Treatment and effect of noise modelling in Bayesian operational modal analysis

Xinda Ma*, Zuo Zhu, Siu-Kui Au

School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

* Corresponding author, E-mail: xinda001@e.ntu.edu.sg

Abstract

Operational modal analysis (OMA) identifies the modal properties, e.g., natural frequencies, damping ratios and mode shapes, of a structure using 'output-only' ambient vibration data. Instrument noise need not be negligible in ambient vibration data, and it is often modelled statistically. Simple noise models, e.g., independent and identically distributed (i.i.d.) among data channels, are often used and are found to give reasonable results in typical applications, although there may be concerns for data with, e.g., low signal-to-noise (s/n) ratio, large difference in noise intensities or significant correlation among data channels. This work aims at investigating the effect of noise models on OMA performed with a Bayesian approach in the frequency domain. In addition to modal identification results, noise models are also assessed from a Bayesian evidence perspective. To enable the study, algorithms for efficient calculation of Bayesian statistics (most probable value and covariance matrix) are developed to account for general noise models that have not been considered in existing algorithms. As a further contribution to OMA theory, it is shown that, by a suitable transformation of data, an OMA problem with general noise model can be converted to one with i.i.d. noise model. Based on this analogy, asymptotic formulae for identification uncertainty of modal parameters, i.e., 'uncertainty law', have been developed. The theory reveals a definition for the modal s/n ratio that is an intuitive yet nontrivial generalisation of the existing formula for i.i.d. noise. The proposed objectives and methodology are investigated in a comprehensive study through synthetic, laboratory and field data.

Keywords

Operational modal analysis; BAYOMA; Noise disparity; Model class selection; Ambient modal identification; Uncertainty law

1. Introduction

Operational modal analysis (OMA) is becoming a popular economical means for identifying the modal properties (modal frequencies, damping ratios, mode shapes, etc.) of a constructed structure based on 'output-only' vibration measurements under ambient conditions [1–4]. In the absence of input excitation information, the theory and computations for modal identification are more complicated than their counterparts based on 'free' or 'forced' vibration data [5], where the excitation that dominates the measured response is either zero or measured, respectively. Modal identification using ambient vibration data often requires statistical concepts, in the first place, to model the unmeasured input excitation. The instrument noise, which arises from the sensor and data transmission medium (e.g., electronics, cables), also needs to be modelled statistically because it is no longer negligible especially when the ambient vibration is low. Modelling assumptions can differ, depending on whether it operates in the time or frequency domain. In time-domain methods, temporally white noise assumption is often applied to simplify theory and computation, where the statistics of excitation and noise can be described by a covariance matrix. Examples include Stochastic Subspace Identification (SSI) [6–8], Second-Order Blind Identification (SOBI) [9,10], and Bayesian time-domain method [11]. How noise is treated can differ from one method to another. Noise is assumed to be uncorrelated in SSI so that it can be averaged out statistically when obtaining the estimate of modal parameters. It also allows uncertainty to be derived by means of perturbation. Statistical characteristics of noise are explicitly quantified in SOBI and Bayesian time-domain method. In frequency-domain methods, the excitation and noise are commonly characterised via the power spectral density (PSD) matrix [3,12], which is often assumed to be constant at least near the natural frequencies of the modes of interest. Examples include the Frequency Domain Decomposition (FDD, and variants) methods [13,14] and Bayesian OMA methods in the frequency domain (BAYOMA) [3], where the noise is assumed to be independent among data channels. Some OMA methods do not model noise explicitly, or rather eliminate them by heuristic regression/averaging, e.g., the methods based on Auto-Regressive (AR) model (see Section 9.2 in [1] and the discussion on AR models in [8]), and random decrement method [15–17].

While the primary target of OMA is on structural modal properties, the statistical modelling of excitation and noise may affect the identification results, and hence deserve attention especially in challenging situations, e.g., 1) when the signal-to-noise (s/n) ratio is low; 2) when there is a significant 'disparity' in noise intensity among different data channels; 3) when the contribution

from unaccounted modes or unknown dynamics is not negligible. When the s/n ratio is low, the noise is no longer negligible and hence its statistical properties will influence those of the measured vibration data. Disparity in noise intensity refers to the difference in the noise level, e.g., in terms of PSD, between different measured data channels. While it is inevitable and does not matter in typical situations when the s/n ratio is high, disparity by orders of magnitude can be a significant violation of modelling assumptions and can lead to bias in modal identification results. Noise disparity is present in modern sensing networks with a variety of sensors of different types and noise qualities. For example, the ambient vibration test of Humber Bridge [18] utilised both piezoelectric and servo-accelerometers, of which the self-noise levels can differ by one or two orders of magnitude. For the main span and south tower of the Golden Gate Bridge [19], a low-cost MEMS accelerometer (ADXL 202E, noise $200\ \mu g/\sqrt{Hz}$) was used to monitor strong shaking while the SiliconDesigns 1221L accelerometer (MEMS, noise $32\mu g/\sqrt{Hz}$) was used for capturing low-amplitude responses. Besides accelerometers, data from velocimeters [20–22] and gyroscopes [23] are sometimes incorporated for analysis, which is another source of noise disparity, i.e., velocimeters tend to have lower noise for lower frequencies than accelerometers.

Despite the fact that the instrument noise at different data channels can often be considered statistically uncorrelated, the same cannot be taken for granted for the 'modelled noise', i.e., the residual difference between the measured data and the modelled counterpart. One common source is unmodelled dynamics. For example, in frequency-domain methods making use of FFT in a selected frequency band for modal identification (BAYOMA) [3,24], the smearing contributions from modes not modelled in the band will tend to increase the residual difference between the data and best fitted response, and hence appear as noise in the identification process and results. Arising from unaccounted modes with a spatial pattern due to mode shapes, such 'noise' contributions are correlated among different measured degrees of freedom (DOFs).

Although not the mainstream topic in OMA, there are studies that investigated issues related to noise or its modelling. For example, simple methods have been developed for determining instrument noise PSD in a 'huddle test' using three sensors with different requirements on alignment [25,26]. The effect of noise with different sensors on modal parameters identified by Eigensystem Realisation Algorithm (ERA) was investigated based on ambient data from the Golden Gate Bridge [27]. It was reported that lower sensor noise level leads to more accurate results, which agrees with intuition. In [28], a series of numerical examples based on FDD reveal that damping ratio estimates can be significantly biased when the s/n ratio is low. Besides noise

level, [29] investigated the effect of noise disparity in covariance-based SSI and BAYOMA. The study suggests that the quality of estimates can be affected when the noise disparity is high, especially for BAYOMA where i.i.d. noise is assumed.

Motivated by the above and related considerations, this paper aims at investigating the effect of noise model on OMA. The focus is on Bayesian OMA approach in the frequency domain (BAYOMA) [3,24], where the Fast Fourier Transform (FFT) of data within a selected frequency band is used for identifying the mode(s) dominating the band. In this context, existing formulations assume that the instrument noise at different DOFs are independent and identically distributed (i.i.d.) with a common value of PSD within the selected band around the mode(s) of interest. Such simple model has been found to be robust in typical applications with reasonable s/n ratios [30–32], although recent applications [23,29,33] motivated detailed investigation in potentially challenging situations with regard to the aforementioned issues.

The key contributions and organisation of this work are outlined as follows. Beyond the existing noise model, we consider in Section 2 two noise models with increasing complexity, i.e., one with independent but possibly different noise PSDs to account for noise disparity, and another characterised by a general covariance matrix to account for correlation as well. The noise models are assessed in terms of modal identification results, as well as from the (abstract) perspective of Bayesian model selection [34–36] (Section 3.5). As a further contribution to OMA theory, we establish a mathematical analogy (Section 4) that bridges an OMA problem based on general noise model with the existing one based on i.i.d. noise. This allows us to obtain explicit analytical expressions for the asymptotic identification uncertainty of modal parameters in terms of test configuration, i.e., ‘uncertainty law’, which is otherwise unlikely to be possible based on existing techniques applicable only for i.i.d. noise [37,38]. Indispensable to this work, efficient algorithms for Bayesian OMA with the general noise models are developed to enable the investigation of their identification results and Bayesian evidence, which is otherwise computationally prohibitive. These include the Expectation-Maximisation (EM) algorithm (Sections 3.1 - 3.3 and Appendix B) for the most probable values (MPV) of modal parameters; and analytical expressions for the Hessian of the log-likelihood function (Section 3.4 and Appendix C), and hence the posterior covariance matrix of modal parameters. The objectives and proposed methodology will be investigated through a comprehensive study based on synthetic data (Section 5) and laboratory/field data (Section 6).

2. Noise models for frequency-domain OMA

In this section, we introduce and discuss the noise models that will be focused on in this work. As an outline, Section 2.1 reviews briefly the context of Bayesian frequency domain OMA, where the role of the noise model is introduced. Section 2.2 discusses the conventional noise model based on i.i.d. assumption, which motivates more general ones to be investigated. The methodology for assessing the noise models is then outlined, which will be in terms of modal identification results as well as Bayesian model class evidence. Computational tools that enable the proposed study will be developed in Section 3 later.

2.1. Bayesian frequency domain OMA

Let $\{\widehat{\boldsymbol{y}}_j\}_{j=0}^{N-1}$ (each $n \times 1$) be the ambient acceleration data measured at n DOFs of a structure, where N is the number of samples per data channel. Its (one-sided) scaled FFT at frequency $\mathrm{f}_k = k/N\Delta t$(Hz) (up to Nyquist frequency) is defined as $\widehat{\boldsymbol{\mathcal{F}}}_k = \sqrt{2\Delta t/N}\sum_{j=0}^{N-1}\widehat{\boldsymbol{y}}_j e^{-2\pi\mathbf{i}jk/N}$, where Δt (s) is the sampling time interval and $\mathbf{i}^2 = -1$. Consider identifying the modes of interest using the scaled FFT within a selected frequency band, denoted by $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$. Within the band, the scaled FFT is modelled to have contributions from modal response and noise, i.e., $\widehat{\boldsymbol{\mathcal{F}}}_k = \boldsymbol{\Phi}\boldsymbol{\eta}_k + \boldsymbol{\varepsilon}_k$, where $\boldsymbol{\varepsilon}_k$ $(n \times 1)$ is the scaled FFT of noise with PSD matrix $\boldsymbol{S}_e$; $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1, \dots, \boldsymbol{\varphi}_m]$ and $\boldsymbol{\varphi}_i$ $(n \times 1)$ is the i-th mode shape covering the measured DOFs and assumed to have a unit norm, i.e., $\boldsymbol{\varphi}_i^T\boldsymbol{\varphi}_i = 1$; $\boldsymbol{\eta}_k = \boldsymbol{h}_k\boldsymbol{p}_k$ $(m \times 1)$ is a vector of the scaled FFT of modal acceleration responses, with $\boldsymbol{h}_k$ being a diagonal matrix of frequency response functions (FRF) $h_{ik} = (1 - \beta_{ik}^2 - 2\zeta_i\beta_{ik}\mathbf{i})^{-1}$, $\beta_{ik} = f_i/\mathrm{f}_k$, f_i and ζ_i are respectively the natural frequency and damping ratio (assuming classical damping); $\boldsymbol{p}_k$ $(m \times 1)$ is a vector of the scaled FFT of modal excitations (per unit mass, same throughout). The modal excitations are assumed to be stochastic stationary with a constant PSD matrix of $\boldsymbol{S}$ within the selected band. They are also independent from the noise.

Let $\boldsymbol{\theta} = \{\boldsymbol{f}, \boldsymbol{\zeta}, \boldsymbol{\Phi}, \boldsymbol{S}, \boldsymbol{S}_e\}$ denote the set of modal parameters to be identified; $\boldsymbol{f}$ and $\boldsymbol{\zeta}$ are vectors collecting the frequencies and damping ratios, respectively. Using Bayes' theorem and assuming a uniform prior distribution for $\boldsymbol{\theta}$, the posterior probability density function (PDF) of $\boldsymbol{\theta}$ given the scaled FFT $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ in the band is proportional to the likelihood function, i.e.,

$$p\big(\boldsymbol{\theta}|\{\widehat{\boldsymbol{\mathcal{F}}}_k\}\big) \propto p\big(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}|\boldsymbol{\theta}\big) = \frac{\pi^{-nN_f}}{\prod_k|\boldsymbol{E}_k|}\exp\big[-\textstyle\sum_k \widehat{\boldsymbol{\mathcal{F}}}_k^*\boldsymbol{E}_k^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k\big] \tag{1}$$

where $|\cdot|$ denotes the matrix determinant and a superscribed $*$ denotes a complex conjugate; the summation and product are over all FFT points in the selected band, whose number is N_f; and

$$\boldsymbol{E}_k = \boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^T + \boldsymbol{S}_e \tag{2}$$

is the theoretical PSD matrix of data; $\boldsymbol{H}_k = \boldsymbol{h}_k\boldsymbol{S}\boldsymbol{h}_k^*$. The likelihood function $p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}|\boldsymbol{\theta}\right)$ in (1) follows from the long-data asymptotic result that the FFTs of a stationary stochastic process follow a (circular symmetric) complex Gaussian distribution and are independent at different frequencies (see, e.g., Section 4.8 in [3]). The centre of the posterior PDF, often called the 'most probable value' (MPV) $\widehat{\boldsymbol{\theta}}$, gives the 'best estimate' of modal parameters. The covariance matrix $\widehat{\boldsymbol{C}}$ of the PDF quantifies the remaining uncertainty of modal parameters. Mathematically, the MPV maximises the likelihood function; the covariance matrix is equal to the inverse of the Hessian of the negative log-likelihood function at the MPV.

As a literature note, Bayesian OMA based on FFT of a selected band was first formulated in [24]. Efficient algorithms for determining the MPV and covariance matrix were later developed in a series of papers for well-separated/closely spaced modes, and single/multiple setups. See the first paper in the series [39]; and a recent monograph [3] that introduces the topic with rationale for the context and summarises the major developments. Early algorithms for MPV were based on iterations with partial solutions derived by taking advantage of the i.i.d. noise model. EM algorithm [40] has been exploited recently to produce simpler algorithms that are easier to computer-code, e.g., [41] for multi-modes with single setup data, and [42] for multi-modes with multi-setup data. Field applications can be found in, e.g., [30–32] for buildings, [43,44] for footbridges and [45,46] for vehicle bridges. Asymptotic expressions have been derived to provide insights for how the identification uncertainty of modal parameters depends on test configurations; see the latest findings in [47].

2.2. Noise models

In the context of Section 2.1, the modelling of noise affects the PSD matrix $\boldsymbol{S}_e$ in $\boldsymbol{E}_k$ of (2), which is the only term through which the modal parameters affect the likelihood function. Conventionally, the noise at different measured DOFs are assumed to be i.i.d. with the same (scalar-valued) PSD S_e, i.e.,

$$\boldsymbol{S}_e = S_e\boldsymbol{I}_n \qquad \text{(Model 0, i.i.d. noise)} \tag{3}$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. While the independence among data channels can be

justified physically, mathematical simplicity is the key reason for a common PSD S_e, as it turns out to simplify theory and computation. In particular, a single parameter model is the simplest possible form. Having $\boldsymbol{S}_e$ as a multiple of the identity matrix allows the determinant and inverse of $\boldsymbol{E}_k$ in (2) to be expressed via an eigenvector expansion, which led to the early iterative algorithms for MPV [39,48]. It also allows asymptotic expressions for posterior covariance to be derived in explicit closed form, i.e., 'uncertainty law' [37], providing understanding of identification uncertainty.

Physically, a common PSD S_e for all channels is an assumption that is likely to be violated. Despite this apparent violation, typical applications with moderate to higher s/n ratios reveal that the identification results of the parameters of primary interest (frequency, damping, mode shape) are still reasonable and somewhat insensitive to the noise level. Naturally, doubts remain when the s/n ratio is not high, and more generally, what the effects are on the identification results when the i.i.d. noise model is violated. For example, when there is a large disparity between the noise PSDs at different channels, a bias may result when (3) effectively enforces an equality constraint among the PSDs. Does this bias smear into other modal parameters?

As an illustration, consider acceleration data with two measured DOFs synthetically generated from a single mode with a natural frequency of 1 Hz and damping ratio of 1%. Channels 1 and 2 are contaminated with noise of PSDs of $500(\mu g)^2/\text{Hz}$ and $100(\mu g)^2/\text{Hz}$, respectively. The squared modulus of the scaled FFT of Channels 1 and 2 are plotted as grey lines in Fig. 1 (a) and (b), respectively. The black line shows the theoretical ('exact') PSD values according to (2). It fits through the grey lines as expected. The dashed line shows the PSD of noise modelled. Using the FFT in the band $0.9 - 1.1$ Hz to identify the modal properties (Section 2.1) with an i.i.d. noise model in (3), the MPV of modal parameters are calculated. Substituting the MPV into the theoretical PSD in (2) gives the red line in Fig. 1. The dashed red line shows the noise level corresponding to the MPV. Visually, the red lines based on a common S_e model tend to 'merge' the different noise levels, i.e., under-estimating the noise level in Channel 1 and over-estimating in Channel 2. Of higher relevance is the smearing effect (if any) of this bias on other modal parameters, i.e., no apparent bias in frequency, but possibly in damping ratio, mode shape and modal force PSD.

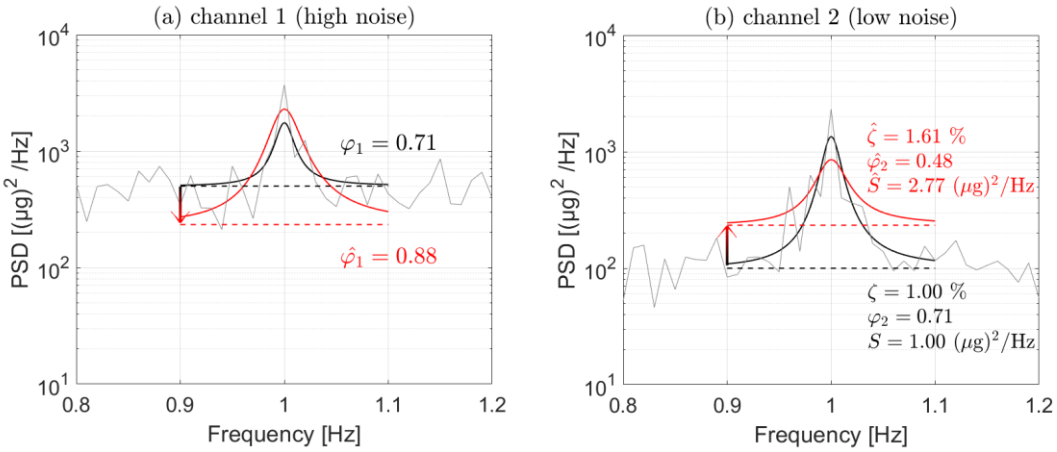


Fig. 1 Illustration of potential bias with i.i.d. noise model for channels with significant noise disparity. (a) Channel 1 (high noise); (b) Channel 2 (low noise). Black and red solid lines show the theoretical ('exact') and 'best fitted' PSD through FFT data (grey lines). Dashed lines show the noise PSD levels.

To investigate the effect of noise model, in this work we consider two models that may be considered as successive generalisations of the i.i.d. model. They are summarised in Table 1. Model 1 allows for non-uniform noise PSDs at different channels, i.e.,

$$\boldsymbol{S}_e = \begin{bmatrix} S_{e1} & & \\ & \ddots & \\ & & S_{en} \end{bmatrix} \qquad \text{(Model 1, independent noise)} \qquad (4)$$

where S_{ei} is the noise PSD of data channel i. Model 2 further generalises by allowing correlation among data channels, with a general real-symmetric matrix for the noise PSD matrix:

$$\boldsymbol{S}_e = \begin{bmatrix} S_{e11} & & \\ \vdots & \ddots & \\ S_{en1} & \dots & S_{enn} \end{bmatrix} \qquad \text{(Model 2, correlated noise)} \qquad (5)$$

Table 1 Summary of noise models considered in this work; n = number of measured DOFs, m = number of modes in the selected frequency band

Model	Assumption	Noise PSD $\boldsymbol{S}_e$	No. of parameters in noise model	Total no. of free parameters in Bayesian OMA
0	i.i.d.	$S_e \boldsymbol{I}_n$	1	$(m+1)^2 + m(n-1)$
1	independent	Diagonal matrix of $\{S_{e1}, \dots, S_{en}\}$	n	$(m+1)^2 + m(n-1) - 1 + n$
2	correlated	Positive-definite symmetric matrix	$\frac{n(n+1)}{2}$	$(m+1)^2 + m(n-1) - 1 + \frac{n(n+1)}{2}$

From first glance, as independence of noise among different channels can be justified physically, Model 1 seems to be sufficiently general and Model 2 seems unnecessary. Model 2 is considered rather to allow for the 'apparent' noise in the residual (difference) between the theoretical model and data that may arise from modelling error, besides instrument noise stemming physically from hardware. In particular, within the selected frequency band around the mode of interest, there can be contributions from other modes that are not modelled. Such contributions will generally be smeared as 'noise' in the modal identification process. As this 'noise' need not be consistent with the assumed noise statistics, the effect on identification results is generally unknown. Arising from modal response, this 'noise' will have a spatial correlation because of mode shape, hence calling for a general noise PSD matrix in (5). Note that 'correlation' is used in a general context for its intuitive meaning, as strictly speaking it is the 'coherence' (in the frequency domain) that is introduced in Model 2, rather than the correlation (in the time domain).

As a remark, with a little loss of generality, the noise PSD matrix in Model 2 is assumed to be real symmetric rather than Hermitian, although theoretically the latter is the most general form. This is intended to strike a balance between generality and identifiability. In particular, a Hermitian $\boldsymbol{S}_e$ has additional parameters (the imaginary parts of off-diagonal entries) that may not be identifiable, depending on the amount and quality of available data. Numerical studies with synthetic and experimental data revealed that when the parameters are identifiable, i.e., when the iterative algorithm (see Section 3.1 later) for MPV converges properly, the identification results assuming a real symmetric $\boldsymbol{S}_e$ are qualitatively similar to their counterparts assuming a Hermitian $\boldsymbol{S}_e$.

Towards the scientific objectives of this work, we shall investigate the noise models (0, 1, 2) in the context of Bayesian OMA (Section 2.1). The appraisal will primarily be in terms of modal identification results and Bayesian model class evidence. Theoretical background for the latter is postponed to Section 3.5 later. The first perspective assesses the noise models in a pragmatic way through case-specific results with a focus on potential bias, which is of direct concern in applications. The second perspective provides a way to compare the noise models in a more holistic, though abstract, manner based on probabilistic evidence from data formulated in a Bayesian context. These two perspectives are complementary. For example, modal identification results for a particular case, though directly relevant to the parameters of interest, do not alone provide high level information about the quality of the model that is being used. On the other hand, Bayesian evidence informs holistically about the model, but not a particular set of modal identification results, or a particular parameter of interest. For example, one may be more

concerned about the bias in the estimates of natural frequency than noise PSD, but both (as well as other) parameters are reflected in the value of Bayesian evidence.

Assessing the noise models in the proposed framework requires efficient calculation of the posterior statistics of Bayesian OMA and model class evidence. This will be addressed next in Section 3.

3. Bayesian OMA and model class evidence for general noise models

In this section, computational tools will be developed for efficient Bayesian OMA and model class evidence calculation with general noise models (Table 1). Of central focus is the determination of posterior MPV and covariance matrix. An EM algorithm is developed in Sections 3.1-3.3 for the MPV. Analytical formulae are developed in Section 3.4 (with further details in Appendix C) for the Hessian matrix of log-likelihood function and hence posterior covariance matrix. The calculation of Bayesian modal class evidence is also outlined in Section 3.5.

To clarify contributions, the idea of EM for determining MPV is not new, but the new noise models (1 and 2) demand analytical work and verification to produce a proper algorithm and computer code that can be applied in a robust manner. The resulting algorithms share similarity with existing ones based on i.i.d. noise (Model 0) [41], but it has important modifications and generalisations, e.g., (17) and (18), that make it applicable for general noise models (1 and 2). Similar comments apply to the formulae for the derivatives of the log-likelihood function in Section 3.4 and Appendix C.

3.1. EM algorithm for MPV

EM algorithm [40,49] is an iterative approach in statistical inference for determining the MPV that maximises the likelihood function. In OMA, EM algorithms have been previously developed in the time-domain [50] (based on white noise) and in the frequency domain [41,42] (based on Model 0). The EM algorithm to be developed here generalises over [41] to be applicable for general noise models (1 and 2). The theory and algorithm are outlined for completeness, but otherwise the focus will be on new contributions that are necessary to accommodate the general noise models.

Starting from an initial guess $\boldsymbol{\theta}^{(0)}$, the EM algorithm finds the MPV $\widehat{\boldsymbol{\theta}}$ by iterative updates $\boldsymbol{\theta}^{(1)}$, $\boldsymbol{\theta}^{(2)}$, …, until convergence. Given the current estimate $\boldsymbol{\theta}^{(t)}$ at iteration t (an integer), the next

estimate $\boldsymbol{\theta}^{(t+1)}$ is determined as the one that maximises the so-called Q-function:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}) = \mathrm{E}_{\{\boldsymbol{\eta}_k\}|\{\widehat{\boldsymbol{\mathcal{F}}}_k\},\boldsymbol{\theta}^{(t)}}\left[L\left(\boldsymbol{\theta}|\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\right)\right] \tag{6}$$

As indicated in the notation, this is the conditional expectation of the 'complete-data log-likelihood' $L\left(\boldsymbol{\theta}|\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\right) = \ln p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}|\boldsymbol{\theta}\right)$ where $\{\boldsymbol{\eta}_k\}$ is distributed conditional on $\{\widehat{\boldsymbol{\mathcal{F}}}_k\}$ and $\boldsymbol{\theta}^{(t)}$. Here, $\boldsymbol{\eta}_k$ is a 'latent' variable which has been chosen so that the Q-function can be calculated efficiently (preferably in analytical form) and reasonable convergence pace can be achieved. The former is important because each iteration requires optimisation of the Q-function with respect to (w.r.t.) $\boldsymbol{\theta}$. In addition to basic iterations, 'parabolic extrapolation' [51,52] is used to speed up convergence.

3.2. Q-function and update formulae

Indispensable to the EM algorithm is the formulation of the Q-function as well as an efficient maximisation strategy to give the next iterative estimate. The basic logic to be presented here follows that for i.i.d. noise [41]. Key formulae in the algorithm will be outlined first. New contributions for general noise models beyond the previous work for i.i.d. noise will be highlighted later.

As suggested by (6), deriving the Q-function involves first obtaining the complete-data likelihood and then taking conditional distribution. Using $p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}|\boldsymbol{\theta}\right) = p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}|\{\boldsymbol{\eta}_k\}, \boldsymbol{\theta}\right)p(\{\boldsymbol{\eta}_k\}|\boldsymbol{\theta})$ and analysing the individual conditional distributions, the complete-data log-likelihood can be obtained in the following form that separates the effects of mode shapes ($\boldsymbol{\Phi}$) from those of the remaining parameters (see Appendix B.1)

$$\begin{aligned} L\left(\boldsymbol{\theta}|\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\right) &= \ln p\left(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}|\boldsymbol{\theta}\right) \\ &= -2nN_f \ln\pi - N_f \ln|\boldsymbol{S}_e| - \textstyle\sum_k \boldsymbol{\eta}_k^* \mathbf{H}_k^{-1} \boldsymbol{\eta}_k + \sum_k \ln|\mathbf{H}_k^{-1}| \\ &\quad - \textstyle\sum_k \left[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi}\boldsymbol{\eta}_k\right]^* \boldsymbol{S}_e^{-1} \left[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi}\boldsymbol{\eta}_k\right] \end{aligned} \tag{7}$$

On the other hand, given $\left\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\theta}\right\}$, it can be shown that (see Appendix B.1) the conditional mean and covariance matrix of latent variable $\boldsymbol{\eta}_k$ are given by, respectively:

$$\mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)}}[\boldsymbol{\eta}_k] = \boldsymbol{P}_k^{-1}\boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k \tag{8}$$

$$\mathrm{Cov}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)}}[\boldsymbol{\eta}_k] = \boldsymbol{P}_k^{-1} \tag{9}$$

where

$$\boldsymbol{P}_k = \boldsymbol{H}_k^{-1} + \boldsymbol{\Phi}^T \boldsymbol{S}_e^{-1} \boldsymbol{\Phi} \tag{10}$$

Taking conditional expectation $\mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)}}[\cdot]$ on (7), expanding the quadratic terms and using (8) to (10) gives the Q-function:

$$Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = -2nN_f \ln\pi - N_f \ln|\boldsymbol{S}_e| - Q_1\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) + Q_2\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) \tag{11}$$

where

$$Q_1\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = \mathrm{tr}\left[\boldsymbol{S}_e^{-1}\left[\sum_k \widehat{\boldsymbol{\mathcal{F}}}_k \widehat{\boldsymbol{\mathcal{F}}}_k^* - 2\boldsymbol{\Phi}\sum_k \mathrm{Re}(\boldsymbol{w}_{1k}\widehat{\boldsymbol{\mathcal{F}}}_k^*) + \boldsymbol{\Phi}\sum_k \mathrm{Re}(\boldsymbol{w}_{2k})\,\boldsymbol{\Phi}^T\right]\right] \tag{12}$$

$$Q_2\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = \sum_k \ln|\boldsymbol{H}_k^{-1}| - \mathrm{tr}(\sum_k \boldsymbol{H}_k^{-1}\boldsymbol{w}_{2k}) \tag{13}$$

Here, $\boldsymbol{w}_{1k} = \mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)}}[\boldsymbol{\eta}_k]$ and $\boldsymbol{w}_{2k} = \mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}^{(t)}}[\boldsymbol{\eta}_k\boldsymbol{\eta}_k^*]$ are respectively the first and second conditional moments of $\boldsymbol{\eta}_k$ given $\widehat{\boldsymbol{\mathcal{F}}}_k$ and $\boldsymbol{\theta}^{(t)}$; $\boldsymbol{w}_{1k}$ can be obtained directly from (8); $\boldsymbol{w}_{2k}$ can be obtained as the sum of covariance in (9) and $\boldsymbol{w}_{1k}\boldsymbol{w}_{1k}^*$.

The update formulae for $\boldsymbol{S}$, $\boldsymbol{\Phi}$ and $\boldsymbol{S}_e$ can be obtained analytically by zeroing the corresponding partial derivatives of the Q-function in (11). The resulting formulae for $\boldsymbol{\Phi}$ and $\boldsymbol{S}$ are given by

$$\boldsymbol{\Phi} = \sum_k \mathrm{Re}(\widehat{\boldsymbol{\mathcal{F}}}_k \boldsymbol{w}_{1k}^*)\,[\mathrm{Re}(\sum_k \boldsymbol{w}_{2k})]^{-1} \tag{14}$$

$$\boldsymbol{S} = \frac{1}{N_f}\sum_k \boldsymbol{h}_k^{-1}\boldsymbol{w}_{2k}\boldsymbol{h}_k^{-*} \tag{15}$$

For $\boldsymbol{S}_e$, it should be updated as $\boldsymbol{S}_e = \boldsymbol{S}_{e\mathrm{i}}$, where $\mathrm{i} = 0, 1$ or 2, depending on the noise model (see Appendix B.2 for proof):

$$\boldsymbol{S}_{e0} = \frac{1}{n}\mathrm{tr}(\boldsymbol{S}_{e1})\boldsymbol{I}_n \qquad \text{(Model 0, i.i.d. noise)} \tag{16}$$

$$\boldsymbol{S}_{e1} = \boldsymbol{S}_{e2} \circ \boldsymbol{I}_n \qquad \text{(Model 1, independent noise)} \tag{17}$$

$$\boldsymbol{S}_{e2} = \frac{1}{N_f}\left[\sum_k \mathrm{Re}\left(\widehat{\boldsymbol{\mathcal{F}}}_k \widehat{\boldsymbol{\mathcal{F}}}_k^*\right) - \boldsymbol{\Phi}\mathrm{Re}(\sum_k \boldsymbol{w}_{2k})\boldsymbol{\Phi}^T\right] \qquad \text{(Model 2, correlated noise)} \tag{18}$$

where '$\circ$' in (17) denotes the Hadamard (elementwise) product. In (16) to (18), the update formulae for different models have been written in a way that highlights their relationship. Equation (18) for correlated noise (Model 2) is the most general one, which is always real symmetric. Equation (17) for independent noise (Model 1) only retains the diagonal entries of (18). Equation (16) for i.i.d. noise (Model 0) takes the average of diagonal entries in Model 1.

As the natural frequencies and damping ratios appear in the Q-function in a nontrivial nonlinear manner, their updated values need to be obtained by numerical optimisation, e.g., '*fminsearch*' in Matlab. Unit norm constraints on the mode shapes are enforced by dividing each updated mode shape in (14) by its Euclidean norm, i.e., square root of the sum of squares. Correspondingly, the (i, j) entry of $\boldsymbol{S}$ in (15) is multiplied by the product of the norms in (14) of the i-th and j-th mode shape (before scaling).

Some remarks are in order on the difference that general noise models make in the above formulae and derivation. Equations (11) to (13), (17) and (18) are intuitive generalisations of their counterparts in [41] for i.i.d. models, where $\boldsymbol{S}_e$ now needs to be retained as a general matrix; simplification that took advantage of the identity matrix in $S_e\boldsymbol{I}_n$ of i.i.d. noise (Model 0) before is no longer possible. The matrix $\boldsymbol{P}_k$ in (10) is defined differently from that in [41], essentially because the inverse and determinant of the data PSD matrix $\boldsymbol{E}_k$ in (2) now involves the Matrix-Inverse Lemma in (56) with no further simplification possible. Equation (13) is the same as before as it does not depend on $\boldsymbol{S}_e$, and so are the update formulae for $\boldsymbol{\Phi}$ and $\boldsymbol{S}$ in (14) and (15), respectively. Regarding the update formulae of $\boldsymbol{S}_e$, (16) for i.i.d. noise (Model 0) is the same as before and recalled for completeness. Equations (17) for independent noise (Model 1) and (18) for correlated noise (Model 2) are nontrivial new formulae. They are derived in Appendix B.2, where matrix algebra identities (Appendix A) have been used to produce compact expressions. The derivation also needs to account for the fact that $\boldsymbol{S}_e$ is a real-symmetric matrix.

3.3. Initial guess of $\boldsymbol{S}_e$

In principle, the initial guess of modal parameters for the EM algorithm can be assigned in the same manner as the existing work for i.i.d. noise (Model 0), e.g., [3] (Section 13.6.1), or [48], where the common noise PSD S_e is estimated in an averaged manner from all measured channels. For general noise models, an initial guess that accounts for noise disparity in the diagonal entries of $\boldsymbol{S}_e$ is developed here, by estimating the noise PSD of each channel separately. The same initial guess can be used for Model 2 (correlated noise).

Consider the case when only the j-th channel is used for modal identification. Then $\boldsymbol{\Phi}_j = [\Phi_{j1}, \dots, \Phi_{jm}]$ and the data PSD is now a real-valued scalar given by $E_k = \boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T + S_{ej}$, where S_{ej} is the j-th diagonal entry in $\boldsymbol{S}_e$. Using (1), the log-likelihood for the j-th channel data is

$$L_j(\boldsymbol{\theta}) = -nN_f \ln\pi - \sum_k \ln(\boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T + S_{ej}) - \sum_k (\boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T + S_{ej})^{-1} |\hat{\mathcal{F}}_{jk}|^2 \tag{19}$$

Zeroing the derivative of L_j w.r.t. S_{ej} and simplifying using the approximation $\boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T \gg S_{ej}$ (high s/n ratio) gives

$$S_{ej} \sim \frac{\sum_k (\boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T)^{-2} (|\hat{\mathcal{F}}_{jk}|^2 - \boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T)}{\sum_k (\boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T)^{-2}} \quad \text{(initial guess for } S_{ej}\text{)} \tag{20}$$

This is an explicit expression that can be used as an initial guess for S_{ej}, given the data and initial guess of other parameters. Intuitively, it is a weighted sum of the difference between the experimental PSD ($\left|\hat{\mathcal{F}}_{jk}\right|^2$) and theoretical PSD ($\boldsymbol{\Phi}_j \boldsymbol{H}_k \boldsymbol{\Phi}_j^T$), with a weight proportional to the latter.

3.4. Posterior covariance matrix

The posterior covariance matrix of modal parameters is equal to the inverse of the Hessian matrix of negative log-likelihood function (NLLF) at the MPV. The latter requires the derivatives of the NLLF. For i.i.d. noise (Model 0), analytical formulae for the derivatives can be found in existing work, e.g., [3] (Section 13.7), or [53]. For general noise models (1 and 2), in view of the remarks in Section 3.2, the derivation of the derivatives of NLLF has been revisited, which results in new formulae applicable for the general noise models. The key results will be outlined first, followed by remarks on the similarity and difference.

For the purpose of calculating derivatives, the NLLF is written as

$$L = L_E + L_H + L_P - L_Q \tag{21}$$

where

$$L_E = nN_f \ln\pi + N_f \ln|\boldsymbol{S}_e| + \sum_k \hat{\boldsymbol{\mathcal{F}}}_k^* \boldsymbol{S}_e^{-1} \hat{\boldsymbol{\mathcal{F}}}_k \tag{22}$$

$$L_H = \sum_k \ln|\boldsymbol{H}_k| \tag{23}$$

$$L_P = \sum_k \ln|\boldsymbol{P}_k| \tag{24}$$

$$L_Q = \sum_k \boldsymbol{r}_k^* \boldsymbol{P}_k^{-1} \boldsymbol{r}_k \tag{25}$$

$$\boldsymbol{r}_k = \boldsymbol{\Phi}^T \boldsymbol{S}_e^{-1} \hat{\boldsymbol{\mathcal{F}}}_k \tag{26}$$

Except for L_H, all terms on the RHS of (21) depend on $\boldsymbol{S}_e$. The derivatives of L_H can be obtained

from previous work; see Table 13.3 of [3] or Appendix I of [53]. The derivatives of other terms are outlined as follows.

The term L_E only depends on $\boldsymbol{S}_e$. Direct differentiation accounting for the symmetry of $\boldsymbol{S}_e$ gives

$$L_E^{(S_{eij}S_{ers})} = N_f \mathrm{tr}\left[-\boldsymbol{S}_e^{-1}\boldsymbol{S}_e{}^{(S_{eij})}\boldsymbol{S}_e^{-1}\boldsymbol{S}_e{}^{(S_{ers})}\right] + \sum_k \widehat{\boldsymbol{\mathcal{F}}}_k^* \boldsymbol{S}_e^{-1}{}^{(S_{eij},S_{ers})} \widehat{\boldsymbol{\mathcal{F}}}_k \tag{27}$$

$$\boldsymbol{S}_e{}^{(S_{eij})} = \mathbf{e}_{ij} + (1-\delta_{ij})\mathbf{e}_{ij}^T \tag{28}$$

$$\boldsymbol{S}_e^{-1}{}^{(S_{eij},S_{ers})} = \boldsymbol{S}_e^{-1}\left[\boldsymbol{S}_e{}^{(S_{eij})}\boldsymbol{S}_e^{-1}\boldsymbol{S}_e{}^{(S_{ers})} + \boldsymbol{S}_e{}^{(S_{ers})}\boldsymbol{S}_e^{-1}\boldsymbol{S}_e{}^{(S_{eij})}\right]\boldsymbol{S}_e^{-1} \tag{29}$$

where $\mathrm{tr}[\cdot]$ denotes the trace (sum of diagonal entries) of the argument matrix; a superscribed symbol in parenthesis denotes a differentiation w.r.t. it; δ_{ij} denotes the Kronecker delta, i.e., equal to 1 when $i = j$ and zero otherwise; S_{eij} is the (i,j) entry of $\boldsymbol{S}_e$, and $\mathbf{e}_{ij}$ denotes a square matrix (with appropriate dimension) with the (i,j) entry being the only nonzero entry equal to 1. Note that the second term $(1-\delta_{ij})\mathbf{e}_{ij}^T$ in (28) stems from the symmetry of $\boldsymbol{S}_e$, i.e., an off-diagonal variable appears twice, above and below the diagonal.

The derivatives of L_P can be obtained in terms of those of $\boldsymbol{P}_k$:

$$L_P^{(S_{eij}S_{ers})} = \sum_k \mathrm{tr}\left[\boldsymbol{P}_k^{-1}\boldsymbol{P}_k^{(S_{eij}S_{ers})} - \boldsymbol{P}_k^{-1}\boldsymbol{P}_k^{(S_{eij})}\boldsymbol{P}_k^{-1}\boldsymbol{P}_k^{(S_{ers})}\right] \tag{30}$$

The derivatives $\boldsymbol{P}_k$ w.r.t. $\boldsymbol{S}_e$ are summarised in Table 3 of Appendix C.

The term L_Q depends on all parameters, through $\boldsymbol{P}_k$ and $\boldsymbol{r}_k$. Table 4 in Appendix C gives the derivatives of L_Q in terms of those of $\boldsymbol{P}_k$ and $\boldsymbol{r}_k$, whose derivatives in turn can be found in Table 3 and Table 5, respectively.

Some remarks are in order on the difference that general noise models make in the above formulae. Except for L_H, the terms in (22) to (26) are different from those in existing work, i.e., [3] (Section 13.7), or [53]. This change arises from the different definition of $\boldsymbol{P}_k$ in (10) from before, and the need to retain $\boldsymbol{S}_e$ as a matrix without further simplification, which was possible before with i.i.d. noise model. As a result, (27) to (29), and the derivatives in Table 3 to Table 5 in Appendix C are newly derived and different from existing formulae for i.i.d. noise.

3.5. Bayesian model class selection

Besides direct investigation of modal identification results, the noise models can also be assessed from the perspective of Bayesian model class evidence [34]. The basic theory and computation are outlined here; detailed investigation is presented later in Sections 5 and 6.

A 'model class' refers to a set of modelling assumptions, based on which the set of parameters is defined. From a model selection perspective, given the data D, a model M is quantified by the conditional probability $p(M|D)$. This probability is difficult to obtain from first principle, but Bayes' theorem allows one to approach it by swapping the roles of M and D, i.e., $p(M|D) = p(D|M)p(M)/p(D)$. Here, $p(D)$ is the probability of data without any model. As it is difficult to quantify, it is often not possible to assess the plausibility of a model in an absolute sense [54]. However, it is possible to compare different models, because then $p(D)$ cancels out in the ratio of $p(M|D)$ for different models. Assuming no prior preference over models, i.e., $p(M) =$ constant, one obtains $p(M|D) \propto p(D|M)$. The probability $p(D|M)$ is called the 'evidence'. For globally identifiable problems where the posterior PDF can be approximated by a Gaussian PDF (Section 2.1), the evidence can be approximated in terms of the posterior MPV $\widehat{\boldsymbol{\theta}}$ and covariance matrix $\widehat{\boldsymbol{C}}$:

$$p(D|M) \approx (2\pi)^{n_p/2}|\widehat{\boldsymbol{C}}|^{1/2}p\big(\widehat{\boldsymbol{\theta}}|M\big)p\big(D|\widehat{\boldsymbol{\theta}},M\big) \tag{31}$$

where n_p is the number of free parameters (see the last column in Table 1). The terms in (31) reveal a combined effect of data fit and model complexity. The former is reflected by the likelihood term $p\big(D|\widehat{\boldsymbol{\theta}},M\big)$, which is higher for models that can give a better fit to data. Model complexity is reflected by the term $(2\pi)^{n_p/2}|\widehat{\boldsymbol{C}}|^{1/2}p\big(\widehat{\boldsymbol{\theta}}|M\big)$, which is often called the 'Ockham factor' [55]. This term tends to decrease exponentially with the number of model parameters, representing a penalty to more complex models [34]. Without much loss of generality, the prior PDF is taken to be a constant, as typically it is the terms that depend on data that matters. Due to the high-dimensional PDF nature of the evidence, its values for different models can differ by orders of magnitude. For this reason, communication is often based on the log of the evidence value, i.e., 'log-evidence'.

4. Equivalent i.i.d. noise OMA analogy and uncertainty law

The previous sections provide the theoretical and computational framework for OMA with general noise models, where a study based on various types of data will be presented later in Sections 5

and 6. This section develops an additional dimension to the understanding of the effect of general noise models (1 and 2) in OMA, by establishing a mathematical analogy with the i.i.d. noise (Model 0), which is the simplest possible model with much experience already generated from previous studies.

As a scientific contribution, we show in Section 4.1 that, by a suitable linear transformation of data, an OMA problem with general noise PSD can be mathematically converted to an equivalent OMA problem with i.i.d. noise. Although in reality such transformation (which depends on noise PSD) is unknown a priori, the mathematical equivalence allows the theory and experience accumulated with i.i.d. noise model to be shared for general noise models. In particular, based on the 'uncertainty law' [37] already developed for OMA with i.i.d. models, the equivalence allows us to obtain asymptotic formulae for the identification uncertainty of modal parameters with general noise models, which is otherwise nontrivial. That will be presented in Section 4.2.

4.1. Mathematical equivalence

Let $\boldsymbol{R}$ be a given $n \times n$ real invertible matrix. Given the measured acceleration data $\{\widehat{\boldsymbol{y}}_j\}_{j=0}^{N-1}$, consider a transformed set of data given by $\widehat{\boldsymbol{y}}'_j = \boldsymbol{R}^{-1}\widehat{\boldsymbol{y}}_j$ $(j = 0, \dots, N-1)$. Correspondingly, the scaled FFT and PSD of the transformed data are respectively given by

$$\widehat{\boldsymbol{\mathcal{F}}}'_k = \boldsymbol{R}^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k = \boldsymbol{R}^{-1}\boldsymbol{\Phi}\boldsymbol{h}_k\boldsymbol{p}_k + \boldsymbol{R}^{-1}\boldsymbol{\varepsilon}_k \tag{32}$$

$$\boldsymbol{E}'_k = \boldsymbol{R}^{-1}\boldsymbol{\Phi}\boldsymbol{h}_k\boldsymbol{S}\boldsymbol{h}_k^*\boldsymbol{\Phi}^T\boldsymbol{R}^{-T} + \boldsymbol{R}^{-1}\boldsymbol{S}_e\boldsymbol{R}^{-T} \tag{33}$$

The mathematical equivalence to be established is essentially concerned with a special choice of $\boldsymbol{R}$ (see (43) later) so that the noise PSD term $\boldsymbol{R}^{-1}\boldsymbol{S}_e\boldsymbol{R}^{-T}$ in (33) can be reduced to a matrix proportional to the identity matrix. Mode shapes and modal force PSD will also need to be rescaled in a consistent manner.

To proceed, we first rewrite $\boldsymbol{E}'_k$ in (33) in the following form that is analogous to (2):

$$\boldsymbol{E}'_k = \boldsymbol{\Phi}'\boldsymbol{h}_k\boldsymbol{S}'\boldsymbol{h}_k^*\boldsymbol{\Phi}'^T + \boldsymbol{R}^{-1}\boldsymbol{S}_e\boldsymbol{R}^{-T} \tag{34}$$

where

$$\boldsymbol{\Phi}' = [\boldsymbol{\varphi}'_1, \dots, \boldsymbol{\varphi}'_m] = \boldsymbol{R}^{-1}\boldsymbol{\Phi}\boldsymbol{B}^{-1} \qquad \boldsymbol{\varphi}'_i = \frac{\boldsymbol{R}^{-1}\boldsymbol{\varphi}_i}{||\boldsymbol{R}^{-1}\boldsymbol{\varphi}_i||} \tag{35}$$

is a matrix containing the transformed mode shapes scaled to have unit norm; and

$$\boldsymbol{S}' = \boldsymbol{BSB} \tag{36}$$

is the corresponding transformed modal force PSD matrix, with

$$\boldsymbol{B} = \begin{bmatrix} ||\boldsymbol{R}^{-1}\boldsymbol{\varphi}_1|| & & \\ & \ddots & \\ & & ||\boldsymbol{R}^{-1}\boldsymbol{\varphi}_m|| \end{bmatrix} \tag{37}$$

Equation (34) can be verified by substituting (36) and (37) to give (33), noting that $\boldsymbol{h}_k\boldsymbol{B} = \boldsymbol{B}\boldsymbol{h}_k$ because the two matrices are diagonal. It indicates that OMA with the original data with noise PSD $\boldsymbol{S}_e$, mode shape matrix $\boldsymbol{\Phi}$ and modal force PSD $\boldsymbol{S}$, is mathematically equivalent to OMA with the transformed data, where the counterparts are now $\boldsymbol{R}^{-1}\boldsymbol{S}_e\boldsymbol{R}^{-T}$ (noise PSD), $\boldsymbol{\Phi}'$ (mode shape) and $\boldsymbol{S}'$ (modal force PSD). We will next show how $\boldsymbol{R}$ can be chosen so that the transformed noise PSD is proportional to the identity matrix, thereby establishing the equivalence with the case of i.i.d. noise.

Let the eigenvector representation of the real symmetric positive definite matrix $\boldsymbol{S}_e$ be

$$\boldsymbol{S}_e = \boldsymbol{V}\boldsymbol{S}'_e\boldsymbol{V}^T \tag{38}$$

where $\boldsymbol{S}'_e$ is a $n \times n$ diagonal matrix of eigenvalues $\{S'_{ei}\}_{i=1}^n$, and $\boldsymbol{V}$ $(n \times n)$ contains in its columns the corresponding orthonormal eigenvectors. It is then clear that choosing $\boldsymbol{R} = \boldsymbol{V}$ will give $\boldsymbol{R}^{-1}\boldsymbol{S}_e\boldsymbol{R}^{-T} = \boldsymbol{S}'_e$ a diagonal matrix, i.e., the transformed problem has independent noise. To further transform the problem to have i.i.d. noise, let S'_{e0} be a chosen reference noise PSD, based on which the diagonal entries $\{S'_{ei}\}_{i=1}^n$ in $\boldsymbol{S}'_e$ are written as

$$S'_{ei} = r_i^2 S'_{e0} \tag{39}$$

where r_i^2 are the corresponding ratios. Correspondingly, $\boldsymbol{S}'_e$ can be written as

$$\boldsymbol{S}'_e = S'_{e0}\boldsymbol{r}\boldsymbol{r}^T \tag{40}$$

where

$$\boldsymbol{r} = \begin{bmatrix} r_1 & & \\ & \ddots & \\ & & r_n \end{bmatrix} \tag{41}$$

Substituting (40) into (38) gives

$$\boldsymbol{S}_e = S'_{e0}(\boldsymbol{V}\boldsymbol{r})(\boldsymbol{V}\boldsymbol{r})^T \tag{42}$$

It is now clear that choosing

$$\boldsymbol{R} = \boldsymbol{V}\boldsymbol{r} \qquad \text{(condition for equivalence)} \qquad (43)$$

will reduce the noise PSD term $\boldsymbol{R}^{-1}\boldsymbol{S}_e\boldsymbol{R}^{-T}$ in (34) to $S'_{e0}\boldsymbol{I}_n$ so that

$$\boldsymbol{E'}_k = \boldsymbol{\Phi'}\boldsymbol{h}_k\boldsymbol{S'}\boldsymbol{h}_k^*\boldsymbol{\Phi'}^T + S'_{e0}\boldsymbol{I}_n \qquad (44)$$

That is, the OMA problem with transformed data now has i.i.d. noise with a common noise PSD of S'_{e0}, hence establishing the mathematical equivalence.

The role of reference noise PSD S'_{e0}

As a remark, in the transformed problem the common reference noise PSD S'_{e0} is arbitrary and its choice is up to the analyst. Conceptually, S'_{e0} affects $\boldsymbol{r}$ by a scalar multiplication through $\boldsymbol{S}_e = S'_{e0}(\boldsymbol{V}\boldsymbol{r})(\boldsymbol{V}\boldsymbol{r})^T$ in (42) ($\boldsymbol{S}_e$ is fixed), which in turn affects $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{r}$ in (43). This will affect the modal force PSD $\boldsymbol{S'} = \boldsymbol{B}\boldsymbol{S}\boldsymbol{B}$ in (36), as $\boldsymbol{B}$ in (37) is a diagonal matrix of $\|\boldsymbol{R}^{-1}\boldsymbol{\varphi}_i\|$. The mode shape $\boldsymbol{\varphi'} = \boldsymbol{R}^{-1}\boldsymbol{\varphi}/\|\boldsymbol{R}^{-1}\boldsymbol{\varphi}\|$ is not affected, however, because it is invariant to scalar multiplication of $\boldsymbol{R}$. Clearly, the natural frequencies and damping ratios are not affected. See the end of Section 4.2, which reveals that identification uncertainty (at least under asymptotic conditions) is not affected either. These observations indicate that the value of S'_{e0} is immaterial to the quality of identification results. In applications it may be chosen to be some nominal value for convenience.

4.2. Uncertainty law for general noise models

One empirical use of the mathematical analogy established in the last section is that, if one knew $\boldsymbol{S}_e$ and hence $\boldsymbol{R}$ in (43), the OMA problem for general noise can be approached via the transformed problem that can make use of existing theory and computational tools developed for i.i.d. noise. In reality, $\boldsymbol{S}_e$ is unknown a priori, but this idea allows an approximate treatment based on an educated guess, e.g., a diagonal matrix of noise PSDs estimated from individual data channels. See, e.g., an illustration with laboratory data in Section 6.1 later.

Beyond the basic objective of OMA, we will next make use of the analogy and existing 'uncertainty law' for i.i.d. noise model to derive explicit formulae for the identification uncertainty of modal parameters. The resulting formulae allow one to understand and plan ambient vibration tests taking into account potential effects from noise disparity and correlation. As a background, 'uncertainty law' in OMA aims at explicit formulae for the posterior coefficient of variation (i.e., standard deviation/mean, abbreviated as 'c.o.v.') of modal parameters, which turns out to be

possible under asymptotic conditions of long data, high s/n ratio and small damping. See [56] for well-separated modes, [47] for close modes, and [57] for multiple setups. The presentation here focuses on the case of well-separated modes, where the effect of noise is encapsulated by the modal s/n ratio.

We first recall briefly the existing results for i.i.d. noise [56]. Consider a single mode with natural frequency f, damping ratio ζ, mode shape $\boldsymbol{\varphi}$ (scaled to unit norm), modal force PSD S and i.i.d. noise PSD S_e. The mode is identified using FFT in the frequency band $f(1 \pm \kappa\zeta)$, where κ is a bandwidth factor, e.g., $\kappa = 1$ for half-power band. Then it has been shown that, for long data, high modal s/n ratio and small damping, the posterior c.o.v. of f, ζ and S are given by, respectively,

$$\delta_f^2 = \frac{\zeta}{2\pi N_c B_f}\left(1 + \frac{a_f}{\gamma}\right) \qquad \delta_\zeta^2 = \frac{1}{2\pi\zeta N_c B_\zeta}\left(1 + \frac{a_\zeta}{\gamma}\right) \qquad \delta_S^2 = \frac{1}{N_f B_S}\left(1 + \frac{a_S}{\gamma}\right) \tag{45}$$

The posterior covariance matrix of $\boldsymbol{\varphi}$ and the posteiorr c.o.v. of S_e are given by, respectively,

$$\boldsymbol{C}_{\boldsymbol{\varphi}} = \frac{2S_e\zeta}{\pi N_c S B_\varphi}(\boldsymbol{I}_n - \boldsymbol{\varphi}\boldsymbol{\varphi}^T) \qquad \delta_{Se}^2 = \frac{1}{(n-1)N_f} \tag{46}$$

In the above, $N_c =$ (data duration)/(natural period) is a dimensionless measure of data length; $N_f = 2\kappa\zeta N_c$ is the number of FFT points in the band $f(1 \pm \kappa\zeta)$; B_f, B_ζ and B_S are 'data length factors' that reflect the amount of usable information in data for identifying the mode; a_f, a_ζ and a_S are sensitivity coefficients associated with the 'modal s/n ratio' γ. The B's and a's above are a function of κ; detailed expressions are referred to (7)-(9) and (14)-(16) of [56]. The modal s/n ratio γ is defined as the ratio of modal response PSD at resonance ($S/4\zeta^2$) to the noise PSD (S_e), which turns out to encapsulate the effect of noise and environment:

$$\gamma = \frac{S}{4S_e\zeta^2} \qquad \text{(i.i.d. noise)} \tag{47}$$

Empirically, moderate to high modal s/n ratio may be referred as $\gamma > 100$. In applications, the c.o.v. of damping ratio is often the most critical one to control, although the expressions for other parameters have been presented here for completeness. In (45) to (47), the parameters represent the 'actual' value of the structure. When the formulae are used for planning ambient vibration tests in the absence of data, the parameters are substituted by best guess.

Equations (45) and (46) were derived based on i.i.d. noise. In what follows, we will apply them with the analogy established in Section 4.1 to derive the uncertainty law for general noise model

so that noise disparity and correlation can be accounted for in ambient vibration test planning. For this purpose, we assign the noise PSD matrix $\boldsymbol{S}_e$ and hence $\boldsymbol{R}$ in (43) based on prior knowledge or some pre-test calibration. Using the analogy, the c.o.v. of f and ζ can be assessed by simply replacing γ in the first two formulae of (45) by the modal s/n ratio γ' (say) of the (conceptually) transformed OMA problem:

$$\gamma' = \frac{S'}{4S'_{e0}\zeta^2} \qquad \text{(general noise)} \tag{48}$$

where, for the transformed OMA problem, S' is the modal force PSD and S'_{e0} is the common reference noise PSD; see (36) and (39), respectively. Equation (48) reveals the correct but nontrivial way to quantify the modal s/n ratio for the general noise models, without going through the tedious derivation of uncertainty law (if possible). Note that γ' is invariant to any scalar multiplication of $\boldsymbol{R}$ because the scaling appears in both S' and S'_{e0}, and hence is cancelled out.

For the mode shape, replacing S_e by S'_{e0}, and $\boldsymbol{\varphi}$ by $\boldsymbol{\varphi}'$, in the first formula of (46) gives the covariance matrix of $\boldsymbol{\varphi}'$, i.e., $\boldsymbol{C}'_{\boldsymbol{\varphi}}$, for the transformed problem. The covariance matrix of $\boldsymbol{\varphi}$ in the original problem can be obtained by perturbation in terms of $\boldsymbol{C}'_{\boldsymbol{\varphi}}$ as follows. Recall that $\boldsymbol{\varphi} = \boldsymbol{R}\boldsymbol{\varphi}'/\|\boldsymbol{R}\boldsymbol{\varphi}'\|$. Applying chain rule gives

$$\frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\varphi}'} = \frac{\partial \boldsymbol{\varphi}}{\partial (\boldsymbol{R}\boldsymbol{\varphi}')}\frac{\partial (\boldsymbol{R}\boldsymbol{\varphi}')}{\partial \boldsymbol{\varphi}'} = ||\boldsymbol{R}\boldsymbol{\varphi}'||^{-1}(\boldsymbol{I}_n - \boldsymbol{\varphi}\boldsymbol{\varphi}^T)\boldsymbol{R} \tag{49}$$

where we have used the result that, for unit vector $\overline{\boldsymbol{u}} = \boldsymbol{u}/\|\boldsymbol{u}\|$, the derivative is $\partial\overline{\boldsymbol{u}}/\partial\boldsymbol{u} = \|\boldsymbol{u}\|^{-1}(\boldsymbol{I}_n - \overline{\boldsymbol{u}}\overline{\boldsymbol{u}}^T)$. The uncertain perturbations $\Delta\boldsymbol{\varphi}$ and $\Delta\boldsymbol{\varphi}'$ are then related by

$$\Delta\boldsymbol{\varphi} = \frac{\partial \boldsymbol{\varphi}}{\partial \boldsymbol{\varphi}'}\Delta\boldsymbol{\varphi}' = ||\boldsymbol{R}\boldsymbol{\varphi}'||^{-1}(\boldsymbol{I}_n - \boldsymbol{\varphi}\boldsymbol{\varphi}^T)\boldsymbol{R}\Delta\boldsymbol{\varphi}' \tag{50}$$

Taking covariance gives

$$\boldsymbol{C}_{\boldsymbol{\varphi}} = ||\boldsymbol{R}\boldsymbol{\varphi}'||^{-2}(\boldsymbol{I}_n - \boldsymbol{\varphi}\boldsymbol{\varphi}^T)\boldsymbol{R}\boldsymbol{C}_{\boldsymbol{\varphi}'}\boldsymbol{R}^T(\boldsymbol{I}_n - \boldsymbol{\varphi}\boldsymbol{\varphi}^T) \tag{51}$$

It can be verified that $\boldsymbol{C}_{\boldsymbol{\varphi}}$ is invariant to any scaling of $\boldsymbol{R}$.

For the modal force PSD, replacing γ by γ' in the third formulae of (45) yields the c.o.v. of S' for the transformed problem. Similar to $\boldsymbol{\varphi}$, the c.o.v. of S in the original OMA problem can be obtained by perturbation as follows. From (36), for a single mode we have $S = \|\boldsymbol{R}^{-1}\boldsymbol{\varphi}\|^{-2}S'$. This can be written as $S = \|\boldsymbol{R}\boldsymbol{\varphi}'\|^2 S'$, where we have used $\|\boldsymbol{R}^{-1}\boldsymbol{\varphi}\|^{-1} = \|\boldsymbol{R}\boldsymbol{\varphi}'\|$. The latter can be verified by multiplying both sides of $\boldsymbol{\varphi}' = \boldsymbol{R}^{-1}\boldsymbol{\varphi}/\|\boldsymbol{R}^{-1}\boldsymbol{\varphi}\|$ by $\boldsymbol{R}$, taking norm and noting that

$\|\boldsymbol{\varphi}\| = 1$. For small perturbations $\Delta S'$ and $\Delta\boldsymbol{\varphi}'$, we have $\Delta S \approx \|\boldsymbol{R}\boldsymbol{\varphi}'\|^2\Delta S' + 2S'\Delta\boldsymbol{\varphi}'^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\varphi}'$. Taken expectation of the square over uncertainties $\Delta S'$ and $\Delta\boldsymbol{\varphi}'$ gives:

$$\sigma_S^2 \sim \|\boldsymbol{R}\boldsymbol{\varphi}'\|^4\sigma_{S'}^2 + 4S'^2\boldsymbol{\varphi}'^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{C}_{\boldsymbol{\varphi}'}\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\varphi}' \tag{52}$$

where σ_S^2 and $\sigma_{S'}^2$ denote the variances of S (original problem) and S' (transformed problem), respectively; $\boldsymbol{C}_{\boldsymbol{\varphi}'}$ is the covariance matrix of $\boldsymbol{\varphi}'$; and we have made use of the fact that S' and $\boldsymbol{\varphi}'$ are asymptotically uncorrelated (see Table 1 of [37]). Dividing both sides of (52) by $S^2 = \|\boldsymbol{R}\boldsymbol{\varphi}'\|^4S'^2$ gives

$$\delta_S^2 \sim \delta_{S'}^2 + 4\|\boldsymbol{R}\boldsymbol{\varphi}'\|^{-4}\boldsymbol{\varphi}'^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{C}_{\boldsymbol{\varphi}'}\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\varphi}' \tag{53}$$

where δ_S and $\delta_{S'}$ denote the c.o.v. of S (original problem) and S' (transformed problem), respectively. It can be shown further that $\delta_S^2 \sim \delta_{S'}^2(1 + b/\gamma')$, where $b > 0$, with $b = 0$ if there is no noise disparity ($r_i =$ constant). See Appendix D for details. This suggests theoretically that the c.o.v. of S is higher when there is noise disparity, although the effect is negligible for high modal s/n ratio.

As a note, the uncertainty law formulae developed here for general noise models have not accounted for the uncertainty (lack of knowledge) in the disparity or correlation in the noise PSD, as evidenced by the fact that the transformation matrix $\boldsymbol{R}$ has been assumed to be fixed. Nevertheless, in ambient vibration test planning one is often concerned with the order of the magnitude of the uncertainty given a particular test configuration rather than the uncertainty in a precise form (which cannot be computed without data), so this treatment is often sufficient, providing a simple way to account for noise disparity and correlation in the assessment of identification uncertainty before the actual test.

The role of S_{e0}', and alternative forms of γ'

Complementing the remark near the end of Section 4.1, in the transformed problem the choice of the common reference noise PSD S_{e0}' does not affect the modal s/n ratio γ'. To see this, note that $S' = \|\boldsymbol{R}^{-1}\boldsymbol{\varphi}\|^{-2}S = [\boldsymbol{\varphi}^T(\boldsymbol{R}\boldsymbol{R}^T)^{-1}\boldsymbol{\varphi}]S = S_{e0}'(\boldsymbol{\varphi}^T\boldsymbol{S}_e^{-1}\boldsymbol{\varphi})S$, where the last equality follows from $\boldsymbol{S}_e = S_{e0}'\boldsymbol{R}\boldsymbol{R}^T$. Substittuting into (48) gives

$$\gamma' = \frac{(\boldsymbol{\varphi}^T\boldsymbol{S}_e^{-1}\boldsymbol{\varphi})S}{4\zeta^2} \qquad \text{(general noise)} \tag{54}$$

This formula depends only on the quantities in the original problem, and hence is invariant to the

choice of S'_{e0}. As the c.o.v.s of natural frequency and damping ratio in (45) depends on S'_{e0} only through γ', they are invariant to S'_{e0}. The invariance also applies to the covariance matrix of $\boldsymbol{\varphi}$ in (51) and the c.o.v. of S in (53), which can be reasoned from the cancelling of the scaling effect.

For the case of independent noise (Model 1), (54) can be expressed in an intuitive form in terms of the s/n ratio of individual channels. In this case, $\boldsymbol{\varphi}^T \boldsymbol{S}_e^{-1} \boldsymbol{\varphi} = \sum_{i=1}^{n} \varphi_i^2 / S_{ei}$ where φ_i is the i-th entry of $\boldsymbol{\varphi}$. Equation (54) then reduces to

$$\gamma' = \sum_{i=1}^{n} \frac{S}{4 S_{ei} \zeta^2} \varphi_i^2 \qquad \text{(Independent noise)} \qquad (55)$$

Since $\sum_{i=1}^{n} \varphi_i^2 = 1$, this formula shows that γ' can be considered a weighted average of the modal s/n ratio of individual channels ($S/4S_{ei}\zeta^2$), where the weight is proportional to the square of mode shape value.

5. Investigation with synthetic data

In this section, the proposed methodology and the effect of noise models (see Table 1) are investigated using synthetic data. Section 5.1 investigates the effect of noise disparity by considering data generated from a single mode and contaminated with independent noise with different PSDs. Section 5.2 investigates the effect of noise correlation arising from unaccounted modal dynamics, based on synthetic data from a ten-storey shear building. For both examples, in addition to basic modal identification results, the noise models are also assessed in terms of Bayesian model class evidence. The analogy of OMA with transformed data (Section 4.1) and the resulting uncertainty law (Section 4.2) will also be illustrated in Section 5.1.

5.1. Single mode data (noise disparity)

Consider synthetic acceleration data generated from a single mode with natural frequency $f = 1$ Hz, damping ratio $\zeta = 1\%$, mode shape $\boldsymbol{\varphi} = [1\ 1\ 1\ 1]^T/2$, and subjected to white noise modal force with PSD $S = 1\ (\mu g)^2/\text{Hz}$. The data is 1000 seconds long and sampled at 100 Hz. It is contaminated by white noise with PSD matrix $\boldsymbol{S}_e = S'_{e0} \boldsymbol{r} \boldsymbol{r}^T$, where S'_{e0} is a reference PSD and $\boldsymbol{r}$ is a diagonal matrix of $\{r, r, 1, 1\}$. In other words, the noise is independent among different channels, but the ratio of noise PSD of the first two channels are r^2 times that of the last two, i.e., a noise disparity when $r \neq 1$. This corresponds to Model 1 in Table 1. The value of r will be varied to investigate the effect of noise disparity. Fig. 2 shows the PSD and SV (singular value) spectrum of a typical set of data when $S'_{e0} = 1\ (\mu g)^2/\text{Hz}$ and $r^2 = 300$.

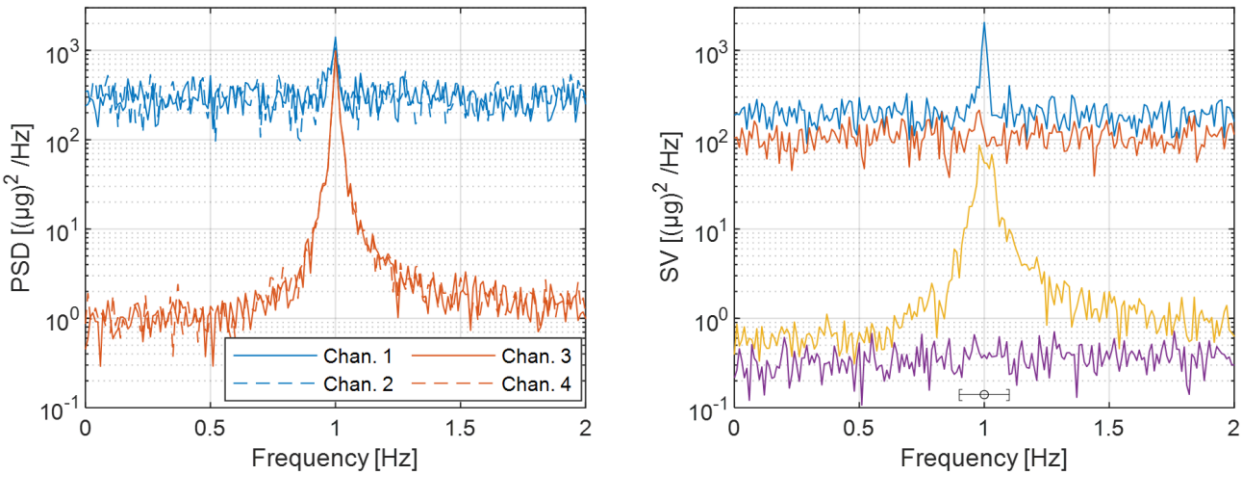


Fig. 2 PSD (left) and SV spectrum (right) of a typical dataset when $S'_{e0} = 1$ (μg)^2/Hz and $r^2 = 300$. Synthetic single mode data. The bracket in the SV plot shows the frequency band for modal identification

Potential bias (histogram of MPV)

We first investigate the effect of noise disparity on potential bias in MPV of modal parameters. For this purpose, we fix $S'_{e0} = 1$ (μg)^2/Hz and vary $r^2 = \{1,100,300\}$. In each case, 1000 i.i.d. datasets are generated. For each dataset, the posterior MPV (Sections 3.1 to 3.3) and covariance matrix (Section 3.4) are determined assuming a particular noise model. Fig. 3 shows in the first two columns the histograms of MPVs of the natural frequency and damping ratio. The histograms of the MAC (modal assurance criterion) between the MPV of mode shape and the exact value are also shown in the third column. As r^2 increases from 1 (top row, no disparity) to 100 and 300 (bottom row), noise disparity increases. It can be seen from the first three columns that, regardless of noise disparity, the histograms for Models 1 and 2 generally overlap and gather around the exact value ($f = 1, \zeta = 1\%$, MAC = 1), and there is no evidence of bias. This is expected, because the data is generated with a noise PSD of Model 1 (independent noise), which is a special case of Model 2 (correlated noise). The histograms of Model 0 (i.i.d. noise) agree with those of Model 1 and 2 when $r^2 = 1$ (no disparity). As disparity increases, however, they depart with a shift (bias) and larger spread in the damping ratio (bias high) and MAC.

As a note, the equivalent modal s/n ratio in (54) is calculated to be $\gamma' = \{2500,1263,1254\}$ for $r^2 = \{1,100,300\}$, respectively. These values are of the same order of magnitude and hence the effect of modal s/n ratio in the results discussed above is considered negligible.

Bayesian evidence

The last column of Fig. 3 shows the histogram of the log-evidence of different models (the higher the better); see Section 3.5 for background theory. When there is no disparity ($r^2 = 1$, top row), the values rank as Model 0 (100%) > Model 1 (0%) > Model 2 (0%), where the number in parenthesis indicates the percentage of time the log-evidence of the model is highest. This is expected, because in this case the three noise models are capable of producing a similar fit to data, and hence their log-evidence is governed by the number of parameters (the lower the better). With significant noise disparity ($r^2 = 100{,}300$), the log-evidence values rank as Model 1 (100%) > Model 2 (0%) > Model 0 (0%). This is also expected, and can be reasoned as follows. Model 0 is worse because it scores low on data fit, despite the smallest number of parameters, i.e., model is too simple. Model 1 can produce a similar fit to data as Model 2, but it has a smaller number of parameters, and hence becomes the preferred one based on log-evidence.

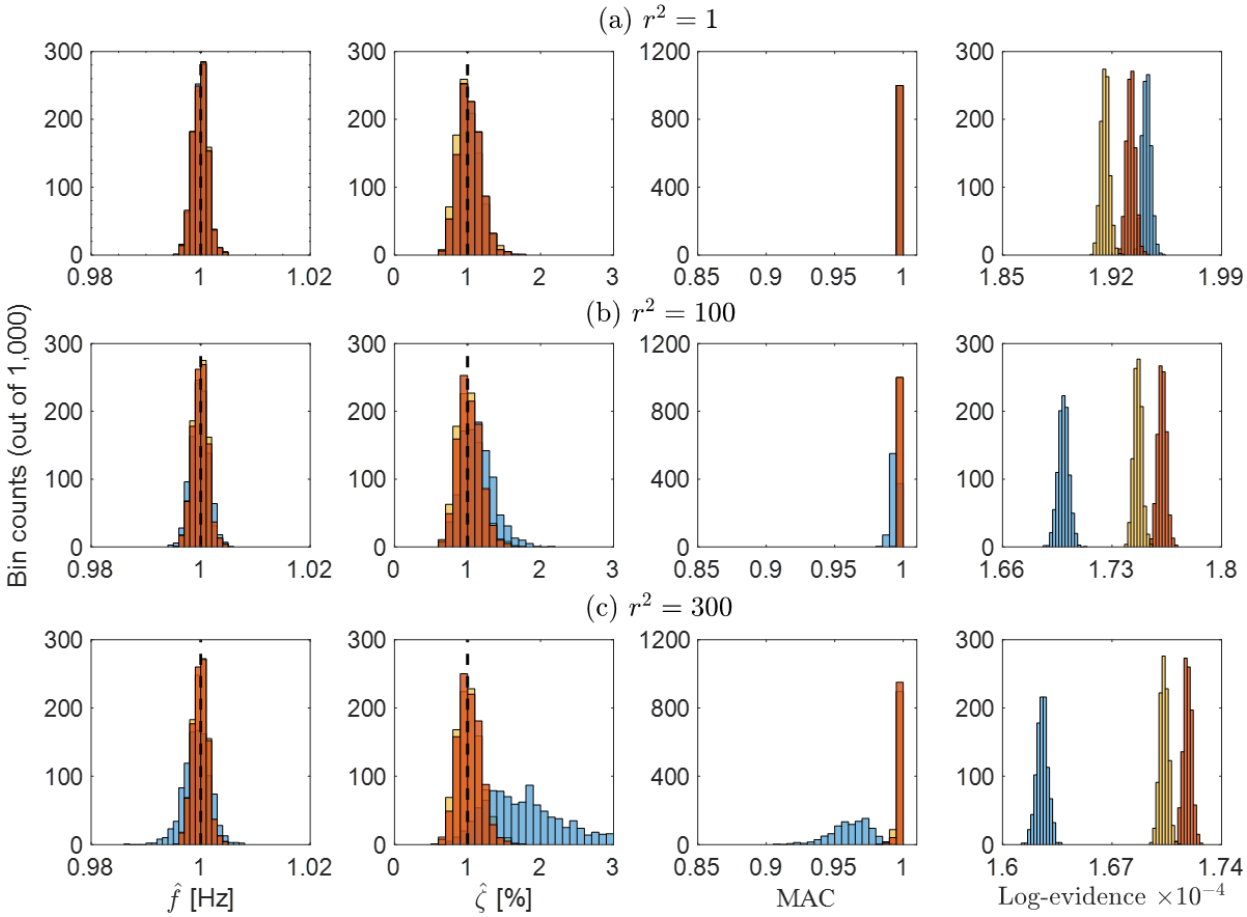


Fig. 3 Histogram of MPVs identified assuming different noise models in Table 1. Synthetic single mode data. Blue: Model 0 (i.i.d. noise), red: Model 1 (independent noise), yellow: Model 2 (correlated noise). A different value of r^2 corresponds to different noise disparity. Dashed line: exact value of parameter

Empirical bias correction through transformed data

The results just presented suggested that using the existing Bayesian OMA algorithms based on i.i.d. noise (Model 0) may lead to potential bias when the noise disparity is significant. Here, we demonstrate how the bias can be reduced empirically via the transformed OMA problem as discussed in Section 4.1. For a given set of data, we can estimate the ratios of noise PSDs among different channels. Together with a nominal (otherwise arbitrary) choice of reference noise PSD S'_{e0}, we can set $\boldsymbol{r}$ in (41). Assuming zero correlation, (43) gives $\boldsymbol{R} = \boldsymbol{r}$ because $\boldsymbol{V} = \boldsymbol{I}_n$. We can then apply existing Bayesian OMA algorithm (Model 0) to the transformed data $\{\hat{\boldsymbol{y}}'_j = \boldsymbol{R}^{-1}\hat{\boldsymbol{y}}_j\}$ and convert the results back to the original problem. As an illustration, consider the dataset for $r^2 = 300$, whose PSD and SV have been shown in Fig. 2. For purpose here, we take $S'_{e0} = 1\ (\mu g)^2/\text{Hz}$ and $\boldsymbol{R}$ as the diagonal matrix of $\{\sqrt{300}, \sqrt{300}, 1, 1\}$, i.e., same as the ones that generated the data. This is perhaps the best one can do, although trial results with ratios actually estimated from PSD plots are qualitatively similar. This choice is considered as the results can serve as a verification of the mathematical analogy established in Section 4.1.

Fig. 4 shows the PSD and SV of the transformed data, which should be compared with Fig. 2 for the original data (before transformation). The SV plot in Fig. 4 now exhibits features similar to that of a single mode with i.i.d. noise, i.e., a single line exhibiting dynamic amplification, and the remaining ones staying at the common noise level. Fig. 5 shows the histograms of MPVs identified assuming Model 0 and based on the transformed data (1000 sets), which may be compared with Fig. 3. The histograms are now clustered around the exact values and with a spread similar to those of Models 1 and 2 in Fig. 3. The bias due to noise disparity has been avoided by OMA on the transformed data.

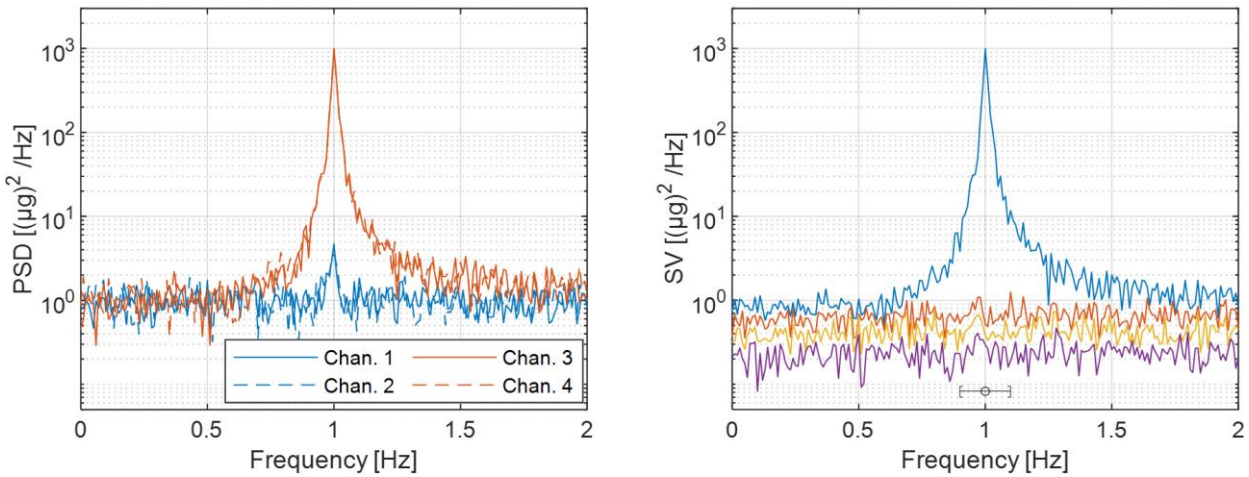


Fig. 4 PSD (left) and SV (right) of data after transformation. Synthetic single mode data. See Fig. 2 for the plots of the original data (before transformation)

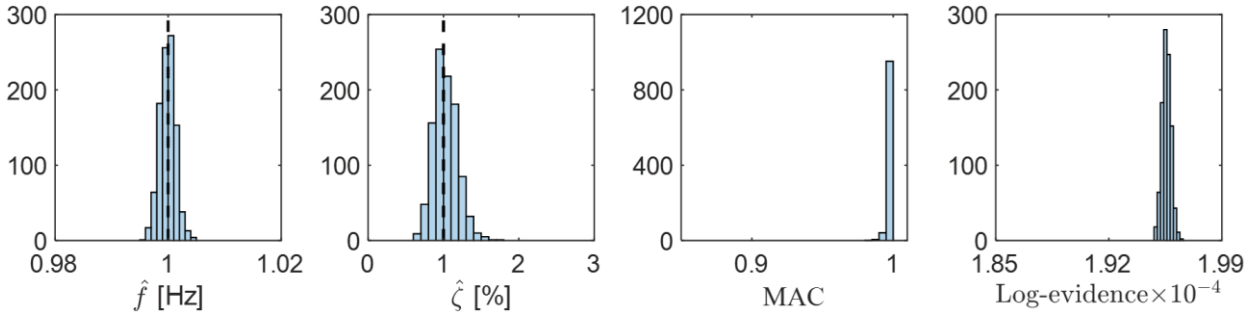


Fig. 5 Histograms of MPV identified using Model 0 from transformed data (1000 sets). Synthetic single mode data. Dashed line: exact value of parameter

Verification of uncertainty law

We next verify the uncertainty law developed in Section 4.2 for general noise models. For this purpose, we consider 11 data sets generated with $S'_{e0} = 100$ (μg)^2/Hz and r^2 ranging from 10^{-3} to 10. These values are considered so that the datasets span over a wide range of modal s/n ratios. For each dataset, two sets of c.o.v.s of modal parameters are calculated, one based on uncertainty law ((45) and (46) substituted with 'dashed' parameters, meanwhile use (51) and (53); all parameters set to theoretical values) and the other based on the posterior covariance matrix calculated from OMA with the original data (Section 3.4), assuming independent noise (Model 1). The results are plotted in Fig. 6 (f, ζ, S) and Fig. 7 ($\boldsymbol{\varphi}$) versus the modal s/n ratio γ'. The latter is calculated using (54) with parameters set to theoretical values. It can be seen from the figures that the two sets of c.o.v. (circle and asterisk) generally agree with each other especially for moderate to high s/n ratios ($\gamma' > 100$), verifying the mathematical correctness of the uncertainty law. As expected, the c.o.v.s converge to a constant non-zero value for high s/n ratio (e.g., $\gamma' >$

1000).

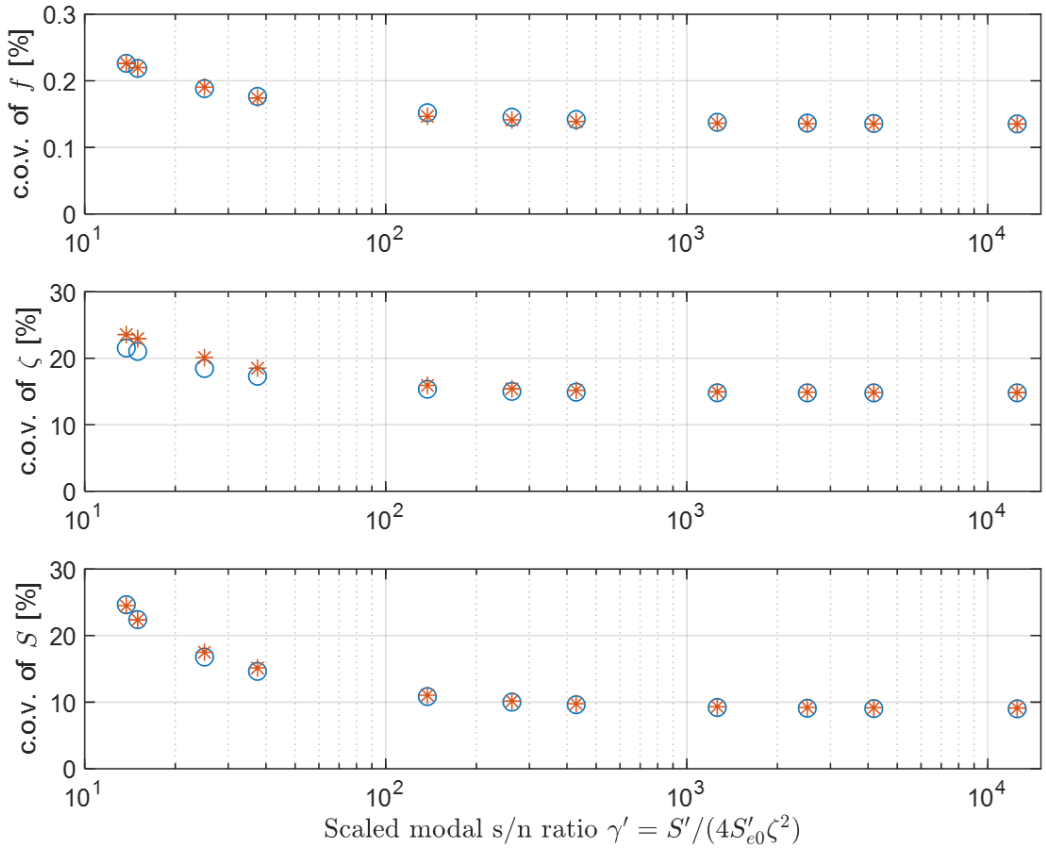


Fig. 6 Posterior c.o.v. of natural frequency (f), damping ratio (ζ) and modal force PSD (S). Synthetic single mode data. Circle: based on posterior covariance matrix; Asterisk: uncertainty law

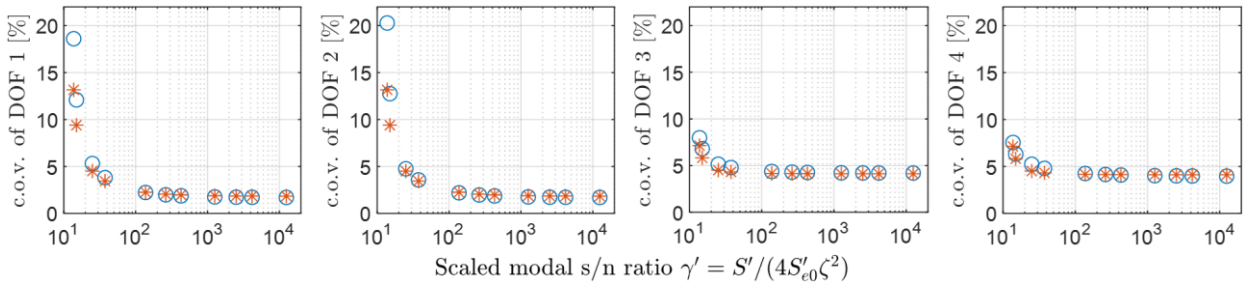


Fig. 7 Posterior c.o.v. of mode shape values at different DOFs. Synthetic single mode data. Circle: based on posterior covariance matrix; Asterisk: uncertainty law

5.2. Ten-storey building with multi-mode data (noise correlation)

This example investigates the effect of noise correlation that arises from unmodelled dynamics. Consider the horizontal vibration of a ten-storey shear building with a uniform mass of 1000 tonnes per floor, inter-storey stiffness of 3678 kN/mm and classical damping ratio of 1% for all

modes. The structure is subjected to i.i.d. white noise excitation at all floors, each with PSD $S_w = 96.2\ \mathrm{N}^2/\mathrm{Hz}$. The accelerations on 1/F, 4/F, 7/F and the roof are measured for 1000 seconds at 256 Hz. The data is contaminated by i.i.d. noise with a common PSD $S_e = 1\ (\mu\mathrm{g})^2/\mathrm{Hz}$. Fig. 8 shows the PSD and SV of a typical dataset. The study here focuses on Modes 1, 7 and 10, which are near 1.5 Hz, 16 Hz and 19 Hz, respectively. They are identified using the FFT in the selected band indicated in Fig. 8. Although the data is generated with i.i.d. instrument noise, the frequency bands in Fig. 8 may contain contributions from unmodelled neighbouring modes, which will appear as correlated noise in the OMA process.

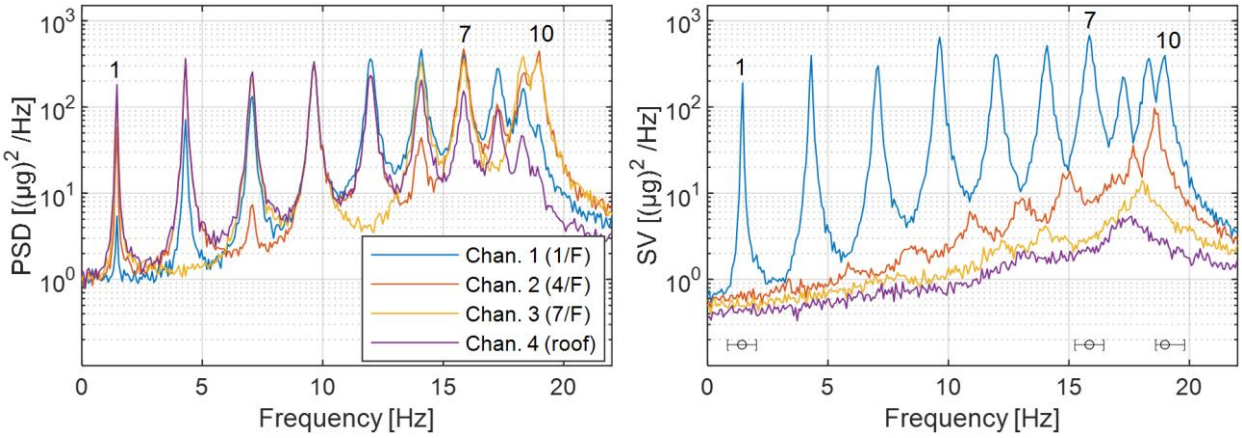


Fig. 8 PSD and SV spectrum. Synthetic ten-storey shear building data. Bracket in SV plot indicates the frequency band for modal identification.

Potential bias in MPV

The potential bias on MPV from noise correlation and Bayesian evidence of different noised models are investigated in a manner similar to Section 5.1, based on 1000 datasets. Fig. 9 shows the results for the damping ratio, which is the most affected parameter. It is seen that the histograms of different models are similar for Mode 1, somewhat different for Mode 7, and quite different for Mode 10. This is consistent with the SV plot in Fig. 8, where the mode to the left of Mode 10 is seen to have a significant contribution to the band indicated, smearing in the OMA process as a correlated noise. For Mode 10, the histogram of Model 0 (i.i.d.) is biased low (-17% in the mean), while the histogram of Model 1 (independent noise) is biased high (+10% in the mean). The histogram of Model 2 (correlated noise) is comparatively unbiased (+1% in the mean), although there is some distortion of the bell shape. The histograms for natural frequencies and mode shapes have been examined, which indicate that there is no significant bias. Details are omitted.

Bayesian evidence

The right column of Fig. 9 provides another perspective on the noise models in terms of log-evidence. For Mode 1, the log-evidence ranks Model 0 (100%) > Model 1 (0%) > Model 2 (0%), justifying the use of existing Bayesian OMA algorithms with i.i.d. noise. For Mode 7, Model 2 (100%) > Model 1 (0%) > Model 0 (0%). Clearly, the log-evidence favours Model 2 (correlated noise), although this is not trivial from the SV in Fig. 8 near 16 Hz. Model 1 performs better than Model 0 in the log-evidence values, which is likely attributed to its better data fitting capability. For Mode 10, the log-evidence values tell clearly that Model 2 (100%) > Model 1 (0%) > Model 0 (0%), indicating a strong preference for a model that can account for noise correlation at the expense of a larger number of parameters.

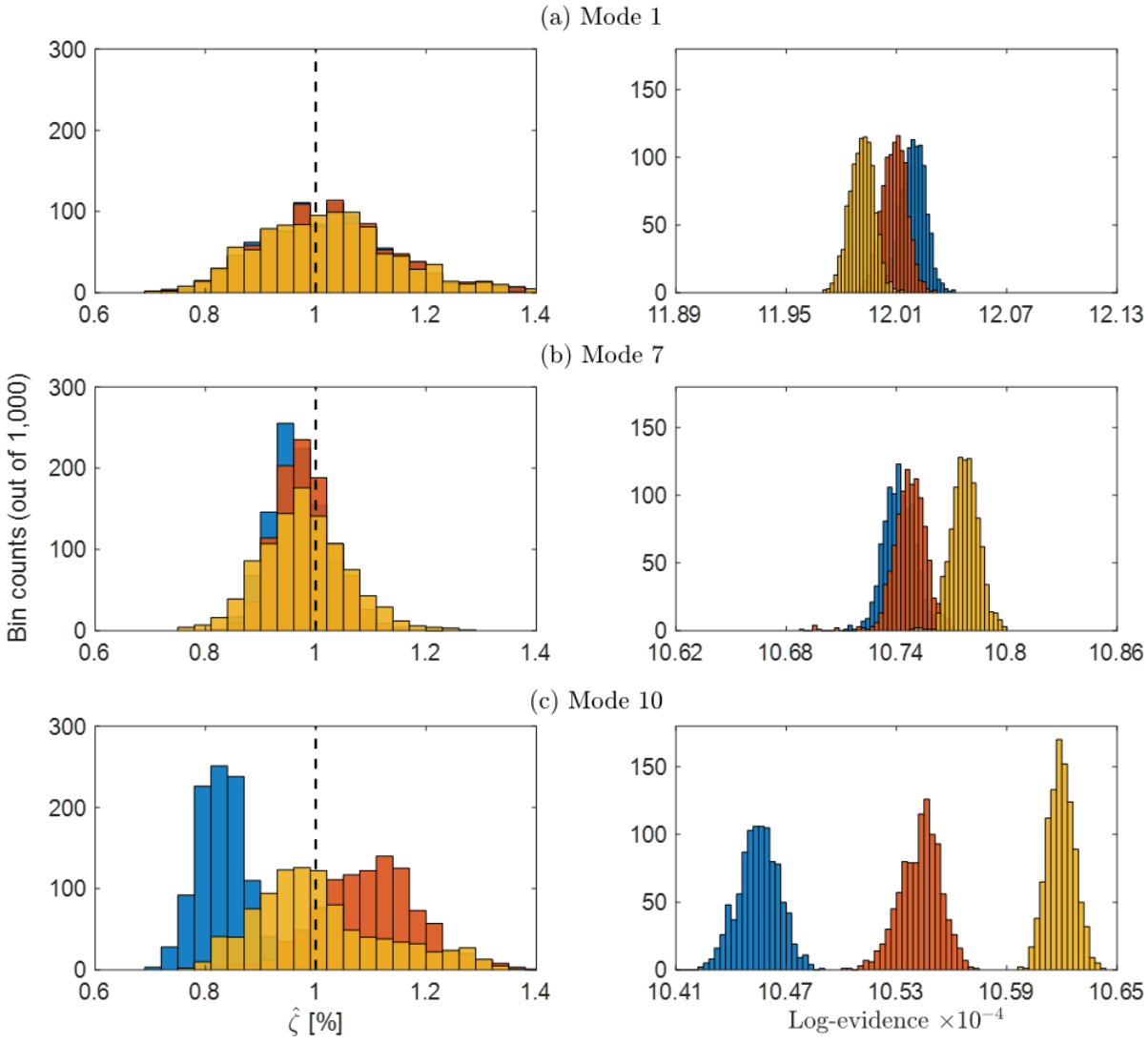


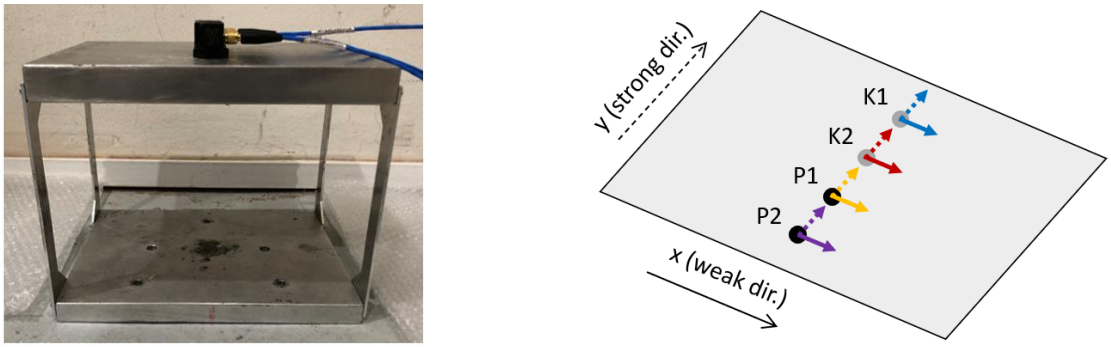
Fig. 9 Histograms of MPVs of damping ratio (left) and log-evidence (right). Synthetic ten-storey shear building data. Blue: Model 0 (i.i.d. noise), red: Model 1 (independent noise), yellow: Model 2 (correlated noise).

6. Investigation with laboratory and field data

In this section, we extend the investigation of noise models to laboratory and field data. The former introduces complications from real experiments in a controlled manner, while the latter provides a context for full-scale structures under complex environment. Noise disparity is investigated in the laboratory data using sensors with different noise levels. In the field data, potential noise correlation arising from unaccounted modes or unknown dynamics will be investigated.

6.1. Laboratory data of SDOF model (noise disparity)

Consider an SDOF aluminium model made with a 'roof' mass of 1.6 kg supported on four columns measuring 148mm × 30mm, as shown in Fig. 10. To investigate noise disparity, the vibration of the roof mass is measured by four tri-axial piezoelectric accelerometers, i.e., K1, K2, P1, P2, where the letter indicates the sensor model. The investigation here is based on a set of ambient data measured for 600 seconds sampled at 2048Hz. Fig. 11 shows the PSD and SV spectrum of the data, which is similar to Fig. 2. As seen from the y-direction data which receives little contribution from the x-direction resonance, the noise PSD drops with frequency. Near 15 Hz the noise PSDs of the K and P sensors are about $700\,(\mu g)^2/Hz$ and $20\,(\mu g)^2/Hz$, respectively.



(a) (b)

Fig. 10 Laboratory SDOF model (left) and sensor layout (right)

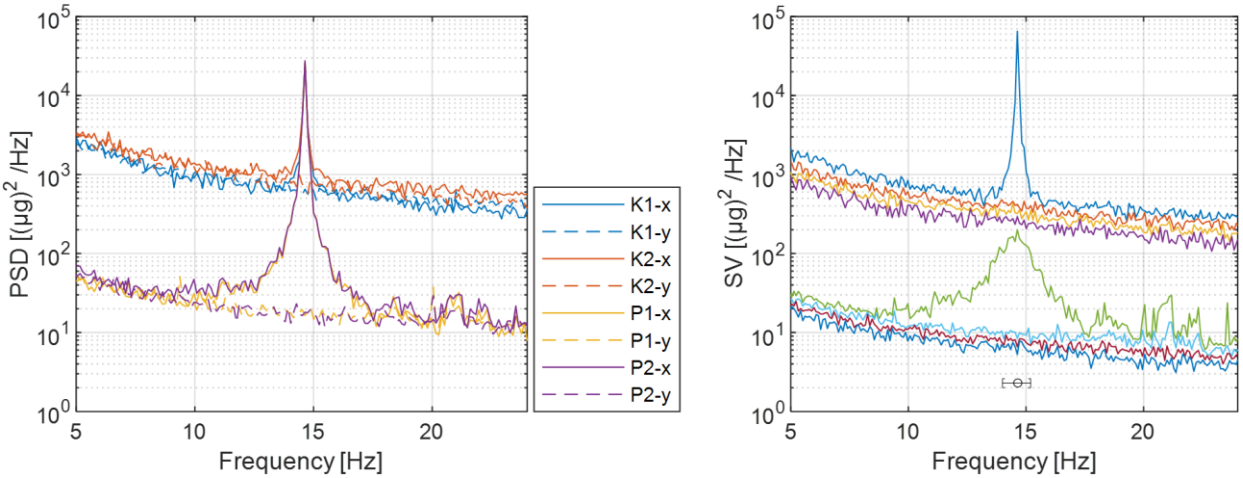


Fig. 11 PSD and SV spectrum. Laboratory data. Bracket in SV plot indicates the frequency band for modal identification

Effect of noise model on modal identification results

Table 2 summarises the modal identification results in terms of posterior MPV and c.o.v. (in parenthesis), and log-evidence based on different noise models. From the MPV and c.o.v., it is seen that the identification results for $\{f, \zeta, S\}$ are generally similar for different models, although a slight difference is seen in S. In particular, the differences between the MPVs (including S) for different models are within two posterior standard deviations of Model 0. The mode shapes identified based on different noise models are all very close, with a MAC of 0.9999 (M0 & M1), 0.9998 (M0 & M2) and 1.0000 (M1 & M2). The similarity of results based on different models may be explained by the fact that the modal s/n ratio is sufficiently high, as evidenced from the SV plot in Fig. 11.

Bayesian evidence

In terms of log-evidence, the last column of Table 2 shows that Model 1 > Model 2 > Model 0. The preference over Model 1 is intuitive because the K and P sensors have quite different levels of noise PSD. This noise disparity is also significant enough to prefer Model 2 over Model 0, despite the fact that Model 2 has a lot more parameters than Model 0. To supplement, when only the channels from P sensor are used, the log-evidence values rank as Model 0 (6.1982×10^4) > Model 1 (6.1878×10^4) > Model 2 (6.1735×10^4). This is intuitive, as in this case a reasonable data fit can already be attained with i.i.d. noise (Model 0), and so the ranking basically follows a descending order of the number of parameters (the more the worse).

Table 2 Summary of identification results (MPV, c.o.v. in parenthesis) and log-evidence. Laboratory data

Model	f [Hz]	ζ [%]	S [(μg)^2/Hz]	log-evidence [$\times 10^5$]
0	14.624 (0.02%)	0.19 (10.8%)	3.2512 (4.9%)	1.1214
1	14.624 (0.02%)	0.19 (10.5%)	3.2246 (4.2%)	1.1688
2	14.625 (0.02%)	0.18 (12.0%)	2.9477 (7.1%)	1.1622

Transformed data OMA and Uncertainty law

We next investigate modal identification based on i.i.d. noise (Model 0) using transformed data in Section 4.1. For this purpose, the reference noise PSD in (39) is set to be $S'_{e0} = 1(\mu g)^2/\text{Hz}$. Based on a rough estimate of noise PSDs from Fig. 11 for the P channels ($20(\mu g)^2/\text{Hz}$) and K channels ($700(\mu g)^2/\text{Hz}$), we set r_i^2 in (39) to be 20 for the P channels and 700 for the K channels. Fig. 12 shows the PSD and SV plot of the transformed data, which may be compared with Fig. 11 for the original laboratory data. The SV plot is now similar to the one with identical noise PSD.

Based on i.i.d. noise (Model 0) and applied to the transformed data, the identification results are (MPV, followed by c.o.v. in parenthesis), 14.624 Hz (0.02%) for frequency f, 0.19% (10.5%) for damping ζ and 3.2181 $(\mu g)^2/\text{Hz}$ (4.2%) for modal force PSD S (after transforming back to the original problem). The MAC of mode shape $\boldsymbol{\varphi}$ with its counterpart based on original data is also very high (0.9999). It is seen that the results for f and ζ are almost identical to their counterparts for Model 0 before transformation (see the second row of Table 2). The results for S is now closer to those for Model 1 (see the third row of Table 2). Of course, this is just for the sake of comparison. As mentioned before, the results in Table 2 for the three noise models do not make a practical difference, because the modal s/n ratio is high.

Finally, based on MPV, the modal s/n ratio of the transformed data in (48) is $\gamma' = 5269$. Correspondingly, the c.o.v.s based on uncertainty law in (45) with γ replaced by γ' are 0.019% for f, 10.5% for ζ and 4.1% for S. These values are quite close to those for Model 1 in the third row of Table 2.

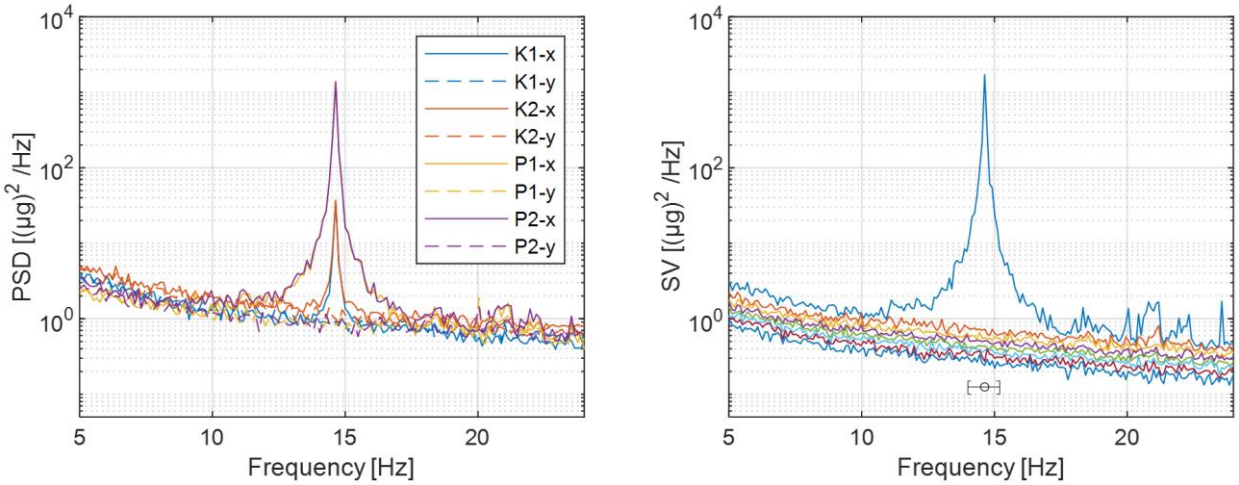


Fig. 12 PSD and SV spectrum. Transformed laboratory data. Bracket in SV plot indicates the frequency band for modal identification. See Fig. 11 for the plots of the original data (before transformation)

6.2. Queen's Park suspension footbridge (noise correlation)

Consider a series of ambient vibration data from multiple setups collected from Queen's Park footbridge across the River Dee in Chester, UK. The data is available for public access in [58]. The test series covered 66 locations distributed on two sides of the bridge in 16 setups with two reference sensors and four roving sensors. Details of the test and its investigation on multi-setup OMA can be found in [42]. All sensors are tri-axial servo-accelerometers and their data are synchronised by GPS. The data in each setup is 15 minutes long and sampled at 200 Hz.

The present study considers the data measured at the two reference sensors. Since an additional single setup was performed at the end of the multiple-setup test with the measurement time being 30 minutes, there are total 18 datasets (15 minutes each) available from the reference sensors, whose locations were fixed near the quarter and mid-span in all setups. Focusing on the band below 4Hz, the PSD and SV spectra of the data are shown in Fig. 13. Although the noise PSDs of the sensors in a laboratory environment were calibrated to be in the order of 1 $(\mu g)^2/\text{Hz}$, the PSD and SV plots suggest a higher noise floor with possible disparity. The brackets in the SV plot indicate the frequency bands around the modes that will be investigated here. These modes are selected as their SV display a reasonable dynamic amplification that allows proper modal identification, but at the same time their noise floors, i.e., lines below the top SV line, display complications (deviating from i.i.d. noise) that serve the intended scope of the study.

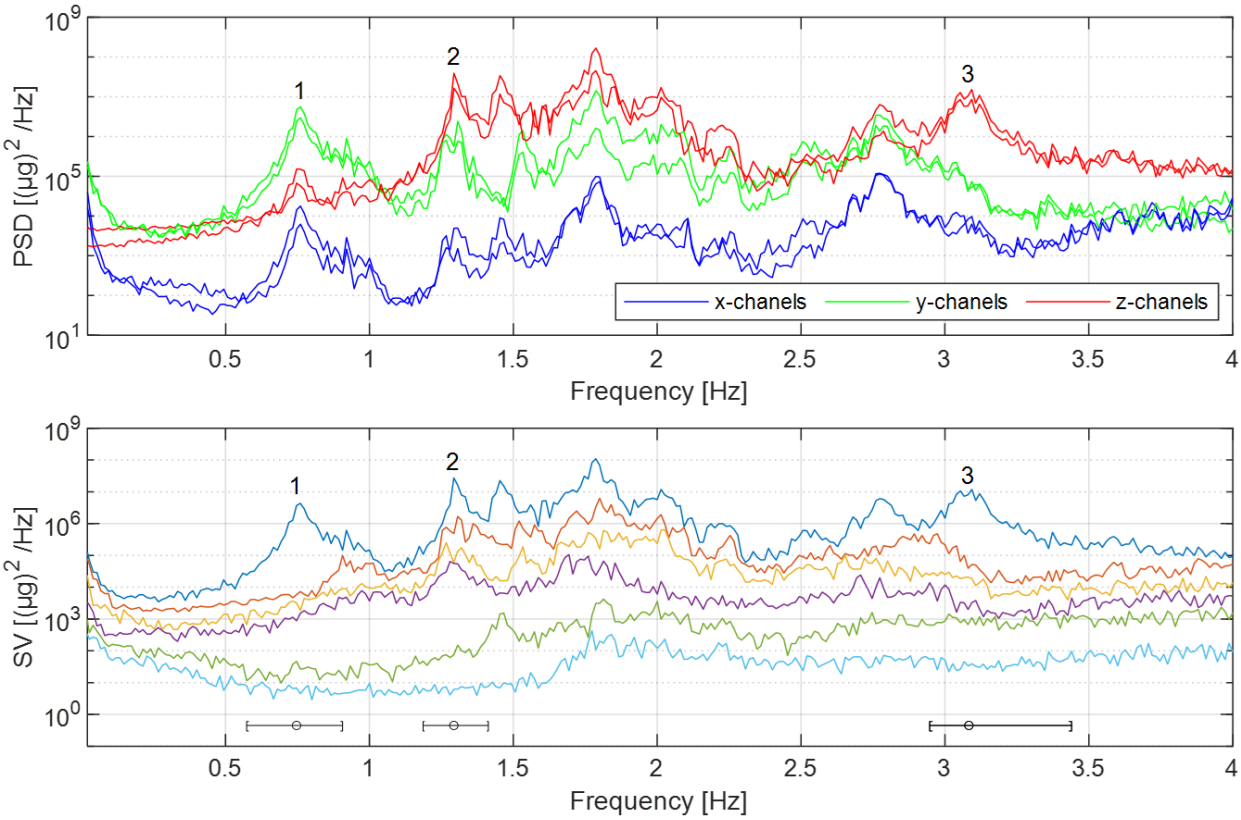


Fig. 13 PSD and SV of data in Setup 16. Queen's Park bridge data. Bracket in SV plot indicates frequency band for modal identification

Effect of noise model on MPV

Using the data in each setup, the posterior MPV and covariance matrix of modal parameters have been calculated for different noise models (0, 1, 2). Fig. 14 shows the MPVs based on Model 0 (blue dot) and Model 1 (red dot) versus those based on Model 2 (x-axis). The (dashed) 1:1 line is shown to facilitate comparison. The values from Model 2 are shown on the x-axis to serve as a reference for comparison, as it turns out to have the highest log-evidence (see later) for the modes investigated here. For the natural frequency of Mode 1, it is seen that the MPVs of Models 0 and 1 are slightly above the 1:1 line, i.e., higher than those of Model 2. For Modes 2 and 3, the values of Models 0 and 1 are scattered around the 1:1 line with no significant offset. For damping ratios, the MPVs tend to lie off the 1:1 line, noticeable in Modes 1 and 2 for Model 0, Modes 1 and 3 for Model 1. Roughly speaking, the damping ratios are more significantly affected by noise models than the frequencies.

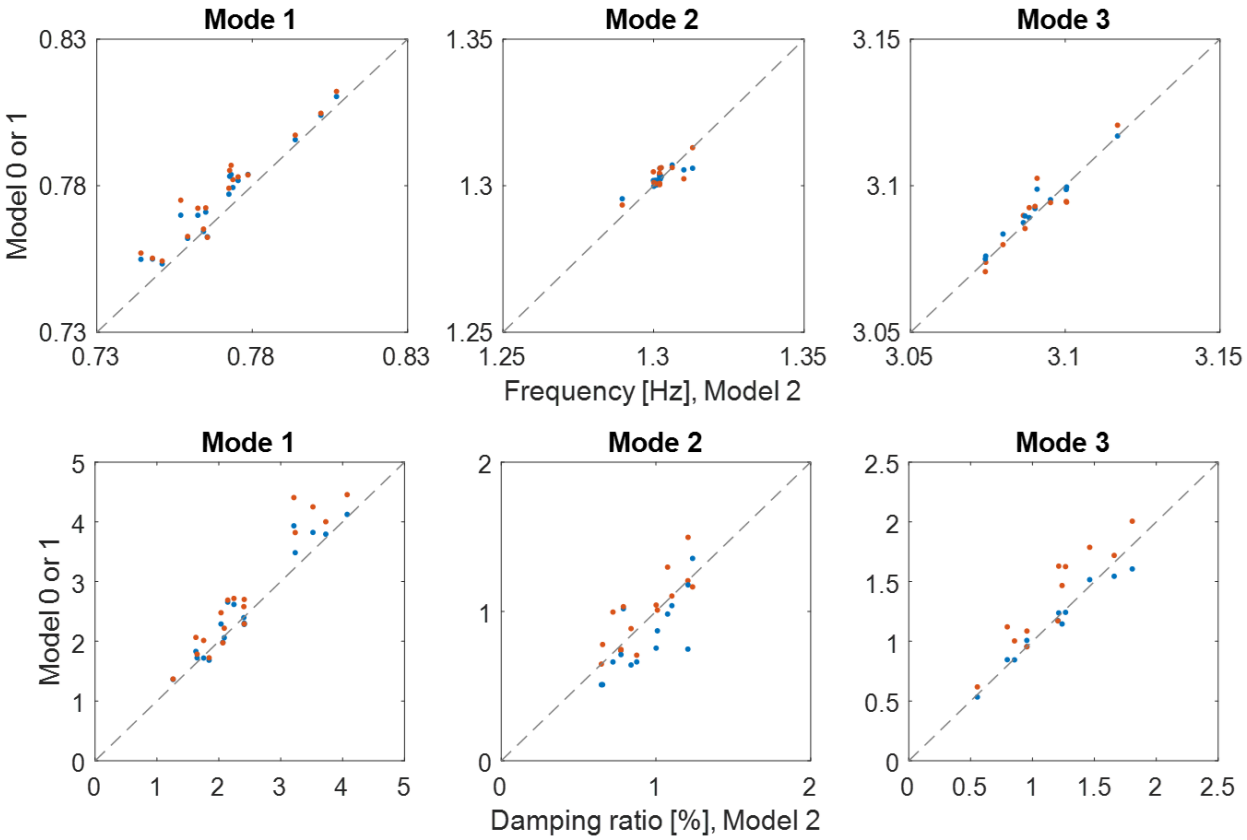


Fig. 14 MPV of Model 0 (blue dot) and Model 1 (red dot) versus those of Model 2 (x-axis). Queen's Park bridge data. Each dot represents the result of a setup.

Bayesian evidence

Fig. 15 shows the log-evidence of different models. Similar to Fig. 14, the values of Model 0 (blue dot) and Model 1 (red dot) are plotted on the y-axis; the values of Model 2 are plotted on the x-axis as a reference for comparison. It is seen that for these three modes the ranking is, consistently for all setups, Model 2 > Model 1 > Model 0. This is not surprising, considering the complications in the 'noise floors' (lines other than the top line) in the SV plot of Fig. 13 near the modes. The margin by which Model 1 (independent noise) is preferred over Model 0 (i.i.d. noise) is bigger than that of Model 2 (correlated noise) over Model 1. This may suggest that, on the basis of log-evidence, the complicated noise floors demand at least a model to account for noise disparity (Model 1). A further complication in Model 2 to account for correlation is still justified, although marginally.

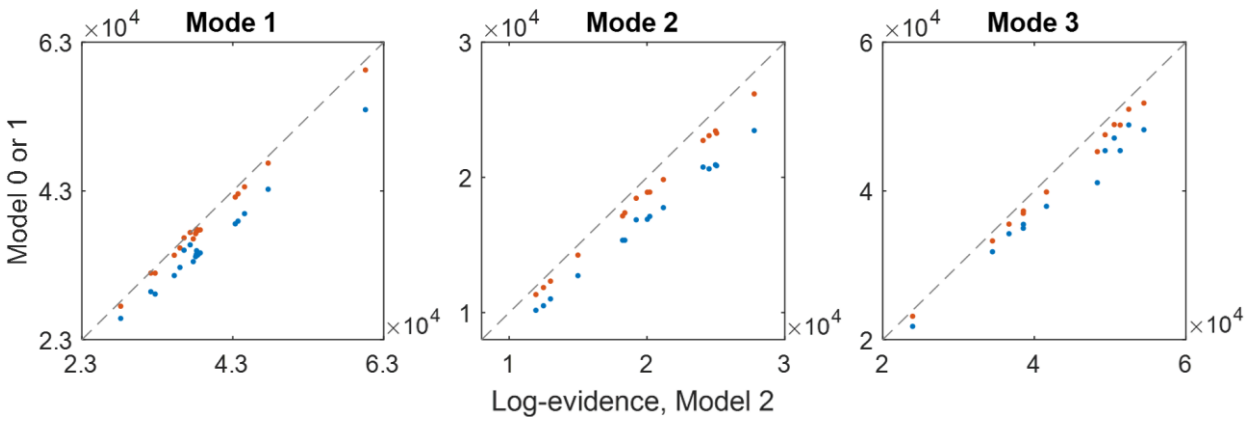


Fig. 15 Log-evidence of Model 0 (blue dot) and Model 1 (red dot) versus those of Model 2 (x-axis). Queen's Park bridge data. Each dot represents the result of a setup.

7. Conclusions

This work has presented a comprehensive investigation on the effect of noise models in Bayesian frequency-domain OMA. Key contributions are multi-folded in terms of theory, computation and insights. Beyond the i.i.d. noise model that is conventionally used, two generalised noise models (see Table 1) have been considered to account for noise disparity and correlation that may be encountered in applications. To enable the study, new efficient computational algorithms applicable for general noise models have been developed for the posterior (i.e., given data) most probable value (Section 3.1) and covariance matrix (Section 3.4; quantifying uncertainty). In addition to basic modal identification results, the noise models have been assessed from the perspective of Bayesian model class evidence (Section 3.5). As a contribution to OMA theory, it has been discovered (Section 4.1) that by a suitable transformation of data, OMA with general noise model is mathematically equivalent to OMA with i.i.d. noise model. Besides allowing one to empirically account for general noise model based on existing tools of i.i.d. noise (see, e.g., Section 5.1), this analogy allows the development of asymptotic formulae for identification uncertainty, i.e., 'uncertainty law', of OMA for general noise models (Section 4.2), which is otherwise nontrivial and has not been obtained so far. The results reveal that for general noise models the modal s/n ratio can be quantified by (48), which is an intuitive yet nontrivial extension of the existing formula (47) for i.i.d. noise.

The proposed objectives and methodology have been investigated and verified using synthetic (Section 5), laboratory (Section 6.1) and field data (Section 6.2). Some comments on the effects of noise and their practical treatment are in order. Generally, whether a noise model matters to the

modal identification results depend on the modal s/n ratio (the higher the less), the extent of deviation (e.g., disparity and correlation), as well as the parameter of interest. The study reveals that the noise models do not have a material effect on the identification result of natural frequency, but the same cannot be taken for granted for damping ratio; see, e.g., Fig. 3 (with noise disparity) and Fig. 9 (with noise correlation). The laboratory data example in Section 6.1 reveals that, when the modal s/n ratio is sufficiently high ($\gamma' > 5000$ in the example), the modal identification results may not be significantly affected.

In the examples, the ranking of noise models based on Bayesian model class evidence generally agree with observations of identification results and PSD/SV spectra. While it provides a sound scientific perspective, it should be noted that the model ranking need not have a direct bearing on the quality of modal identification results of parameters of interest. In reality the choice of a noise model may be decided by more practical considerations such as experience gained with existing noise model and computational effort/robustness. For the former, modal identification results are not 'fool-proof' and still need to be interpreted by knowledgeable personnel. For the latter, a complex noise model may take longer time for identification, or its MPV may not even converge when there is not enough information in data. When existing algorithm based on i.i.d. noise model is adopted, an OMA based on transformed data (Section 4) may provide a simple way to empirically account for noise disparity and correlation.

8. Acknowledgements

The research presented in this work is supported by grant SUG/4 (04INS000618C120) from the Nanyang Technological University (NTU), Singapore. The first author would like to acknowledge the graduate research scholarship offered by NTU.

Appendix A Matrix identities used in this work

This appendix summarises the matrix identities that are useful in developing the paper. A good reference can be found in [59].

Matrix Inverse Lemma

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1}\boldsymbol{UC}=\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1} \tag{56}$$

$$(\boldsymbol{A}+\boldsymbol{UCV})^{-1}=\boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\boldsymbol{U}(\boldsymbol{C}^{-1}+\boldsymbol{VA}^{-1}\boldsymbol{U})^{-1}\boldsymbol{VA}^{-1} \tag{57}$$

Derivatives of Trace

When $\boldsymbol{X}(n \times m)$ has functionally independent elements (i.e., $\boldsymbol{X}$ is not symmetric, etc.)

$$\frac{\partial \mathrm{tr}(\boldsymbol{XA})}{\partial \boldsymbol{X}} = \boldsymbol{A}^T \tag{58}$$

when $\boldsymbol{X}$ is symmetric

$$\frac{\partial \mathrm{tr}(\boldsymbol{XA})}{\partial \boldsymbol{X}} = 2\boldsymbol{A} - \boldsymbol{A} \circ \boldsymbol{I} \tag{59}$$

where '$\circ$' represents Hadamard (elementwise) product.

Derivatives of log-determinant

For nonsingular $\boldsymbol{X}$ with functionally independent elements

$$\frac{\partial \ln|\boldsymbol{X}|}{\partial \boldsymbol{X}} = \boldsymbol{X}^{-T} \tag{60}$$

If $\boldsymbol{X}$ is symmetric

$$\frac{\partial \ln|\boldsymbol{X}|}{\partial \boldsymbol{X}} = 2\boldsymbol{X}^{-1} - \boldsymbol{X}^{-1} \circ \boldsymbol{I} \tag{61}$$

Appendix B Derivations related to EM algorithm

B.1 Complete-data likelihood and conditional statistics in (7) - (9)

This appendix provides the proof of the equations in the section title for general noise models (1 and 2). It is similar to the one in previous work for i.i.d. noise (Model 0) [41], except that the noise PSD matrix $\boldsymbol{S}_e$ now needs to be retained without further simplification that was possible before.

For (7), we first write $p\big(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}|\boldsymbol{\theta}\big) = p\big(\{\widehat{\boldsymbol{\mathcal{F}}}_k\}|\{\boldsymbol{\eta}_k\}, \boldsymbol{\theta}\big)p(\{\boldsymbol{\eta}_k\}|\boldsymbol{\theta})$. Given $\boldsymbol{\theta}$ (hence $\boldsymbol{\Phi}$) and $\boldsymbol{\eta}_k$, $\widehat{\boldsymbol{\mathcal{F}}}_k$ is complex Gaussian with mean $\boldsymbol{\Phi\eta}_k$ and covariance matrix $\boldsymbol{S}_e$. Also, given $\boldsymbol{\theta}$, $\boldsymbol{\eta}_k$ is complex Gaussian with mean zero and covariance matrix $\boldsymbol{H}_k$. These give

$$\begin{aligned} L\big(\boldsymbol{\theta}|\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}\big) &= \ln p\big(\{\widehat{\boldsymbol{\mathcal{F}}}_k, \boldsymbol{\eta}_k\}|\boldsymbol{\theta}\big) = \sum_k \ln p\big(\hat{\mathcal{F}}_k|\boldsymbol{\eta}_k, \boldsymbol{\theta}\big) \; + \sum_k \ln p(\boldsymbol{\eta}_k|\boldsymbol{\theta}) \\ &= -2nN_f \ln\pi - N_f \ln|\boldsymbol{S}_e| - \sum_k \boldsymbol{\eta}_k^* \mathbf{H}_k^{-1} \boldsymbol{\eta}_k + \sum_k \ln|\mathbf{H}_k^{-1}| \\ &\quad - \sum_k \big[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi\eta}_k\big]^* \boldsymbol{S}_e^{-1} \big[\widehat{\boldsymbol{\mathcal{F}}}_k - \boldsymbol{\Phi\eta}_k\big] \end{aligned} \tag{62}$$

For (8) and (9), we make use of the standard results of conditional Gaussian variates. Let $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ be jointly complex Gaussian vectors with respectively auto-covariance matrices $\boldsymbol{C}_{11}$ and $\boldsymbol{C}_{22}$, and cross covariance matrix $E[\boldsymbol{X}_1\boldsymbol{X}_2^*] = \boldsymbol{C}_{12}$. Then, given $\boldsymbol{X}_1$, the conditional mean and covariance matrix of $\boldsymbol{X}_2$ are respectively $\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{X}_1$ and $\boldsymbol{C}_{21}\boldsymbol{C}_{11}^{-1}\boldsymbol{C}_{12}$. Applying this result with $\boldsymbol{X}_1 = \widehat{\boldsymbol{\mathcal{F}}}_k$ and $\boldsymbol{X}_2 = \boldsymbol{\eta}_k$, $\boldsymbol{C}_{11} = E\left[\widehat{\boldsymbol{\mathcal{F}}}_k\widehat{\boldsymbol{\mathcal{F}}}_k^*|\boldsymbol{\theta}\right] = \boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^T + \boldsymbol{S}_e$, and $\boldsymbol{C}_{12} = \boldsymbol{C}_{21}^* = E\left[\widehat{\boldsymbol{\mathcal{F}}}_k\boldsymbol{\eta}_k|\boldsymbol{\theta}\right] = \boldsymbol{\Phi}\boldsymbol{H}_k$, the conditional mean and covariance matrix of $\boldsymbol{\eta}_k$ are given by

$$\begin{aligned}\mathrm{E}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}}[\boldsymbol{\eta}_k] &= \boldsymbol{H}_k\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^T + \boldsymbol{S}_e)^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k \\ &= (\boldsymbol{H}_k^{-1} + \boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1}\widehat{\boldsymbol{\mathcal{F}}}_k\end{aligned} \tag{63}$$

$$\begin{aligned}\mathrm{Cov}_{\boldsymbol{\eta}_k|\widehat{\boldsymbol{\mathcal{F}}}_k,\boldsymbol{\theta}}[\boldsymbol{\eta}_k] &= \boldsymbol{H}_k - \boldsymbol{H}_k\boldsymbol{\Phi}^T(\boldsymbol{\Phi}\boldsymbol{H}_k\boldsymbol{\Phi}^T + \boldsymbol{S}_e)^{-1}\boldsymbol{\Phi}\boldsymbol{H}_k \\ &= (\boldsymbol{H}_k^{-1} + \boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1}\boldsymbol{\Phi})^{-1}\end{aligned} \tag{64}$$

The identities in (56) and (57) have been used in arriving at the last equalities in (63) and (64), respectively.

B.2 Update formulae for $\boldsymbol{S}_e$ in (17) and (18)

Here, we derive the update formula for $\boldsymbol{S}_e$ in the EM algorithm for general noise models. The case of Model 2 where $\boldsymbol{S}_e$ is a general real-symmetric matrix is considered first. The case of Model 1 where $\boldsymbol{S}_e$ is a diagonal matrix will be considered later.

In principle, the update formula for $\boldsymbol{S}_e$ can be derived by solving $\partial Q/\partial \boldsymbol{S}_e = \boldsymbol{0}$, where the Q-function is from (11). To facilitate derivation, we let $\boldsymbol{X} = \boldsymbol{S}_e^{-1}$ and consider the stationary condition $\partial Q/\partial \boldsymbol{X} = \boldsymbol{0}$, which is equivalent to $\partial Q/\partial \boldsymbol{S}_e = \boldsymbol{0}$ as long as $\boldsymbol{S}_e$ is invertible. Also, let $\boldsymbol{G}$ denote the term in the inner bracket of (12) that does not depend on S_e, so that $Q_1\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = \mathrm{tr}(\boldsymbol{S}_e^{-1}\boldsymbol{G})$. Then we can write

$$Q\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = N_f \ln|\boldsymbol{X}| - \mathrm{tr}(\boldsymbol{X}\boldsymbol{G}) + \text{constant of } \boldsymbol{S}_e \tag{65}$$

Using (59) and (61) respectively for the derivatives of trace and log-determinant for symmetric $\boldsymbol{X}$,

$$\frac{\partial Q}{\partial \boldsymbol{X}} = N_f(2\boldsymbol{X}^{-1} - \boldsymbol{X}^{-1} \circ \boldsymbol{I}_n) - (2\boldsymbol{G} - \boldsymbol{G} \circ \boldsymbol{I}_n) \tag{66}$$

where '$\circ$' denotes the Hadamard (element-wise) product. By observation, $\partial Q/\partial \boldsymbol{X} = \boldsymbol{0}$ when $\boldsymbol{X}^{-1} = \boldsymbol{G}/N_f$ and so, after substituting $\boldsymbol{G}$, the update formula for real-symmetric $\boldsymbol{S}_e$ is given by

$$\boldsymbol{S}_e = \frac{1}{N_f}\left[\sum_k \widehat{\boldsymbol{\mathcal{F}}}_k \widehat{\boldsymbol{\mathcal{F}}}_k^* - 2\boldsymbol{\Phi}\sum_k \mathrm{Re}(\boldsymbol{w}_{1k}\widehat{\boldsymbol{\mathcal{F}}}_k^*) + \boldsymbol{\Phi}\sum_k \mathrm{Re}(\boldsymbol{w}_{2k})\,\boldsymbol{\Phi}^T\right] \tag{67}$$

Substituting the update formula for $\boldsymbol{\Phi}$ from (14) into (67) and simplifying gives the final update formula for $\boldsymbol{S}_e$ in (18).

For independent noise (Model 1), $\boldsymbol{S}_e$ is a diagonal matrix and so is $\boldsymbol{X}$. In this case, only the equations corresponding to the diagonal entries of $\partial Q/\partial \boldsymbol{X} = \boldsymbol{0}$ in (66) are relevant. Correspondingly, diagonal entries of $\boldsymbol{S}_e$ that solve these equations are simply given by the diagonal entries in (67). Writing it in terms of Hadamard product gives the update formula in (17).

Appendix C Derivatives of $\boldsymbol{P_k}$, $\boldsymbol{L_Q}$ and $\boldsymbol{r_k}$

The derivatives of $\boldsymbol{P}_k$ in (10), L_Q in (25) and $\boldsymbol{r}_k$ in (26), are given in Table 3, Table 4 and Table 5, respectively. A superscribed variable in parenthesis denotes a derivative w.r.t. it. The derivatives w.r.t. the entries of $\boldsymbol{S}_e$ have already taken account of its symmetry. As such, the double derivatives $(\cdot/\partial S_{eij}\partial S_{ers})$ should not be used for $(i,j) = (s,r)$ for an off-diagonal entry (i,j). For independent noise (Model 1), only the derivatives w.r.t. the off-diagonal entries of $\boldsymbol{S}_e$ are irrelevant.

Table 3 Derivatives of $\boldsymbol{P}_k = \boldsymbol{H}_k^{-1} + \boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1}\boldsymbol{\Phi}$

$\boldsymbol{P}_k{}^{(xy)}$		$y =$		
		$f_j, \zeta_j, S_{jj}, U_{rs}, V_{rs}$	S_{ers}	Φ_{rs}
$x =$	$f_i, \zeta_i, S_{ii}, U_{ij}, V_{ij}$	$(\boldsymbol{H}_k^{-1})^{(xy)}$	0	0
	S_{eij}		$\boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1(xy)}\boldsymbol{\Phi}$	$\boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1(x)}\mathbf{e}_{rs} + (\cdot)^T$
	Φ_{ij}	Sym.		$\mathbf{e}_{ij}^T\boldsymbol{S}_e^{-1}\mathbf{e}_{rs} + (\cdot)^T$
$\boldsymbol{P}_k{}^{(y)}$		$(\boldsymbol{H}_k^{-1})^{(y)}$	$\boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1(y)}\boldsymbol{\Phi}$	$\boldsymbol{\Phi}^T\boldsymbol{S}_e^{-1}\mathbf{e}_{rs} + (\cdot)^T$

See Table 13.3 in [3] for derivatives of $\boldsymbol{H}_k^{-1}$. Φ_{ij} is the (i,j) entry in $\boldsymbol{\Phi}$, $U_{ij} = \mathrm{Re}S_{ij}$, $V_{ij} = \mathrm{Im}S_{ij}$, and $(\cdot)^T$ denotes a term equal to the transpose of the previous one.

Table 4 Derivatives of $L_Q = \sum_k \boldsymbol{r}_k^* \boldsymbol{P}_k^{-1} \boldsymbol{r}_k$

$L_Q{}^{(xy)}$		$y =$	
		$f_j, \zeta_j, S_{jj}, U_{rs}, V_{rs}$	S_{ers}, Φ_{rs}
$x =$	$f_i, \zeta_i, S_{ii}, U_{ij}, V_{ij}$	$\sum_k \boldsymbol{r}_k^* (\boldsymbol{P}_k^{-1})^{(xy)} \boldsymbol{r}_k$	$\sum_k \boldsymbol{r}_k^* (\boldsymbol{P}_k^{-1})^{(xy)} \boldsymbol{r}_k + 2\mathrm{Re}\left[\sum_k \boldsymbol{r}_k^{*(y)} \boldsymbol{P}_k^{-1(x)} \boldsymbol{r}_k\right]$
	S_{eij}, Φ_{ij}	Sym.	$\sum_k \boldsymbol{r}_k^* (\boldsymbol{P}_k^{-1})^{(xy)} \boldsymbol{r}_k + 2\mathrm{Re}\left[\sum_k \boldsymbol{r}_k^{*(x)} (\boldsymbol{P}_k^{-1})^{(y)} \boldsymbol{r}_k\right] + 2\mathrm{Re}\left[\sum_k \boldsymbol{r}_k^{*(xy)} \boldsymbol{P}_k^{-1} \boldsymbol{r}_k\right] + 2\mathrm{Re}\left[\sum_k \boldsymbol{r}_k^{*(y)} (\boldsymbol{P}_k^{-1})^{(x)} \boldsymbol{r}_k\right] + 2\mathrm{Re}\left[\sum_k \boldsymbol{r}_k^{*(y)} \boldsymbol{P}_k^{-1} \boldsymbol{r}_k{}^{(x)}\right]$
$L_Q{}^{(y)}$		$\sum_k \boldsymbol{r}_k^* (\boldsymbol{P}_k^{-1})^{(y)} \boldsymbol{r}_k$	$\sum_k \boldsymbol{r}_k^* (\boldsymbol{P}_k^{-1})^{(y)} \boldsymbol{r}_k + 2\mathrm{Re}\left[\sum_k \boldsymbol{r}_k^{*(y)} \boldsymbol{P}_k^{-1} \boldsymbol{r}_k\right]$

Table 5 Derivatives of $\boldsymbol{r}_k = \boldsymbol{\Phi}^T \boldsymbol{S}_e^{-1} \widehat{\boldsymbol{\mathcal{F}}}_k$

$\boldsymbol{r}_k{}^{(xy)}$		$y =$	
		S_{ers}	Φ_{rs}
$x =$	$S_{e_{ij}}$	$\boldsymbol{\Phi}^T \boldsymbol{S}_e^{-1(xy)} \widehat{\boldsymbol{\mathcal{F}}}_k$	$\mathbf{e}_{rs}^T \boldsymbol{S}_e^{-1(y)} \widehat{\boldsymbol{\mathcal{F}}}_k$
	Φ_{ij}	Sym.	0
$\boldsymbol{r}_k{}^{(y)}$		$\boldsymbol{\Phi}^T \boldsymbol{S}_e^{-1(y)} \widehat{\boldsymbol{\mathcal{F}}}_k$	$\mathbf{e}_{rs}^T \boldsymbol{S}_e^{-1} \widehat{\boldsymbol{\mathcal{F}}}_k$

Appendix D Further uncertainty law expression for δ_S^2

In this appendix we show that the uncertainty law expression for δ_S^2 in (53) can be further expressed as

$$\delta_S^2 \sim \delta_{S'}^2 \left(1 + \frac{b}{\gamma'}\right) \tag{68}$$

Here, for the equivalent problem (indicated by a dash), $\gamma' = S'/4\zeta^2 S_e'$ is the modal s/n ratio in (48);

$$\delta_{S'}^2 = \frac{1}{N_f B_S}\left(1 + \frac{a_S}{\gamma'}\right) \tag{69}$$

is the uncertainty law c.o.v. (squared) of S' obtained by replacing γ with γ' in the rightmost expression in (45); and

$$b = \frac{4\kappa}{\pi}\frac{B_S}{B_\varphi}\left(1+\frac{a_S}{\gamma'}\right)^{-1}\frac{1}{\mu^2}\sum_{i=1}^{n}(r_i^2-\mu)^2\varphi_i'^2 \tag{70}$$

where φ_i' is the mode shape value at measured DOF i; $\{r_i^2\}_{i=1}^n$ are the ratios of noise PSD defined in (39), and

$$\mu = \sum_{i=1}^{n} r_i^2\varphi_i'^2 \tag{71}$$

is their $\varphi_i'^2$-weighted mean (recall that $\sum_{i=1}^n \varphi_i'^2 = 1$). It is clear from (70) that $b > 0$, and $b = 0$ if $r_i =$ constant, because then $r_i^2 = \mu$ for all $i = 1, \dots, n$. This verifies the claim mentioned after (53).

For convenience, let

$$F = 4\|\boldsymbol{R}\boldsymbol{\varphi}'\|^{-4}\boldsymbol{\varphi}'^T\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{C}_{\boldsymbol{\varphi}'}\boldsymbol{R}^T\boldsymbol{R}\boldsymbol{\varphi}' \qquad \text{(2nd term in (53))} \tag{72}$$

be the second term in (53), so that $\delta_S^2 = \delta_{S'}^2 + F = \delta_{S'}^2\left(1 + F/\delta_{S'}^2\right)$. The proof of (68) then essentially involves showing

$$\frac{F}{\delta_{S'}^2} = \frac{b}{\gamma'} \qquad \text{(to show)} \tag{73}$$

To begin the proof, first note from (43) that $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{r}$, where $\boldsymbol{V}$ is a matrix of orthonormal basis ($\boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}_n$, see (38)) and $\boldsymbol{r}$ is a diagonal matrix of $\{r_i\}_{i=1}^n$ (see (39)). This gives $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{r}^T\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{r} = \boldsymbol{r}^T\boldsymbol{r}$, which is a diagonal matrix of $\{r_i^2\}_{i=1}^n$. On the other hand, using (46) for the equivalent problem gives $\boldsymbol{C}_{\boldsymbol{\varphi}'} = \left(2S_e'\zeta/\pi N_c S' B_\varphi\right)\left(\boldsymbol{I}_n - \boldsymbol{\varphi}'\boldsymbol{\varphi}'^T\right)$. Substituting these into (72) and noting that $\|\boldsymbol{R}\boldsymbol{\varphi}'\|^2 = \boldsymbol{\varphi}'^T\boldsymbol{r}^T\boldsymbol{V}^T\boldsymbol{V}\boldsymbol{r}\boldsymbol{\varphi}' = \sum_{i=1}^n r_i^2\varphi_i'^2 = \mu$ (see (71)), one obtains, after simplification,

$$F = \frac{8S_e'\zeta}{\pi N_c S' B_\varphi}\left(\frac{\sum_{i=1}^n r_i^4\varphi_i'^2}{\mu^2} - 1\right) \tag{74}$$

By viewing $\left\{\varphi_i'^2\right\}_{i=1}^n$ as weights ($\sum_{i=1}^n \varphi_i'^2 = 1$), the sums $\sum_{i=1}^n r_i^4\varphi_i'^2$ and $\mu = \sum_{i=1}^n r_i^2\varphi_i'^2$ can be

viewed respectively as the $\varphi_i'^2$-weighted second moment and mean of $\{r_i^2\}_{i=1}^n$. Using the standard result that the second moment is equal to the sum of variance and square of mean,

$$\sum_{i=1}^{n} r_i^4 \varphi_i'^2 = \sum_{i=1}^{n} (r_i^2 - \mu)^2 \varphi_i'^2 + \left(\sum_{i=1}^{n} r_i^2 \varphi_i'^2 \right)^2 = \sum_{i=1}^{n} (r_i^2 - \mu)^2 \varphi_i'^2 + \mu^2 \tag{75}$$

Substituting into (74) and simplifying gives

$$F = \frac{8 S_e' \zeta}{\pi N_c S' B_\varphi} \cdot \frac{1}{\mu^2} \sum_{i=1}^{n} (r_i^2 - \mu)^2 \varphi_i'^2 \tag{76}$$

Dividing F in (76) by $\delta_{S'}^2$ in (69) with $N_f = 2\kappa\zeta N_c$ gives, after simplification,

$$\frac{F}{\delta_{S'}^2} = \frac{16\kappa S_e' \zeta^2 B_S}{\pi S' B_\varphi} \left(1 + \frac{a_S}{\gamma'}\right)^{-1} \frac{1}{\mu^2} \sum_{i=1}^{n} (r_i^2 - \mu)^2 \varphi_i'^2 \tag{77}$$

Equation (73) then follows by recognising $\gamma' = S'/4\zeta^2 S_e'$ and the definition of b in (70).

References

[1] R. Brincker, C. Ventura, Introduction to Operational Modal Analysis, Wiley, Chichester, 2015.

[2] Helmut. Wenzel, Dieter. Pichler, Ambient Vibration Monitoring, John Wiley, Chichester, 2005.

[3] S.K. Au, Operational Modal Analysis, Springer, Singapore, 2017.

[4] J.M.W. Brownjohn, Structural health monitoring of civil infrastructure, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 365 (2007) 589–622.

[5] D.J. Ewins, Modal Testing: Theory, Practice and Application, Research Studies Press LTD., Baldock, Hertfordshire, England, 2000.

[6] P. van Overschee, B. de Moor, Subspace Identification for Linear Systems: Theory, Implementation, Applications, Kluwer Academic Publishers, 1996.

[7] B. Peeters, G. de Roeck, Stochastic system identification for operational modal analysis: a review, Journal of Dynamic Systems, Measurement, and Control. 123 (2001) 659–667.

[8] E. Reynders, System identification methods for (operational) modal analysis: review and comparison, Archives of Computational Methods in Engineering. 19 (2012) 51–124.

[9] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, E. Moulines, A blind source separation technique

using second-order statistics, IEEE Transaction on Signal Processing. 45 (1997) 434–444.

[10] F. Poncelet, G. Kerschen, J.C. Golinval, D. Verhelst, Output-only modal analysis using blind source separation techniques, Mechanical Systems and Signal Processing. 21 (2007) 2335–2358.

[11] K.V. Yuen, L.S. Katafygiotis, Bayesian time-domain approach for modal updating using ambient data, Probabilistic Engineering Mechanics. 16 (2001) 219–231.

[12] R. Pintelon, J. Schoukens, System identification: a frequency domain approach, Wiley-IEEE Press, 2001.

[13] R. Brincker, L. Zhang, P. Andersen, Modal identification of output-only systems using frequency domain decomposition, Smart Materials and Structures. 10 (2001) 441–445.

[14] Ç. Hızal, Modified frequency and spatial domain decomposition method based on maximum likelihood estimation, Engineering Structures. 224 (2020) 111007.

[15] S.R. Ibrahim, Random decrement technique for modal identification of structures, Journal of Spacecraft and Rockets. 14 (1977) 596–700.

[16] J.C. Asmussen, R. Brincker, S.R. Ibrahim, Statistical theory of the vector random decrement technique, Journal of Sound and Vibration. 226 (1999) 329–344.

[17] R. Brincker, S. Amador, On the theory of random decrement, Mechanical Systems and Signal Processing. 173 (2022) 109060.

[18] J. Brownjohn, M. Bocciolone, A. Curami, M. Falco, A. Zasso, Humber bridge full-scale measurement campaigns 1990-1991, Journal of Wind Engineering and Industrial Aerodynamics. 52 (1994) 185–218.

[19] S. Kim, S. Pakzad, D. Culler, J. Demmel, G. Fenves, S. Glaser, M. Turon, Health monitoring of civil infrastructures using wireless sensor networks, in: Proceedings of the 6th International Conference on Information Processing in Sensor Networks, Association for Computing Machinery, 2007: pp. 254–263.

[20] J.H. Weng, C.H. Loh, J.P. Lynch, K.C. Lu, P.Y. Lin, Y. Wang, Output-only modal identification of a cable-stayed bridge using wireless monitoring systems, Engineering Structures. 30 (2008) 1820–1830.

[21] S. Sumitroa, Y. Matsuia, M. Koflo', T. Okamotoa, K. Fujii, Long span bridge health monitoring system in Japan, in: The 6th Annual International Symposium on NDE for Health Monitoring and Diagnostics, Society of Photo-Optical Instrumentation Engineers (SPIE), Newport Beach, CA, United States, 2001: pp. 517–524.

[22] H. Katsuchi, H. Yamada, M. Asce, S. Kusuhara, Structural monitoring and design verification of Akashi Kaikyo Bridge, in: 11th Biennial ASCE Aerospace Division International Conference on Engineering, Science, Construction, and Operations in Challenging Environments, 2008: pp. 1–8.

[23] S.H. Sung, J.W. Park, T. Nagayama, H.J. Jung, A multi-scale sensing and diagnosis system combining accelerometers and gyroscopes for bridge health monitoring, Smart Materials and

Structures. 23 (2014) 015005.

[24] K.V. Yuen, L.S. Katafygiotis, Bayesian fast Fourier transform approach for modal updating using ambient data, Advances in Structural Engineering. 6 (2003) 81–95.

[25] R. Sleeman, A. van Wettum, J. and Trampert, Three-channel correlation analysis: A new technique to measure instrumental noise of digitizers and seismic sensors, Bulletin of the Seismological Society of America. 96 (2006) 258–271.

[26] Z. Zhu, S.K. Au, X. Wang, Instrument noise calibration with arbitrary sensor orientations, Mechanical Systems and Signal Processing. 117 (2019) 879–892.

[27] S. Dorvash, S.N. Pakzad, Effects of measurement noise on modal parameter identification, Smart Materials and Structures. 21 (2012) 065008.

[28] Ç. Hızal, E. Aktaş, Probabilistic investigation of error propagation in frequency domain decomposition-based operational modal analysis, Structural Control and Health Monitoring. 28 (2021) e2759.

[29] Ç. Hızal, H. Ceylan, Influence of non-uniform noise levels on modal identification procedures, Measurement: Sensors. 18 (2021) 100129.

[30] F. Zhang, H. Xiong, W. Shi, X. Ou, Structural health monitoring of Shanghai Tower during different stages using a Bayesian approach, Structural Control and Health Monitoring. 23 (2016) 1366–1384.

[31] Y. Ni, X. Lu, W. Lu, Operational modal analysis of a high-rise multi-function building with dampers by a Bayesian approach, Mechanical Systems and Signal Processing. 86 (2017) 286–307.

[32] H.F. Lam, F.L. Zhang, Y.C. Ni, J. Hu, Operational modal identification of a boat-shaped building by a Bayesian approach, Engineering Structures. 138 (2017) 381–393.

[33] J.M.W. Brownjohn, S.K. Au, Y. Zhu, Z. Sun, B. Li, J. Bassitt, E. Hudson, H. Sun, Bayesian operational modal analysis of Jiangyin Yangtze River Bridge, Mechanical Systems and Signal Processing. 110 (2018) 210–230.

[34] J.L. Beck, K.V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, Journal of Engineering Mechanics. 130 (2004) 192–203.

[35] T. Yin, H. Zhu, S. Fu, Model selection for dynamic reduction-based structural health monitoring following the Bayesian evidence approach, Mechanical Systems and Signal Processing. 127 (2019) 306–327.

[36] J.H. Yang, Q.Z. Kong, H.J. Liu, H.Y. Peng, Efficient Bayesian model class selection of vector autoregressive models for system identification, Structural Control and Health Monitoring. 28 (2021) e2780.

[37] S.K. Au, Uncertainty law in ambient modal identification - Part I: Theory, Mechanical Systems and Signal Processing. 48 (2014) 15–33.

[38] S.K. Au, B. Li, J.M.W. Brownjohn, Achievable precision of close modes in operational modal

analysis: Wide band theory, Mechanical Systems and Signal Processing. 147 (2021) 107016.

[39] S.K. Au, Fast Bayesian FFT method for ambient modal identification with separated modes, Journal of Engineering Mechanics. 137 (2011) 214–226.

[40] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, Journal of the Royal Statistical Society: Series B (Methodological). 39 (1977) 1–22.

[41] B. Li, S.K. Au, An expectation-maximization algorithm for Bayesian operational modal analysis with multiple (possibly close) modes, Mechanical Systems and Signal Processing. 132 (2019) 490–511.

[42] Z. Zhu, S.K. Au, B. Li, Y.L. Xie, Bayesian operational modal analysis with multiple setups and multiple (possibly close) modes, Mechanical Systems and Signal Processing. 150 (2021) 107261.

[43] Y.C. Ni, F.L. Zhang, Bayesian operational modal analysis of a pedestrian bridge using a field test with multiple setups, International Journal of Structural Stability and Dynamics. 16 (2016) 1550052.

[44] J. Brownjohn, S.K. Au, B. Li, J. Bassitt, Optimised ambient vibration testing of long span bridges, Procedia Engineering. 199 (2017) 38–47.

[45] Y. Xie, B. Li, J. Guo, Bayesian operational modal analysis of a long-span cable-stayed sea-crossing bridge, Journal of Zhejiang University-Science A. 21 (2020) 553–564.

[46] X. Zhou, C.W. Kim, F.L. Zhang, K.C. Chang, Vibration-based Bayesian model updating of an actual steel truss bridge subjected to incremental damage, Engineering Structures. 260 (2022) 114226.

[47] S.K. Au, J.M.W. Brownjohn, B. Li, A. Raby, Understanding and managing identification uncertainty of close modes in operational modal analysis, Mechanical Systems and Signal Processing. 147 (2021) 107018.

[48] S.K. Au, Fast Bayesian ambient modal identification in the frequency domain, Part I: Posterior most probable value, Mechanical Systems and Signal Processing. 26 (2012) 60–75.

[49] G.J. McLachlan, T. Krishnan, The EM algorithm and extensions, John Wiley & Sons, New York, 1998.

[50] T.J. Matarazzo, S.N. Pakzad, STRIDE for structural identification using expectation maximization: iterative output-only method for modal identification, Journal of Engineering Mechanics. 142 (2016) 04015109.

[51] A. Berlinet, C. Roland, Parabolic acceleration of the EM algorithm, Statistics and Computing. 19 (2009) 35–47.

[52] A.F. Berlinet, C. Roland, Acceleration of the EM algorithm: P-EM versus epsilon algorithm, Computational Statistics and Data Analysis. 56 (2012) 4122–4137.

[53] S.K. Au, Fast Bayesian ambient modal identification in the frequency domain, Part II: Posterior uncertainty, Mechanical Systems and Signal Processing. 26 (2012) 76–90.

[54] S.K. Au, Model validity and frequency band selection in operational modal analysis, Mechanical Systems and Signal Processing. 81 (2016) 339–359.

[55] S.F. Gull, Bayesian Inductive inference and maximum entropy, Maximum-Entropy and Bayesian Methods in Science and Engineering. (1988) 53–74.

[56] S.K. Au, J.M.W. Brownjohn, J.E. Mottershead, Quantifying and managing uncertainty in operational modal analysis, Mechanical Systems and Signal Processing. 102 (2018) 139–157.

[57] Y.L. Xie, S.K. Au, B. Li, Asymptotic identification uncertainty of well-separated modes in operational modal analysis with multiple setups, Mechanical Systems and Signal Processing. 152 (2021) 107382.

[58] Y.L., Xie, Z. Zhu, X.R. Wang, B. Li, S.K. Au, 2020, Queen's Park Bridge ambient vibration data with multiple setups, Harvard Dataverse, V1.

[59] Shayle R. Searle, André I. Khuri, Matrix algebra useful for statistics, 2nd ed., WILEY, Hoboken, 2017.