighting matrix'. Pre-multiplying (18) by $E[\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M]^{-1}$, rearranging and using (24), one obtains

$$E[\mathbf{W}\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_0,M] \approx \boldsymbol{\theta}_0 \tag{25}$$

This result can be interpreted as saying that the ensemble average of the 'weighted MPV' is equal to the actual parameter.

To obtain a similar result for the posterior covariance matrix, write (22) as

$$E\{\hat{\mathbf{C}}^{-1}[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T-\hat{\mathbf{C}}]\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M\} \approx \mathbf{0} \tag{26}$$

Pre- and post-multiplying by $E[\hat{\mathbf{C}}^{-1}|\boldsymbol{\theta}_0,M]^{-1}$ and using (23) gives

$$E[\mathbf{W}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T\mathbf{W}^T|\boldsymbol{\theta}_0,M] \approx E[\mathbf{W}\hat{\mathbf{C}}\mathbf{W}^T|\boldsymbol{\theta}_0,M] \tag{27}$$

This result means that the ensemble covariance matrix of the weighted MPV is equal to the ensemble average of the (doubly-) weighted posterior covariance matrix.

It should be noted that the weighting matrix is defined merely to give an intuitive interpretation of the results from the second order theory. Its probabilistic characteristic can be very complex and not amenable to further analysis. We advocate that the weighting matrix is of little substance in practice, as we shall argue heuristically and through numerical studies that our intuitions may also provide an acceptable approximation.

2.4. Heuristic arguments

More intuitive forms result from some heuristic arguments. If one ignores the variability of $\mathbf{W}$ or assumes that it is uncorrelated from $\hat{\boldsymbol{\theta}}$, then (25) and (27) reduce to what are expected intuitively:

$$E[\hat{\boldsymbol{\theta}}|\boldsymbol{\theta}_0,M] \approx \boldsymbol{\theta}_0 \tag{28}$$

$$E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)^T|\boldsymbol{\theta}_0,M] \approx E[\hat{\mathbf{C}}|\boldsymbol{\theta}_0,M] \tag{29}$$

Thus, our intuitions, though need not exactly correct, could still be a good approximation. These equations provide a mathematical argument connecting the frequentist and Bayesian description of parameter uncertainty. That is, in the absence of modeling error, the trial-to-trial (frequentist) covariance of the MPV from identically distributed experiments is approximately equal to the ensemble mean of the posterior covariance (Bayesian). Since the posterior covariance depends primarily on the data size, its value in a particular trial may be used as a proxy for the trial-to-trial covariance over different setups. The utility of this statement to a frequentist is that it is not necessary to perform repeated trials to estimate the variability of the identified parameters, assuming that there is little or no modeling error and further trials are performed under similar conditions.

2.5. A simple illustrative example

We use a simple example to consolidate the foregoing results and to illustrate that the second order theory or intuition need not be exact, as the situation is problem-dependent. Let $\{z_i \in R: i=1, \ldots, N\}$ be observations (data) of a scalar quantity μ to be identified from N trials of experiment. They are modeled as $z_i=\mu+\varepsilon_i$ where $\{\varepsilon_i: i=1, \ldots, N\}$ are independent and identically distributed (i.i.d.) Gaussian prediction error with zero mean and variance S_e. In this case, given μ and S_e, $\{z_1, \ldots, z_N\}$ are i.i.d. Gaussian with mean μ and variance S_e. The NLLF is then given by

$$L = \frac{N}{2}\ln 2\pi + \frac{N}{2}\ln S_e + \frac{1}{2}S_e^{-1}\sum_{k=1}^{N}(z_i-\mu)^2 \tag{30}$$

The derivatives of L are given by

$$\frac{\partial L}{\partial \mu} = -S_e^{-1}\sum_{k=1}^{N}(z_i-\mu), \quad \frac{\partial L}{\partial S_e} = \frac{N}{2}S_e^{-1} - \frac{S_e^{-2}}{2}\sum_{k=1}^{N}(z_k-\mu)^2 \tag{31}$$

$$\frac{\partial^2 L}{\partial \mu \partial S_e} = S_e^{-2}\sum_{k=1}^{N}(z_i-\mu), \quad \frac{\partial^2 L}{\partial \mu^2} = NS_e^{-1}, \quad \frac{\partial^2 L}{\partial S_e^2} = -\frac{N}{2}S_e^{-2} + S_e^{-3}\sum_{k=1}^{N}(z_k-\mu)^2 \tag{32}$$

If $S_e=S_{e0}$ (say) is known, then the only parameter to be identified is μ, i.e., $\boldsymbol{\theta}\equiv\mu$. In the context of the model $M: z_i=\mu+\varepsilon_i$, suppose the data corresponds to the actual parameter $\boldsymbol{\theta}_0=\mu_0$. Setting $\partial L/\partial\mu=0$ and solving for μ, the MPV is $\hat{\mu}=\sum_{i=1}^{N} z_i/N$. Its posterior variance is $V_\mu=\partial^2 L/\partial\mu^2(\mu=\hat{\mu})=S_{e0}/N$. Thus, $E[\hat{\mu}|\boldsymbol{\theta}_0,M]=\sum_{i=1}^{N}E[z_i|\boldsymbol{\theta}_0,M]/N=\mu_0$. Also, $\mathrm{var}[\hat{\mu}|\boldsymbol{\theta}_0,M]=\mathrm{var}[z_i|\boldsymbol{\theta}_0,M]/N=S_{e0}/N$, which is equal to the expectation of the posterior variance, since $E[V_\mu|\boldsymbol{\theta}_0,M]=S_{e0}/N$. The weighting factor in this case is given by $W_\mu=E[N/S_{e0}|\boldsymbol{\theta}_0,M]^{-1}N/S_{e0}=1$. Thus, when the prediction error variance S_e is known, the second order theory and intuition give identical results which are exact.

The picture is different when S_e is unknown and is included in the set of parameters to be identified, i.e., $\boldsymbol{\theta}\equiv[\mu,S_e]$. In this case the MPVs are $\hat{\mu}=\sum_{i=1}^{N} z_i/N$ (same as before) and $\hat{S}_e=\sum_{i=1}^{N}(z_i-\hat{\mu})^2/N$. It can be readily shown that $\partial^2 L/\partial\mu\,\partial S_e=0$ at the MPV and so μ and S_e are uncorrelated according to the posterior distribution. The posterior variance of μ and S_e are given by $V_\mu=\hat{S}_e/N$ and $V_{S_e}=2\hat{S}_e^2/N$, respectively. Their weighting factors are therefore $W_\mu=E[\hat{S}_e^{-1}|\boldsymbol{\theta}_0,M]^{-1}\hat{S}_e^{-1}$ and $W_{S_e}=E[\hat{S}_e^{-2}|\boldsymbol{\theta}_0,M]^{-1}\hat{S}_e^{-2}$. Thus, $E[\hat{\mu}|\boldsymbol{\theta}_0,M]=\mu_0$ but $E[W_\mu\hat{\mu}|\boldsymbol{\theta}_0,M]=E[\hat{S}_e^{-1}|\boldsymbol{\theta}_0,M]^{-1}E[\hat{S}_e^{-1}\hat{\mu}|\boldsymbol{\theta}_0,M]$, which is generally not equal to μ_0. On the other hand, $\mathrm{var}[\hat{\mu}|\boldsymbol{\theta}_0,M]=S_{e0}/N$ (same as before) but the expectation of the posterior variance is given by $E[\hat{S}_e/N|\boldsymbol{\theta}_0,M]=(N-1)S_{e0}/N^2$, and so they are not exactly equal. Similar observations can be made on the weighted mean quantities, which show that $E[W_\mu^2(\hat{\mu}-\mu_0)^2|\boldsymbol{\theta}_0,M]\neq E[W_\mu^2 V_\mu|\boldsymbol{\theta}_0,M]$. It can be seen that the second order theory and intuition are not exact when the prediction error is one of the parameters to be identified. This is essentially due to the nonlinear nature of S_e in the NLLF which deviates from a quadratic function. Similar situation may occur in other problems since the prediction error is generally unknown.

3. Modal identification problem

The second order theory developed in the last section applies to Bayesian inference problems in general. In the remaining of this work we shall focus on modal identification problems with ambient (output-only) data (e.g., [7–9]). In this class of problems the uncertainty in the identified modal properties is of important concern [5,10]. A Bayesian approach offers a clear advantage by providing a fundamental quantification of the posterior uncertainties. A frequency

domain approach is adopted here, whose original formulation is due to [11]. Fast algorithms allowing practical implementation are recently developed in [12] for well-separated modes and in [13,14] for general (possibly close) modes. The basic formulation is outlined here; details can be found in the references.

Let the digitally measured ambient acceleration time history data at sampling interval Δt be $\{\hat{\ddot{\mathbf{x}}}_j \in R^n : j=1,\ldots,N\}$. The FFT of $\{\hat{\ddot{\mathbf{x}}}_j\}$ is defined as

$$\mathcal{F}_k = \sqrt{\frac{2\Delta t}{N}} \sum_{j=1}^{N} \hat{\ddot{\mathbf{x}}}_j \exp\left[-2\pi \mathrm{i} \frac{(k-1)(j-1)}{N}\right] \tag{33}$$

Note that $\mathbf{D}_k = \mathcal{F}_k \mathcal{F}_k^*$ is the spectral density matrix at frequency $\mathrm{f}_k=(k-1)/N\Delta t$ (Hz) for $k=1, \ldots, N_q$, where N_q is the index corresponding to the Nyquist frequency, equal to the integer part of $N/2+1$. A plot of the eigenvalues of $\mathbf{D}_k$ for different k form the 'singular value spectrum'. In practice, a frequency band is selected based on the singular value spectrum, whose FFT data shall be used for identifying the modes within the band. The FFT of the response within the band is modeled as the sum of the contribution from the resonant modes in the band and the prediction error. The set of model parameters, in this case being the set of 'modal parameters', $\boldsymbol{\theta}$, includes the natural frequencies $\{f_i : i=1, \ldots, m\}$, damping ratios $\{\zeta_i : i=1, \ldots, m\}$, mode shape matrix $\boldsymbol{\Phi} \in R^{n\times m}$, the power spectral density (PSD) matrix of modal force $\mathbf{S} \in C^{m\times m}$ and the PSD of prediction error (assumed to be i.i.d.) S_e; n and m are the number of measured dofs and contributing modes, respectively. For small Δt and long duration of data (compared to the natural periods), the NLLF is given by

$$L(\boldsymbol{\theta}) = \frac{1}{2}\sum_k \ln \det \mathbf{C}_k(\boldsymbol{\theta}) + \frac{1}{2}\sum_k \mathbf{Z}_k^T \mathbf{C}_k(\boldsymbol{\theta})^{-1} \mathbf{Z}_k \tag{34}$$

where the sum is over all frequencies in the selected band for identification; $\mathbf{Z}_k = [\mathrm{Re}\,\mathcal{F}_k; \mathrm{Im}\,\mathcal{F}_k] \in R^{2n}$ is an augmented vector consisting of the real and imaging part of $\mathcal{F}_k$;

$$\mathbf{C}_k = \frac{1}{2}\begin{bmatrix} \boldsymbol{\Phi}\,\mathrm{Re}\,\mathbf{H}_k\boldsymbol{\Phi}^T & -\boldsymbol{\Phi}\,\mathrm{Im}\,\mathbf{H}_k\boldsymbol{\Phi}^T \\ \boldsymbol{\Phi}\,\mathrm{Im}\,\mathbf{H}_k\boldsymbol{\Phi}^T & \boldsymbol{\Phi}\,\mathrm{Re}\,\mathbf{H}_k\boldsymbol{\Phi}^T \end{bmatrix} + \frac{S_e}{2}\mathbf{I}_{2n} \tag{35}$$

is the covariance matrix of $\mathbf{Z}_k$; $\mathbf{I}_{2n} \in R^{2n\times 2n}$ denotes the identity matrix; $\mathbf{H}_k \in C^{m\times m}$ is the transfer matrix whose (i,j)-entry is given by

$$\mathbf{H}_k(i,j) = S_{ij}[(\beta_{ik}^2-1)+\mathbf{i}(2\zeta_i\beta_{ik})]^{-1}[(\beta_{jk}^2-1)-\mathbf{i}(2\zeta_j\beta_{jk})]^{-1} \tag{36}$$

$\beta_{ik}=f_i/\mathrm{f}_k$ is a frequency ratio; S_{ij} is the (i,j)-entry of $\mathbf{S}$.

Computing the MPV and posterior covariance matrix is highly non-trivial, due to the nonlinear nature of the NLLF and the large number of modal parameters involved. The latter grows linearly with the number of measured dofs and in a quadratic manner with the number of modes. Nevertheless fast solution methods have been developed recently, allowing Bayesian identification to be performed in typically a few seconds to a few minutes [12–14]. Essentially, it is found that the mode shapes can be found almost analytically in terms of the remaining parameters and so the growth of computational effort with the number of measured dofs can be effectively suppressed.

It is important to note that the Bayesian FFT method, although operating in the frequency domain, is distinct from frequency domain decomposition (FDD) or its variants [15]. In particular no statistical averaging or spectral smoothing is required. Rather, the FFT data in (33) directly enters the NLLF in (34). The selected frequency band does not need to be confined to one inside which the Modal Assurance Criterion (MAC) values among the eigen vectors of the power spectral density matrix are large. The Bayesian FFT method yields directly physical mode shapes rather than operating deflection shapes. For example, the identified mode shapes need not be orthogonal. As a result of the one–one correspondence between the time domain data and FFT, the method is capable of making inference using potentially all relevant information contained in the data consistent with modeling assumptions used in structural dynamics.

4. Investigation with synthetic data

We first consider two examples with synthetic data to illustrate and check the second order theory (Eqs. (25) and (27)) and intuitions (Eqs. (28) and (29)). The first example considers a single-degree-of-freedom (SDOF) structure with a single measured dof, which is a simple case for verification without complication by modeling error. The second example considers a 4-DOF shear building structure with 4 measured dofs. It allows the effect of modeling error related to unmodeled modes to be examined.

4.1. SDOF structure

Consider an SDOF structure subjected to white noise excitation

$$\ddot{x}(t)+2\zeta_1\omega_1\dot{x}(t)+\omega_1^2 x(t) = W(t) \tag{37}$$

where $\omega_1=2\pi f_1, f_1=1$ Hz is the natural frequency, $\zeta_1=1\%$ is the damping ratio and $W(t)$ has a one-sided spectral density of $S_1=1(\mu g)^2/\mathrm{Hz}$. The acceleration response is simulated digitally at a sampling rate of 20 Hz. The acceleration data is

contaminated by i.i.d. Gaussian white noise with a one-sided spectral density of $S_{e1}=100(\mu g)^2/Hz$. The assumption of excitation and channel noise here leads to modal signal-to-noise (s/n) ratio in the response that is typical (or on the low side) of full-scale ambient tests [12,14]. In particular, the modal s/n ratio (at resonance) is equal to $\gamma=S_1/4S_{e1}\zeta_1^2=1/4(100)(0.01)^2=25$. The set of modal parameters $\boldsymbol{\theta}$ to be identified consists of the natural frequency f_1, damping ratio ζ_1, spectral density of excitation S_1 and prediction error S_{e1}. The mode shape $\boldsymbol{\Phi}$ is trivially equal to 1.

Fig 1 shows the singular value spectrum of a typical set of 300 s data. The horizontal bar indicates the selected frequency band whose FFT data shall be used for modal identification. Using this set of data the MPV and posterior covariance matrix of modal parameters are calculated. The posterior marginal distribution of the modal parameters is plotted as solid line in Fig 2. The histograms (shaded bar) in the same figure shall be explained later and can be ignored at the moment. The marginal distribution is obtained by numerically integrating the joint posterior distribution with respect to the remaining (three) parameters. The dashed line in the figure shows the Gaussian approximation of the marginal distribution, which has a mean at the MPV and a variance equal to the posterior variance obtained from the corresponding diagonal entry of the posterior covariance matrix. It is seen that the Gaussian approximation is quite good, especially for the natural frequency and damping ratio. The posterior marginal distribution need not center at the exact value, since it is obtained based on a single set of data with finite length. Nevertheless its support region covers well the exact value, in a manner consistent with its spread. Note that in practice one does not calculate the exact marginal distribution (solid line) because it is time-consuming to perform the multi-dimensional numerical integration. Instead, only the Gaussian approximation (dashed line) is used, since the MPV and the posterior covariance matrix can be efficiently computed.

To examine the ensemble (frequentist) statistics of the MPVs among statistically identical experiment trials, we generate 1000 i.i.d. sets of data (300 s each). Correspondingly, 1000 'samples' of the MPVs are calculated. The histogram of the MPVs is shown in Fig. 2 as shaded bar. It is scaled to have a unit area in the figure, to be consistent with the marginal distribution. The histogram has a peak near the exact value of the modal parameter, as expected. It is not the same as the posterior marginal distribution based on a particular data set (solid line); it need not be. Nevertheless it has a spread similar to that of the posterior marginal distribution. It is remarkable that, *in the absence of modeling error, although the*

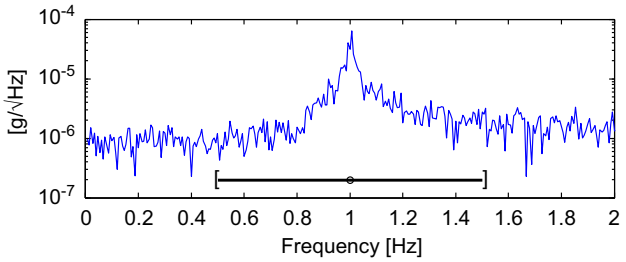

Fig. 1. Singular value spectrum, SDOF synthetic data.

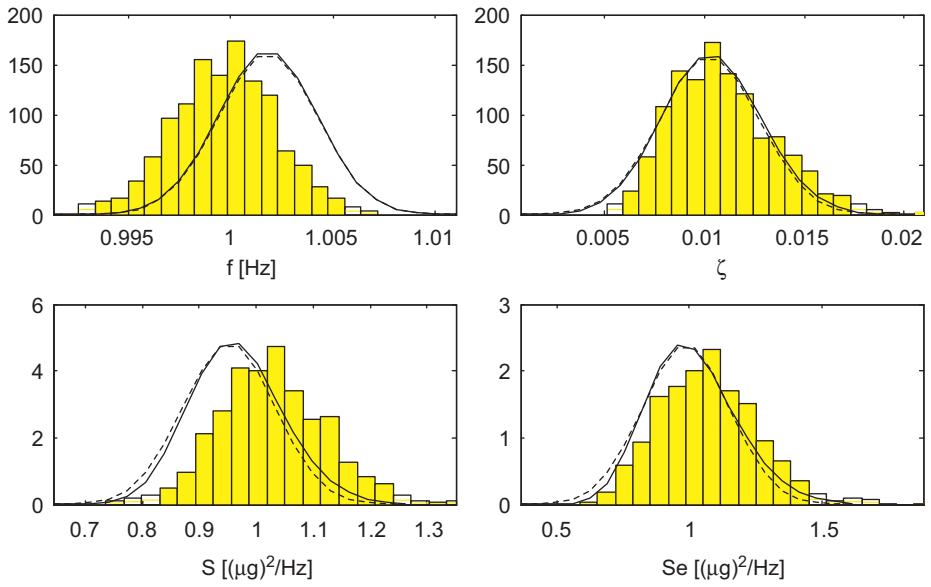

Fig. 2. Posterior marginal distribution calculated using a typical data set (solid line); Gaussian approximation (dashed line); the histogram (shaded bar) of MPVs among 1000 trials. SDOF example.

posterior distribution is based on data from a single data set, its spread reflects the ensemble variability of the MPV among statistically identical experiment trials. This aspect is next investigated quantitatively.

Table 1 compares the frequentist and Bayesian statistics of the modal identification results among the 1000 trials. The second to fifth column shows the exact parameter value that generated the data, the MPV calculated using a single data set, sample mean of MPV (LHS of (28)) and weighted sample mean (LHS of (25)) of each modal parameter. It is seen that both the sample mean and the weighted mean are quite close to the exact value. The sixth column titled 'Freq.' (short for 'frequentist') shows the sample coefficient of variation (c.o.v.) of the MPV among the 1000 trials, defined as the sample standard deviation divided by the sample mean of the MPV. The seventh column titled 'Bay.' (short for 'Bayesian') shows the sample root mean square (r.m.s.) value of the posterior standard deviation divided by the sample mean of the MPV. For discussion purpose we refer this quantity as the 'equivalent mean posterior c.o.v.'. It can be readily verified that the sample variance of the MPV (LHS of (29)) is equal to the sample average of the posterior variance (RHS of (29)) if and only if the sample c.o.v. of the MPV (sixth column) is equal to the equivalent mean posterior c.o.v. (seventh column). The c.o.v. quantities are presented for their intuitive (non-dimensional) nature. The ratio of the frequentist to Bayesian quantity is shown in the column titled 'A/B'. It is seen that they are all close to 1. The counterparts for the weighted statistics based on second order theory, i.e., LHS and RHS of (27), are shown in the last three columns of Table 1. Again the ratio between the frequentist and the Bayesian quantity is quite close to 1, in a manner similar to the unweighted statistics.

The number of data sets used here (1000) is large enough so that statistical estimation error is negligible in the results. Specifically, the standard deviation of the sample mean is reduced down to $1/(1000)^{1/2}\sim 3\%$ of that of the quantity being averaged. Repeated runs have also been performed, whose results are qualitatively the same as those presented in the table here. In this context, it is fair to say that the sample c.o.v. of MPV (frequentist) and the equivalent mean posterior c.o.v. (Bayesian) are approximately equal for both the unweighted (intuition) or weighted (second order theory) statistics.

4.2. Four-DOF structure

We next consider a four-storied lab-scale shear frame with synthetic data. The idea is to check the extent to which the second order theory or intuition is consistent in the presence of model error related to un-modeled modes excluded in the identification model that is applied on the selected frequency band. Each floor weighs 25.2 kg. The interstory stiffnesses for the first to fourth story are 79.4, 10.6, 10.4 and 11.3 N/mm. These properties are similar to the lab frame considered later with real experimental data. The structure is assumed to be classically damped with a damping ratio of 1% in all modes. Each floor is subjected to i.i.d. Gaussian white noise excitation with a one-sided spectral density of 10^{-6} N^2/Hz. In each data set (experiment trial), the measured data is assumed to consist of 300 s acceleration time history sampled at 256 Hz at all the four floors, contaminated by i.i.d. Gaussian white channel noise with a one-sided spectral density of $100(\mu g)^2$/Hz. The channel noise assumed here is typical of capacitive (miniature) sensors.

Fig 3 shows the singular value spectrum computed using a typical set of 300 sec data. The four modes are identified using FFT data on separate frequency bands indicated in the figure. Similar to the first example, sample statistics of modal identification results are calculated using 1000 i.i.d. data sets.

Table 1
Sample and Bayesian statistics, SDOF synthetic data.

	Exact	Single Set MPV	Sample mean of MPV	Weighted sample mean of MPV	Freq.[a] (%) A	Bay.[b] (%) B	A/B	Weighted freq. (%) C	Weighted bay. (%) D	C/D
f_1 (Hz)	1	0.997	1.000	1.000	0.253	0.251	1.00	0.249	0.243	1.02
ζ_1 (%)	1	1.285	1.096	1.026	24.11	23.63	1.02	24.49	24.32	1.01
S_1 ($(\mu g)^2$/Hz)	1	1.081	1.028	1.011	9.31	8.63	1.08	9.60	8.59	1.12
S_{e1} ($(\mu g)^2$/Hz)	1	0.921	1.072	1.038	17.53	16.96	1.03	17.95	16.91	1.06

[a] Frequentist = sample c.o.v. of MPV = (sample std. of MPV)/(sample mean of MPV).
[b] Bayesian = (r.m.s. of posterior std.)/(sample mean of MPV).

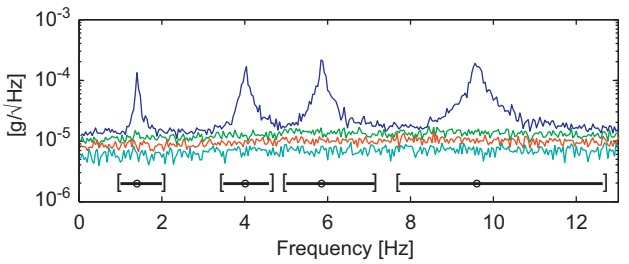

Fig. 3. Singular value spectrum, four-DOF synthetic data.

Table 2
Sample and Bayesian statistics, four-DOF synthetic data.

	Exact	Single set MPV	Sample mean of MPV	Weighted sample mean of MPV	Freq. [a] (%) A	Bay. [a] (%) B	A/B	Weighted freq. (%) C	Weighted bay. (%) D	C/D
f_1 (Hz)	1.401	1.401	1.401	1.401	0.213	0.219	0.98	0.210	0.213	0.99
f_2	4.024	4.028	4.024	4.024	0.123	0.126	0.98	0.124	0.124	1.00
f_3	5.872	5.883	5.873	5.873	0.097	0.102	0.95	0.097	0.101	0.96
f_4	9.597	9.593	9.597	9.597	0.080	0.079	1.02	0.080	0.078	1.02
ζ_1 (%)	1	1.218	1.067	1.009	20.6	21.1	0.98	20.9	21.6	0.97
ζ_2	1	0.906	1.000	0.973	13.4	13.4	1.00	13.3	13.6	0.98
ζ_3	1	0.974	0.965	0.949	11.1	11.1	1.00	11.0	11.1	0.99
ζ_4	1	1.030	0.942	0.933	8.40	8.59	0.98	8.43	8.63	0.98
S_1 ((μg)^2/Hz)	16.36	15.79	16.97	16.58	10.04	9.65	1.04	10.30	9.70	1.06
S_2	16.36	15.43	16.16	15.91	7.54	7.36	1.02	7.57	7.37	1.03
S_3	16.36	16.45	15.46	15.33	5.82	5.64	1.03	5.86	5.65	1.04
S_4	16.36	15.27	14.97	14.91	3.92	4.09	0.96	3.97	4.09	0.97
S_{e1} ((μg)^2/Hz)	100	94.8	100	100	3.26	3.26	1.00	3.25	3.26	1.00
S_{e2}	100	108	113	113	3.04	3.16	0.96	3.05	3.16	0.97
S_{e3}	100	126	129	129	2.30	2.27	1.01	2.30	2.27	1.02
S_{e4}	100	125	127	127	1.50	1.48	1.02	1.51	1.48	1.02

[a] See Table 1.

Table 2 summarizes the results for all the modes. All modal quantities are subscripted by their mode number. The columns in the table have the same meaning as those in Table 1. For the natural frequencies, the unweighted and weighted mean of MPV are identical to the exact value to within the digits shown. The sample c.o.v. and equivalent mean posterior c.o.v. also agree very well for both the unweighted (A/B) and weighted statistics (C/D), regardless of mode. For the damping ratio ζ_i, the sample mean and weighted mean of MPV agree with the exact value, although to a lower precision compared to the natural frequencies.

The sample mean of the MPV of ζ_i tends to bias towards the low side as the mode number increases. The same can also be said for S_i. This is believed to arise from modeling error in the identification model that ignores non-resonant modes in the frequency band. This modeling error is not significant in the lower modes because the contribution in acceleration response from higher modes is second order small in the low frequency band [12]. In contrast, the contribution from low frequency modes in the high frequency band is quasi-static, and the number of contributing modes accumulates with the mode order. Modeling error biases the sample mean of MPV away from the exact value. This increase in modeling error with mode is also reflected in the prediction error S_{ei}. For S_{ei} the deviation of the sample mean of MPV from the exact value should not be seen as bias, because the exact value only accounts for the channel noise. Despite the bias in MPV for ζ_i, S_i and S_{ei}, the sample c.o.v. of MPV and the equivalent mean posterior c.o.v. agree well for the unweighted (A/B) and weighted statistics (C/D). Note that the sample c.o.v. (unweighted or weighted) are calculated with respect to the sample mean instead of the exact value, since the latter is not known in reality.

5. Investigation with laboratory/field data

The foregoing theoretical and numerical observations with synthetic data show that the frequentist statistics agree well with their Bayesian counterparts. An important feature of the synthetic data examples is that the data indeed results from models which remain invariant in all the experiment trials. In reality this should never be taken for granted since the stochastic environment and system parameters can change (even if modeling mechanisms are correct), and modeling error can exist, e.g., in the damping mechanism, all in an unknown manner.

In this section we investigate the frequentist and Bayesian statistics with real data collected in the laboratory and in the field. In the laboratory example a shear frame is considered where the environmental condition is well-controlled with the intention of maintaining a stationary stochastic loading environment. In contrast, no control is possible in the field test examples with full-scale structures, in which case the results serve to give a realistic picture of the proximity between the frequentist and Bayesian statistics, where all effects are present naturally. We shall focus on the unweighted statistics, i.e., (28) and (29), rather than the weighted statistics derived from the second order theory. The unweighted statistics is more intuitive and is the quantity often used in practice. Nevertheless the weighted statistics has been investigated (details omitted here) and was found to give similar results as the unweighted statistics. We also omit details of the prediction error variance, as it is of secondary importance in practice. The study here should be distinguished from works that monitor modal properties over an extended period, in which case systematic effects from temperature, humidity, etc. are anticipated [16,17].

Table 3 summarizes the examples investigated; the synthetic data examples are included for completeness. In the real data examples, the duration of each data set was determined so that the posterior c.o.v. in the damping ratio was acceptable and the data was not excessively long to incur significant modeling error risk (judged by intuition).

5.1. Laboratory shear frame

The objective in this example is to compare the frequentist and Bayesian statistics under a well-controlled condition. For this purpose effort was made to maintain a steady stochastic loading environment. The subject structure is a four-storied steel frame which can be considered to have shear building behavior. Fig 4 shows the experimental setup. Each floor weighs 25.2 kg. The horizontal acceleration along the weak direction at the four floors are measured by Kistler K8330 accelerometers. The overall channel noise is about 1 $\mu g/(Hz)^{1/2}$ in the frequency range of interest (1–10 Hz). Digital data was originally sampled at 2048 Hz and latter decimated to 256 Hz for modal identification. The frame was kept in an air-conditioned room. Data was collected in the midnight where there was no human activities nearby. Four hours of data were recorded, which were divided into 30 data sets, each of 10 min. The duration of each data set is chosen to trade off identification precision and time-resolution. If the duration is too short the posterior uncertainty of modal parameters, which is often controlled by that of the damping ratio, will be too large. On the other hand, a large data duration reduces the time-resolution and weakens the stationary assumption in the loading and response.

The singular value spectrum calculated using a typical 10-min data set is shown in Fig. 5, where the frequency bands for modal identification are also shown in '[–]'. The response is richer in the high frequencies because the base excitation (micro-tremor) has richer frequency content there. Table 4 summarizes the frequentist and Bayesian statistics. First note that the sample c.o.v. of the MPV of S_i (fourth column) is a few tens of percentage. This c.o.v. largely reflects the variability in the environment because the posterior c.o.v. (fifth column) is relatively small. Although not large, this level of variability may be a bit surprising considering the efforts made in controlling the experimental environment. Of course, one never knows what level of variability can be maintained in the stochastic loading without a real experiment. As evidenced in the

Table 3
Summary of examples.

No.	Description	No. of measured dofs	Duration of time window (s)	No. of data sets	Duration spanned by test
1	SDOF, synthetic data	1	300	1000	Irrelevant
2	4DOF shear building, synthetic data	4	300	1000	
3	Lab-scale frame (4 stories)	4	600	30	5 h
4	Factor Building (16 stories)	16	900	24	6 h
5	Super tall building (> 300 m)	3	1800	36	18 h

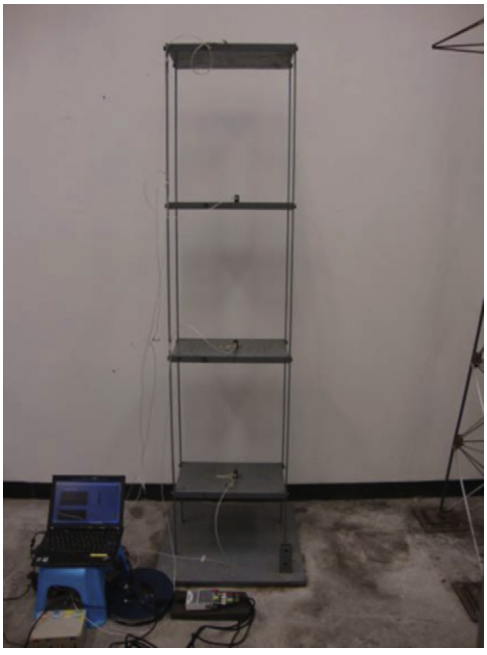

Fig. 4. Laboratory shear frame.

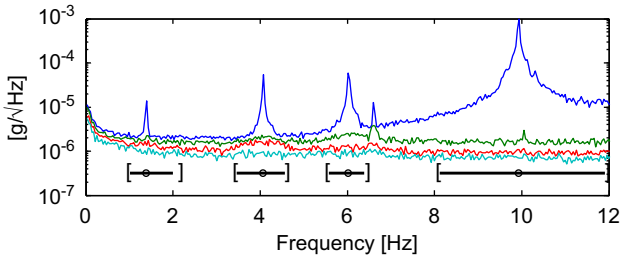

Fig. 5. Singular value spectrum, lab frame.

Table 4
Sample and Bayesian statistics, lab-scale frame.

Parameter	Single set MPV	Sample mean of MPV	Freq.[a] (%) A	Bay.[a] (%) B	A/B
f_1 (Hz)	1.383	1.383	0.0606	0.0703	0.86
f_2	4.064	4.064	0.0259	0.0270	0.96
f_3	6.028	6.028	0.0187	0.0193	0.97
f_4	9.930	9.932	0.0201	0.0138	1.46
ζ_1 (%)	0.115	0.213	57.7	35.0	1.65
ζ_2	0.075	0.107	33.8	25.5	1.32
ζ_3	0.107	0.082	22.8	23.8	0.96
ζ_4	0.077	0.071	36.5	19.6	1.86
S_1 ((μg)^2/Hz)	0.0094	0.015	54.5	15.9	3.43
S_2	0.031	0.036	11.5	6.0	1.91
S_3	0.055	0.051	27.8	5.6	5.00
S_4	6.417	4.990	20.2	2.2	9.26

[a] See Table 1.

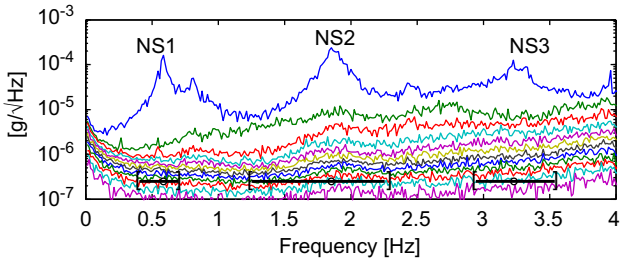

Fig. 6. Singular value spectrum, Factor building.

last column (A/B), for the natural frequencies and damping ratios, the sample c.o.v. may be considered similar to the equivalent mean posterior c.o.v. One exception, however, is the fourth (highest) mode.

One may wonder whether the large ratio of A/B may be due to significant variability in the value of posterior variance over different data sets. This has been investigated but no significant pattern is found. For reference, the c.o.v. of posterior variance among different data sets is calculated to be about 80% for f_1, ζ_1 and S_1. Nevertheless, their A/B values are quite different. For other modes, the c.o.v. is not large, about 20–40%, although a variety of A/B values are seen in different parameters.

5.2. Factor building

This example contributes to investigation related to a medium rise building under a typical ambient environment. The subject structure in this example is the UCLA Doris and Louis Factor Health Science Building. It has a rectangular plan of 126 m-by-73 m and a height of 66 m above ground (long side is oriented NS). A dense array of 71 Kinemetrics FBA-11 accelerometers has been permanently installed. Acceleration data is acquired at 100 Hz on 24-bit resolution Quanterra 4128 recorders. The dynamic characteristics of the building have been investigated previously [18–20]. The focus here is on the first three NS translational modes (NS1, NS2, NS3), to be identified using ambient data of 16 channels measuring the NS component of the building on the East Wall from the ground level up to the roof. Six hours of data are used, starting from midnight to early morning. The data is divided into 24 sets, each of 15 min.

Fig 6 shows the singular value spectrum calculated using a typical data set. The peak near the first mode is contributed by the fundamental EW translational mode and so is not included in the frequency band for identification.

Table 5
Sample and Bayesian statistics, Factor building.

Parameter	Single set MPV	Sample mean of MPV	Freq.[a] (%) A	Bay.[a] (%) B	A/B
f_1 (Hz)	0.582	0.582	0.477	0.320	1.49
f_2	1.838	1.856	0.551	0.129	4.27
f_3	3.182	3.185	0.453	0.180	2.51
ζ_1 (%)	3.08	2.84	13.2	12.5	1.06
ζ_2	1.50	1.59	10.2	8.5	1.19
ζ_3	3.47	3.20	10.4	7.6	1.37
S_1 ((μg)^2/Hz)	113	98.4	18.9	7.7	2.46
S_2	68.3	67.2	13.3	3.8	3.53
S_3	37.8	34.3	13.1	8.0	1.65

[a] See Table 1.

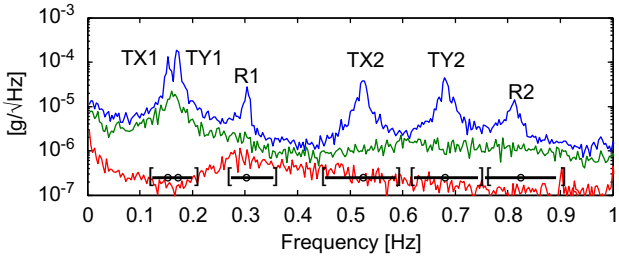

Fig. 7. Singular value spectrum, super tall building.

Table 5 summarizes the results of the frequentist and Bayesian statistics. It is seen that the MPV of S_i does not change significantly (even smaller than the laboratory example). The corresponding sample c.o.v. (A) is larger than the equivalent mean posterior c.o.v. (B). This is also true for the natural frequency, but less so for the damping ratios. The former may be understood as an under-estimation of ensemble variability by the posterior c.o.v. arising from modeling error, e.g., time-invariance. A theoretical interpretation of this observation is suggested in the Appendix.

As a reference, for all modes the c.o.v. of posterior variance among different data sets is 10–20% for f_i and ζ_i; 30–40% for S_i. No pattern is found between these values and the departure of A/B values from 1.

5.3. Super tall building

The subject structure in this example is a super tall building in Hong Kong, measuring 50 m-by-50 m in floor plan and over 300 m high. Ambient data on a normal day was obtained on the roof using a tri-axial force-balanced accelerometer (Guralp CMG5T). Eighteen hours of data sampled at 50 Hz are considered, from the evening to the next day morning. The data is divided into 36 data sets, each of 30 min. The duration of each data set is relatively long because the natural frequencies of the building are quite small (below 1 Hz). In fact this imposes a practical challenge on studying the insitu modal properties of super tall buildings. The data duration cannot be too short for identification precision reasons. On the other hand, extending the data duration runs the risk of modeling error related to response stationarity, especially during, e.g., a typhoon event.

Fig. 7 shows the singular value spectrum of a typical data set. The first two peaks (TX1, TY1) correspond to fundamental lateral translational modes. They are identified on the same frequency band with two modes assumed. The symbols R1 and R2 denote the torsional modes. Table 6 summarizes the frequentist and Bayesian statistics. The variability in the MPV of S_i (fourth column) is moderate. The large value of A/B for S_i is likely attributed to environmental variability over different data sets. The frequentist and Bayesian statistics are similar for the natural frequencies and damping ratios, although the frequentist statistics is consistently higher.

As a reference, the c.o.v. of posterior variance of f_i and ζ_i is about 50% for the first two modes, and 30% for the remaining modes. For S_i, the c.o.v. is about 100% for all modes. Although no significant pattern is found between these values and the departure of A/B from 1, the coincidence of large c.o.v. and A/B value for S_i is worth-noting.

6. Conclusions

This paper has investigated the relationship between the frequentist and Bayesian quantification of parameter uncertainty in system identification. It is motivated by the question of whether the Bayesian posterior covariance matrix

Table 6
Sample and Bayesian statistics, super tall building.

Parameter	Single set MPV	Sample mean of MPV	Freq.[a] (%) A	Bay.[a] (%) B	A/B
f_1 (Hz)	0.154	0.153	0.276	0.212	1.30
f_2	0.171	0.171	0.200	0.218	0.92
f_3	0.303	0.302	0.277	0.162	1.71
f_4	0.524	0.524	0.158	0.103	1.54
f_5	0.681	0.680	0.106	0.082	1.30
f_6	0.808	0.808	0.205	0.107	1.92
ζ_1 (%)	0.72	0.73	35.7	29.5	1.21
ζ_2	0.63	0.86	47.7	26.3	1.82
ζ_3	0.64	0.77	29.1	21.9	1.33
ζ_4	0.55	0.58	20.1	18.4	1.09
ζ_5	0.52	0.46	22.4	18.4	1.21
ζ_6	0.89	0.80	18.4	14.3	1.29
S_1 ($(\mu g)^2$/Hz)	6.63	20.80	66.5	10.6	6.25
S_2	14.32	32.47	73.7	11.1	6.61
S_3	0.17	0.48	57.8	11.6	4.98
S_4	0.29	0.56	51.9	8.1	6.41
S_5	0.25	0.36	59.1	9.2	6.44
S_6	0.049	0.09	57.5	10.2	5.61

[a] See Table 1.

in an ensemble average sense is equal to the ensemble covariance matrix of the most probable value of identified parameters. A general theory has been presented, which is based on the fundamental property of the likelihood function as a legitimate probability distribution. It makes use of a second order Taylor approximation of the log-likelihood function and is therefore approximate. The second order theory says that, in the absence of modeling error, the ensemble average of the posterior covariance is equal to the ensemble covariance of the MPV but in a weighted sense. Under some heuristic arguments the weighting can be removed, leading to the unweighted mean that is expected intuitively.

The exact relationship between the Bayesian and frequentist measure of uncertainty is highly non-trivial and depends on the particular identification problem. Theoretical results that can be sought are thus likely to be approximate. The second order theory developed in the paper on one hand allows a theoretical statement to be obtained, while retaining some important features, including the first two statistics of parameters. Of course, it is only applicable to identifiable situations. In most applications of modal identification, structural model updating or reliability analysis, it is sufficient to have an order of magnitude knowledge (e.g., 1%, 10%) about parameter uncertainty.

Focusing on modal identification, the weighted (second order theory) and unweighted (intuition) statistics have been examined using synthetic and experimental data. The synthetic data examples show that neither the weighted nor the unweighted statistics are exact, but they provide a good approximation with similar quality when there is little or no modeling error. In the examples with real data, the frequentist and Bayesian statistics are generally consistent to within an order of magnitude. Nevertheless, significant inconsistency can exist and it is believed that modeling error play an important role in this regard. We make no position as to whether the frequentist and Bayesian statistics are close in applications with real data, as it depends on the identification model, the environmental conditions, or even the nature of the parameter under question. Rather, the experimental observations serve to give an idea of the order of magnitude the two statistics in the test conditions stated (however qualitative). A frequentist explanation of the observations in the real data examples has been suggested in the Appendix, assuming that the modeling assumptions are still valid but the actual system properties may vary in different experiment trials. It should be borne in mind that this is by no means the only interpretation.

The experimental results highlight the role of the posterior covariance matrix of model parameters in Bayesian system identification. It reflects the residual uncertainty in the model parameters for a given data set and modeling assumptions. It can provide useful information for quality control, e.g., for deciding whether additional data is required to improve accuracy. Nevertheless it does not guarantee that additional sets of data will yield consistent identification results, even in the ensemble average sense. In reality testing conditions can change, or some of the modeling assumptions are simply wrong. In this regard, stating the model used in the Bayesian identification result statement is essential.

Acknowledgments

This paper is supported by General Research Fund 9041550 (CityU 110210) from the Research Grants Council of the Hong Kong Special Administrative Region, China. Mr Feng-Liang Zhang, PhD candidate at CityU, assisted in the testing of the laboratory frame and the super tall building. The author would like to thank Dr Alex To, Associate at Ove Arup &

Partners Hong Kong Ltd., for logistics help on the field test of super tall building. The author would also like to thank Dr. Kohler for providing the UCLA Factor Building data and generous assistance during the process. Constructive comments from anonymous reviewers are gratefully acknowledged.

Appendix. Interpreting observed variability by system change

It is observed in the real data examples in Section 5 that the sample c.o.v. of MPV (frequentist) tends to be greater than the equivalent mean posterior c.o.v. (Bayesian), especially for the PSD of modal force S_i. On the other hand the agreement is better in the damping ratio than in the natural frequency. At first glance this is somewhat surprising because the damping ratio is often much more difficult to estimate than the natural frequency and it has more modeling error with respect to, e.g., time-invariance, physical mechanism. In this Appendix we suggest a theoretical argument that attempts to explain the discrepancy between the frequentist and Bayesian statistics observed in the real data examples. The argument assumes that the discrepancy arises from the variability of the modal properties among different data sets, which violates the conditioning statement in our theoretical results. It is essentially a frequentist perspective, although it also makes use of the approximate equality between the frequentist and Bayesian statistics. It is by no means the only interpretation, however, for 'nature cannot be fooled' [21].

Assume that the model M is still correct but the set of 'true' system parameters $\boldsymbol{\theta}_0$ may vary in different experiment trials (data sets). That is, $\boldsymbol{\theta}_0$ is now a random variable that may take on a different value in different experiment trials. In a frequentist perspective there are thus two sources of uncertainty associated with the MPV of a particular (scalar) model parameter, say, $\hat{\theta}(\mathbf{z})$. The first comes from the fact that even if $\boldsymbol{\theta}_0$ and M are given, the data $\mathbf{z}$ is still uncertain and so is $\hat{\theta}(\mathbf{z})$. Mathematically this variability is $\mathrm{var}[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M]$ and is the frequentist quantity we have studied so far. The overall variability observed in real data is interpreted as $\mathrm{var}[\hat{\theta}(\mathbf{z})|M]$. It need not be equal to $\mathrm{var}[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M]$ because $\boldsymbol{\theta}_0$ is no longer deterministic. The uncertainty in $\boldsymbol{\theta}_0$ among different experiment trials adds to the second source of uncertainty since $\mathbf{z}$ is distributed as $p(\mathbf{z}|\boldsymbol{\theta}_0,M)$. To connect these two sources of uncertainty with the overall uncertainty observed in real data, we make use of the 'conditional variance formula', which is a standard result in classical probability [22]

$$\mathrm{var}[\hat{\theta}(\mathbf{z})|M] = E\{\mathrm{var}[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M]|M\} + \mathrm{var}\{E[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M]|M\} \tag{38}$$

To simplify, note from (28) that $E[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M] \approx \theta_0$ and so

$$\mathrm{var}\{E[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M]|M\} \approx \mathrm{var}[\theta_0|M] \tag{39}$$

On the other hand, note from (29) that $\mathrm{var}[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M] \approx E[\hat{V}(\mathbf{z})|\boldsymbol{\theta}_0,M]$, where $\hat{V}(\mathbf{z})$ is the posterior variance of θ given the data $\mathbf{z}$ and model M. This means that

$$E\{\mathrm{var}[\hat{\theta}(\mathbf{z})|\boldsymbol{\theta}_0,M]|M\} \approx E\{E[\hat{V}(\mathbf{z})|\boldsymbol{\theta}_0,M]|M\} = E[\hat{V}(\mathbf{z})|M] \tag{40}$$

Consequently, (38) becomes

$$\mathrm{var}[\hat{\theta}(\mathbf{z})|M] \approx E[\hat{V}(\mathbf{z})|M] + \mathrm{var}[\theta_0|M] \tag{41}$$

That is, in the case of uncertain $\boldsymbol{\theta}_0$, the ensemble variance of the MPV among different experiment trials is approximately equal to the sum of the mean posterior variance and the variance of θ_0. Note that $\mathrm{var}[\hat{\theta}(\mathbf{z})|M]$ and $E[\hat{V}(\mathbf{z})|M]$ do not depend on the unknown $\boldsymbol{\theta}_0$ and so they can be estimated in real experiments. In particular, $\mathrm{var}[\hat{\theta}(\mathbf{z})|M]$ can be estimated by the sample variance of $\hat{\theta}(\mathbf{z})$ over different experiment trials; $E[\hat{V}(\mathbf{z})|M]$ by the sample mean of $\hat{V}(\mathbf{z})$. In contrast, $\mathrm{var}[\theta_0|M]$ can never be directly estimated from real experiments.

Eq. (41) is in fact quite intuitive. In the current context, it says that neither the (Bayesian) posterior variance nor the (frequentist) sample variance describes directly the variability of the parameter θ_0 among the repeated experiments. The posterior variance largely reflects estimation error. The sample variance of MPV over-estimates the variability of θ_0 because it is contaminated by estimation error. The correct estimation is simply the difference between the sample variance and the mean of the posterior variance.

If we normalize (41) by the square of $\overline{\theta} = E[\hat{\theta}|M]$, then it says that

$$(c.\text{o.v. of MPV})^2 = (\text{equiv. mean posterior c.o.v.})^2 + (c.\text{o.v. of } \theta_0)^2 \tag{42}$$

That is, in the context of Tables 4–6,

$$A^2 \approx B^2 + (\text{c.o.v. of } \theta_0)^2 \tag{43}$$

The above arguments are approximate and is still conditional on the model M, which is necessary to maintain the legitimacy of $\boldsymbol{\theta}$.

Based on (42), we now attempt to explain the observations in the real data examples. Generally, the sample c.o.v. of MPV (A) tends to be greater than the equivalent mean posterior c.o.v. (B) because of the nonzero variability in θ_0. For the natural frequency f_i, note that the posterior c.o.v. is quite small. For the factor building and super tall building example, it is likely that the posterior c.o.v. (B) is of similar order of magnitude as the c.o.v. of θ_0 and so A/B is significantly greater than 1. A similar argument can be applied to the PSD of modal force S_i. Although by absolute magnitude the posterior c.o.v. of S_i is not considered small, it could be of similar magnitude or even small than the c.o.v. of θ_0, rendering again A/B significantly

greater than 1. For the damping ratio, it could be that the posterior c.o.v. is not small even compared to the c.o.v. of θ_0, and so it dominates A, rendering $A/B \approx 1$. Thus, the fact that the frequentist and Bayesian statistics appear to agree for the damping ratio could just be a direct consequence of its relatively lower identification precision compared to the underlying variability. Last but not the least, we remark that these explanations are all conditional on M, reminding that they are only as good as their assumptions and validity of their context.

References

[1] J.L. Beck, Bayesian system identification based on probability logic, Struct. Control Health Monit. 17 (7) (2010) 825–847.
[2] K.V. Yuen, Bayesian Methods for Structural Dynamics and Civil Engineering, John Wiley & Sons, 2010.
[3] C. Papadimitriou, J.L. Beck, L.S. Katafygiotis, Updating robust reliability using structural test data, Probab. Eng. Mech. 16 (2) (2001) 103–113.
[4] E.T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.
[5] S.K. Au, F.L. Zhang, On assessing posterior mode shape uncertainty in ambient modal identification, Probab. Eng. Mech. 26 (3) (2011) 427–434.
[6] S.K. Au, F.L. Zhang, Ambient modal identification of a primary–secondary structure by fast Bayesian FFT method, Mech. Syst. Signal Process. (2011), doi:10.1016/j.ymssp.2011.07.007.
[7] B. Peeters, G. De Roeck, Stochastic system identification for operational modal analysis: a review, J. Dyn. Syst. Meas. Control 123 (2001) 659–667.
[8] G.H. James, T.G. Carne, J.P. Lauffer, The natural excitation technique (NExT) for modal parameter extraction from operating structures, J. Anal. Exp. Modal Anal. 10 (2) (1995) 260–277.
[9] J.M.W. Brownjohn, F. Magalhaes, A. Cunha, Ambient vibration re-testing and operational modal analysis of the Humber Bridge, Eng. Struct. 32 (8) (2010) 2003–2018.
[10] E. Reynders, R. Pintelon, G. De Roeck, Uncertainty bounds on modal parameters obtained from stochastic subspace identification, Mech. Syst. Signal Process. 22 (2007) 948–969.
[11] K.V. Yuen, L.S. Katafygiotis, Bayesian Fast Fourier, Transform approach for modal updating using ambient data, Adv. Struct. Eng. 6 (2) (2003) 81–95.
[12] S.K. Au, Fast Bayesian FFT method for ambient modal identification with separated modes, J. Eng. Mech. ASCE 137 (3) (2011) 214–226.
[13] S.K. Au, Fast Bayesian ambient modal identification in the frequency domain. Part I: posterior most probable value, Mech. Syst. Signal Process. 26 (1) (2012) 60–75.
[14] S.K. Au, Fast Bayesian ambient modal identification in the frequency domain. Part II: posterior uncertainty, Mech. Syst. Signal Process. 26 (1) (2012) 76–90.
[15] R. Brincker, L. Zhang, P. Anderson, Modal identification of output-only systems using frequency domain decomposition, Smart Mater. Struct. 10 (3) (2001) 441–455.
[16] K.V. Yuen, S.C. Kuok, Ambient interference in long-term monitoring of buildings, Eng. Struct. 32 (8) (2010) 2379–2386.
[17] P. Moser, B. Moaveni, Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge, Mech. Syst. Signal Process. 25 (7) (2011) 2336–2357.
[18] M.D. Kohler, P.M. Davis, E. Safak, Earthquake and ambient vibration monitoring of the steel-frame UCLA Factor building, Earthquake Spectra 21 (3) (2005) 715–736.
[19] M.D. Kohler, T.H. Heaton, S.C. Bradford, Propagating waves in the steel, moment-frame factor building recorded during earthquakes, Bull. Seismol. Soc. Amer. 97 (4) (2007) 1334–1345.
[20] D. Skolnik, Y. Lei, E. Yu, J.W. Wallace, Identification, model updating, and response prediction of an instrumented 15-story steel-frame building, Earthquake Spectra 22 (3) (2006) 781–802.
[21] R.P. Feynman, The Pleasure of Finding Things Out, Helix Books, 1999.
[22] S.M. Ross, Introduction to Probability Models, 8th edition, Academic Press, San Diego, USA, 2003.