Model Validity and Frequency Band Selection in Operational Modal Analysis

Siu-Kui Au

Institute for Risk and Uncertainty and Center for Engineering Dynamics

University of Liverpool

E-mail: siukuiau@liverpool.ac.uk

Abstract

Experimental modal analysis aims at identifying the modal properties (e.g., natural frequencies, damping ratios, mode shapes) of a structure using vibration measurements. Two basic questions are encountered when operating in the frequency domain: Is there a mode near a particular frequency? If so, how much spectral data near the frequency can be included for modal identification without incurring significant modeling error? For data with high signal-to-noise (s/n) ratios these questions can be addressed using empirical tools such as singular value spectrum. Otherwise they are generally open and can be challenging, e.g., for modes with low s/n ratios or close modes. In this work these questions are addressed using a Bayesian approach. The focus is on operational modal analysis, i.e., with 'output-only' ambient data, where identification uncertainty and modeling error can be significant and their control is most demanding. The approach leads to 'evidence ratios' quantifying the relative plausibility of competing sets of modeling assumptions. The latter involves modeling the 'what-if-not' situation, which is non-trivial but is resolved by systematic consideration of alternative models and using maximum entropy principle. Synthetic and field data are considered to investigate the behavior of evidence ratios and how they should be interpreted in practical applications.

Keywords: Bandwidth, modal identification, maximum entropy, model class selection, operational modal analysis

1 Introduction

Ambient modal identification, conventionally known as 'operational modal analysis' (OMA), aims at identifying the modal properties (primarily, natural frequencies, damping ratios, mode shapes) of a structure using 'output-only' vibration data under ambient conditions [1][2][3]. The input excitation is not measured (often impractical to do so) but assumed to be 'broadband random' so that the statistical characteristics of measured response reflect primarily the properties of vibration modes rather than excitation. High economy and feasibility in data collection is a major advantage. The approach is promising for response and model updating [4]; and more generally health monitoring of civil structures [5][6][7][8]. OMA can be performed in the time or frequency domain, Bayesian or non-Bayesian.

Conventional methods are mostly non-Bayesian, e.g., stochastic subspace identification [9] (time domain) and frequency domain decomposition [10] (frequency domain). Bayesian methods have been formulated in the time and frequency domain, e.g., [11][12]. Application was limited until computational problems were solved in the frequency domain [13][14]. Reported cases are more recent, perhaps because the approach is less intuitive and derivation is more mathematically involved; see review in [15].

Identifying modal properties without input information is theoretically more involved. Results have much higher variability/uncertainty and sensitivity to algorithmic parameters compared to their counterparts identified with free or forced (known input) vibration data. Frequency domain methods make use of spectral quantities in a selected frequency band for identifying the modes within it. In doing so the identification (ID) model only needs to account for the modes dominating the band and so can be significantly simplified. For well-separated modes the band can be selected to cover only one mode. In general the number of close modes rarely exceeds three. ID results are insensitive to activities in other bands because their spectral data (e.g., FFT) do not enter into the calculation process (e.g., likelihood function in Bayesian approach). This is especially attractive for OMA since ambient data contains a variety of activities in different bands which are irrelevant to identifying the mode(s) of interest or difficult to model.

1.1 Basic questions in frequency domain modal ID

Performing modal ID in the frequency domain, among the first few questions is where the potential modes are, or equivalently, whether a mode exists near a particular frequency. Existing means are empirical. Stabilization diagram [16] observes whether the eigenvalues of system matrix consistently appear near a particular frequency as the model order of a time-domain state-space model increases. Another common tool is singular value (SV) spectrum [17], i.e., a plot of the singular values of sample PSD with frequency. Sample PSD matrix is positive semi-definite Hermitian and so its singular values and eigenvalues are the same. The number of lines significantly above the remaining ones indicates the dimension of the space spanned by the contributing measured mode shapes and their variation with frequency is similar to the dynamic amplification factor.

Given that a mode exists near a particular frequency, the next question is to select the band whose spectral data shall be used for modal ID. This is a trade-off between ID precision and modelling error risk [15]. As the band widens more information in data is included for modal ID, hopefully leading to higher ID precision. This however increases the risk that the ID model may not hold in the additional band, creating modeling error (bias) in results. For

well-separated modes with high modal signal-to-noise (s/n) ratios a reasonable band is not difficult to decide based on SV spectrum; ID results are insensitive to the choice. Otherwise, a prudent choice currently relies on experience. Examining ID results with different choice of band is a straight-forward strategy but involves repeated calculations. The variation of results is also subjected to interpretation. For example, even when there is no modeling error, results can still fluctuate as more data is used. Although not directly related, see other issues on modal estimation [18][19][20][21]; and achievable precision limits [22][23].

1.2 Problem statement

This paper aims at developing quantitative means for addressing the above concerns for OMA so that modeling error can be effectively suppressed. The problem is approached in a 'Bayesian model class selection' context [28][32], using the FFT of data for probabilistic inference. Bayesian approach provides the basic principle but questions need to be properly formulated to yield useful conclusions while admitting a legitimate mathematical analysis. In this work, the concerns are addressed via the following two questions:

Question 1. Is the modal ID model valid in a band near a particular frequency?

Question 2. If the answer to Question 1 is positive, is the model valid in a wider band?

These questions are illustrated in Figure 1. Focusing near 0.17Hz, Question 1 asks if there is a mode in the grey band. If the answer is positive, Question 2 follows up with how wide the band can be expanded to incorporate more FFT data to improve modal ID precision. If one includes the FFT data in the yellow band, will it cause significant modeling error, e.g., with regard to single mode? To keep the questions well-posed, the band in Question 1 (grey area) is assumed to surround a potential natural frequency. Although this band is often narrow, it is assumed to be wide enough for modal ID, in the sense that the search algorithm for the most probable value (MPV) of Bayesian ID method converges. Otherwise Questions 1 and 2 are no longer relevant or their answers are trivially negative at least in a practical sense.

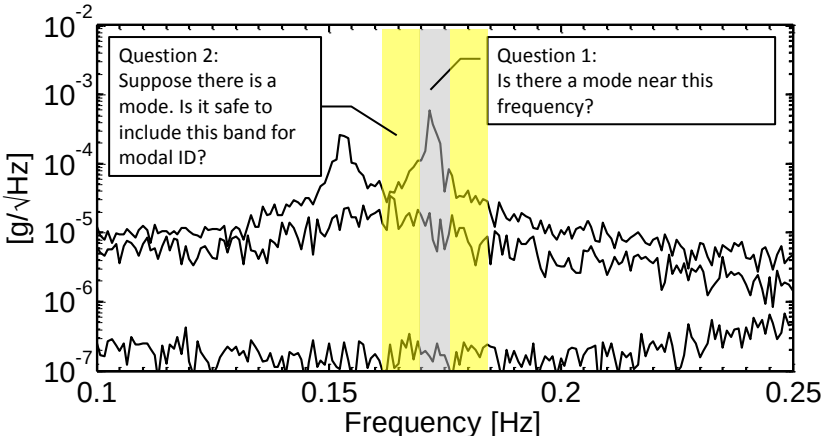

Figure 1. Illustration of Questions 1 and 2, showing the sample SV spectrum (averaged for visualization) of one hour triaxial ambient acceleration data measured on a tall building roof.

2 Organization and outline of results

For the ease of reading, the organization of this paper is outlined with contributions highlighted first. Section 3 presents the basic assumptions about the data and statistical properties of their FFT in a general OMA setting. Bayesian concepts are reviewed in Section 4. Difficulties in assessing model validity are discussed. It is concluded that decision making is inevitably based on comparing competitive models, which can be quantified by 'evidence ratios'. Section 5 addresses Question 1, where it is found that assessing the validity of modal ID model in a band requires statistical modeling of the PSD of data at multiple frequencies when it does not hold, i.e., noise model, which is highly non-trivial. Allowing the PSD to be independently parameterized at different frequencies leads to an inferior model due to the large number of parameters growing with the amount of data. A feasible strategy is to consider noise models with constant statistical properties in the band, which is justified for narrow band. The 'band evidence ratio' assesses the plausibility of modal ID model to that of noise model in the subject band:

$$R = \exp\left[(\hat{L}_j - \hat{L}_0) + \frac{n_j - n_0}{2} \ln N_f \right] \tag{1}$$

where N_f (>>1) is the number of FFT points in the band; $\hat{L}_0$ and $\hat{L}_j$ denote the value of the 'negative log-likelihood function' (NLLF) of modal ID model (with n_0 parameters) and noise model (with n_j parameters), respectively, evaluated at the most probable value (MPV) of parameters. The values of NLLF at MPV for modal ID and noise models are given in Table 1. The MPV of modal ID model can be found by fast Bayesian algorithms [15]. A ratio of $R > 1$ prefers modal ID model over noise model; and vice-versa. Analytical and numerical analysis reveal that, for typically large amount of data, R is either exponentially large or small, a common phenomenon in model class selection with vibration data.

A similar difficulty with PSD modeling recurs in addressing Question 2 in Section 6. To bypass this difficulty, 'point evidence ratio' is developed to assess model validity at a single frequency. This allows the use of maximum entropy principle to determine a representative noise model class, which is found to be the complex Wishart distribution. A closed form expression for the point evidence ratio has been obtained (see also Figure 2):

$$r_k = e^{-\xi_k} \frac{(n\xi_k)^n}{2n!} \left[\sum_{i=1}^{n} \frac{(-\sqrt{n\xi_k})^{n+i}}{[(i-1)!]^2 (n-i)!} K_{i+n-2}(2\sqrt{n\xi_k}) \right]^{-1} \tag{2}$$

where n is the number of measured degrees of freedom; $\xi_k = \mathcal{F}_k^* \hat{\mathbf{E}}_k^{-1} \mathcal{F}_k$ (real positive scalar); $\mathcal{F}_k$ ($n \times 1$) is the FFT of data at the subject frequency (indexed by k); $K_{i+n-2}(\cdot)$

denotes the modified Bessel function of the second kind with order $i+n-2$; and $\hat{\mathbf{E}}_k$ is the theoretical PSD matrix of data at MPV when modal ID model holds in the reference band (see (13)). By virtue of the maximum entropy principle, the point evidence ratio can be practically taken as an 'absolute' measure of validity. For its point-wise nature, however, its values for different frequencies should be viewed as a group so that interdependence of model validity at multiple frequencies can be incorporated in decision making. General guidance is given in Section 7 and illustrative examples are given in Section 8.

3 Stochastic stationary OMA data

Conventionally, the data in OMA is assumed to follow a stochastic stationary process. This is justified on a reasonable time scale over which structural properties and statistical characteristics of ambient excitation are constant. Let $\{\hat{\mathbf{y}}_j\}_{j=0}^{N-1}$ (each $n\times 1$) be the data (acceleration, velocity or displacement) at n DOFs (degrees of freedom) of the structure. Its (scaled) FFT is defined as

$$\mathcal{F}_k = \sqrt{\frac{2\Delta t}{N}} \sum_{j=0}^{N-1} \hat{\mathbf{y}}_j e^{-2\pi \mathbf{i} jk/N} \qquad k = 0,\ldots, N-1 \qquad (3)$$

where Δt (sec) is the sampling interval. Up to the Nyquist frequency ($1/2\Delta t$ Hz), $\mathcal{F}_k$ corresponds to frequency $\mathrm{f}_k = k/N\Delta t$ (Hz). The scaling factor $\sqrt{2\Delta t/N}$ is applied so that $E[\mathcal{F}_k\mathcal{F}_k^*]$ (ensemble expectation, '*' denotes complex conjugate) gives the theoretical one-sided PSD matrix. As a standard result in spectral analysis [24], for long data length, $\mathcal{F}_k$ s at different frequencies are independent. Each has a 'circularly symmetric complex Gaussian distribution', with PDF (probability density function) given by

$$p(\mathcal{F}_k \mid \mathbf{Q}_k) = \pi^{-n} \mid \mathbf{Q}_k \mid^{-1} \exp\left(-\mathcal{F}_k^* \mathbf{Q}_k^{-1} \mathcal{F}_k\right) \qquad (4)$$

where $\mathbf{Q}_k = E[\mathcal{F}_k\mathcal{F}_k^*]$. Equivalently, $\mathrm{Re}\,\mathcal{F}_k$ and $\mathrm{Im}\,\mathcal{F}_k$ are independent zero-mean Gaussian with $E[\mathcal{F}_k\mathcal{F}_k^T \mid \mathbf{Q}_k] = \mathbf{0}$ and $E[\mathcal{F}_k\mathcal{F}_k^* \mid \mathbf{Q}_k] = \mathbf{Q}_k$. In modal ID, data is considered sufficiently long when the number of FFT points in the resonance band is large compared to 1 [22].

Equation (4) is valid as long as data is long and stationary. Whether a mode exists in the band affects only the structure of $\mathbf{Q}_k$ and the parameters it involves. Since $\{\mathcal{F}_k\}$ (collection of FFTs in the band) are independent, their joint PDF is the product of marginal PDFs. For

convenience in analysis, the PDF is written as $p(\{\mathcal{F}_k\}|\{\mathbf{Q}_k\}) = e^{-L}$ where L is the 'negative log likelihood function' (NLLF):

$$L = -\ln p(\{\mathcal{F}_k\}|\{\mathbf{Q}_k\}) = nN_f \ln \pi + \sum_k \ln|\mathbf{Q}_k| + \sum_k \mathcal{F}_k^* \mathbf{Q}_k^{-1} \mathcal{F}_k \quad (5)$$

The sums are over all frequencies in the band, whose number is N_f.

4 Bayesian system identification and model class selection

Bayesian approach [25][26][27][28] provides a rational means for making inference based on real observations and modeling assumptions. Model class selection aims at assessing the validity of a set of modeling assumptions rather than identifying the parameters of a model based on such assumptions (as in system ID). Akaike's Information Criterion (AIC) [29][30][31] is a celebrated example providing a simple measure for comparing regression models with different number of parameters and data fitting power. Reference [32] is one early work in structural dynamics. Subsequent development can be found in, e.g., [33][34]. See also [35] for a tutorial in statistics literature. In this section, basic concepts are reviewed. The difficulties of assessing model validity in an absolute sense are then discussed.

A 'model class' M refers to the collection of all models with the same set of assumptions but different possible parameter values, collected in a vector $\boldsymbol{\theta}$. Without loss of generality, $\boldsymbol{\theta}$ is assumed to be continuous-valued. System ID aims at identifying $\boldsymbol{\theta}$ using data D within M. The 'likelihood function' $p(D|\boldsymbol{\theta},M)$ is the PDF of D for given $\boldsymbol{\theta}$ derived according to the assumptions in M. Given D, the ID results for $\boldsymbol{\theta}$ are encapsulated in the 'posterior' (i.e., given data) PDF $p(\boldsymbol{\theta}|D,M)$. Using Bayes' Theorem it can be expressed in terms of $p(D|\boldsymbol{\theta},M)$ by swapping the roles of $\boldsymbol{\theta}$ and D:

$$p(\boldsymbol{\theta}|D,M) = p(D|M)^{-1} p(D|\boldsymbol{\theta},M) p(\boldsymbol{\theta}|M) \quad (6)$$

As the RHS is viewed as a PDF of $\boldsymbol{\theta}$, $p(D|M)$ is a normalizing constant of no substance in the ID of $\boldsymbol{\theta}$. The term $p(\boldsymbol{\theta}|M)$ is called the 'prior PDF', reflecting the analyst's state of knowledge about $\boldsymbol{\theta}$ in the absence of data. Typically, $p(\boldsymbol{\theta}|D,M)$ is only implicitly known. It need not belong to any standard distribution (e.g., normal, exponential) because $p(D|\boldsymbol{\theta},M)$ can depend on $\boldsymbol{\theta}$ in a complicated ('black-box') manner. Determining the posterior statistics of $\boldsymbol{\theta}$ (e.g., mean, covariance) is one major task in Bayesian system ID.

Assessing the validity of model class M is more challenging than identifying $\boldsymbol{\theta}$ assuming M, in both modeling and computation. From first principle, the validity of M in the presence of D is measured by $P(M \mid D)$. Using Bayes' Theorem,

$$P(M \mid D) = p(D \mid M)P(M) / p(D) \tag{7}$$

Throughout this work upper-case 'P' and lower-case 'p' denote probability and PDF, respectively. The term $p(D \mid M)$ is called 'evidence'. It was immaterial in (6) but is now a deciding factor. It can be further expressed using the theorem of total probability as

$$p(D \mid M) = \int p(D \mid \boldsymbol{\theta}, M)\, p(\boldsymbol{\theta} \mid M) d\boldsymbol{\theta} \tag{8}$$

In (7), $P(M)$ reflects the 'plausibility' of the model class being valid based on prior knowledge. 'Plausibility' is referred instead of 'probability' to distinguish from the classical probability (frequentist's, non-Bayesian) notion of relative frequency of occurrence in repeated experiments, which is absurd here. The term $p(D)$ is an unknown quantity with philosophical significance. Using the theorem of total probability,

$$p(D) = p(D \mid M)P(M) + p(D \mid \overline{M})P(\overline{M}) \tag{9}$$

where $\overline{M}$ denotes the 'complement' (as set) or 'negation' (as logic) of M, meaning 'all model classes other than M'. The terms $p(D \mid M)$, $P(M)$ and $P(\overline{M}) = 1 - P(M)$ can be determined or assigned. However, $p(D \mid \overline{M})$ requires quantifying the plausibility of obtaining D when M does not hold. It can be intractable because possibilities can be vast and unimaginable. One may not be able to tell in 'absolute terms' the plausibility of a model class being valid without specifying/modeling the likelihood of data when it is not valid.

4.1 Comparing model classes

Determining whether one model class is better than another is a more tractable problem. Model class M_1 is more plausible than M_2 if $P(M_1 \mid D) / P(M_2 \mid D) > 1$. This ratio does not involve $p(D)$ because it is cancelled out:

$$\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{p(D \mid M_1)P(M_1) / p(D)}{p(D \mid M_2)P(M_2) / p(D)} = \frac{p(D \mid M_1)P(M_1)}{p(D \mid M_2)P(M_2)} \tag{10}$$

One takes $P(M_1) = P(M_2)$ when there is no prior preference or when the ratio is to reflect data effect only. The ratio is then proportional to the ratio of evidence:

$$\frac{P(M_1 \mid D)}{P(M_2 \mid D)} = \frac{p(D \mid M_1)}{p(D \mid M_2)} \qquad (P(M_1) = P(M_2)) \tag{11}$$

The above discussion reveals that it is difficult (if not impossible) to assess model validity in an absolute sense. The validity of modal ID model shall be assessed in a relative sense against competing model classes, which must be carefully considered for results to be useful.

5 Model validity in potential resonance band (Question 1)

In this section the evidence ratio for assessing the validity of modal ID model in a band around the potential mode(s) of interest is developed. Difficulties in the general context are discussed and legitimate competing model classes in the 'what-if-not' situation are recommended and analyzed. Here, $D = \{\mathcal{F}_k\}$ comprises the FFTs in the band, i.e., the grey band in Figure 1. For each model class, the most probable value (MPV, peak location of posterior PDF) and the value of NLLF at MPV will be presented, as they are involved in the calculation of evidence (Section 5.3). Results are summarized later in Table 1.

5.1 Modal ID model class (M0)

When there is at least one mode in the subject band, the FFT of data is assumed to comprise the contributions from m classically damped modes and measurement noise:

$$\mathcal{F}_k = -\sum_{i=1}^{m} \boldsymbol{\varphi}_i h_{ik} p_{ik} + \boldsymbol{\varepsilon}_k \tag{12}$$

where $\boldsymbol{\varphi}_i$ ($n \times 1$) is the (partial) mode shape confined to measured DOFs; p_{ik} (complex scalar) and $\boldsymbol{\varepsilon}_k$ ($n \times 1$ complex) are FFT of the i-th modal force and measurement noise, respectively; $h_{ik} = (2\pi \mathrm{f}_k)^{-\mathrm{q}}[(\beta_{ik}^2 - 1) + \mathbf{i}(2\zeta_i \beta_{ik})]^{-1}$, $\beta_{ik} = f_i / \mathrm{f}_k$, q = 0, 1 or 2 for acceleration, velocity or displacement data, respectively; f_i (Hz) and ζ_i are the natural frequency and damping ratio of the i-th mode, respectively. In the band, modal forces are assumed to have constant PSD matrix $\mathbf{S}$ ($m \times m$ positive definite Hermitian), whose (i, j)-entry is $S_{ij} = E[p_{ik} p_{jk}^* \,|\, \boldsymbol{\theta}]$. Measurement noise is assumed to be independent of modal forces and i.i.d. (independent and identically distributed) at different measured DOFs with constant PSD S_e in the band, i.e., $E[\boldsymbol{\varepsilon}_k \boldsymbol{\varepsilon}_k^* \,|\, \boldsymbol{\theta}] = S_e \mathbf{I}_n$, where $\mathbf{I}_n$ denotes $n \times n$ identity matrix. Consequently, the PSD of data is given by

$$\mathbf{Q}_k = \mathbf{E}_k(\boldsymbol{\theta}) = \sum_{i=1}^{m}\sum_{j=1}^{m} S_{ij} h_{ik} h_{jk}^* \boldsymbol{\varphi}_i \boldsymbol{\varphi}_j^T + S_e \mathbf{I}_n = \boldsymbol{\Phi} \mathbf{H}_k \boldsymbol{\Phi}^T + S_e \mathbf{I}_n \qquad \text{(M0, modal)} \tag{13}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\varphi}_1, ..., \boldsymbol{\varphi}_m]$ ($n \times m$) and $\mathbf{H}_k(i,j) = S_{ij} h_{ik} h_{jk}^*$. The modal ID model class, where $\mathbf{Q}_k$ is given by (13), will be referred as M0. The set of parameters is $\boldsymbol{\theta} = \left\{ \{f_i\}_{i=1}^m, \{\zeta_i\}_{i=1}^m, \mathbf{S}, S_e, \{\boldsymbol{\varphi}_i\}_{i=1}^m \right\}$. Recognizing that $\mathbf{S}$ is Hermitian and there are m scaling constraints on mode shapes, the number of free parameters in M0 is $n_0 = (m+1)^2 + m(n-1)$. The above model applies to well-separated modes ($m = 1$) and close modes ($m > 1$). The MPV of $\boldsymbol{\theta}$ can be determined efficiently using fast Bayesian algorithms [13][14].

5.2 Competing model classes

As long as data is long and stationary the PDF of FFT $\mathcal{F}_k$ is given by (4), which is fully determined by $\mathbf{Q}_k$. When M0 does not hold, $\mathbf{Q}_k$ is no longer given by (13). Its possible form is an open question and affects model validation conclusions [36]. Competing model classes must not be too restrictive but at the same time be properly constrained to keep the problem well-posed. The least restrictive class allows $\mathbf{Q}_k$ to have 'any distribution', but this renders the likelihood function improper with arbitrarily small PDF and drives the evidence ratio to trivially prefer M0. Too much constraint eliminates possibilities that can be plausible, undermining the representativeness of results.

Competing model classes are motivated by removing modal contributions in M0 and generalizing the form of noise PSD. Taking out the first term in (13), $\mathbf{Q}_k = S_e \mathbf{I}_n$. This is referred as M1. It assumes i.i.d. noise with constant PSD S_e in the band, which is the only parameter to identify. Substituting this $\mathbf{Q}_k$ into (5) and minimizing with respect to (w.r.t.) S_e, the MPV is $\hat{S}_e = n^{-1} \sum_{i=1}^n \tilde{\mathbf{Q}}(i,i)$, where $\tilde{\mathbf{Q}}(i,i)$ denotes the i-th diagonal element of the sample PSD matrix averaged over the band, $\tilde{\mathbf{Q}} = N_f^{-1} \sum_k \mathcal{F}_k \mathcal{F}_k^*$. A slightly more general model class allows the measured DOFs to have possibly different noise PSDs, say, $\{S_{ei}\}_{i=1}^n$. The PSD matrix is given by $\mathbf{Q}_k = \mathrm{diag}\{S_{e1}, ..., S_{en}\}$ (diagonal matrix with entries $\{S_{ei}\}_{i=1}^n$). This is referred as M2. It has n free parameters, $\{S_{ei}\}_{i=1}^n$. Substituting this $\mathbf{Q}_k$ into (5) and minimizing w.r.t. S_{ei}, the MPV is $\hat{S}_{ei} = \tilde{\mathbf{Q}}(i,i)$ ($i = 1, ..., n$). As long as $\mathbf{Q}_k$ is assumed to be constant in the band, the most general form is $n \times n$ positive definite Hermitian matrix, which has n^2 free parameters. This is referred as M3. Substituting $\mathbf{Q}_k = \mathbf{Q}$ and considering perturbation of (5) w.r.t. $\mathbf{Q}$, it can be shown that the MPV is $\hat{\mathbf{Q}} = \tilde{\mathbf{Q}}$.

Further generalizing beyond M3 will allow $\mathbf{Q}_k$ to depend on frequency f_k. This requires modeling the variation of PSD w.r.t. frequency. Possibilities are vast. For specific context, e.g., wind- or microtremor-driven ambient conditions, modeling may proceed in a (physics-based) parametric manner, which is out of the present scope. In a non-parametric manner the most general form allows $\mathbf{Q}_k$ s at different frequencies to be individually parameterized, involving $n^2 N_f$ free parameters. Accounting for continuity/differentiability of PSD w.r.t. frequency still leads to a number growing with N_f. As can be deduced from (15) in the next subsection, for large N_f this renders the model class trivially inferior to those whose number of parameters does not grow with N_f. This observation eliminates many models that are otherwise appealing. For example, parameterizing modal force and noise PSD individually for different frequencies produces a very general model class for modal ID but it will be trivially inferior to M0.

5.3 Bayesian evidence

Evaluating the evidence $p(D|M_i)$ via the integral in (8) is a computational problem. Let $\boldsymbol{\theta}_i$ contains the parameters of M_i. The difficulty depends on the 'identifiability' of $\boldsymbol{\theta}_i$, i.e., the topology of its posterior PDF $p(\boldsymbol{\theta}_i|D,M_i) \propto p(D|\boldsymbol{\theta}_i,M_i)p(\boldsymbol{\theta}_i|M_i)$ [37]. In modal ID, the amount of data is sufficiently large that $p(\boldsymbol{\theta}_i|M_i)$ is slowly varying compared to $p(D|\boldsymbol{\theta}_i,M_i)$. Also, the posterior PDF is 'globally identifiable', having a unique peak at the MPV $\hat{\boldsymbol{\theta}}_i$ that minimizes the NLLF $L_i = -\ln p(D|\boldsymbol{\theta}_i,M_i)$. It can be approximated by a Gaussian PDF with mean $\hat{\boldsymbol{\theta}}_i$ and covariance matrix $\hat{\mathbf{C}}_i$ equal to the Hessian of NLLF at MPV. Noting that $p(D|\boldsymbol{\theta}_i,M_i)$ is concentrated at $\hat{\boldsymbol{\theta}}_i$ and ignoring the variation of $p(\boldsymbol{\theta}_i|M_i)$ near $\hat{\boldsymbol{\theta}}_i$, (8) implies $p(D|M_i) \approx p(\hat{\boldsymbol{\theta}}_i|M_i)\int p(D|\boldsymbol{\theta}_i,M_i)d\boldsymbol{\theta}_i$. The integral can be further approximated by writing $p(D|\boldsymbol{\theta}_i,M_i) = \exp(-L_i)$ and using the second order approximation $L_i \approx \hat{L}_i + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_i)^T \hat{\mathbf{C}}_i^{-1}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_i)/2$. Substituting yields an integrand proportional to Gaussian PDF that can be integrated analytically. Consequently,

$$p(D|M_i) \approx K_i \; p(D|\hat{\boldsymbol{\theta}}_i,M_i) \qquad \text{(globally identifiable)} \qquad (14)$$

where $K_i = (2\pi)^{n_i/2}|\hat{\mathbf{C}}_i|^{1/2}\; p(\hat{\boldsymbol{\theta}}_i|M_i)$ is called the 'Ockham factor' and is found to have a penalizing effect on model complexity in terms of the number of parameters [38][39].

Equation (14) was derived previously in [32] and its large data behavior was investigated. In the present context, for large N_f, $K_i \sim C_i \exp[-n_i(\ln N_f)/2]$ where C_i is $O(1)$ constant that depends on the prior PDF. Substituting into (14), the evidence ratio of two model classes M_i and M_j is asymptotically of the form

$$\frac{p(D|M_i)}{p(D|M_j)} = \frac{C_i}{C_j} \exp\left[(\hat{L}_j - \hat{L}_i) + \frac{n_j - n_i}{2} \ln N_f \right] \qquad \text{(large } N_f\text{)} \qquad (15)$$

Roughly speaking $\hat{L}_i$ reflects the best model fit with data among all possible values of $\boldsymbol{\theta}_i$. It can be reasoned that $\hat{L}_i$ generally grows with N_f. The same is also true for the magnitude of the exponent in (15). The evidence ratio then asymptotically becomes a 'zero-infinity law', i.e., either zero (exponentially small) when the exponent is negatively large, or infinity (exponentially large) when the exponent is positively large. The relative plausibility of model class becomes a 'zero-one law', implying either 'No' (zero) or 'Yes' (one) conclusion with practically 100% confidence. The factor C_i / C_j becomes immaterial and can be omitted in calculations, i.e., prior PDF asymptotically plays no role.

Table 1 summarizes the competing model classes. They can be representative of possibilities other than M0 for narrow band where the PSD of data is likely to be constant. The last column shows the value of NLLF at MPV, obtained by substituting $\mathbf{Q}_k$ at MPV into (5). It can be shown that $\hat{L}_1 \geq \hat{L}_2 \geq \hat{L}_3$. This agrees with the general fact that a lower minimum can be achieved under less constraint. The first inequality $\hat{L}_1 \geq \hat{L}_2$ can be shown by noting that $\hat{S}_e$ is the arithmetic mean of $\{\hat{S}_{ei}\}_{i=1}^n$, which is bounded below by geometric mean. The second inequality $\hat{L}_2 \geq \hat{L}_3$ can be shown using Hadamard's inequality [40], i.e., the determinant of a positive definite matrix is bounded above by the product of its diagonal entries.

For pure noise data the MPV search algorithm for modal ID will not converge and so the answer to Question 1 is negative at least in a practical sense. Conceptually this can also be seen from (15). In this case the modal ID model M0 will not be able to reduce the value of NLLF below that of the pure noise model, i.e., $\hat{L}_0 \approx \hat{L}_j$ (j for noise model). Since M0 typically has a larger number of parameters than a pure noise model, the evidence ratio will be asymptotically small for large N_f and so M0 will be rejected.

Table 1. Summary of competing model classes

Model Class	PSD of data ($\mathbf{Q}_k$)	No. of free Parameters	MPV	NLLF at MPV ($\hat{L}_i$)
M0 (modal ID)	$\boldsymbol{\Phi}\mathbf{H}_k\boldsymbol{\Phi}^T + S_e\mathbf{I}_n$	$(m+1)^2 + m(n-1)$	Found by fast algorithm [14]	No simplified expression
M1 (i.i.d. noise)	$S_e\mathbf{I}_n$	1	$\hat{S}_e = \frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{Q}}(i,i)$	$nN_f[1+\ln\hat{S}_e]$
M2 (ind. noise)	$\mathrm{diag}\{S_{e1},...,S_{en}\}$	n	$\hat{S}_{ei} = \tilde{\mathbf{Q}}(i,i)$	$nN_f\left[1+\frac{1}{n}\sum_{i=1}^{n}\ln\hat{S}_{ei}\right]$
M3 (corr. noise)	Positive definite Hermitian $\mathbf{Q}$	n^2	$\hat{\mathbf{Q}} = \tilde{\mathbf{Q}}$	$nN_f\left[1+\frac{1}{n}\ln\lvert\hat{\mathbf{Q}}\rvert\right]$

Note: $\tilde{\mathbf{Q}} = \sum_k \mathcal{F}_k\mathcal{F}_k^* / N_f$ is the sample PSD averaged over the subject band

6 Frequency band selection (Question 2)

This section addresses the second question in Section 1, i.e., the validity of modal ID model M0 in a questionable band, given that M0 is valid in a 'reference band'. Typically the reference band is a narrow band near the natural frequency (grey area in Figure 1); while the questionable band appears at the tail (yellow area). Recall that the competing model classes (i.e., other than M0) in the last section all assumed constant PSD matrix in the reference band. Otherwise they are trivially inferior or require modeling the variation of PSD with frequency, which is out of the present scope. The same difficulty recurs in this section. While the narrow band assumption was reasonable when addressing Question 1, it is unlikely to be so here because the questionable band can contain frequencies far from the reference band. To resolve this, 'point evidence ratio' for a particular frequency rather than a band is developed. This eliminates the need for modeling the variation of PSD. It also allows the competing model class in the 'what-if-not' situation to be determined more fundamentally using maximum entropy principle. Closed form expression for the point evidence ratio shall be obtained, allowing for convenient implementation. Although the point evidence ratio does not directly address the 'band-wise' question originally posed, it shall be seen later that gathering the answers for different frequencies does lead to a satisfactory answer.

6.1 General context

General equations governing model validity in a questionable band for Question 2 shall be formulated first. The band shall be later limited to a single frequency to derive the final result. Let D_1 and D_2 comprise the FFTs over the reference band and questionable band, respectively. Let B_1 and B_2 denote the propositions that M0 is valid on D_1 and D_2,

respectively. The total data set for inference is $D=\{D_1, D_2\}$. Given that M0 is valid in D_1, the validity of M0 in D_2 is determined by

$$r = \frac{P(B_2 \mid D, B_1)}{P(B_2' \mid D, B_1)} = \frac{p(D \mid B_1, B_2)P(B_2 \mid B_1)}{p(D \mid B_1, B_2')P(B_2' \mid B_1)} \tag{16}$$

where B_2' denotes the competing model class representative of all possibilities other than M0. Taking $P(B_2 \mid B_1) = P(B_2' \mid B_1)$, i.e., the validity of M0 in D_1 provides no information on validity in D_2, one obtains $r = p(D \mid B_1, B_2) / p(D \mid B_1, B_2')$. This can be further simplified by recognizing $D=\{D_1, D_2\}$ and conditioning on D_1:

$$r = \frac{p(D_1, D_2 \mid B_1, B_2)}{p(D_1, D_2 \mid B_1, B_2')} = \frac{p(D_2 \mid D_1, B_1, B_2)p(D_1 \mid B_1, B_2)}{p(D_2 \mid D_1, B_1, B_2')p(D_1 \mid B_1, B_2')} \tag{17}$$

Assuming that the validity of M0 in D_2 has no influence on the distribution of D_1, $p(D_1 \mid B_1, B_2) = p(D_1 \mid B_1, B_2')$. Substituting into (17) gives

$$r = \frac{P(D_2 \mid D_1, B_1, B_2)}{P(D_2 \mid D_1, B_1, B_2')} \tag{18}$$

6.2 Point-wise context

The numerator in (18) can be obtained using the likelihood function of M0, which was presented in Section 5.1. The denominator involves modeling the variation of PSD with frequency when M0 does not hold. This is bypassed by limiting D_2 to contain the FFT at a single frequency only, which leads to the point evidence ratio:

$$r_k = \frac{p(\mathcal{F}_k \mid D_1, B_1, B_2)}{p(\mathcal{F}_k \mid D_1, B_1, B_2')} \tag{19}$$

The subscript k indicates explicit dependence on frequency f_k of the FFT $\mathcal{F}_k$ under question. The next two subsections discuss the determination of the numerator and denominator. The treatment is quite different from Sections 5.1 and 5.2 due to the difference in conditioning information and now point-wise nature of the evidence.

6.3 Modal ID model class

The evidence $p(\mathcal{F}_k \mid D_1, B_1, B_2)$ in (19) can be obtained via the likelihood function of M0 (Section 5.1). Let $\boldsymbol{\theta}$ be the set of parameters associated with M0. Conditioning on $\boldsymbol{\theta}$,

$$p(\mathcal{F}_k \mid D_1, B_1, B_2) = \int p(\mathcal{F}_k \mid \boldsymbol{\theta}, D_1, B_1, B_2)\, p(\boldsymbol{\theta} \mid D_1, B_1, B_2) d\boldsymbol{\theta} \tag{20}$$

Given B_2 (i.e., M0 holds in D_2) and $\boldsymbol{\theta}$, the PDF of $\mathcal{F}_k$ is already determined to be $p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0)$ and the information $\{D_1, B_1\}$ is redundant. Thus

$$p(\mathcal{F}_k \mid \boldsymbol{\theta}, D_1, B_1, B_2) = p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0) \tag{21}$$

On the other hand, given $\{D_1, B_1\}$, i.e., D_1 is available and M0 holds in D_1, $\boldsymbol{\theta}$ is distributed as the posterior PDF $p(\boldsymbol{\theta} \mid D_1, M_0)$ and the validity of M0 in D_2 (i.e., B_2) provides no further information. Thus,

$$p(\boldsymbol{\theta} \mid D_1, B_1, B_2) = p(\boldsymbol{\theta} \mid D_1, B_1) = p(\boldsymbol{\theta} \mid D_1, M_0) \tag{22}$$

Substituting (21) and (22) into (20),

$$p(\mathcal{F}_k \mid D_1, B_1, B_2) = \int p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0)\, p(\boldsymbol{\theta} \mid D_1, M_0) d\boldsymbol{\theta} \tag{23}$$

In the integrand, $p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0)$ is given by (4) with $\mathbf{Q}_k = \mathbf{E}_k(\boldsymbol{\theta})$ in (13):

$$p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0) = \pi^{-n} \mid \mathbf{E}_k(\boldsymbol{\theta}) \mid^{-1} \exp[-\mathcal{F}_k^* \mathbf{E}_k(\boldsymbol{\theta})^{-1} \mathcal{F}_k] \tag{24}$$

As mentioned in Section 5.3, the posterior PDF $p(\boldsymbol{\theta} \mid D_1, M_0)$ can be approximated by a Gaussian PDF centered at the MPV $\hat{\boldsymbol{\theta}}$:

$$p(\boldsymbol{\theta} \mid D_1, M_0) \approx (2\pi)^{-n_0/2} \mid \hat{\mathbf{C}} \mid^{-1/2} \exp\left[-\frac{1}{2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \hat{\mathbf{C}}^{-1} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \tag{25}$$

where n_0 is the number of parameters in $\boldsymbol{\theta}$; and $\hat{\mathbf{C}}$ is the inverse of Hessian of NLLF $L = -\ln p(D_1 \mid \boldsymbol{\theta}, M_0)$ w.r.t. $\boldsymbol{\theta}$ at MPV.

Having obtained the integrands in (23), the next question is to evaluate the integral. In the present case, analytical approximation can be developed by exploiting the slowly varying nature of $p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0)$ (w.r.t. $\boldsymbol{\theta}$) compared to $p(\boldsymbol{\theta} \mid D_1, M_0)$, as the latter is the posterior PDF based on a lot more data than the former. The simplest approximation ignores the variation of $p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0)$, taking it as a constant over the effective support of $p(\boldsymbol{\theta} \mid D_1, M_0)$ that gives the main contribution to the integral. This gives, since $\int p(\boldsymbol{\theta} \mid D_1, M_0) d\boldsymbol{\theta} = 1$,

$$p(\mathcal{F}_k \mid D_1, B_1, B_2) \approx p(\mathcal{F}_k \mid \hat{\boldsymbol{\theta}}, M_0) = \pi^{-n} \mid \hat{\mathbf{E}}_k \mid^{-1} \exp\left(-\mathcal{F}_k^* \hat{\mathbf{E}}_k^{-1} \mathcal{F}_k \right) \quad \text{(zeroth order)} \tag{26}$$

where $\hat{\mathbf{E}}_k$ is given by (13) at MPV. A slightly more sophisticated approximation accounts for the (up to) second order variation of $-\ln p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0)$ near $\hat{\boldsymbol{\theta}}$. This gives (see Section 11)

$$p(\mathcal{F}_k \mid D_1, B_1, B_2) \approx A \; p(\mathcal{F}_k \mid \hat{\boldsymbol{\theta}}, M_0) \qquad \text{(second order)} \qquad (27)$$

$$A = \mid \mathbf{I} + H_1^{-1} H_2 \mid^{-1/2} \exp[\, g_2^T (H_1 + H_2)^{-1} g_2 \,] \qquad (28)$$

$g_2 = -\nabla^T \ln p(\mathcal{F}_k \mid \hat{\boldsymbol{\theta}}, M_0)$; $H_1 = -\nabla^2 \ln p(D_1 \mid \hat{\boldsymbol{\theta}}, M_0)$ and $H_2 = -\nabla^2 \ln p(\mathcal{F}_k \mid \hat{\boldsymbol{\theta}}, M_0)$; $\mathbf{I}$ is $n_0 \times n_0$ identity matrix. This approximation is good when $g_2^T (H_1 + H_2)^{-1} g_2 << 1$ and $H_1^{-1} H_2 << \mathbf{I}$ (in the sense of eigenvalues), for which $A \approx 1$. Note that if $g_2 = \mathbf{0}$ and $H_2 = \mathbf{0}$, then $A = 1$ and (27) reduces to (26), as expected. Since $g_2 = O(1)$, $H_1 = O(N_f)$ and $H_2 = O(1)$, it can be reasoned using (28) that $A = 1 + O(N_f^{-1})$. Equation (26) is therefore asymptotically correct for large N_f.

6.4 Competing model class

Determining $p(\mathcal{F}_k \mid D_1, B_1, B_2')$ in (19) involves modeling when modal ID model does not hold. As in Section 5.2, for long stationary data the PDF of $\mathcal{F}_k$ is given by (4), which depends on $\mathbf{Q}_k$ only. Conditioning on $\mathbf{Q}_k$,

$$p(\mathcal{F}_k \mid D_1, B_1, B_2') = \int p(\mathcal{F}_k \mid \mathbf{Q}_k, D_1, B_1, B_2') \, p(\mathbf{Q}_k \mid D_1, B_1, B_2') d\mathbf{Q}_k \qquad (29)$$

Here, $d\mathbf{Q}_k$ denotes symbolically the differential element in the space of $n \times n$ positive definite Hermitian matrices. Given $\mathbf{Q}_k$, the PDF of $\mathcal{F}_k$ is fixed and other information $\{D_1, B_1, B_2'\}$ is irrelevant. The first term in the integrand of (29) is then given by

$$p(\mathcal{F}_k \mid \mathbf{Q}_k, D_1, B_1, B_2') = \pi^{-n} \mid \mathbf{Q}_k \mid^{-1} \exp\left(-\mathcal{F}_k^* \mathbf{Q}_k^{-1} \mathcal{F}_k\right) \qquad (30)$$

The PDF $p(\mathbf{Q}_k \mid D_1, B_1, B_2')$ in (29) involves modeling $\mathbf{Q}_k$ under the information of $\{D_1, B_1, B_2'\}$. This shall be investigated next.

6.4.1 Maximum entropy distribution

If M0 holds at f_k (i.e., B_2) and $\boldsymbol{\theta}$ is given then $\mathbf{Q}_k$ is simply $\mathbf{E}_k(\boldsymbol{\theta})$ in (13). Under $\{D_1, B_1, B_2'\}$, however, M0 does not hold at f_k and so knowing $\boldsymbol{\theta}$ does not fix $\mathbf{Q}_k$. The

matrix $\mathbf{Q}_k$ need not even be equal to $\mathbf{E}_k(\boldsymbol{\theta})$ for some $\boldsymbol{\theta}$. Since M0 holds in D_1, D_2 can still contain modal contributions from M0, although it can as well contain contributions from other unmodeled dynamics. The FFT of data at f_k in D_2 can be viewed as

$$\mathcal{F}_k = -\sum_{i=1}^{m} h_{ik}\, p_{ik}\boldsymbol{\varphi}_i + \text{meas. noise} + \text{unmodeled dynamics} \tag{31}$$

Correspondingly, its PSD can be viewed as

$$\mathbf{Q}_k = \sum_{i=1}^{m}\sum_{j=1}^{m} \underbrace{S_{ijk}}_{\substack{\text{Unknown cross PSD} \\ \text{of } p_{ik} \,\&\, p_{jk}}} \underbrace{h_{ik}h_{jk}^{*}\boldsymbol{\varphi}_i\boldsymbol{\varphi}_j^{T}}_{\substack{\text{depends on } \boldsymbol{\theta}, \\ \text{relatively small} \\ \text{uncertainty}}} + \begin{matrix}\text{Unknown PSD} \\ \text{of meas. noise}\end{matrix} + \begin{matrix}\text{Unknown PSD of} \\ \text{unmodeled dynamics}\end{matrix} \tag{32}$$

The cross PSD of modal forces, S_{ijk}, can depend on f_k, though in an unknown manner that can be difficult to model. Balancing the generality of competing model class and the well-posedness of problem, the influence of D_1 on $\mathbf{Q}_k$ is modeled through an expectation:

$$E[\mathbf{Q}_k \mid D_1, B_1, B_2'] = \hat{\mathbf{E}}_k \tag{33}$$

where $\hat{\mathbf{E}}_k = \mathbf{E}_k(\hat{\boldsymbol{\theta}})$ is defined for convenience. This corresponds to 1) ignoring the posterior uncertainty of $\boldsymbol{\theta}$ compared to other sources; 2) assuming $E[S_{ijk} \mid D_1, B_1, B_2'] = \hat{S}_{ij}$ (MPV of modal force PSD in D_1) and 3) assuming the expectation of PSD due to measurement noise and unmodeled dynamics to be equal to $\hat{S}_e\mathbf{I}_n$ (MPV of noise PSD in D_1).

Under the expectation constraint (33), the distribution of $\mathbf{Q}_k$ is determined as the one that maximizes its information entropy. That is, $p(\mathbf{Q}_k \mid D_1, B_1, B_2')$ is the stationary point of the following functional of PDF p:

$$J(p, \chi, \boldsymbol{\Lambda}) = -\int p(\mathbf{Q}) \ln p(\mathbf{Q}) d\mathbf{Q} + \chi\left[1 - \int p(\mathbf{Q}) d\mathbf{Q}\right] + tr\left[\boldsymbol{\Lambda}\left(\hat{\mathbf{E}}_k - \int \mathbf{Q}\; p(\mathbf{Q}) d\mathbf{Q}\right)\right] \tag{34}$$

where $tr(\cdot)$ denotes the trace of argument matrix (sum of diagonal elements); χ (scalar) and $\boldsymbol{\Lambda}$ ($n \times n$) are Lagrange multipliers that enforce normalization of PDF and expectation constraint, respectively. This functional optimization problem has been investigated in details in [41][42], where the objective was to obtain the distribution of wireless data

channels under various forms of information; see [43] for a similar mathematical problem in uncertainty modeling in structural dynamics. It can be shown that $\mathbf{Q}_k$ has a complex Wishart distribution with n DOFs and covariance matrix $\hat{\mathbf{E}}_k / n$, i.e.,

$$p(\mathbf{Q}_k \mid D_1, B_1, B_2') = \frac{\exp\{-tr[(\hat{\mathbf{E}}_k / n)^{-1}\mathbf{Q}_k]\}}{|\hat{\mathbf{E}}_k / n|^n \, \pi^{n(n-1)/2} \prod_{i=1}^{n}(i-1)!} \tag{35}$$

The proof follows standard arguments in calculus of variation; see Section 12.

6.5 Bayesian evidence

The point evidence of the modal ID model class M0 has been derived in (26). Computing the point evidence $p(\mathcal{F}_k \mid D_1, B_1, B_2')$ in (29) for the competing model class involves evaluating the integral where the two terms in the integrand are given by (30) and (35). The integral appears intractable but analytical expression can be obtained using the results in [41] (details omitted):

$$p(\mathcal{F}_k \mid D_1, B_1, B_2') = \frac{2n!}{\pi^n (n\xi_k)^n \, |\hat{\mathbf{E}}_k|} \sum_{i=1}^{n} \frac{(-\sqrt{n\xi_k})^{n+i}}{[(i-1)!]^2 (n-i)!} K_{i+n-2}(2\sqrt{n\xi_k}) \tag{36}$$

where $K_{i+n-2}(\cdot)$ denotes the modified Bessel function of the second kind with order $i+n-2$; and $\xi_k = \mathcal{F}_k^* \hat{\mathbf{E}}_k^{-1} \mathcal{F}_k$ is a real positive scalar, since $\hat{\mathbf{E}}_k^{-1}$ is positive definite Hermitian. Substituting (26) and (36) into (19), the point evidence ratio at frequency f_k reads

$$r_k(\xi_k) = \frac{p(\mathcal{F}_k \mid D_1, B_1, B_2)}{p(\mathcal{F}_k \mid D_1, B_1, B_2')} = e^{-\xi_k} \frac{(n\xi_k)^n}{2n!} \left[\sum_{i=1}^{n} \frac{(-\sqrt{n\xi_k})^{n+i}}{[(i-1)!]^2 (n-i)!} K_{i+n-2}(2\sqrt{n\xi_k}) \right]^{-1} \tag{37}$$

Deriving (36) in the general case is quite technical, requiring special techniques. The case for $n=1$ is presented in Section 13 to offer some insights.

7 Interpretation of point evidence ratio

The point evidence ratio r_k in (37) is a function of $\xi_k = \mathcal{F}_k^* \hat{\mathbf{E}}_k^{-1} \mathcal{F}_k$ only and the number of measured DOFs n. Figure 2(a) shows its numerator $p(\mathcal{F}_k \mid D_1, B_1, B_2)$ and denominator $p(\mathcal{F}_k \mid D_1, B_1, B_2')$ w.r.t. ξ_k. It is seen that $p(\mathcal{F}_k \mid D_1, B_1, B_2')$ has a wider spread and heavier tail than $p(\mathcal{F}_k \mid D_1, B_1, B_2)$, the extend of which increases with n. Figure 2(b) shows r_k. A value above 1 indicates that M0 is more likely to hold at f_k than otherwise. The higher the

value, the stronger the evidence. There is a bounded interval of ξ_k where $r_k > 1$, whose extent increases with n. A value of ξ_k too low or too high will reject M0. The ratio r_k has a bounded maximum (which increases with n) but its minimum can approach zero. For a given n it is thus possible to reject M0 with 100% confidence. It is not possible to accept with 100% confidence, however, especially when n is small. For example, the maximum value of r_k for $n = 2$ is about 3 at $\xi_k = 2.46$ (3 digits). Ignoring other possibilities, the 'most optimistic' plausibility of M0 being valid at f_k is about 3/(3+1)=0.75 (when ξ_k happens to be 2.46). This feature agrees with the intuition that it requires an infinite amount of information to confirm the validity of a model, but only a finite amount to reject it.

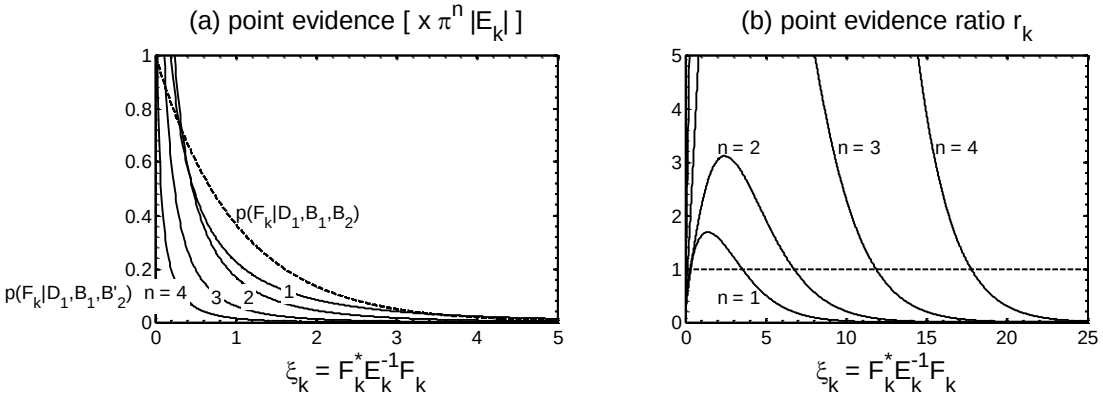

Figure 2. (a) $p(\mathcal{F}_k \mid D_1, B_1, B_2')$ (dashed line) and $p(\mathcal{F}_k \mid D_1, B_1, B_2)$ (solid line); (b) Point evidence ratio $r_k = p(\mathcal{F}_k \mid D_1, B_1, B_2) / p(\mathcal{F}_k \mid D_1, B_1, B_2')$

Strictly speaking, r_k in (37) is not applicable when $\mathcal{F}_k$ belongs to D_1 because the derivation assumed that $\mathcal{F}_k$ does not come from D_1. To use the formula one should exclude $\mathcal{F}_k$ from D_1 when calculating r_k. Effectively, $\hat{\mathbf{E}}_k = \mathbf{E}_k(\hat{\boldsymbol{\theta}})$ should be evaluated with the MPV $\hat{\boldsymbol{\theta}}$ calculated based on a slightly reduced data set, i.e., D_1 with $\mathcal{F}_k$ excluded. The resulting value of MPV, however, is similar to that based on D_1 because their data sets only differ by one FFT point, negligible compared to the total N_f points in D_1. For this reason, practically r_k can be used for frequencies not in D_1 as well as those in D_1, i.e., all frequencies.

Potential fallacy is worth-noting. Ideally, Question 2 in Section 1 should be answered using the band evidence ratio $p(D_2 \mid D_1, B_1, B_2) / p(D_2 \mid D_1, B_1, B_2')$ where $D_2 = \{\mathcal{F}_k\}$ contains the FFTs in the questionable band. This band-wise ratio is not available; only point-wise ratio has been developed. Note that

$$p(D_2 \mid D_1, B_1, B_2) = p(\{\mathcal{F}_k\} \mid D_1, B_1, B_2) \neq \prod_k p(\mathcal{F}_k \mid D_1, B_1, B_2) \qquad (38)$$

because $\mathcal{F}_k$ s are only independent when conditional on $\boldsymbol{\theta}$. The product of point evidence ratios r_k s does not give the evidence ratio for the band either:

$$\prod_k r_k = \prod_k \frac{P(\mathcal{F}_k \mid D_1, B_1, B_2)}{P(\mathcal{F}_k \mid D_1, B_1, \overline{B}_2)} \neq \frac{P(\{\mathcal{F}_k\} \mid D_1, B_1, B_2)}{P(\{\mathcal{F}_k\} \mid D_1, B_1, B_2')} = \frac{P(D_2 \mid D_1, B_1, B_2)}{P(D_2 \mid D_1, B_1, B_2')} \tag{39}$$

Just evaluating r_k s for different frequencies does not allow one to directly assess the validity of M0 in D_2. However, they can be viewed collectively to detect bands of significant modeling error risk. A band dominated by high r_k s (e.g., >100) suggests a valid band; a band dominated by low r_k s (e.g., <0.01) suggests an invalid band. A band with a mix of high and low r_k s still suggests an invalid band when taking continuity of spectrum with frequency into account. More discussion on interpreting the point evidence ratio shall be given through examples in the next section.

Increasing data length has little systematic effect on the point evidence ratio because it only depends on the FFT at the subject frequency through the term $\xi_k = \mathcal{F}_k^* \hat{\mathbf{E}}_k^{-1} \mathcal{F}_k$. The plot of the ratio with frequency, however, will contain more information for making decision because there will be more FFT points in the band under investigation. The number increases linearly because the frequency resolution is 1/T, where T (sec) is the data duration.

8 Illustrative examples

In this section synthetic and field data examples are presented to illustrate the behavior of evidence ratios and how they can be used for decision making. In all cases, the posterior statistics (posterior MPV and variance) of modal parameters are calculated using fast Bayesian algorithms [15].

8.1 Synthetic data

Consider synthetic data $\{\hat{\mathbf{y}}_j\}$ comprising two modes generated at 100Hz (Δt =0.01sec) for 1200sec according to $\hat{\mathbf{y}}_j = \boldsymbol{\varphi}_1 \ddot{\eta}_1(t_j) + \boldsymbol{\varphi}_2 \ddot{\eta}_2(t_j) + \boldsymbol{\varepsilon}(t_j)$ where $t_j = j\Delta t$ ($j = 0,1,2,\ldots$); $\ddot{\eta}_i(t)$ ($i = 1,2$) is modal acceleration satisfying $\ddot{\eta}_i(t) + 2\zeta_i\omega_i\dot{\eta}_i(t) + \omega_i^2\eta_i(t) = p_i(t)$ with $\omega_i = 2\pi f_i$ (rad/sec); $f_1 = 1$Hz and $f_2 = 1.5$Hz (natural frequencies); $\zeta_1 = \zeta_2 = 1\%$ (damping ratios); $\boldsymbol{\varphi}_1 = \begin{bmatrix}1 & 2 & 2\end{bmatrix}^T/3$ and $\boldsymbol{\varphi}_2 = \begin{bmatrix}-2 & 1 & 2\end{bmatrix}^T/3$ (mode shapes); and $p_i(t)$ s are modal

excitations, stationary Gaussian with PSDs $S_1 = S_2 = 1(\mu g)^2/\text{Hz}$; $\boldsymbol{\varepsilon}(t)$ (3×1) is measurement noise, i.i.d. Gaussian white with PSD $S_e = 1(\mu g)^2/\text{Hz}$. Figure 3 shows the root PSD and root SV spectrum of data, averaged for visualization. The two modes are well-separated and they can be identified separately based on M0 with a single mode (m=1) and FFT data on a band around the peak. The question is how the contribution from the other mode affects the validity of M0; and how wide the band could be expanded for modal ID without incurring significant modeling error with regard to single mode assumption.

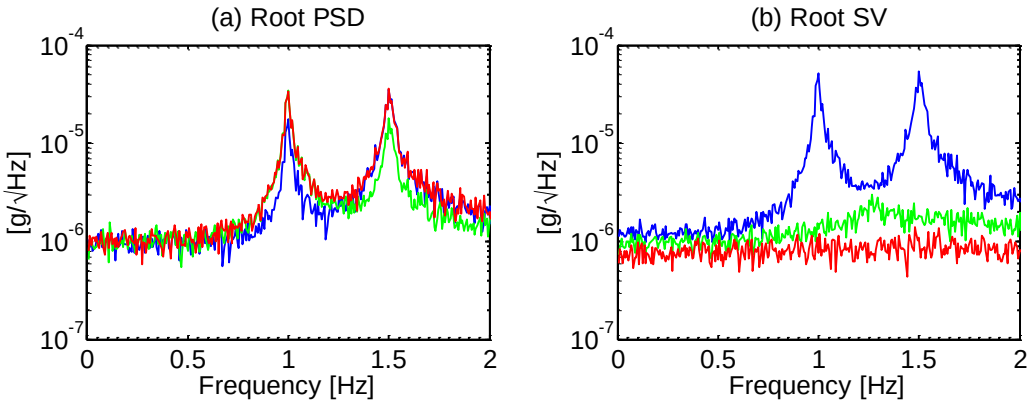

Figure 3. (a) Root PSD and (b) SV spectrum of synthetic data, well-separated modes. In (a), blue, green, red = measured DOF 1, 2, 3. In (b), blue, green, red = largest, second largest and smallest eigenvalue.

Band evidence ratio

Figure 4 shows a series of plots that investigate the use of evidence ratios to address the two questions above. Part (a) shows the root SV spectra zoomed around the spectral peaks. The shaded region indicates the reference band, i.e., D_1 (Section 6), taken to be [0.97,1.03], where it looks reasonable that M0 with a single mode is valid. The choice of the reference band does not affect qualitatively the results presented here. For this band, the log band evidence ratios of M0 to other competing models in Section 5.2 are shown at the left and right boundary of the shaded region in Part (b). The band evidence ratio is calculated using (1). The values of log evidence ratios for M0/M1 (dot), M0/M2 (circle) and M0/M3 (cross) are 750, 718 and 26, respectively. Addressing the first question, these values indicate that M0 is more plausible than all competing models M1, M2 and M3, with a plausibility of practically 1; although M3 is the most competing one among the non-modal models.

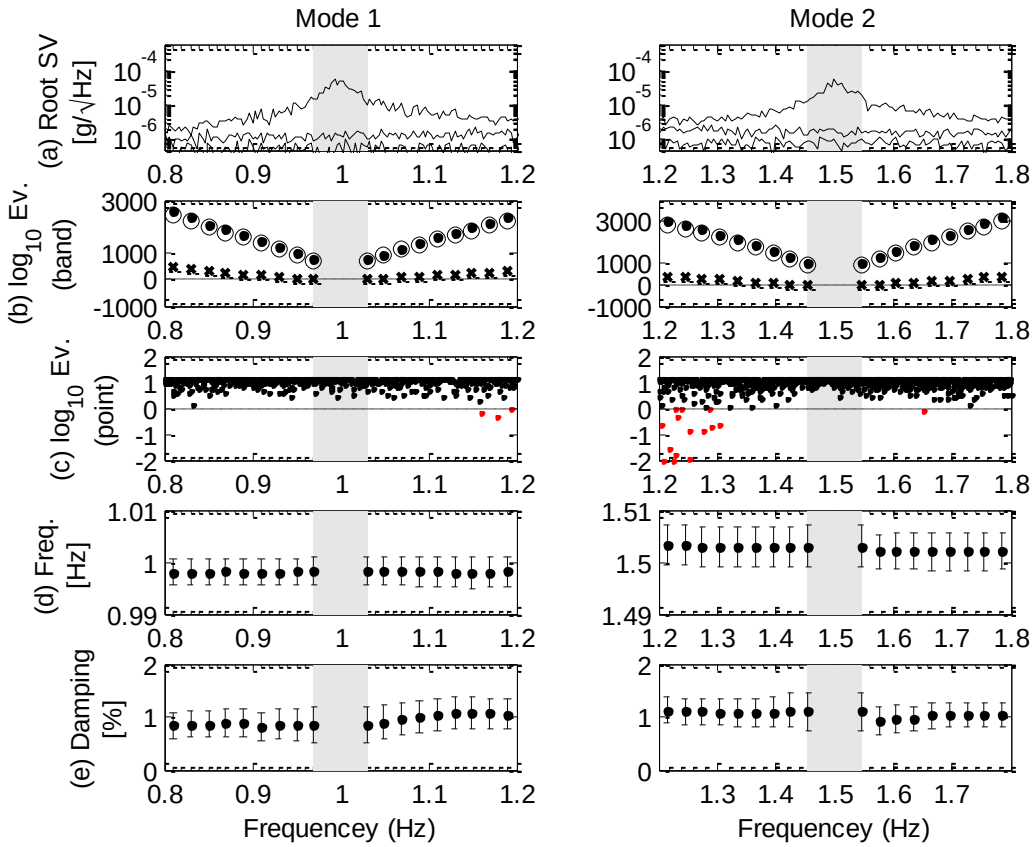

Figure 4. Results for synthetic data with 3 measured DOFs, well-separated modes. In (b), dot, circle, cross = log evidence ratio (band) of M0 to M1, M2 and M3, respectively. In (c), point evidence ratios are plotted at each frequency, assuming that M0 holds in the shaded reference band. In (b) and (c), negative values are colored red. In (c), values smaller than -2 are plotted at -2. (d) and (e) show the natural frequency and damping ratio identified using different bands; dot = MPV, error bar = ± 2 posterior standard deviations.

Suppose now the band [0.97,1.03] is expanded and the validity of M0 is re-investigated for the new band. If one keeps the upper frequency to be the same but reduces the lower frequency to **0.95**Hz, the new band is [**0.95**,1.03]Hz. Using the FFTs in this band the modal properties are identified (i.e., posterior MPV and covariance matrix calculated), from which the band evidence ratio can be calculated. The evidence ratios of M0 to M1-M3 are plotted at **0.95**Hz in Part (b). Similarly, if one keeps the lower frequency to be the same but increases the upper frequency to be **1.05**Hz, the new band is [0.97,**1.05**]Hz and its evidence ratio is plotted at **1.05**Hz. The two figures of Part (b) indicate that as the band widens on either side the evidence ratio of M0 increases over all competing models.

The band evidence ratios in Figure 4(b) was originally intended for narrow band, as M1-M3 all assumed constant spectral properties. They are shown here for wider bands to illustrate issues with their behavior. As the band widens, M1-M3 become less plausible compared to

those model classes that allow the PSD to vary with frequency. The latter are not considered here and hence their effect is not reflected in the results. One implication is that when the band is wide the band evidence ratios become less indicative of the validity of M0 compared to 'all other possibilities'. In absolute terms the validity of M0 need not increase with bandwidth.

In Figure 4(b), the band evidence ratios M0/M1 and M0/M2 look very similar but they are in fact different. They are similar because the data was generated with noise assumption consistent with M1 and M2, giving similar values of NLLF at MPV. Their difference is mainly attributed to the difference in the number of parameters, whose effect is relatively small in this example (M2 has 3 parameters and M1 has 1 parameter).

Point evidence ratio

Figure 4(c) shows the log point evidence ratio in (2). It gives the plausibility of M0 at a single frequency, assuming that M0 holds in the shaded reference band. Negative values are shown in red. The values are mostly above zero, consistent with intuition. For both modes there are some negative values, which correctly indicate the lowering validity of M0 near the natural frequency of the other (unmodeled) mode. This is in contrast with the increasing trend of the band evidence ratios in Figure 4(b).

Identified modal properties

Figure 4(d) and (e) show the identified natural frequencies and damping ratios as the band for modal ID expands on either side. The results are presented with the same convention as Figure 4(b), i.e., the value plotted at 0.95Hz indicates the result when the band is [0.95,1.03]Hz. The dot shows the MPV and the error bar shows ± 2 posterior standard deviations. The ID results are relatively insensitive to the band, which generally agrees with the validity of M0. In practice Figure 4(d) and (e) are usually not produced as they involve repeated applications of modal ID algorithm, which is unnecessary when the band is reasonable. These figures are presented to counter-check the behavior of the evidence ratios with the identified modal properties.

8.1.1 Effect of measured DOFs

The ability to assess model validity depends on the amount of available information. To investigate the effect of measured DOFs, Figure 5 shows the results analogous to Figure 4 but now only DOFs 1 and 2 are measured. In Figure 6 only DOF 1 is measured. The values in

Figure 5(b) are still above zero, advocating M0 over M1-M3 with a plausibility of practically 1, although their values are less than those in Figure 4(b). The values in Figure 6(b) are further reduced. For the log point evidence ratios in Figure 5(c), although most values are still above zero, compared to Figure 4(c) there are more values below zero. Figure 6(c) sees even more. This results directly from the reduction in the amount of information for making inference. Further results (omitted here) with a larger number of measured DOFs show a similar picture in Figure 5(b) for the band ratios with proportionally larger values and a similar picture in Figure 5(c) for the point ratios. In Figure 6(b) the ratios of M0 to M1-M3 are all the same because for one measured DOF M1-M3 are identical.

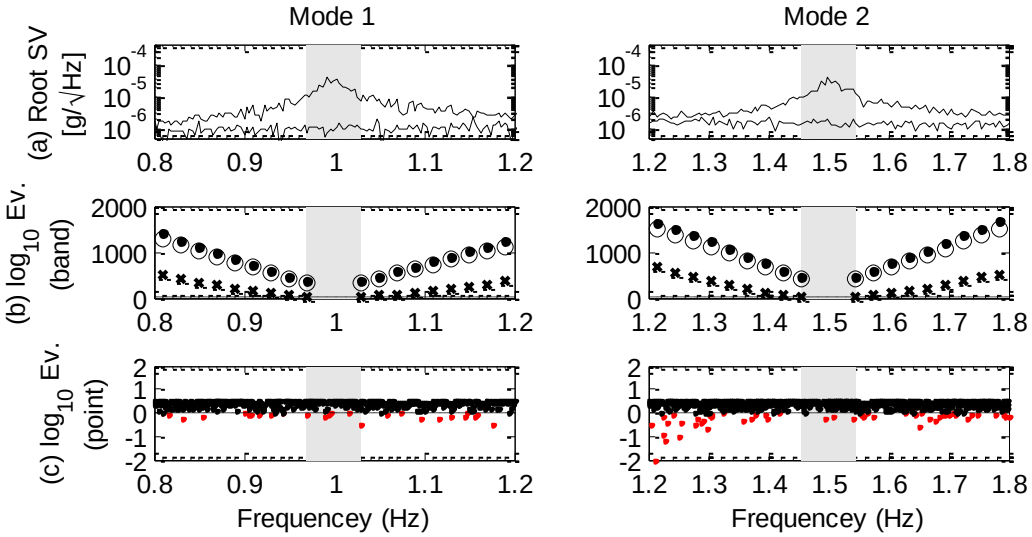

Figure 5. Results for synthetic data when only DOFs 1 and 2 are measured, well-separated modes. Same convention as Figure 4

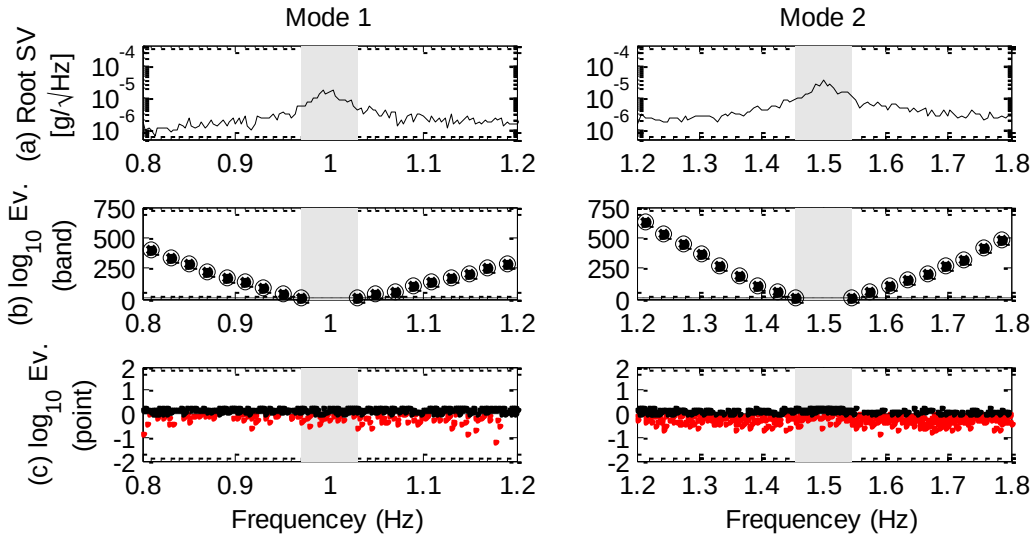

Figure 6. Results for synthetic data when only DOF 1 is measured, well-separated modes. Same convention as Figure 4

8.1.2 Effect of unmodeled mode

The behavior of evidence ratios in detecting model violation due to unmodeled modes is next investigated. Keeping everything else the same, synthetic data is now generated with a second mode frequency of 1.1Hz. The root PSD and root SV spectrum are shown in Figure 7. The two modes are now visually much closer. Although there is still a band around each peak where M0 with a single mode can be considered valid, the band is narrower than before. The question here is how the interaction of the two modes affects the validity of M0 near the natural frequency; and how wide the band can be expanded.

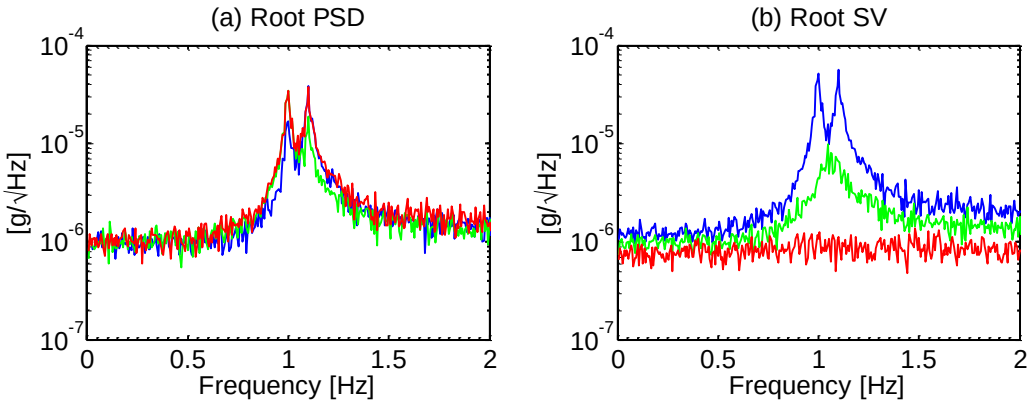

Figure 7. (a) Root PSD and (b) SV spectrum of synthetic data, close modes. Same convention as Figure 3.

Figure 8 shows results analogous to Figure 4. In (b), the band ratios tend to reduce as the band extends towards the other mode. This agrees with intuition because the FFT data in the expanding band is 'contaminated' to an increasing extent by unmodeled contribution from the other mode. Nevertheless, log M0/M1 and log M0/M2 are still above zero, even when the band is so wide that covers the spectral peak of the other mode. This indicates that M0/M1 and M0/M2 are insensitive to the presence of unmodeled modes. On the other hand, log M0/M3 is below zero even for the narrowest band taken (grey area), suggesting that it may be too sensitive. For mode 1, when the upper frequency of the band is 1.11, 1.13 or 1.15 Hz, the MPV search algorithm converges to a value outside the band. Such MPV is clearly unreasonable. A similar issue occurs with mode 2 when the lower frequency of the band is 0.96 or 0.98Hz. These results are not plotted in (b).

Figure 8(c) shows the log point evidence ratios. For both modes the negative values correctly detect the existence of unmodeled contribution from the other mode. Even for bands where M0 is clearly invalid, the point evidence ratio can still be above zero. This stems from the stochastic nature of the point-wise definition. The point evidence ratios for different frequencies should be viewed together rather than individually. Figure 8(d) and (e) show the identified natural frequencies and damping ratios. They are generally insensitive to the band except when it extends significantly into the resonance band of the other mode.

Similar to Figure 8(b), results are not plotted for mode 1 when the upper frequency is 1.11, 1.13 or 1.15Hz; and for mode 2 when the upper frequency is 0.96 or 0.98Hz.

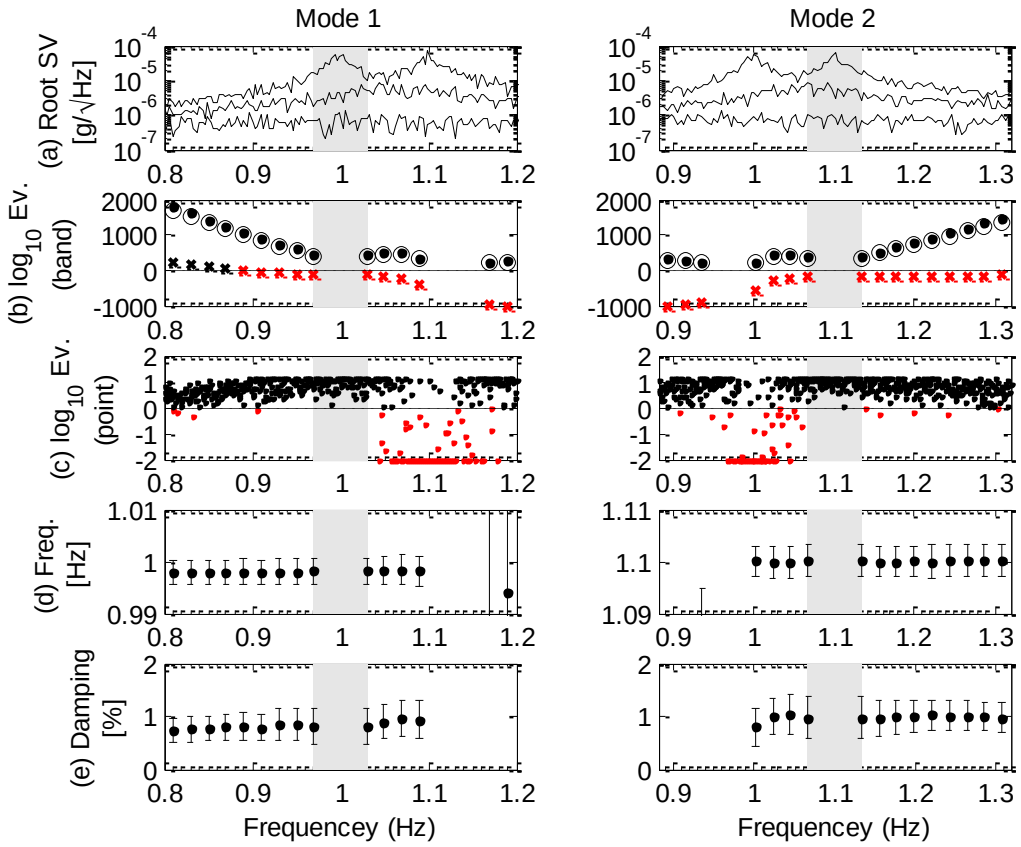

Figure 8. Results for synthetic data with 3 DOFs, close modes. Same convention as Figure 4

8.2 Field data

Consider a set of field data with three DOFs (x, y, z) measured for one hour at 50Hz from a triaxial servo-accelerometer placed at one corner on the roof of a tall building (50m x 50m x 300m) under ambient condition. The horizontal DOFs (x and y) align with the two sides of the building plan. The identified modal properties and modal forces of the building under normal and strong wind conditions were studied in [45]. Figure 9 shows the root PSD and root SV spectrum. The peaks in (b) indicate six modes below 1Hz. The mode numbers are marked next to the peaks. Modes 1 and 4 are translational along the x direction; 2 and 5 are translational along the y direction; 3 and 6 are rotational. Modes 1 and 2 appear close but zooming in their band indicates a reasonable separation. Modes 4 and 5 are well-separated.

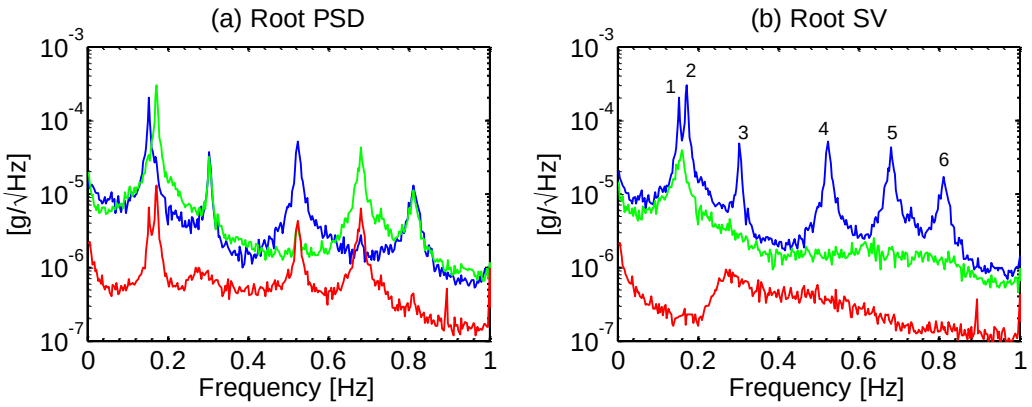

Figure 9. (a) Root PSD and (b) SV spectrum of tall building data. In (a), blue, green, red = x, y, z. In (b), blue, green, red = largest, second largest and smallest eigenvalue.

8.2.1 Modes 1 and 2 (close)

Figure 10 shows the results for modes 1 and 2, with the same convention in Figure 4. The SV spectra in (a) indicate some separation between the two modes, although their dominating band (shaded, hand-picked) can be narrow. In (b), log M0/M3 (cross) is always below zero, rejecting M0. Log M0/M1 (dot) is always above zero. Log M0/M2 (circle) is negative as the band extends towards the other mode. Similar to Section 8.1.2, M0/M1 appears to be insensitive while M0/M3 is too sensitive. M0/M2 lies in between, but its conclusion is still a zero-one law. For mode 1, as the upper frequency increases from 0.157Hz to 0.181Hz, log M0/M2 decreases from -5 to -248. For mode 2, as the lower frequency reduces from 0.155Hz to 0.138Hz, log M0/M2 decreases from -2 to -277. These values all reject M0 with a plausibility of practically 1, although intuition may expect a graduate change depending on how wide the band expands.

Similar to Figure 8(c) in Section 8.1.2, the point evidence ratio in Figure 10(c) correctly detects the region where unmodeled contribution from the other mode is significant. It indicates that for mode 1 the lower frequency may be reduced to, say, 0.14Hz, without incurring much modeling error. This is reinforced by the fact that the identified natural frequency and damping ratio in (d) and (e) are insensitive to the choice of band as the lower frequency is reduced. Reducing the lower frequency has a marginal effect on improving modal ID precision, as evidenced by the similar size of error bars. Figure 10(c) also advises for mode 1 not to increase the upper frequency beyond 0.16Hz (where the red dots start appearing). Doing so may significantly bias the results. As seen in (d), for mode 1 when the upper frequency is 0.172 to 0.181Hz (rightmost four values) the MPV search algorithm converges to the natural frequency of the other mode (value is out of scale). The damping ratios in (e) are significantly biased low as they reflect rather the damping of the unmodeled mode. For mode 2, it is not advised to reduce the lower frequency below 0.163Hz. The

upper frequency may be increased to (say) 0.185Hz, although the improvement in ID precision is only marginal, as evidence in (d) and (e).

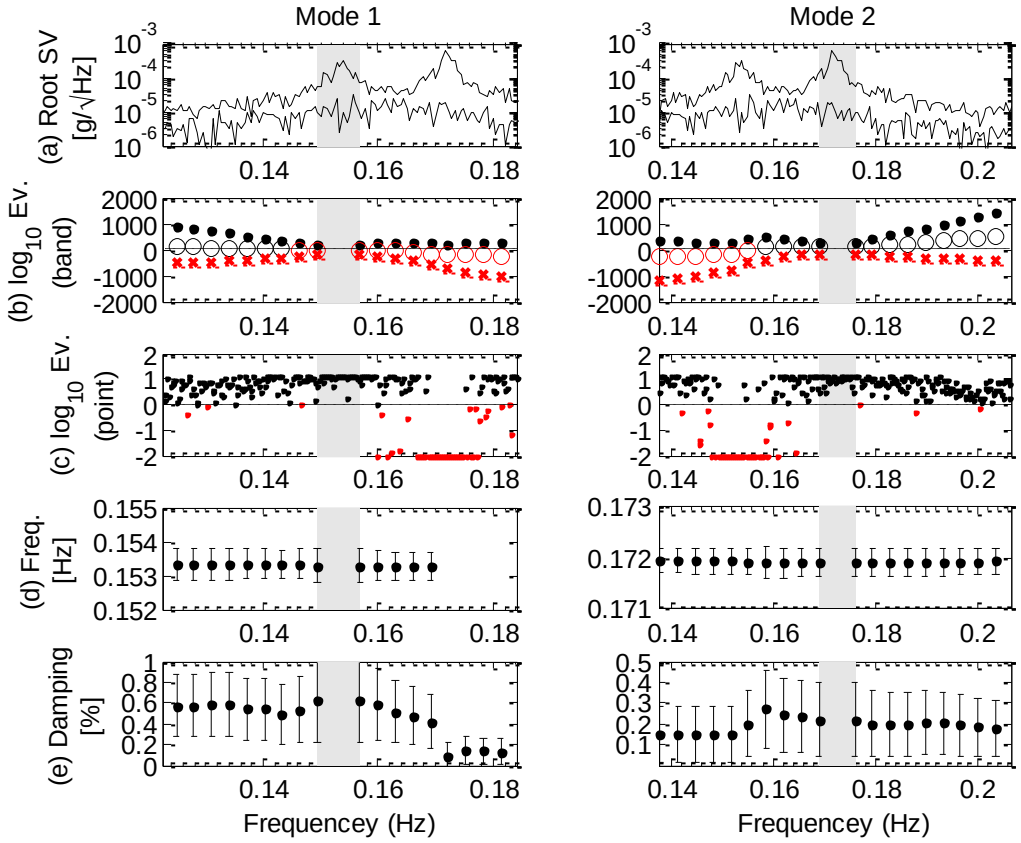

Figure 10. Results for field data, close modes (1 and 2). Same convention as Figure 4

The example presented here corresponds to the scenario relevant and demanding in practice, i.e., where one desires to identify the modes separately (because of simplicity and speed in single mode algorithms) but is not sure how wide the band should be. While the proposed method is applicable to very close modes, they demand less decision making than moderately close modes (as in this example) because the answer is apparent – based on the SV spectrum they should be identified by a multi-mode algorithm.

8.2.2 Mode 4 and 5 (well-separated)

Figure 11 shows the results for mode 4 and 5, which appeared as well-separated in Figure 9(b). In Figure 11(b), log M0/M1 and log M0/M2 are always positive, advocating M0 over M1 and M2 for relatively wide frequency bands. Log M0/M3 detects problems for the narrowest band but becomes positive as the band expands. The log point evidence ratios in Figure 11(c) are mostly above zero for mode 4 and 5 even when the band extends towards one another. This reflects the non-interactive nature of the two separated modes. On the other hand, many values are below zero, for mode 4 when the lower frequency of the band is reduced; and for mode 5 when the upper frequency is increased. While problems with

these bands may not be apparent from the SV spectra in Figure 11(a), the identified modal properties in Figure 11(d) and (e) do confirm their existence. For mode 4 the identified frequencies and damping ratios are significantly biased high as the lower frequency is reduced. For mode 5, as the upper frequency increases the identified modal properties are not as severely affected.

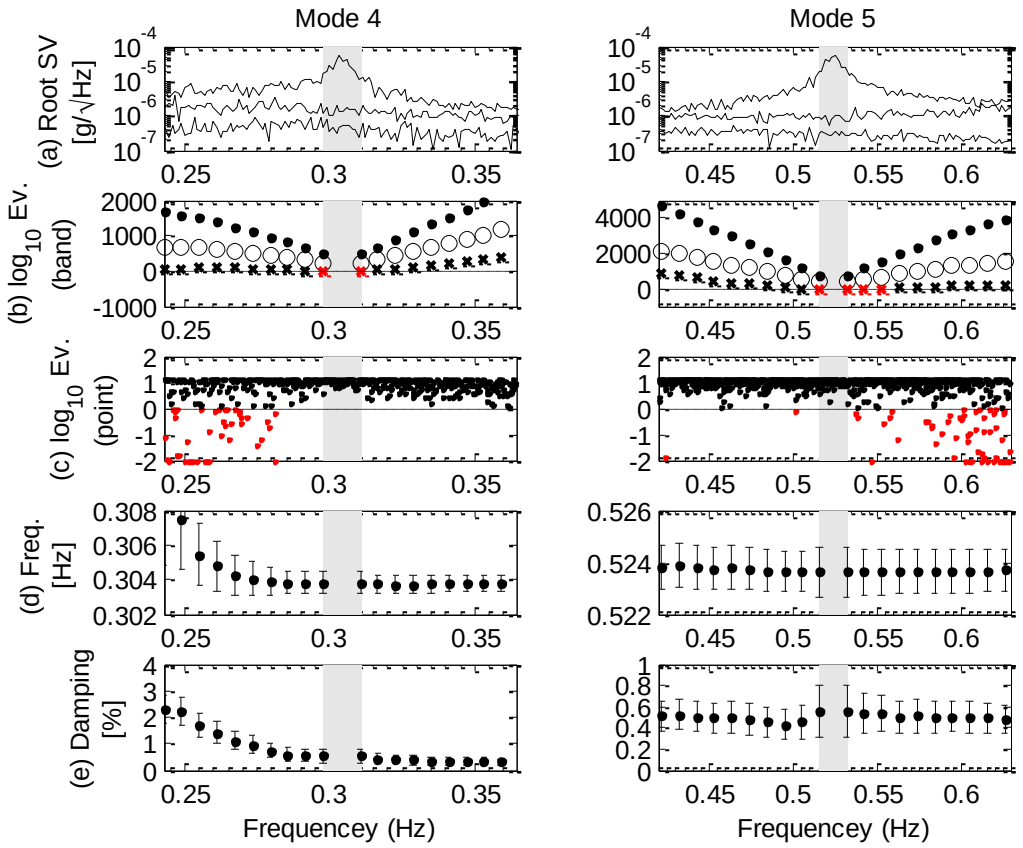

Figure 11. Results for field data, well-separated modes (4 and 5). Same convention as Figure 4

8.2.3 Modes with low s/n ratio

Finally, modes with low s/n ratios are investigated, focusing on two modes near 3.3Hz and 4.1Hz (not shown in Figure 9). For reference, the root modal s/n ratios (= ratio of root PSD of modal response to prediction error) of modes 1, 2, 4 and 5 are, respectively, 23, 66, 46 and 90. The modal s/n ratios of the two modes (referred as 6 and 7) studied here are respectively 4.7 and 2.4, which are substantially lower than the ones studied before and are reaching the lower limit where the modes can be identified. Figure 12 shows the results. As seen in (a), the spectral peaks are less pronounced compared to the modes studied in the previous sections, a reflection of low s/n ratio. Their shape is also less obvious to identify with the variation of dynamic amplification factor. The band evidence ratio in (b) exhibits a zero-one law as before, with M0/M1 and M0/M2 preferring M0 while M0/M3 rejecting it. The point evidence ratio in (c) indicates that it is safe to widen the band by reducing the

lower frequency, say, down to 3.2Hz. It is not advised to increase the upper frequency, however, indicated by the few red dots on the right side of the reference band. For mode 7, (c) indicates that one may increase slightly the upper frequency. These interpretations are consistent with the behavior of the ID results in (d) and (e).

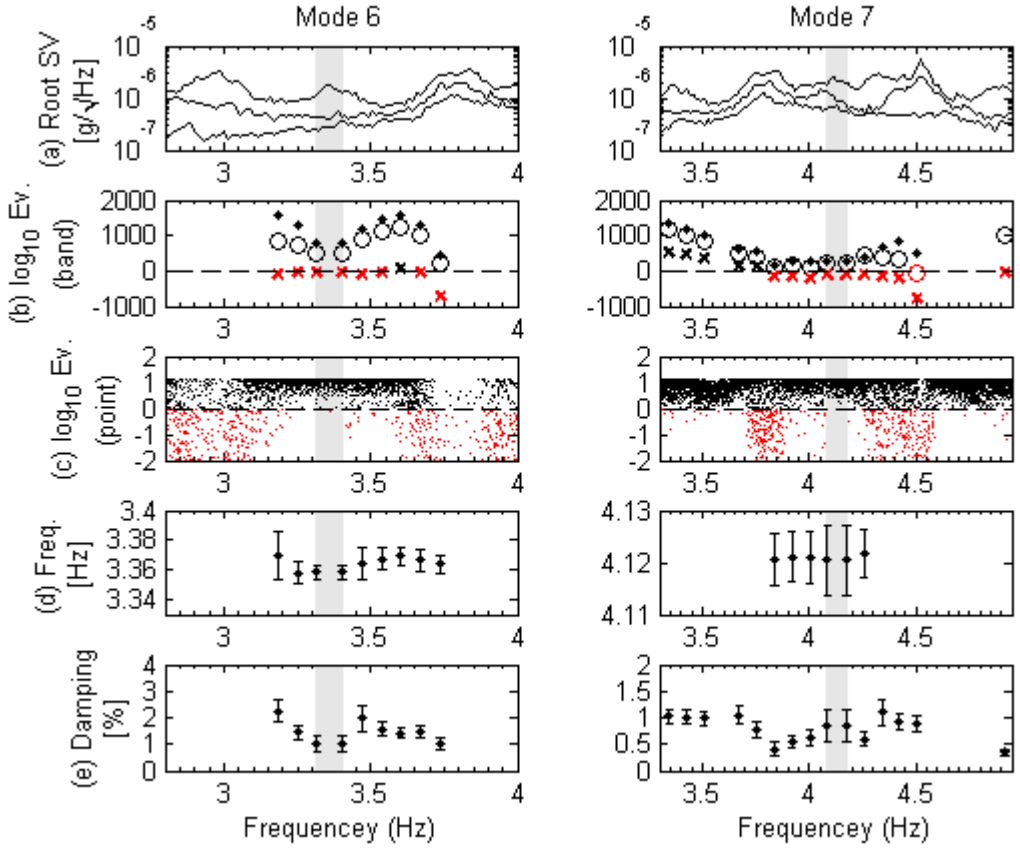

Figure 12. Results for field data, low s/n modes (6 and 7). Same convention as Figure 4

8.3 Discussion

A number of observations and comments can be made from the examples. The band evidence ratios typically lead to zero-one law in the plausibility of the modal model M0, i.e., either 'Yes' or 'No' with practically 100% confidence. This need not reflect reality because their conclusions depend on the competing model class being compared. The model classes considered are not exhaustive. They may not be representative of possibilities other than the modal model when the band is wide, because they all assume constant spectral properties in the subject band. M1 (i.i.d. noise) and M2 (independent noise) are insensitive to unmodeled dynamics, while M3 (correlated noise) can be too sensitive. Using M1 or M2 for validation may fail to detect problems, while using M3 can be too conservative. These characteristics stem from their assumption on constant PSD with frequency. Competing models that account for PSD variation may have better characteristics but they are non-trivial to develop and are out of the present scope.

The point evidence ratio has been found to detect problems correctly. Of course it does not give directly the plausibility of ID model class M0 in a band. Arising from its stochastic nature, the values at different frequencies fluctuate. They should be viewed together rather than individually. The examples provided some benchmarking experience for this. Interpretation based on the point evidence ratio is generally consistent with the bias in the identified modal properties.

Computationally, evaluating the band evidence ratio for a given band requires determining the MPV of modal properties using FFTs of the band. Producing the ratio for a series of expanding bands involves repeated determination of MPVs. Evaluating the point evidence ratio for different frequencies involves determining the MPV only once, for the reference band where the modal model is assumed to be valid.

9 Conclusions

This paper has investigated how to quantify the validity of modal ID model in a narrow frequency band and whether it is safe to expand the band to include more spectral data for operational modal analysis. Band evidence ratio (1) was developed for the former (assuming narrow band) and point evidence ratio (2) for the latter (point-wise but applicable for wide band). Modeling possible model classes when the modal ID model does not hold is an open question that has recurred in limiting what the evidence ratios can tell. Competing model classes and long data behavior of their evidence have been discussed. Trivially inferior ones have been eliminated. For the band evidence ratio, three (non-trivially) competing model classes have been considered for band evidence ratio, which were found to give zero-one law (i.e., 'No' or 'Yes') for model plausibility and were either insensitive or too conservative in detecting unmodeled dynamics. Point evidence ratio was found to strike the middle way, although it does not give a direct quantitative measure for the band in question and still requires human intelligence to interpret. The ratio has a clear Bayesian probabilistic context, allowing for a rigorous development of future decision making tools.

Balancing theoretical rigor and pragmatism, it is advocated that the existence of mode(s) on a narrow band (Question 1) be still judged based on SV spectrum, but the band where spectral data can be safely incorporated (Question 2) can be determined using point evidence ratio (2). The band evidence ratio (1) currently is of limited use. After all, one finds from the examples that the correct conclusions the band evidence ratios were able to give could have been arrived more intuitively and directly from SV spectrum.

Admittedly, the paper does not provide the 'ultimate' solution to the thematic questions. The following discoveries/remarks are however worth-mentioning and contribute to advancing the state-of-the-art. Questions 1 and 2 have been posed in a legitimate way to address model validity and bandwidth selection while admitting a formal mathematical analysis using Bayesian probability logic. Addressing model validity in an absolute sense is conceptually challenging even within the framework of Bayesian model class selection because a representative competing model class is highly non-trivial to construct in the 'what-if-not' situation. It is doubtful whether it is possible to assess model validity in an absolute sense. Properly posing the model validity problem becomes critical for scientific development. Assessing model validity in a band is difficult because it requires modeling the variation of the unknown spectral characteristics of data in the 'what-if-not' situation. In this work this difficulty is bypassed by considering validity at a single frequency. This point-wise consideration also allows the use of maximum entropy principle to determine a representative competitive model class, so that the resulting point evidence ratio can be practically interpreted in an 'absolute sense' for model validity assessment. A closed form expression for the point evidence ratio has been obtained in (2), which allows convenient implementation. Equation (18) provides the general form of the evidence ratio for addressing Question 2, on which future work (in band-wise context) can build.

The proposed method applies to both well-separated and closed modes, whose MPV should be determined by fast algorithms accordingly [13][14]. Although the point evidence ratio is developed in a Bayesian context, it can be directly applied without any Bayesian concept. The ratio can be calculated with the theoretical PSD matrix $\mathbf{E}_k$ in (13) evaluated at the 'best estimate' (non-Bayesian) of modal parameters identified from data or reasonable value determined by other means. As long as the substituted value does not differ significantly from the MPV, results are expected to be qualitatively the same.

10 Acknowledgments

This work is partially supported by the University of Liverpool (Grant EGG10034) and the UK Engineering and Physical Sciences Research Council (Responsive Mode Grant EP/N017897/1 & EP/N017803/1).

11 Appendix A. Derivation of (27)

Let $F(\boldsymbol{\theta}) = -\ln p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0)$ so that $p(\mathcal{F}_k \mid \boldsymbol{\theta}, M_0) = \exp[-F(\boldsymbol{\theta})]$. Approximating $F(\boldsymbol{\theta})$ by a second order Taylor series at $\hat{\boldsymbol{\theta}}$, $F(\boldsymbol{\theta}) \approx F(\hat{\boldsymbol{\theta}}) + g_2^T(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T H_2(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})/2$ where

$g_2 = \nabla^T F(\hat{\boldsymbol{\theta}})$ and $H_2 = \nabla^2 F(\hat{\boldsymbol{\theta}})$ are the (transposed) gradient and Hessian of F at $\hat{\boldsymbol{\theta}}$. Substituting this and (25) into (23), $p(\mathcal{F}_k \mid D_1, B_1, B_2) \approx A\ p(\mathcal{F}_k \mid \hat{\boldsymbol{\theta}}, M_0)$ where

$A = \int (2\pi)^{-n_0/2} \mid H_1 \mid^{1/2} e^{-q(\boldsymbol{\theta})} d\boldsymbol{\theta}$, $H_1 = \hat{\mathbf{C}}^{-1}$, n_0 is the dimension of $\boldsymbol{\theta}$; and

$$q(\boldsymbol{\theta}) = g_2^T(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T (H_1 + H_2)(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \tag{40}$$

It remains to show that A is given by (28). Writing $q(\boldsymbol{\theta})$ in complete square form,

$$q(\boldsymbol{\theta}) = \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^T (H_1 + H_2)(\boldsymbol{\theta} - \boldsymbol{\mu}) - \frac{1}{2} g_2^T (H_1 + H_2)^{-1} g_2 \tag{41}$$

where $\boldsymbol{\mu} = \hat{\boldsymbol{\theta}} - (H_1 + H_2)^{-1} g_2$. Substituting gives,

$$\begin{aligned} A = & \frac{\mid H_1 \mid^{1/2}}{\mid H_1 + H_2 \mid^{1/2}} e^{g_2^T (H_1 + H_2)^{-1} g_2} \times \\ & \int (2\pi)^{-n_0/2} \mid H_1 + H_2 \mid^{1/2} \exp\left[-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^T (H_1 + H_2)(\boldsymbol{\theta} - \boldsymbol{\mu}) \right] d\boldsymbol{\theta} \end{aligned} \tag{42}$$

The integrand is a Gaussian PDF with mean $\boldsymbol{\mu}$ and covariance matrix $(H_1 + H_2)^{-1}$ and so it integrates to 1. Equation (28) then follows by noting that $\mid H_1 \mid^{1/2} / \mid H_1 + H_2 \mid^{1/2} = 1/\sqrt{\mathbf{I} + H_1^{-1} H_2}$.

12 Appendix B. Proof of maximum entropy distribution in (35)

This appendix shows that the maximum entropy distribution of positive definite Hermitian matrix $\mathbf{Q}$ with expectation $\hat{\mathbf{E}}_k$ is given by the complex Wishart PDF with n DOFs and covariance matrix $\hat{\mathbf{E}}_k / n$, as in (35). The solution is the stationary point of the functional $J(p)$ in (34) under the constraint in (33). Taking perturbation of J w.r.t. p,

$$\delta J = -\int \left[\ln p + 1 + \chi + tr(\boldsymbol{\Lambda}\mathbf{Q}) \right] \delta p\ d\mathbf{Q} \tag{43}$$

Setting $\delta J = 0$ to hold for arbitrary δp requires the bracketed term to be identically zero. This gives $p(\mathbf{Q}) = \exp(-1-\chi)\exp[-tr(\boldsymbol{\Lambda}\mathbf{Q})]$. Enforcing $\int p(\mathbf{Q}) d\mathbf{Q} = 1$ gives

$$\exp(-1-\chi) = \left[\int \exp[-tr(\boldsymbol{\Lambda}\mathbf{Q})] d\mathbf{Q} \right]^{-1} = \frac{\mid \boldsymbol{\Lambda} \mid^n}{\pi^{n(n-1)/2} \prod_{i=1}^{n} (i-1)!} \tag{44}$$

The integral in (44) was obtained by noting that a complex Wishart PDF with n DOFs and covariance matrix $\mathbf{\Lambda}^{-1}$ is $\exp[-tr(\mathbf{\Lambda Q})]|\mathbf{\Lambda}|^n/\pi^{n(n-1)/2}\prod_{i=1}^{n}(i-1)!$ and it integrates w.r.t. $\mathbf{Q}$ to 1. Substituting (44) shows that p is a complex Wishart PSD with n DOFs and covariance matrix $\mathbf{\Lambda}^{-1}$. A matrix variate with this PDF has expectation $n\mathbf{\Lambda}^{-1}$, and so enforcing the expectation constraint (33) gives $\mathbf{\Lambda}=\hat{\mathbf{E}}_k/n$.

13 Appendix C. Derivation of (35) and (36) for the case of one measured DOF

To develop intuition into the derivation of (35) and (36), consider the case of one measured DOF ($n=1$). Then $\mathcal{F}_k$ is a circularly symmetric complex Gaussian scalar with zero mean and some (unknown) variance $Q_k>0$. That is, $X_k=\mathrm{Re}\,\mathcal{F}_k$ and $Y_k=\mathrm{Im}\,\mathcal{F}_k$ are independent Gaussian with zero mean and variance $Q_k/2$, so that

$$\begin{aligned}E[\mathcal{F}_k\mathcal{F}_k^T\,|\,Q_k]&=E[X_k^2\,|\,Q_k]-E[Y_k^2\,|\,Q_k]+2\mathbf{i}E[X_kY_k\,|\,Q_k]=\frac{Q_k}{2}-\frac{Q_k}{2}+2\mathbf{i}\times 0=0\\ E[\mathcal{F}_k\mathcal{F}_k^*\,|\,Q_k]&=E[X_k^2\,|\,Q_k]+E[Y_k^2\,|\,Q_k]=\frac{Q_k}{2}+\frac{Q_k}{2}=Q_k\end{aligned}\tag{45}$$

The PDF of $\mathcal{F}_k$ is the joint PDF of X_k and Y_k:

$$p(\mathcal{F}_k\,|\,Q_k,D_1,B_1,B_2')=\frac{e^{-X_k^2/2(Q_k/2)}}{\sqrt{2\pi(Q_k/2)}}\times\frac{e^{-Y_k^2/2(Q_k/2)}}{\sqrt{2\pi(Q_k/2)}}=\frac{1}{\pi Q_k}e^{-|\mathcal{F}_k|^2/Q_k}\tag{46}$$

This agrees with (30). When M0 does not hold at f_k the PDF of Q_k is modeled as the maximum entropy distribution with its mean is equal to $\hat{E}_k$. It is obtained by minimizing the following functional of PDF p with Lagrange multipliers χ and λ:

$$J(p,\chi,\lambda)=-\int_0^\infty p(Q)\ln p(Q)dQ+\chi\left[1-\int_0^\infty p(Q)dQ\right]+\lambda\left[\hat{E}_k-\int_0^\infty Q\;p(Q)dQ\right]\tag{47}$$

Setting $\delta J=0$ (perturbation w.r.t. variation in p) gives

$$-\int_0^\infty(\ln p+1+\chi+\lambda Q)\delta p\;dQ=0\tag{48}$$

This is satisfied for arbitrary δp when $\ln p+1+\chi+\lambda Q\equiv 0$, i.e., $p(Q)=e^{-1-\chi-\lambda Q}$.

Enforcing $\int_0^\infty p(Q)dQ=1$ gives $e^{-1-\chi}=1/\int_0^\infty e^{-\lambda Q}dQ=\lambda$ and so $p(Q)=\lambda e^{-\lambda Q}$, i.e., exponential PDF with mean $1/\lambda$. Further enforcing $\int_0^\infty Qp(Q)dQ=\hat{E}_k$ gives $\lambda=1/\hat{E}_k$. Thus,

$$p(Q_k \mid D_1, B_1, B_2') = \frac{1}{\hat{E}_k} \exp\left(-\frac{Q_k}{\hat{E}_k}\right) \qquad (n=1) \qquad (49)$$

This agrees with (35). The evidence $p(\mathcal{F}_k \mid D_1, B_1, B_2')$ is obtained by marginalizing out Q_k:

$$\begin{aligned} p(\mathcal{F}_k \mid D_1, B_1, B_2') &= \int_0^\infty p(\mathcal{F}_k \mid Q_k, D_1, B_1, B_2')\, p(Q_k \mid D_1, B_1, B_2')\, dQ_k \\ &= \int_0^\infty \frac{1}{\pi Q_k} \exp\left(-\frac{|\mathcal{F}_k|^2}{Q_k}\right) \times \frac{1}{\hat{E}_k} \exp\left(-\frac{Q_k}{\hat{E}_k}\right) dQ_k \\ &= \frac{1}{\pi \hat{E}_k} \int_0^\infty \frac{1}{Q_k} \exp\left(-\frac{|\mathcal{F}_k|^2}{Q_k} - \frac{Q_k}{\hat{E}_k}\right) dQ_k \end{aligned} \qquad (50)$$

For $a, b > 0$, $\int_0^\infty x^{-1} e^{-a/x - x/b} dx = 2K_0(2\sqrt{a/b})$. Applying this result,

$$p(\mathcal{F}_k \mid D_1, B_1, B_2') = \frac{2}{\pi \hat{E}_k} K_0\left(2\sqrt{\frac{|\mathcal{F}_k|^2}{\hat{E}_k}}\right) \qquad (n=1) \qquad (51)$$

This agrees with (36) where $\xi_k = |\mathcal{F}_k|^2 / \hat{E}_k$.

For general $n \geq 1$, the derivation is much more complicated because $\mathbf{Q}_k$, being positive definite Hermitian, is a 'structured' matrix (i.e., entries are related). The associated Jacobian in the integration over $\mathbf{Q}_k$ needs to be evaluated using special techniques [44]. Demarginalizing $\mathbf{Q}_k$ is not trivial either; details are referred to [41].

14 References

[1] C. Rainieri, G. Fabbrocino, Operational Modal Analysis of Civil Engineering Structures: An Introduction and Guide for Applications, Springer, New York, 2014.

[2] R. Brincker, C. Ventura, Introduction to Operational Modal Analysis, Wiley, London, 2015.

[3] D.J. Ewins, Modal testing: Theory and Practice, Research Studies Press, PA, USA, 2000.

[4] M.I. Friswell, J.E. Mottershead, Finite Element Model Updating in Structural Dynamics, Kluwer Academic Publishers, Dordrecht, 1995.

[5] H. Wenzel, D. Pichler, Ambient vibration monitoring, Wiley, UK, 2005.

[6] C.R. Farrar, K. Worden, Structural Health Monitoring: A Machine Learning Perspective, Wiley, London, 2012.

[7] S.W. Doebling, C.R. Farrar, M.B. Prime, D.W. Shevitz, A review of damage identification methods that examine changes in dynamic properties, Shock and Vibration Digest, 30 (2) (1998) 91–105.

[8] J.M.W. Brownjohn, Structural health monitoring of civil infrastructure, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 365(1851) (2007) 589-622.

[9] B. Peeters, G. De Roeck, Stochastic system identification for operational modal analysis: a review, Journal of Dynamical Systems, Measurement, and Control, 123 (2001) 659-667.

[10] R. Brincker, L. Zhang, P. Anderson, Modal identification of output-only systems using frequency domain decomposition, Smart Materials and Structures, 10 (3) (2001) 441-455.

[11] K.V. Yuen, L.S. Katafygiotis, Bayesian time-domain approach for modal updating using ambient data, Probabilistic Engineering Mechanics, 16(3) (2001) 219-231.

[12] K.V. Yuen, L.S. Katafygiotis, Bayesian Fast Fourier Transform approach for modal updating using ambient data, *Advances in Structural Engineering*, 6(2) (2003) 81-95.

[13] S.K. Au, Fast Bayesian FFT method for ambient modal identification with separated modes, Journal of Engineering Mechanics, ASCE, 137(3) (2011) 214-226.

[14] S.K. Au, Fast Bayesian ambient modal identification in the frequency domain, Part I: posterior most probable value, Mechanical Systems and Signal Processing, 26(1) (2012) 60-75.

[15] S. K. Au, F. L. Zhang, Y. C. Ni, Bayesian operational modal analysis: theory, computation, practice, Computers and Structures, 126 (2013) 3-14.

[16] R.S. Pappa, K.B. Elliott, A. Schenk, Consistent-mode indicator for the eigensystem realization algorithm, Journal of Guidance, Control and Dynamics, 16(5) (1993) 852-858.

[17] J.B. Elsner, A.A. Tsonis, *Singular Spectrum Analysis. A New Tool in Time Series Analysis*, Plenum Press, 1996.

[18] W. Fladung, R. Rost, Application and correction of the exponential window for frequency response functions, Mechanical Systems and Signal Processing, 11(1) (1997) 23–36.
[19] H. Jokinen, J. Ollila, O. Aumala, On windowing effects in estimating averaged periodograms of noisy signals, Measurement, 28 (2000) 197–207.
[20] P.J. Torvik, On estimating system damping from frequency response bandwidths, Journal of Sound and Vibration, 330 (2011) 6088–6097.
[21] S.K. Au, Insights on the Bayesian spectral density approach for operational modal analysis, Mechanical Systems and Signal Processing, 66-67 (2015) 1-12.
[22] S.K. Au, Uncertainty law in ambient modal identification. Part I: theory, Mechanical Systems and Signal Processing 48(1-2) (2014) 15-33, 2014.
[23] E.M. Hernandez, N.R. Polanco, A lower bound for the variance of frequency and damping ratio identified from noisy vibration measurements, Structural Control and Health Monitoring, 23(1) (2016) 5-19.
[24] D.R. Brillinger, *Time series: Data analysis and theory*, Holden-Day, Inc., San Francisco, 1981.
[25] R.T. Cox, *The Algebra of Probable Inference*, Johns Hopkins Press, Baltimore, 1961.
[26] G.E.P. Box, G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Wiley, New York, 1992.
[27] E.T. Jaynes, *Probability Theory: The Logic of Science*, Cambridge University Press, UK, 2003.
[28] J.L. Beck, Bayesian system identification based on probability logic, *Structural Control and Health Monitoring*, 17(7) (2010) 825–847.
[29] H. Akaike, A new look at the statistical identification model, IEEE Transactions of Automation and Control, 19 (1974) 716–723.
[30] H. Akaike, On entropy maximization principle. Applications of statistics, P. R. Krishnaiah, ed., North-Holland, Amsterdam, The Netherlands, 27–41, 1976.
[31] G. Schwarz, Estimating the dimension of a model.'' Annals of Statistics, 6(2) (1978) 461–464.
[32] J.L. Beck, K.V. Yuen, Model selection using response measurements: Bayesian probabilistic approach, *Journal of Engineering Mechanics*, 130(2) (2004) 192-203.
[33] J. Ching, Y.J. Chen, Transitional Markov Chain Monte Carlo method for Bayesian model updating, model class selection and model averaging, Journal of Engineering Mechanics, ASCE, 133 (2007) 816-832.
[34] S.H. Cheung, J.L. Beck, Calculation of the posterior probability for Bayesian model class assessment and averaging from posterior samples based on dynamic system data, Computer-Aided Civil and Infrastructure Engineering 25 (2010) 304-321.
[35] J.A. Hoeting, D. Madigan, A.E. Raftery, C.T. Volinsky, Bayesian model averaging: a tutorial, Statistical Science, 14(4) (1999) 382-417.
[36] E. Simoen, C. Papadimitriou, G. Lombaert, On prediction error correlation in Bayesian model updating, Journal of Sound and Vibration, 332 (18) (2013) 4136-4152.

[37] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, Journal of Engineering Mechanics, 124(4) (1998) 455–461.

[38] S.F. Gull, Bayesian inductive inference and maximum entropy. In G. J. Erickson & C. R. Smith (eds), *Maximum Entropy and Bayesian Methods*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1988.

[39] W.H. Jefferys, J.O. Berger, Ockham's Razor and Bayesian Analysis, *American Scientist*, 80 (1992) 64-72.

[40] R.A. Horn, C.R. Johnson, Matrix Analysis, Cambridge University Press, UK, 2013.

[41] M. Guillaud, M. Debbah, A.L. Moustakas, Modeling the multiple-antenna wireless channel using maximum entropy methods, International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Saratoga Springs, NY, IEEE, 435-442, 2007.

[42] M. Guillaud, M. Debbah, A.L. Moustakas, A maximum entropy characterization of spatially correlated MIMO wireless channels, Wireless Communications and Networking Conference 2007, IEEE, 1039-1044, 2007.

[43] S. Adhikari, Matrix variate distributions for probabilistic structural dynamics, AIAA Journal, 45(7) (2007) 1748-1762.

[44] A.M. Mathai, Jacobians of Matrix Transformation and Functions of Matrix Arguments, World Scientific, London, 1997.

[45] S.K. Au, P. To, Full-scale validation of dynamic wind load on a super-tall building under strong wind, Journal of Structural Engineering, ASCE, 138(9) (2012) 1161-1172.